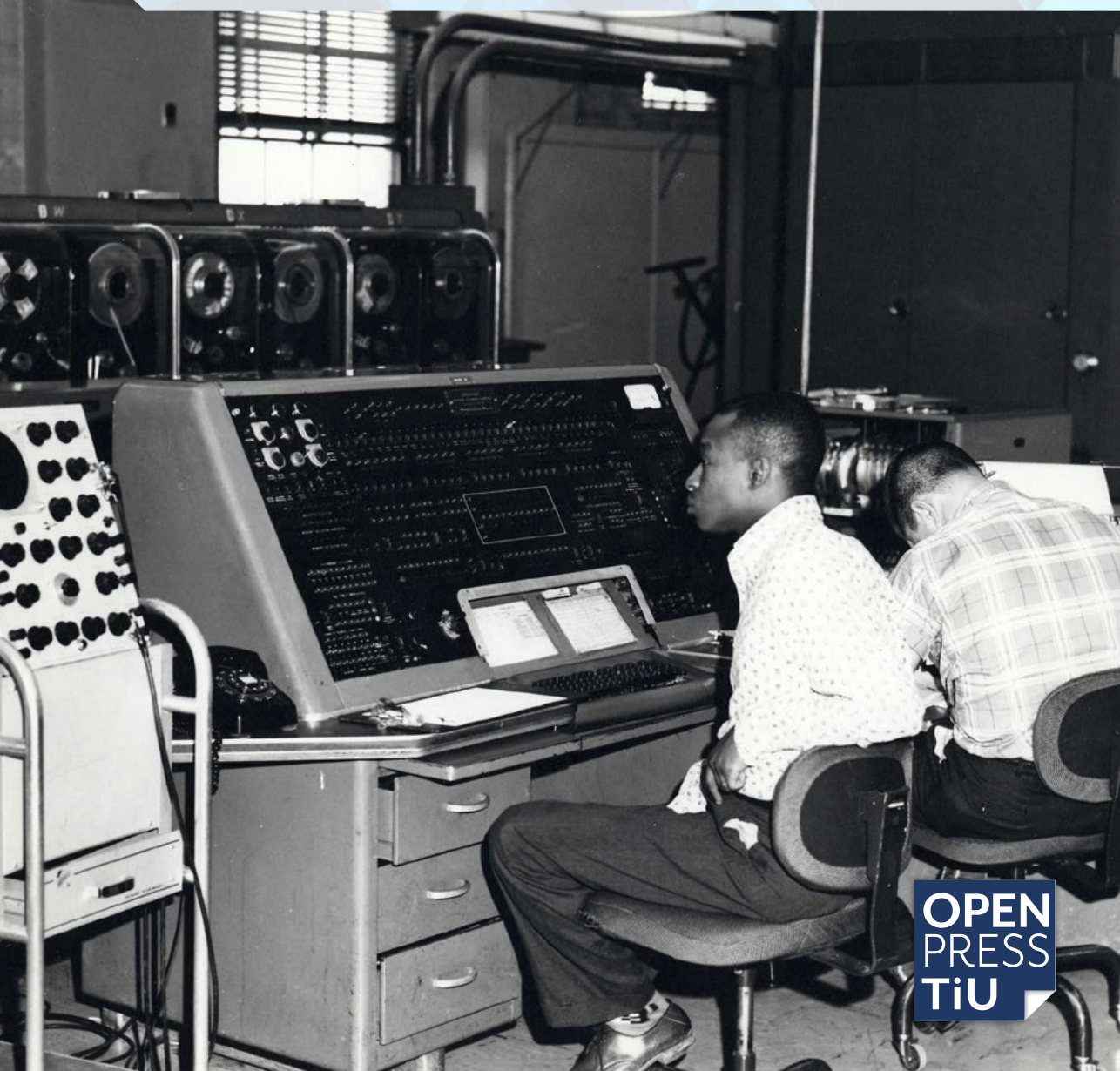
Technology and Regulation

2020
Volume 2



OPEN PRESS TiU

TECHNOLOGY AND REGULATION 2020

Volume 2

DOI: 10.26116/techreg.volume.2020
ISBN: 978-94-6240-671-1 (Interactive PDF)

Technology and Regulation
Tilburg Institute for Law, Technology, and Society (TILT)
Tilburg Law School
P.O. Box 90153
5000 LE Tilburg
The Netherlands
techreg.org

Principal Contact:
Ronald Leenes
Editor-in-Chief
Tilburg Institute for Law, Technology,
and Society (TILT), Tilburg Law School
r.e.leenes@tilburguniversity.edu

Support Contact:
Aaron Martin
a.k.martin@uvt.nl

Published by: Open Press TiU
Contact details: info@openpresstiu.edu
https://www.openpresstiu.org/

Cover Design by: Wolf Publishers, Claudia Tofan

Aims and Scope

Technology and Regulation (TechReg) is an international journal of law, technology and society, with an interdisciplinary identity. TechReg provides an online platform for disseminating original research on the legal and regulatory challenges posed by existing and emerging technologies (and their applications) including, but by no means limited to, the Internet and digital technology, artificial intelligence and machine learning, robotics, neurotechnology, nanotechnology, biotechnology, energy and climate change technology, and health and food technology. We conceive of regulation broadly to encompass ways of dealing with, ordering and understanding technologies and their consequences, such as through legal regulation, competition, social norms and standards, and technology design (or in Lessig's terms: law, market, norms and architecture). We aim to address critical and sometimes controversial questions such as: How do new technologies shape society both positively and negatively? Should technology development be steered towards societal goals, and if so, which goals and how? What are the benefits and dangers of regulating human behaviour through technology? What is the most appropriate response to technological innovation, in general or in particular cases? It is in this sense that TechReg is intrinsically interdisciplinary: we believe that legal and regulatory debates on technology are inextricable from societal, political and economic concerns, and that therefore technology regulation requires a multidisciplinary, integrated approach. Through a combination of monodisciplinary, multidisciplinary and interdisciplinary articles, the journal aims to contribute to an integrated vision of law, technology and society. We invite original, well-researched and methodologically rigorous submissions from academics and practitioners, including policy makers, on a wide range of research areas such as privacy and data protection, security, surveillance, cybercrime, intellectual property, innovation, competition, governance, risk, ethics, media and data studies, and others.

TechReg is double-blind peer-reviewed and completely open access for both authors and readers. TechReg does not charge article processing fees.

Editorial Team

CONTENTS

01

Technology and Regulation

Constructing Commercial Data Ethics

Linnet Taylor* and Lina Dencik**

data ethics, big data, business ethics, corporate social responsibility, information

l.e.m.taylor@tilburguniversity.edu

dencikl@cardiff.ac.uk

The ethics of big data and AI have become the object of much public debate. Technology firms around the world have set up ethics committees and review processes, which differ widely in their organisation and practice. In this paper we interrogate these processes and the rhetoric of firm-level data ethics. Using interviews with industry, activists and scholars and observation of public discussions, we ask how firms conceptualise the purposes and functions of data ethics, and how this relates to core business priorities. We find considerable variation between firms in the way they use ethics. We compare strategies and rhetoric to understand how commercial data ethics is constructed, its political and strategic dimensions, and its relationship to data ethics more broadly.

1. Introduction

The rapid ascent of big data and AI as objects of attention in public debate over the last decade has created acute visibility and demand for both data and AI ethics. Firms engaged in the data economy have had to engage in discussions on ethics that at first took them largely by surprise, and have experienced a steep learning curve as they have been forced to define a moral stance on civil and political rights, freedom of speech, privacy, autonomy, and to justify their research and operational choices beyond concerns of shareholder value. The applied ethics of data and, more recently, AI have been central to how firms have addressed this challenge, bringing the ethics of technology out of the academy and into the corporate world through consulting, advisory boards and the formation of tools, guidelines and assessment services by third parties on an entrepreneurial basis.

This extraction of applied ethics from its origins in academia and its insertion into the high-stakes, high-velocity field of commercial technology development has resulted in a new commercially stimulated data ethics with its own objectives and rhetoric. This commercial ethics aims to shape social expectations of both data technologies and of the firms that create and deploy them: it is an instrumental ethics that aims to have tangible political and economic effects. In order to understand these effects, one starting point is to analyse data ethics as a discourse, separating out the rhetoric and practices involved in commercial data ethics and exploring them as strategic tools in a business environment. This paper aims to interrogate its starting points, its moral stance on data technologies and, most importantly, what kind of work its proponents and stakeholders see it as doing in relation to the technology sector.

The research for this paper was conducted over the period 2014-2019. The methods used consisted of institutional ethnography and elite interviews[1] at technology firms including mobile network operators and data analytics consultancies, observation and follow-up interviews conducted by participating in data ethics and governance events where we participated in discussions with a range of groups. These included academic computer science and data science researchers, specialists in NGOs and international organisations conducting data analytics, and commercial data analytics specialists within firms. We also followed policy discussions over this period through meetings and reports. Finally, we conducted participant observation at various events on the ethics of AI and data analytics in the UK, the Netherlands, Germany and hosted by international organisations such as the World Economic Forum and the United Nations.

In order to further inform our findings, we conducted a series of eight interviews comprising three leaders of civil society organisations working on technology and rights, an ethicist, two corporate employees leading data ethics programs, and one independent member of a corporate ethics committee. These interviews focused on the specific issues we planned to focus on in this paper. It is therefore both these

* Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, the Netherlands.
** School of Journalism, Media and Cultrural Studies, Cardiff University, UK.
This paper was written with the support of the Horizon 2020 program of the European Union, ERC Starting Grant no. 757247 (Linnet Taylor) and ERC Starting Grant no. 759903 (Lina Dencik). We thank our interviewees for their contributions, and also two anonymous reviewers for their advice.

Received 10 Sept 2019, Accepted 23 Mar 2020, Published: 17 Apr 2020.

1 This paper draws on qualitative research conducted during a Marie Curie postdoctoral fellowship at the University of Amsterdam on big data in the development and humanitarian sectors, and during the 'Global Data Justice' ERC project at the University of Tilburg. Together these projects involved 200 formal and informal interviews with users and managers of big data resources, as well as observation in firms dealing with big data in different sectors. Some of the interviews were also conducted as part of the OSF-funded project 'Toward Democratic Auditing: Civic Participation in the Scoring Society'.

interviews and the fieldwork preceding them which form the basis for our analysis of the current state of play in commercial data ethics. Where possible we reference the source of our findings, but we have also reported some findings where the source or interviewee did not wish to be named, or where doing so would expose them to negative consequences. It is worth noting that this final set of nine interviewees were all based in or worked in Europe. In our research we aimed to understand what the actors involved understand by data ethics, what practices and power relations they observe in relation to the practice of data ethics in the corporate environment, and finally, what is ethical about data ethics.

This paper is not written from either the disciplinary perspective of ethics, or the subdiscipline of data ethics. Our aim, rather than establishing principles or advocating a particular agenda for the field of data ethics, is to provide a critical analysis of the commercial sector's development of 'data ethics' as a guiding set of principles, and to interrogate how it opens up possibilities for action and avenues of discussion while closing down others. As such, the starting point for this analysis is the notion that we can identify particular constructs of 'data ethics' and 'AI ethics' existing amongst private-sector developers and implementers of data technology, and that these need to be interrogated to highlight their power dynamics and politics. Which perspectives and aims do these discourses of ethics centre; which actors in tech companies define and articulate them; and what are the political and rhetorical strategies they use to leverage influence and change? As such, this paper does not offer or endorse any particular ethical view on data technologies, but instead provides a critical perspective on the work these constructs are doing in the private sector, and in society more broadly. The paper therefore takes a political economy approach to the phenomenon of technology firms' ethics processes. Our scope does not extend to the intersecting world of public-sector data ethics, though this type of analysis could also be conducted there (and some of our interviews with activists touched on this area in their responses).

2. Typology - What kinds of ethics are appearing in relation to data science?

Overall, commercial perspectives on data ethics are, unsurprisingly, defensive. They are defined by a technologically determinist framing where innovation is axiomatically good and therefore marches on, and the economic value of data must be realised. The big tech and advisory firms focus on ethics as a way to build, maintain or resurrect 'consumer trust'[2], a trust that is also cited as an objective to be achieved through investment in ethics centres and research within academia, such as the Facebook-funded AI ethics research centre at Technical University of Munich.[3] But without strong regulation of the technology sector to create trustworthiness, it may be premature to focus on evoking trust in data technologies. It is worth asking what kind of ethics is at work under this rubric of promoting trust and functionality in a world of inexorable technological expansion? Observations in the field suggest various possibilities: ethical discussion is seen by some as the oil that enables the digital economy to run smoothly without interruption from law and regulation; others pragmatically use ethics discussions for the tactical containment of reputational risk. The bigger firms see data ethics as a kind of insurance: an antidote to moral panic on the part of the public (one of the anxieties driving warnings of 'loss of consumer trust'), while others see it as a variant of corporate social responsibility that is part of a mission statement about promoting certain public values while not doing harm.

Data ethics, as a field, can be thought of as a network of nodes representing frequently entangled and interacting but different streams of thought and practice. First, a philosophical node stemming from the academy, which defines data ethics as the branch of ethics that studies and evaluates moral problems related to data, algorithms and corresponding practices, in order to formulate and support morally good solutions.[4] Second, there is a node of applied ethics conducted by philosophers, computer and social scientists, many of them working within, or in collaboration with, the commercial domain, of which value-sensitive design is one element.[5] Another element within this node continues a long-standing tradition of computer ethics while changing the level of abstraction of ethical enquiries from an information-centric to a data-centric one, i.e. from a focus on how to treat information as an input and output of computing to a focus on how people access, analyse and manage data in particular.[6] This node tends to focus not on any specific technology but on what any digital technology manipulates. Key issues concern re-identification or de-anonymization and risks to privacy, forms of discrimination and abuse, trust, transparency, accountability, lack of public awareness and responsible innovation and usage. This node is connected to one of civil society advocacy where data ethics is providing a framework for guidelines to advance data developments 'for good' across a range of contexts (for example Open Data Institute's 'Data Ethics Canvas' and UNI Global Union's call for a 'Global Convention on Ethical AI'). In the UK, the government agreed to set up a 'Council of Data Ethics' in 2016 in response to a report by the Science and Technology Committee on 'The big data dilemma'[7], which became the Centre for Data Ethics and Innovation. This is in parallel to similar councils being created in the US and elsewhere. Finally, there is a node of the network dominated by industry, incorporating advisory services, tech corporations' own operations with regard to ethical review and reflection, and work by specialists that aims to shape these corporate processes.[8]

Whilst we recognize the entanglement of these different nodes, in

2 Accenture, 'Universal Principles of Data Ethics: 12 Guidelines for Developing Ethics Codes' (2016) https://www.accenture.com/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf; World Economic Forum, 'Rethinking Personal Data : A New Lens for Strengthening Trust' (World Economic Forum 2014).

3 See Deutsche Welle, 'Facebook Funds AI Ethics Center in Munich' (DW.com, 2019). Available at https://www.dw.com/en/facebook-funds-ai-ethics-center-in-munich/a-47156591 (last accessed 1 April 2020).

4 Luciano Floridi and Mariarosaria Taddeo, 'What Is Data Ethics?' (2016) What is data ethics? *Phil. Trans. R. Soc. A.* 374: 20160360, https://doi.org/10.1098/rsta.2016.0360.

5 I van de Poel and L Royakkers, 'The Ethical Cycle' (2007) 71 Journal of Business Ethics 1; Jeroen van den Hoven, 'ICT and Value Sensitive Design' in Philippe Goujon and others (eds), *The Information Society: Innovation, Legitimacy, Ethics and Democracy In honor of Professor Jacques Berleur s.j.*, vol 233 (Springer US 2007) https://link.springer.com/10.1007/978-0-387-72381-5_8 (last accessed 1 April 2020); Peter-Paul Verbeek, *What Things Do: Philosophical Reflections on Technology, Agency, and Design*, vol 43 (Penn State Press 2005) https://choicereviews.org/review/10.5860/CHOICE.43-1523 (last accessed 1 April 2020).

6 See e.g. Adams, A. A., Report of a debate on Snowden's actions by ACM members. (2014) *ACM SIGCAS Computers and Society, 44(3), 5–7.* https://doi.org/10.1145/2684097.2684099 *(last accessed 1 April 2020); Jacob Metcalf, and Kate Crawford, Where are human subjects in big data research? The emerging ethics divide. (2016) Big Data & Society, June, 1–34.* https://doi.org/10.1177/2053951716650211 *(last accessed 1 April 2020).*

7 Commons Science and Technology Committee, 'The Big Data Dilemma' (UK House of Commons 2016) https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2015/big-data/ (last accessed 1 April 2020).

8 See e.g. Gry Hasselbalch and Pernille Tranberg, *Data Ethics - the New Competitive Advantage* (Publishare 2016) *11.*

this paper we are particularly interested in the corporate engagement with data ethics, its vision and objectives, and the kind of power it draws on. Perhaps the most recognisable narrative for this agenda is articulated by Hasselbalch and Tranberg, who frame data ethics as a new evolution of the corporate social responsibility agenda, forming 'a new competitive advantage':

> A company's degree of "data ethics awareness" is not only crucial for survival in a market where consumers progressively set the bar, it's also necessary for society as a whole. It plays a similar role as a company's environmental conscience – essential for company survival, but also for the planet's welfare.[9]

This struggle for competitive advantage through data ethics is remarkable for its social scope and penetration. For example, on issues relating to data, law and ethics Microsoft has established a theme within its research arm, Microsoft Research, but also makes gifts to universities and think tanks, sponsors conferences such as the Fairness, Accountability and Transparency in Computer Science series, and offers project sponsorship and individual fellowships for scholars. Google's reach is similar, as is Facebook's, creating a web of funding that touches a substantial proportion of the public intellectuals critical of the power and reach of big tech.

This is perhaps not so surprising considering that one of the challenges of applied data ethics is creating a process that has both moral substance and traction at the operational level. A long list of data ethics principles and codes can be found on the websites of tech firms, civil society organisations and government authorities, but principles lack traction on daily behaviour. If employees are required to 'do the right thing'[10], or to 'be fair'[11], very different ideas of 'right' or 'fair' may come into play.[12] Conversely, if a precise taxonomy of harms is produced and operationalised into guidelines, this potentially creates the feeling that employees may do anything that is not on the list.

In the commercial sphere, negotiating this tension is made more difficult by the fact that 'data ethics' is relatively rarely practiced by ethicists and instead tends to become a flexible and general approach to 'doing no evil', unstructured by the apparatus of ethical reflection built up over thousands of years of philosophical tradition. This approach lends itself to relativism, the belief that nothing is inherently right or wrong, and to a situation where ethical reflection is bounded by the moral norms of the environment in which it is practiced. Where the environment in question is the data technology sector, the task of ethical reflection tends to be framed in terms of making it possible for data to flow within market structures – an approach which constitutes an attempt at capture by industry of the starting point for ethical reflection.

One instance of this capture was publicly surfaced in the debate over specific boundaries and no-go areas for AI, in relation to the European Union's economic strategy for the technology over the coming decade. Rapporteur Thomas Metzinger described in his testimony to the European Parliament how after the European Commission's High Level Expert Group worked for several months to establish 'non-negotiable red lines' in relation to the use of AI, 'industry [participants in the expert group] said the word 'red lines' cannot be in this document any more, at any point [...] and the words 'non-negotiable have to be out of this document.'[13]

This need for flexibility can lead to a situation where instead of a process of reflection guided by a core set of philosophical principles, and where the outcome is decided by that reflection, the outcome is already decided at the start and then ethical reflection is shaped to provide a route to it. As Hannah Couchman of Liberty notes, 'the problem with data ethics is it does mean something different to everyone'[14]. This process can also give rise to 'a checklist approach to ethics', according to Javier Ruiz, Policy Director at Open Rights Group in the UK (hereafter 'ORG'), where 'as long as you can tick all these boxes, you can be sure that what you are doing is ethical'[15].

It is possible to distinguish (at least) two main currents in the emerging field of data ethics. One might be described as a micro-ethical approach which asks how the individual should approach their work with data in research or practice. This approach is the basis for guidelines and codes, and for much of the work of consultants and external advisors working with firms on their ethical profile.[16] Accenture, for example, frames 'universal principles' based in biomedical ethics.[17] These endorse the fundamental principles of research ethics: beneficence, respect for persons and justice, and which focus largely on the individual researcher as responsible for his or her own ethical behaviour. They do not point at the organisational level in terms of ethical duty, but instead (quoting the Association of Computing Machinery (ACM)'s guidance) warn that an individual data scientist has a responsibility to warn their organisation if it is using data science unethically overall.

Ethics codes tend to incorporate requirements for legal compliance (citing privacy, informed consent, security and data ownership), again targeted at the individual. This creates a paradox where individuals may be doing ethical and compliant work for a company that is, in the larger context of its business model, using their work to violate rights. One example of this is the justifications provided by both those employed at Cambridge Analytica and at Global Science Research, the two organisations that collaborated to make Facebook user data available for political microtargeting of US voters in the 2016 presidential election. Each claimed to have been doing their own work with due regard for research ethics, privacy and compliance, while also unwittingly collaborating in actions which were overall unethical

9 Hasselbalch and Tranberg (n 8) 11.

10 Kate Conger, 'Google Removes "Don't Be Evil" Clause From Its Code of Conduct' (*Gizmodo*, 2018) https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393 (last accessed 1 April 2020).

11 Monetary Authority of Singapore, 'Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector' (Monetary Authority of Singapore 2019) https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat (last accessed 1 April 2020).

12 See e.g. Keyes, O., Hutson, J., & Durbin, M., A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry. (2019) *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19, 1–11.* https://doi.org/10.1145/3290607.3310433 on how a formalised idea of fairness could be employed for entirely ethically impermissible on how a formalised idea of fairness could be employed for entirely ethically impermissible purposes.

13 Thomas Metzinger, 'Dialogue Seminar on Artificial Intelligence: Ethical Concerns; European Parliament' (2019) https://www.europarl.europa.eu/streaming/?event=20190319-1500-SPECIAL-SEMINAR1&start=2019-03-19T15:44:53Z&end=2019-03-19T15:56:00Z&language=en (last accessed 1 April 2020).

14 Interview with Hannah Couchman, Advocacy and Policy Officer at Liberty, 1 June 2018

15 Interview with Javier Ruiz, Policy Director at Open Rights Group, 22 June 2018.

16 Luke Stark and Anna Lauren Hoffmann, 'Data Is the New What? PopularMetaphors & Professional Ethics in Emerging Data Culture', 2 May 2019 *Journal of Cultural Analytics.*

17 Accenture (n 2).

in their outcomes.[18] As Wagner points out, corporations' actions may simultaneously be in line with their ethics statement but in conflict with the law on a more general level, leading to a situation where firms simultaneously act both in accordance with ethical guidelines and illegally.[19] He draws on an example of this reported by Powles and Hodson[20], where Google DeepMind processed UK patients' data without a legal basis based on the claim that DeepMind was 'an ethical company developing ethical products'[21].

The second current of ethical thinking that has surfaced in relation to the use of private-sector data technology is a more macro-ethical one that asks how such technologies should be governed, how we should think of their implications across space and time, and what boundaries should be set in relation to their use. This other level of ethical inquiry incorporates a political view on data, and does not always refer to itself as ethical reflection. This work takes place mainly within academia but aims to impact the ways in which data technologies are developed and applied. Examples include the work of Floridi et al. regarding 'Onlife' and its implications for society[22], and the research conducted under the Virt-EU project[23], which includes topics such as how (digital) 'things shape values' and how accountability for data technologies' application should operate. This strand of work also takes in the notion of social justice[24] in relation to data technologies' use and governance.

These two perspectives come into conflict around the tension explored above, where structural market realities limit the space for ethical behaviour. This tension has surfaced in the form of employee resistance, including the 'Tech Won't Build It' movement, where workers at the largest technology firms registered their unwillingness to develop technology that would support human rights violations by US immigration and border enforcement[25] and link technology ethics to labour rights and to the #metoo movement, as occurred with the Google Walkout where tens of thousands of the firm's employees demonstrated over workers' rights at the firm.[26]

The tension is also manifested in the separation observable in the field between the search for guidelines (imagined as a static, durable set of principles to resolve individual-level dilemmas), and the search for more dynamic, flexible processes of reflection and policy-building which are relevant on the collective level and which can provide leverage against damaging corporate practices. Pasquale criticises this formalisation of ethical reflection, where 'firms assume that the demand for accountability must be translated in some way into computer science, statistics, or managerialist frameworks, where concerns can be assuaged by a tweak of a formula or the collection of more data'[27]. Dynamic reflection on ethics is risky for a corporate sponsor. It opens up the possibility that experts may disagree with each other, or worse still, may come to a consensus that the company is wrong. Citing the use of guidelines, however, is a weak response to public criticism and does not remedy reputational damage with immediate activity signifying the potential for change.

These micro and macro approaches tend toward different streams of thinking on ethics. The micro-ethics approach draws on deontological frameworks in terms of discussing duty toward research subjects, as framed in the US Common Rule and bioethics in general, and by doing so offers principles and duties to shape the choices of individuals working with data (rather than, for instance, setting out an explicitly utilitarian requirement that they personally balance costs and benefits). When framed in regard to data science this stream of thinking usually starts from an acknowledgement of the human right to privacy and the related responsibility to practice confidentiality when handling data. Ethics codes aimed at corporate activity, however, do not offer an account of what to do when a firm's business model brings law and ethics into conflict.[28] Nor do they address the complex political questions raised by principles of transparency, accountability or fairness, namely what their operationalisation should achieve and for whom. Instead commercial data ethics might be seen as a kind of branding activity, using discourses shaped to appeal to the corporate client, such as 'competitive advantage'[29]. This commercial ethics does not posit a process that could fundamentally change the course or focus of an organisation's dealings, but instead promises to shape existing activities in accordance with ethical principles. In line with this appeal to the corporate survival instinct, it is often framed as a longer term strategic necessity for foreseeing legal challenges and harms that might lead to customer churn through reputation damage, and a shorter term tactical one for avoiding regulatory action when things go wrong. This branded data ethics also draws strongly on utilitarianism[30] in its claim that negative consequences of data science applications can be predicted and pre-empted through compliance with standard principles. Javier Ruiz of ORG identifies this approach as a 'utilitarian aspect which is also quite problematic because it allows you to justify pretty much everything'[31].

This commercial brand of data ethics is based in the liberal individual model of the individual rights claimant and does not easily take into account notions such as group interests in privacy in response to invisible algorithmic groupings[32], or the collective origins and downstream effects of much data processed today, particularly in

18 Carole Cadwalladr, '"I Made Steve Bannon's Psychological Warfare Tool": Meet the Data War Whistleblower' *(theguardian.com, 17 March 2018)*, 2018. Available at https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump (last accessed 1 April 2020).

19 Ben Wagner, 'Ethics as an Escape from Regulation: From "ethics-Washing" to Ethics-Shopping?' in Emre Bayamlioglu and others (eds), *Being Profiled, Cogitas Ergo Sum* (Amsterdam University Press 2018).

20 Julia Powles and Hal Hodson, H., 'Google DeepMind and healthcare in an age of algorithms' (2017) *Health and Technology*, 7(4), 351–367. https://doi.org/10.1007/s12553-017-0179-1.

21 Wagner (n 19) 84.

22 Luciano Floridi, 'The Onlife Manifesto: Being Human in a Hyperconnected Era' (2015).

23 https://virteuproject.eu/ (last accessed 1 April 2020).

24 Lina Dencik, Arne Hintz and Jonathan Cable, 'Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism' (2016) 3 *Big Data & Society* 1; Linnet Taylor, What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally (June 26, 2017). Available at SSRN https://dx.doi.org/10.2139/ssrn.2918779.

25 Science for the People, 'Solidarity Letter: Tech Won't Build It!' (25 September 2018) https://scienceforthepeople.org/2018/09/25/solidarity-letter-tech-wont-build-it/ (last accessed 1 April 2020).

26 Mar Hicks, 'The Long History behind the Google Walkout' (*The Verge*, 9 November 2018) https://www.theverge.com/2018/11/9/18078664/google-walkout-history-tech-strikes-labor-organizing accessed 16 February 2020.

27 Frank Pasquale, Odd numbers: Algorithms alone can't meaningfully hold other algorithms accountable. (Real Life, 20 August 2018) https://reallifemag.com/odd-numbers/

28 Wagner (n 19).

29 Hasselbalch and Tranberg (n 8).

30 Linnet Taylor, 'The Ethics of Big Data as a Public Good: Which Public? Whose Good?' (2016) 374 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 20160126.

31 Interview with Javier Ruiz, Policy Director at Open Rights Group, 22 June 2018.

32 Linnet Taylor, Luciano Floridi and Bart van der Sloot, *Group privacy: New challenges of data technologies* (Berlin etc: Springer International Publishing, 2017).

machine learning models.[33] Following from the individual nature of its responsibilisation and the claims it can answer, it also relies, like data protection, on the idea that data can be anonymised and that it is rendered harmless by doing so. This focus on compliance and on individual responsibility has the effect of making a strong claim for voluntary self-regulation, and allowing (commercial) data science to proceed with business as usual.

This also suggests that the concern of data ethics is with data that is personally identifiable. Yet in both the fields of law and social justice concern is emerging around the notion that data not attached to a personal identity should not be subject to ethical or legal consideration. As Purtova demonstrates, many forms of data usually considered non-personal may in fact come within the bounds of data protection.[34] One salient example is the case, discussed at a 2018 data protection conference[35], of an AI application on a production line where the system assessed the average speed at which workers performed a particular task, and which then resulted in those judged below average losing their jobs. In this case, at the point where the data affected workers negatively it is judged to have become personal data, and therefore to trigger obligations under the GDPR for the firm in question.

The influence of data protection's individual- and identifiability-focused starting point on data ethics becomes problematic in relation to the main objective of avoiding harm because it permits the data handler to stop at compliance rather than demanding consideration of the public interest. Moreover it demands a clear picture of the consequences of data use, whereas those practicing data science are usually doing so remotely, without a clear idea of the context or the people implicated. A cost-benefit analysis is an accessible form of reasoning for data scientists trained in exact science disciplines, and one that they are comfortable with as a test. Drawing on experiences of teaching data ethics to economics and business students in a university context,[36] each time a group was presented with different framings for ethical reflection and asked to indicate which they used in their own work, they universally indicated consequentialism, and in a majority of cases argued for this to the exclusion of other modes of reasoning.

2.1 Deflecting and repositioning regulation and governance

Floridi, in his review of the misuse of ethical review processes, foregrounds the dual aims of distracting people from what is going wrong, and masking or not changing behaviour that should be changed.[37] In line with this, one main observable characteristic of commercially-targeted data ethics guidelines and principles is that they tend to emerge at moments where reputational damage is occurring and regulatory attempts to change or limit firms' business models are a possibility[38]. Google established, then rapidly disbanded, an AI Ethics Council in 2019 around the same time that employees had protested its work developing AI for weapons systems;[39] Dutch bank ING claimed to have established a 'data ethics council' after a series of highly publicised missteps on customer data reuse[40]. Facebook established an Ethics Working Group in 2016 after several instances where its use of data did not match up with its users' expectations[41], later to be disbanded when the Cambridge Analytica scandal forced the company to justify its actions in political fora. Ethics remained a tool for managing the company's position with regard to regulation, however: interviewed in 2018, Norberto Andrade, Facebook's Privacy and Public Policy Manager, explained that 'ethics is becoming an important platform for legal discussions'[42].

Yet establishing ethics for such discussion may also be part of serving various strategic ends for firms as ORG's Javier Ruiz outlined:

> at the moment a lot of the data ethics debate is really about how do we avoid regulation. It's about saying this is too complex, regulation cannot capture it, we cannot just tell people what to do because we don't really know the detail. Everything is moving too fast so the best thing we can do is to try to give people some more general criteria to allow them to make decisions as best as they can. And also by bringing all these ethical discussions, we can generate trust because if you put the word ethics on something, you automatically make a mental connection with trust and goodness.[43]

The philosopher Thomas Metzinger, serving as rapporteur to the European Commission-convened High Level Group on Artificial Intelligence (2019), noted that industry members of the group had come to the process with a very different motivation from the academic members. Such debates, he said, represent an important tactical weapon for industry:

> You organise and cultivate ethical debates because you want to delay, postpone, avoid or deter people from policymaking or regulation. That is actually the major goal of the industry, to do everything to avoid concrete, enforceable law. For instance, Facebook and Amazon, they like it if we have long ethical debates in Europe, because the longer we have these debates, the longer they have before we can enforce law.[44]

If regulation is something to be avoided in high-income regions such as the EU and US, it is also something to be negotiated and repositioned in regions where the data economy is less regulated. In commercial and research activities conducted in relation to low- and middle-income countries, firms may actively seek a form of trans-

33 Metcalf and Crawford (n 6).
34 Nadezhda Purtova, 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) *Law, Innovation and Technology*, 10:1, 40-81, https://doi.org/10.1080/17579961.2018.1452176.
35 Computers, Privacy and Data Protection (CPDP) conference, Brussels, January 24-26 2018, panel with Peter Hustinx, European Data Protection Supervisor.
36 Observations based on ten academic courses given in the Netherlands in association with Tilburg University, ranging from bachelors' to professional executive level, between 2016 and 2019.
37 Luciano Floridi, 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical' (2019) *Philosophy & Technology*, 32(2), 185–193, 188 https://doi.org/10.1007/s13347-019-00354-x
38 See also Metcalf and Crawford (n 6).
39 Google's AI ethics council was disbanded due to controversy over the appointment of a member from the politically conservative Heritage Foundation, and resulting employee pushback over this appointment (see, e.g., *BBC News*, 'Google's Ethics Board Shut Down' May 4, 2019. Available at https://www.bbc.com/news/technology-47825833 (last accessed 12 April 2020).
40 ING.com, 'Data Ethics' https://www.ing.com/Sustainability/Our-Stance/Data-ethics.htm (accessed 24 June 2019).
41 Anna Lauren Hoffmann, 'Facebook Has a New Process for Discussing Ethics. But Is It Ethical?' *The Guardian* (17 June 2016) https://www.theguardian.com/technology/2016/jun/17/facebook-ethics-but-is-it-ethical (last accessed 1 April 2020).
42 Interviewed 21 May 2018
43 Interview with Javier Ruiz, Policy Director at Open Rights Group, 22 June 2018.
44 https://www.europarl.europa.eu/streaming/?event=20190319-1500-SPECIAL-SEMINAR1&start=2019-03-19T15:44:53Z&end=2019-03-19T15:56:00Z&language=en

parency to authorities through processes of data ethics in an effort to demonstrate that they are not behaving irresponsibly in countries where they hold a licence from the government to do business. When commercial data is extracted from populations where data is un- or under-regulated, as occurs in the fields of international development and humanitarian work[45], reputational risk and contractual repercussions become an issue for multinational firms. These firms can be observed to be using ethics as a basis for their operations where, for example, data protection law or a constitutional right to privacy are missing in a particular national context. In a good scenario, industry incorporates local representatives in its boundary-setting process, as Orange Telecom did when it established an ethical advisory board for its 'Data for Development' challenge in Senegal.[46] In a less good scenario, institutions establish their own boundaries for these environments. This can be problematic when those institutions also enjoy legal immunity with relation to their use of data, such as UN bodies.

The relationship between data protection and data ethics is a tangled one precisely because one deals with what can be pinned down and demanded of those handling data, and the other with what should be. In practice, what Floridi terms 'ethics shopping'[47] is common, with data protection and ethics principles being cherry-picked in the search to retrofit guidelines to behaviour. The risk of this is that firms may frame compliance with data protection law as a complete ethical approach to data and thus miss other important subjects of ethical reflection. Examples would be a concern with only personal data, or the idea that once consent has been acquired from the subject no further problems are possible. It also does not help where legal systems diverge: as Zara Rahman of the Engine Room points out, 'The things that are legal in certain countries are outrageously not ethical'[48].

2.2 De-politicising data's politics

The ethics initiatives observable at big tech firms can also be seen as strategic public relations efforts which allow firms to make public statements about their values without framing it as advertising. For example, statements about ethics are a safe space in which to discuss the fact that technology is not neutral and firms' applications have social and political impacts. Andrade, for example, describes the Facebook review process explicitly in terms of the firm's aim to create social and behavioural change: the firm's aim with ethics, he says, is 'to create ethically responsible outcomes for people on our platform and for society. To empower them to make ethically sound decisions on our mission to bring the world together. It's not a neutral statement, or mission'[49].

Where these political implications and effects have a destabilising internal impact, an ethical review or discussion process can provide scaffolding for resolving disputes and defusing tensions, thus preserving the internal status quo that allows firms to do business. Palantir, the US data analytics giant, is one example of this. The company has come under public criticism for, among other things, accepting core funding from the CIA[50], supporting the Trump regime's effort to separate children of undocumented immigrants from theirfamilies[51], and avoiding public scrutiny when providing potentially discriminatory urban policing systems.[52] Palantir started to publicise its ethical credentials in 2012 when it established a 'Civil Liberties Board' staffed by leading privacy scholars from the US and EU[53]. The company also established a 'privacy and civil liberties engineering team' which offers ethical guidance to employees. Courtney Bowman, co-director of the team, explains that the purpose is to help employees reconcile progressive political views with the work Palantir does:

> Most of the institutions we draw from in terms of CS [computer science] hires are bastions of more left-leaning political views – Stanford, Berkeley, Harvard, MIT, CalTech. The majority of employees come from a general leaning of real interest and concern about the fate of Western liberal democracies and the importance of not undermining and eroding those institutions, so I don't think they would feel comfortable working at a company with the reputation... [...] there's a disconnect between the way Palantir is represented in the media and my experience of working on these issues. [So the ethics process means] that we can get these candidates who otherwise would be unwilling to engage with us. They can see it's not us privacy-washing or paying lip service, there's a real credible effort on the ground.[54]

In setting up its ethics process, Palantir was ahead of the game. Over the 2010's almost all the technology giants experienced employee pushback on a level that threatened their public image. Microsoft employees protested their firm's work with the US border authority at the start of the Trump administration's family separation initiative in 2018[55]; the same year Google experienced a rebellion over providing AI to a Pentagon weapons program, and the year before Silicon Valley employees had protested the Trump administration's banning of travel from certain Muslim-majority countries.[56] Employee unrest also occurred at Facebook when Joel Kaplan, Facebook's vice president for global public policy, sat behind Brett Kavanaugh at the congressional hearing where he was interrogated over accusations of sexual assault.[57] Like Palantir, by 2019 Google and Facebook had both set up expert-led ethics advisory processes, while Microsoft so far has not.

45 Linnet Taylor and Dennis Broeders, 'In the Name of Development: Power, Profit and the Datafication of the Global South' (2015) 64 *Geoforum* 229.
46 Taylor (n 30).
47 'the malpractice of choosing, adapting, or revising ... ethical principles, guidelines, codes, frameworks, or other similar standards (especially but not only in the ethics of AI), from a variety of available offers, in order to retrofit some pre-existing behaviours', Floridi (n 38) 186.
48 Interview with Zara Rahman, Deputy Director of Engine Room, 14 June 2018.
49 Interview with Norberto Andrade, Privacy and Public Policy Manager for Facebook, 21 May 2018
50 Andy Greenberg, 'How A "Deviant" Philosopher Built Palantir, A CIA-Funded Data-Mining Juggernaut' (*Forbes*, 2013) https://www.forbes.com/sites/andygreenberg/2013/08/14/agent-of-intelligence-how-a-deviant-philosopher-built-palantir-a-cia-funded-data-mining-juggernaut/ (last accessed 1 April 2020).
51 Mijente.net, Who's Behind ICE? The tech and data companies fueling deportations (mijente.net, 23 October 2018) https://mijente.net/2018/10/whos-behind-ice-the-tech-companies-fueling-deportations/ (last accessed 1 April 2020).
52 Ali Winston, 'Palantir Has Secretly Been Using New Orleans to Test Its Predictive Policing Technology' (*The Verge*, 27 February 2018) https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd (last accessed 1 April 2020).
53 Palantir.com, 'Announcing the Palantir Council on Privacy and Civil Liberties' (*Palantir*, 2012) https://palantir.com/2012/11/announcing-the-palantir-council-on-privacy-and-civil-liberties (last accessed 1 April 2020).
54 Courtney Bowman, director, privacy and civil liberties engineering team, Palantir. Interviewed 11 October 2018
55 Sheera Frenkel, 'Microsoft Employees Protest Work With ICE, as Tech Industry Mobilizes Over Immigration , *The New York Times* (19 June 2018) https://www.nytimes.com/2018/06/19/technology/tech-companies-immigration-border.html (accessed 21 June 2019).
56 Kenneth P. Vogel, 'New America, a Google-Funded Think Tank, Faces Backlash for Firing a Google Critic', *New York Times (*1 September 2017) https://www.nytimes.com/2017/09/01/us/politics/anne-marie-slaughter-new-america-google.html (last accessed 1 April 2020).
57 New York Times, 'Rifts Break Open at Facebook Over Kavanaugh Hearing', 4 October 2018, https://www.nytimes.com/2018/10/04/technology/facebook-kavanaugh-nomination-kaplan.html (accessed 21 June 2019).

Instead the company claims to focus on business ethics, corporate social responsibility and 'integrity and governance'[58].

Aside from internal ethics processes, firms also use external engagement on ethics-related issues, apparently to support their ethical branding and neutralise protest. Technology giants sponsor academic research, fund think tanks and sponsor both conferences and specific sessions in the domain of law, human rights and privacy studies. In terms of conference support, Palantir, Google and Facebook are commonly found on the list of sponsors of major law and privacy conferences including the Amsterdam Privacy Conference and the Privacy Law Scholars Conference. This process establishes tech companies as a highly visible presence where regulation or the politics of technology are being discussed. For instance, Facebook announced in early 2019 that it would sponsor an AI ethics centre within the Technical University of Munich, run by Professor of Business Ethics Christoph Luetge[59], previously a member of Facebook's 2016-17 ethics review group. Microsoft Research in the US has served as a research hub for many scholars doing critical work on privacy and rights; Google extensively sponsors institutes and academic research projects in the US and EU, as well as independent research projects. The firm received public criticism when it de-funded a research group at the New America Foundation after its lead researcher praised the EU's fining of Google for antitrust violations.[60] Internal criticism led to the dropping of Palantir as a long-time sponsor of the Privacy Law Scholars Conference after the program committee raised objections to its sponsorship[61].

This external engagement has been called 'ethics-washing'[62] where it deflects from actual violations of rights or norms in their everyday activities. It may also, however, represent pre-emptive action in response to growing pressure on firms to engage with public criticism of their work. MariaRosaria Taddeo, a philosopher and ethicist of technology at the Oxford Internet Institute, makes this connection:

> We may see ethics more outside academia because we are starting to see the consequences of company behaviour. Even if it's not for goodwill they will have to deal with ethics. It's easier to stay with compliance but it will be hard, and maybe not safe for committees not to go beyond compliance and seek for ethics.[63]

The aim of this mix of strategies seems to be instrumental: used strategically to preserve the status quo, an ethics advisory process can act to de-politicise highly sensitive concerns around rights and public values by changing the discourse (for example shifting attention from the legitimacy of a particular intervention to privacy compliance), and thus allow contracts to go forward while paying attention to employees' and the public's concerns. Asked if the scholars on Palantir's Civil Liberties Board, or its own internal privacy and civil liberties team, could veto any of the company's activities, Bowman answers, 'I wouldn't characterise it as an explicit veto.' He describes the latter team as being established 'so you can achieve furthering the mission of sovereign nations or organisations in a way that is privacy protective and sensitive to social concerns'[64].

MIT's 'Moral Machine' project illustrates how the rhetoric of ethics can serve to shape the future along particular paths. The researchers asked people around the world to respond to the 'trolley problem' – a classic thought experiment where the subject is asked to decide how to direct an out-of-control vehicle heading for a group of people, but which could be diverted by a lever to a track where it would hit just one person instead.[65] The problem offers different variants (for instance, would you divert the trolley if the one person on the other track was a child? Would you feel different about hitting old people? Overweight people?). The problem is designed to highlight differences in ethical frameworks and ways of thinking. Instead of a trolley, however, MIT frames the problem around a self-driving car. This choice has several potential effects: the existence of self-driving cars becomes normalised as an everyday problem; public anxiety is allayed by the sense that ethical issues are being addressed and thus policymakers' options for allowing such cars into the road are widened; people can be reassured that the governance of this new technology is taking their opinion into account,[66] and they may feel some resulting ownership of the policy decisions that are made to allow such cars into public space. MIT's choice of focus, as an institution working to develop new technologies, is strategic. Created in 2016, at a time when self-driving cars were starting to appear (and malfunction) on roads in the US, the Moral Machine project, though framed as academic research, can also be seen as a pre-emptive political and regulatory play: a statement that automated vehicles are an inevitability.

2.3 Data ethics as a route to technical standardisation

One positive view on data ethics is that of its emergent concerns and responses as the basis for guidance for the field. Silkie Carlo, director of Big Brother Watch, a UK organisation that advocates for a human rights approach to developing technology, describes data ethics as 'a guiding way of thinking about how the law should be shaped', but also 'of growing importance when we come to design new frameworks. For example, if we need to develop, which we probably do, a framework for dealing with artificial intelligence, then clearly some ethical background is going to be absolutely vital'[67].

On the technical level, we might similarly see data ethics as a form of standard-setting, where the local development of principles and guidelines can create opportunities for discussion and training that then may become institutionally embedded into practice, and reflected back to the field through inter-firm collaborations.[68] The

58 Microsoft, 'Microsoft Code of Conduct | Ethics & Compliance' https://www.microsoft.com/en-us/legal/compliance/default.aspx (accessed 21 June 2019).
59 Deutsche Welle (n 3).
60 Vogel (n 56).
61 Lizette Chapman, 'Palantir Dropped by Berkeley Privacy Conference After Complaints' Bloomberg (5 June 2019) https://www.bloomberg.com/news/articles/2019-06-05/palantir-dropped-by-berkeley-privacy-conference-after-complaints (last accessed 1 April 2020).
62 Elettra Bietti, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from within Moral Philosophy', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020); Wagner (n 19).
63 Mariarosaria Taddeo, Research Fellow and Deputy Director, Digital Ethics Lab, Oxford Internet Institute, interviewed 20 April 2018
64 Courtney Bowman, director, privacy and civil liberties engineering team, Palantir. Interviewed 11 October 2018
65 MIT Media Lab, 'Moral Machine' (*Moral Machine*, 2016) https://moralmachine.mit.edu (accessed 19 June 2019).
66 The author of the project writes that one main justification for the work is 'to uncover [people's] biases and know when to anticipate them in order to plan regulations that achieve public acceptance', and later adds that 'a platform to promote public discussion about the ethics of machines [...] to provide one input to policy makers and regulators, highlighting the factors that may raise public concern.' Edmond Awad, 'Moral Machine: Perception of Moral Judgment Made by Machines' (Massachusetts Institute of Technology 2017, MA Thesis) 63.
67 Interview with Silkie Carlo, Director of Big Brother Watch, 26 June 2018.
68 AF Winfield and M Jirotka, 'Ethical Governance Is Essential to Building Trust in Robotics and Artificial Intelligence Systems' (2018) 376 *Phil. Trans.*

potential disadvantage of this approach, however, is that to succeed companies must determine for themselves what is good or right without routing through public discussion or governance processes. This relies heavily on their being able to engage in ethical reflection without being influenced by profit motives, shareholder demands or pressure of competition, and furthermore on a race to the top where ethical principles and practice spread between companies. Instead the current state of play in the technology field involves separate ethical ecosystems, each formed in the image of a company's own business model and each with different standards for what is acceptable.

As well as using ethical thinking to shape new requirements and standards, Norberto Andrade of Facebook describes Facebook's 2017-18 review process as also trying to standardise ethical thinking across the company's different product development teams:

> We were having discussions with product managers and engineers that were ethical. They weren't named that way but were debating ethical questions. I wanted to standardise the ethics discussion around all the products we were developing, and I wanted to do an ethics discussion without intimidating people.[69]

A high-profile example of this is the various governmental and private-sector discussions around ethics for artificial intelligence.[70] This standard-setting process can also involve confirmation and scaffolding of the company's business model. Javier Ruiz from ORG explains how ethics can help make a business model more acceptable:

> [P]art of the problem is they say they are going to carry on business as usual, [...] you're having to use ethics [...] as almost a harmonisation exercise at the end. It's like we're going to do this and we're going to build a nuclear missile system and then at the end, you're going to bring ethics to see how do you minimise so we're [...] going to hit as far away from a school as possible.[71]

In this vein, Palantir's ethics statement[72] emphasises privacy by design, keeping humans in decision-making loops where AI is used, making systems accessible to oversight and not engaging in solutionism (using technology to 'solve' problems where it is inappropriate). These are all credible principles rooted in various approaches to technology and data ethics. None of these, however, addresses the higher-level problem of whether it is ethically permissible to engage with a maleficent system or process, which is the main criticism which has been levelled at Palantir over time.[73] Palantir's website, for example, emphasises ensuring the effective implementation of 'rigorous privacy policies' in the provision of analytic systems for policing. The privacy problem, however, has not been central to debates on the ethics of data-driven policing. Social justice issues including discrimination, racial and economic inequality and issues of using probabilistic analysis in relation to decisions that affect people's freedom and civil rights have been more prominent.[74] Hannah Couchman, Advocacy and Policy Officer at civil rights organisation Liberty, cited this problem of different levels of ethical concern: 'Liberty as an organisation is hesitant, in some senses, to talk about what we need to do to make [a particular technological solution] safe when essentially, we fundamentally object to what's going on'[75]

A micro-ethics of data often points away from the political questions. An individual worker or a group within a technology company may be following the company's ethical code or guidelines, designing for privacy, practicing data minimisation, and generally working on their own level for the betterment of humanity. But if the company as a whole is engaged in providing software for autonomous weapons systems, supporting discriminatory law enforcement, or helping to jail children and separate them from their parents, it is not hard to see how a focus on micro-level privacy and ethics, however necessary, could pull focus from higher-level ethical problems.

Javier Ruiz (ORG) surfaces this tension between data ethics as an instrument for integrating new technological applications into society and data ethics as philosophical inquiry - part of a larger ethics of building a good society. He asks:

> How do you build common values in diverse societies and how do you do it in a way that doesn't mean that you become reactionary, or automatically conservative, where you freeze those values and they can't evolve? The premise of data ethics, it's almost like it sits on top of huge ethical challenges [...] you cannot just tackle data ethics in isolation without having a broader discussion.[76]

An optimistic vision of the theory of change involved in corporate ethics processes might identify Google DeepMind Health's ethics committee as an example of one which had a greater degree of freedom and scope than the classic problem-oriented or guidelines-based processes. Julian Huppert, who was appointed chair of the committee when it was formed, explained that as far as the committee could tell, their brief was 'largely to hold [Mustafa Suleyman, CEO of DeepMind Health] and the organisation to account and to push them in the right directions.' The committee members were under no confidentiality agreements, and could hire external researchers to do investigations or analyses. The committee did publicly express concern about the Google subsidiary's ability to keep Alphabet, Google's parent company, from using health data gathered by DeepMind Health for profit when DeepMind Health was absorbed into Google in 2018[77], moving the analysis of NHS patient data one step closer to the for-profit functions of the company. In 2018, possibly a victim of its own success, the ethics committee was disbanded.

The overall business model of informational capitalism - data extraction and marketing - is itself an ethical minefield and often seen as undemocratic and exploitative.[78] Ethics potentially allows for higher-level questions such as whether people should be treated as a means to an end. These considerations can inform questions such as

R. Soc. A,.

69 Interview with Norberto Andrade, Privacy and Public Policy Manager for Facebook, 21 May 2018

70 Luciano Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, 'AI4People---An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 *Minds and Machines* 689.

71 Interviewed 22 June 2018.

72 Palantir, 'Privacy and Civil Liberties Engineering' (2019) https://www.palantir.com/pcl/ (accessed 5 April 2019).

73 Mijente.net (n 51).

74 J. Angwin, J. Larson, S. Mattu, & L. Kirchner, L., Machine Bias. (ProPublica, 23 May 2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; W. Dieterich, C. Mendoza, & T. Brennan. COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity, (8 July 2016) Northpointe inc.

75 Interview with Hannah Couchman, Advocacy and Policy Officer at Liberty, 1 June 2018.

76 Interviewed 22 June 2018.

77 Sam Shead, 'DeepMind Is Handing DeepMind Health Over To Google' Forbes (13 November 2018) https://www.forbes.com/sites/samshead/2018/11/13/deepmind-is-handing-over-deepmind-health-to-google/#c03e21e2d551 (last accessed 1 April 2020).

78 Shoshana Zuboff, *The Age of Surveillance Capitalism The Fight for a Human Future at the New Frontier of Power* (Profile 2019).

whether it matters if informed consent is only based on partial disclosure, or whether we should define behavioural data as a fundamental component of people's identity or an asset to be traded (and if both, as we do currently, what this means for rules and boundaries). Industry ethics codes and review processes, based on the empirical research conducted for this paper, are not designed to address these questions, nor do they take account of employee unease with exploitative labour practices such as Facebook's use of low-paid workers in lower-income countries to vet content for violence and sexual abuse[79]. The latter is evidenced by the fact that during Facebook's 2016-17 ethics review process the company's labour practices were marked as out of scope.

3. Towards an ethical data ethics

There are good reasons why the notion of ethics in relation to digital data has been subject to corporate capture. We need an ethics of the digital because commercially produced data is becoming the bedrock of many economies around the world. AI, based on huge amounts of data, is forecast to generate 13 trillion dollars in economic activity by 2030, primarily for OECD countries.[80] Data technologies also increasingly play an important role in how people form and exercise their identities, on both the group and the individual levels. Data ethics is a thriving and well-funded field of inquiry within academia and beyond that seeks to inform how data should be used in society for the public good. However, it is exactly this demand that offers opportunities for capture. As Floridi points out just claiming to be engaging in data ethics in no way guarantees that any ethical reflection is happening.[81]

If we address data ethics as a discourse, separately from its existence as a subject of study and a process of reflection, we can see that discourse doing particular work in society. First, an ethical process that focuses on reducing harm from particular technologies, for example autonomous vehicles, also has its own politics. Centring autonomous vehicles sidelines the politics of the car industry, and by extension urban development and industrial policy. While we are deciding how many people autonomous cars can ethically kill, we are not looking at the larger ethical question about whether we should be aiming for a world of cars at all. Similarly if a social media company mandates its workers follow ethical guidelines when they build applications or moderate content, this may serve as a way of distracting attention from the larger problem of an extractive business model.

If an ethics process is used strategically to justify an unjust business model, or if it takes place without consideration of the underlying assumptions about society and justice, then the process is cosmetic. Metzinger's criticism of the EU's High-Level Group on AI takes this view: if the possibility of delineating meaningful boundaries for technology – something the advocates of corporate data ethics interviewed for this paper claim is its function – is off the table, then so is an important part of the task of ethics. At this point, as Metzinger demonstrated by going public with his criticism, politics becomes instrumental in establishing a meaningful space for ethical reflection. This dynamic means that without acquiring traction through an accompanying consideration of politics, much of 'data ethics' may be reduced to selecting new wallpaper for a building that is on fire.

If we wish to promote an approach to the commercial use of data technologies that takes social justice as well as legal compliance into account, then integrating data ethics into business models becomes the central problem that anyone working on this problem academically or commercially must confront. This is also the task of law, but many of the problems highlighted in this paper (including the support of unjust government policy and the development of technologies that have a high likelihood of resulting in rights violations) can be characterised, as Wagner argues, as not illegal but nevertheless unjust.[82] This suggests that ethics needs to concern itself explicitly with not only what constitutes the public good, but the dynamics and power relations in place that shape the processes of such assertions. A data ethics process separate from the decision-making core of a company signals that ethics is an add-on, something that must not come into conflict with the bottom line. Defining ethical reflection as a separate process to the everyday business of the company also runs the risk of demanding too little from management and employees: perhaps the question we should be asking is not how companies should integrate ethics processes into their work, but why those ethics processes need to be integrated in the first place. Adding in a discourse of data ethics to the corporate mission may also, ironically, absolve companies from interrogation about their business models. If another sector with implications for public safety and wellbeing such as airlines or civil engineering, began setting up public-facing ethics review boards, we might take this as a cue to ask whether the planes we fly on and the bridges we walk over were safe.

The activists interviewed for this research suggest that there are several possible ways in which data ethics might facilitate change in corporate practice. Regarding public understanding and behaviour, one is a 'moment of truth'-type strategy[83], where discussions about ethics help to clarify that a problem exists, and the public starts to reject technologies that have been shown to have abusive business models or effects. Routes to change might include smaller technology developers (if ethics really does become 'a new competitive advantage'), a rise in ethical consumption amongst the general public, and finally, the creation of governmental initiatives on data ethics which shape law and regulation. This mechanism is clearly the one envisaged by the German national data ethics commission, whose mission is to create 'suggestions for possible legislation'[84]. Christiane Woopen, the commission's chair, asked:

> What are the alternatives to considering ethics as a basis for regulation? Ethics is often captured but our commission looks at ethics as a basis for regulation, for setting rules. Can you have a different yardstick in a democratic society than ethical and fundamental values?[85]

This is not true everywhere. Despite recommendations from its parlia-

79 Joshua Brustein, 'Facebook Grappling With Employee Anger Over Moderator Conditions' Bloomberg (25 February 2019) https://www.bloomberg.com/news/articles/2019-02-25/facebook-grappling-with-employee-anger-over-moderator-conditions (last accessed 1 April 2020).

80 Bhaskar Chakravorti, Ajay Bhalla and Ravi Shankar Chaturvedi, 'Which Countries Are Leading the Data Economy?' (2019) *Harvard Business Review* https://hbr.org/2019/01/which-countries-are-leading-the-data-economy (last accessed 1 April 2020).

81 Floridi (n 37).

82 Wagner (n 19).

83 Ian (Gus) Hosein, 'A Research Note on Capturing Technology: Toward Moments of Interest' (2003) 110 *IFIP Advances in Information and Communication Technology* 133; Esther Görnemann, 'Digital Privacy Moments of Truth: A Concept of Moral Indignation over Personal Data Usage' (unpublished 2018).

84 German Data Ethics Commission, 'Data Ethics Commission' (*Federal Ministry of the Interior, Building and Community*, 2018) http://www.bmi.bund.de/EN/topics/it-internet-policy/data-ethics-commission/data-ethics-commission.html;jsessionid=29BF9E2D3283A4EDAD9BB756BE0089F5.2_cid295?nn=9385466 (accessed 24 June 2019).

85 Christiane Woopen, closing statement. German Data Ethics Commission open meeting, 9 May 2019, German Ministry of the Interior.

mentary committee[86], the UK's 'Council of Data Ethics' was not created as a regulatory body but as an advisory one instead (named the Centre for Data Ethics and Innovation), bearing out the committee's evaluation that the government was taking a pragmatic pro-business perspective at the expense of protecting individuals from negative impacts. Pasquale[87] sees this privileging of the business perspective as risky because it separates academia from policy. Under these conditions, he says, 'it is easy for academics to give up on trying to influence government policy and seek changes directly from corporate leaders.' This then creates a risk of 'translating one's work into a way of advancing overall corporate goals [...] Such corporate goals may help burnish scholars' reputations at first, but eventually they need to boost the bottom line.'

4. Conclusion

We have made the case that, as well as a branding exercise, commercial processes of data ethics are one forum where the responsibilities of firms toward the public – and therefore what firms may be held accountable for – are negotiated. If this is true, and if we wish to develop a response to corporate (mis)uses of data ethics, we might begin by reframing the question to include other relevant perspectives on what is ethical. These might include a rights-based perspective that focuses on profiles and inferences as well as personal data[88]; approaches to averting harm that go beyond personal identification[89], and an ethics of algorithms[90]. For instance, moving from an individual to a collective anchoring for ethics, as suggested by work on group privacy[91], would suggest direct engagement with impacted communities and social groups, and creating a diverse set of fora where different opinions about what data and uses matter can be heard. This engagement situates data ethics in a social and economic justice framework, where datafication is not a revolution that is drastically changing the structural power and political economy of modern society, but an extension of conditions that have resulted in grievances and injustices towards historically marginalised and politically constructed targets.[92] Similarly, the social stratifications of different (data) classes are an expression of concentration of power and related to a wider trend of privatization and deregulation, along with a shift in decision-making away from the public realm. This perspective is in line with that of Gangadharan and Niklas, who advocate '"see[ing] through" technology, acknowledging its connection to larger systems of institutionalized oppression'[93].

Commercial data ethics processes have multiple functions. They can serve as a political strategy to avoid governmental regulation in favour of self-regulation and to deflect attention from unjust business models, but they are also used pragmatically to manage internal and external expectations. As such, they also serve a purpose in relation to governance, as a claim by corporations about their legitimacy as custodians of the public's data. All of these functions are of importance to technology firms, but none of these bear a clear relation to genuine ethical reflection, which has the essential characteristic of taking place before action is taken, rather than during or afterwards, and in a context where there is some freedom to choose one's actions. Where the path is already set by the company's business model, this freedom is missing: the underlying purpose of data ethics becomes to justify the path and mitigate, rather than avoid, harm, while cultivating trust amongst those affected by the technology in question. If we can better interrogate companies' ethical claims, we may be able to change the demands we make of those companies. Rather than reducing net harm, we could frame harm as unacceptable. Rather than weighing how many people automated cars may kill in comparison to conventional ones, we might engage in a different debate about the kind of world we wish to live in, and the kinds of technology that would help build that world. Finally, rather than aiming to evoke public trust in technology-sector business models as they currently exist, we might move instead towards enforcing greater trustworthiness through regulation and enforcement, and shaping business models in line with the public interest. Moving from a bounded and instrumental data ethics to a more expansive ethics of the digital that takes in the broader social context and aims for justice seems a necessary first step.

86 Commons Science and Technology Committee (n 7) 36.
87 Pasquale (n 27).
88 Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (LawArXiv 2018) preprint https://osf.io/mu2kf accessed 18 June 2019.
89 Purtova (n 34); Taylor, Floridi and van der Sloot (n 32).
90 Louise Amoore, Doubt and the algorithm: On the partial accounts of machine learning. (2019) *Theory, Culture & Society*, 36(6), 147-169.
91 Taylor, Floridi and van der Sloot (n 32).
92 Lina Dencik,Fieke Jansen and Philippa Metcalfe, (2018) 'A conceptual framework for approaching social justice in an age of datafication. datajusticeproject.net. Available at: https://datajusticeproject.net/2018/08/30/a-conceptual-framework-for-approaching-social-justice-in-an-age-of-datafication/
93 Seeta Peña Gangadharan and Jedrzej Niklas, Decentering technology in discourse on discrimination. Information, (2019) *Communication and Society*, 22 (7). 882-899, 883 https://doi.org/10.1080/1369118X.2019.1593484

02

Technology and Regulation

The Impacts of AdTech on Privacy Rights and the Rule of Law

Róisín Áine Costello*

AdTech, rule of law, autonomy, privacy, data protection

costelr@tcd.ie

This article argues that the AdTech market has undermined the fundamental right to privacy in the European Union and that current legislative and fundamental rights protections in the EU have been unsuccessful in restraining these privacy harms. The article further argues that these privacy consequences have imported additional reductions in individual autonomy and have the capacity to harm the Rule of Law.

"Although we feel unknown ignored
As unrecorded blanks
Take heart! Our vital selves are stored
In giant data banks" [1]

1. Introduction

Sarah Igo has speculated that the collision, or collusion, between the disclosure of personal data and the technological capacity to capture, analyse, and harness this data will be the defining feature of the twenty first century privacy landscape. While this tension between what can be known and what should be concealed is an enduring one, individuals' ability to exercise control over the boundaries of their private experience has, in the last decade, receded rather than being augmented by technological advances.[2]

This article argues that the online AdTech market, as currently constituted, has been central to this recession, and has undermined the fundamental right to privacy as it is protected in the European Union.[3] In particular, the article establishes that online markets for personal data are specifically orientated to enable large scale collection of personal data in circumstances where individuals have a limited understanding of the ways in which that information will be used, and offers no functional choice to consumers in seeking to access goods or services which do not operate such data collection practices.

1 Felicia Lamport, 'Deprivacy' *Look Magazine* (1970).
2 Sarah E Igo, *The Known Citizen: A History of Privacy in Modern America* (Harvard University Press 2018), 353.
3 Regulation (EU) 2016/679 (henceforth GDPR).

* Trinity College Dublin, School of Law.
Received 10 Sept 2019, Accepted 23 Mar 2020, Published: 14 Apr 2020

The negative privacy impacts which flow from the large-scale collection of personal data in the AdTech market are also harmful to individual autonomy - and cumulatively harmful to the Rule of Law through the diminution of individual liberty and the associated participatory capacity of individuals to engage in the democratic process. In this respect the article argues that the right to privacy is an essential component of the substantive or 'thick' conception of the Rule of Law endorsed by the Union in as much as it acts as an effective restraint on State overreach and secures a constitutionally mandated zone of individual autonomy.

The article argues that the legislative measures taken by the European Union to combat the development of the AdTech market, while motivated by the ostensible aim of securing fundamental rights, have in fact created a hierarchy in which data protection as a market-orientated right has been elevated above the socially oriented right of privacy.

As part of this development, the contractual practices which enable the AdTech landscape have proliferated largely unopposed on the understanding, only recently challenged, that they satisfy the threshold notice and information requirements required by data protection. Meanwhile, there has been a marked failure to engage in a substantive manner with the normative harms to individual privacy which may subsist alongside the satisfaction of a market orientated vision of data protection.

The article begins, in section two, with an explanation of the operation of the AdTech market and its impacts on individuals' lives before moving in section three to examine the legal landscape in which AdTech operates. Section four then examines how AdTech fits within the legal framework based on Article 8 CFR before moving, in section five, to examine how the right to privacy is impacted by the current legal and practical schema. Finally, in section six, the article expands its examination to consider how AdTech implicates negative harms

not only for privacy but also for autonomy and the rule of law.

2. What is AdTech and How Does It Impact Our Lives

The capacity, and desire, to track consumers is not new. Laurence Fontaine in a study of the notebooks of pedlars working in Europe during the fifteenth through eighteenth centuries documented the extensive, personalised notes they kept not only on their customers but on the relatives of those customers (who would expect similar deals) and the demeanour and the standing of those individuals in their communities.[4] Contemporaneously, sellers have engaged in similar attempts to measure and categorise customers and order patterns – first with simple mechanisms like turnstiles[5] and later through more sophisticated methods such as barcoding.[6]

In this context, criticism of AdTech has been dismissed on the basis that AdTech is merely the most recent evolution in a long-standing market practice of consumer surveillance, whose negative impacts are proportionate to the market efficiencies and thus individual benefits they afford.[7] Yet this is not necessarily the case[8] and a historic overview of consumer surveillance indicates that even in the context of less sophisticated, contextual, [9] consumer surveillance mechanisms, concerns abounded about the individual privacy impacts of such activity.[10]

As advertising markets moved online, such concern diminished, driven not by a reduced concern but by a market design which effectively shielded the surveillance mechanisms of the digital market from consumer scrutiny. Indeed, digital advertising networks like DoubleClick (now a subsidiary of Google) recognised the potential of the internet early on and began developing mechanisms for aggregating large and detailed consumer data sets to assess and map consumer behaviour.

The emergence of this AdTech landscape was enabled, to a significant extent, by the development of the computer cookie in 1993[11] and the subsequent move from contextual and towards behavioural advertising in the AdTech market a move which shifted activity towards the collection and aggregation of consumer data on a large scale and its deployment in a targeted, predictive manner to influence consumer behaviour and attitudes.[12]

2.1 Cookies

Cookies are small text files which are placed on a consumer's hard drive by websites which the user visits and which are accessible only to the consumer and the company or actor who placed them. [13] Cookies allow those placing them to track consumer activity on the website to which the cookie relates (through the use of first party cookies) but can also allow those placing them to track consumer behaviour across the web (through the use of analytics cookies). Crucially, cookies do not operate in a vacuum but can be linked to personally identifiable information such as a name or e-mail address provided to access a platform or service thus enabling the actor who placed the cookie to store that consumer's information so that even where a consumer deletes a cookie if they subsequently visit the site again their previous information can be re-associated with them.[14]

While this alone seems harmful to privacy, in practice analytics services and the analytics cookies on which such services rely are predominantly offered by Google and Facebook with the result that such cookies effectively operate as third-party cookies. Third party cookies are placed on consumer devices, as the name would suggest, by third parties who contract with numerous websites to learn what consumers do on sites across the web.[15]

By offering such analytics services these actors can negotiate further cookie placement agreements with hundreds or thousands of companies thus generating a substantive profile of online activity, personal characteristics and behaviours of individual consumers in an attempt to map their preferences and subsequently to target advertising to influence their preferences or choices.[16]

Currently Google and Facebook take some 65% and 90% of total digital advertising spends respectively and 20% of all advertising spends globally.[17] On Google's part this has been enabled in part by the company's acquisition of DoubleClick (now part of the Google Marketing Platform) whose cookies are found on an estimated 87% of websites.[18] Google's own databases - independent of DoubleClick prior to its absorption into the Platform include information about consumer behaviour across Google's services including the location, time and date a device is turned on, an individual's search history[19] and, controversially, the contents of communications sent via Gmail.[20]

4 Laurence Fontaine, *History of Pedlars in Europe* (Duke University Press 1996), 8 et seq.
5 Joseph Turow, *The Aisles have Eyes: How Retailers Track your Shopping, Strip your Privacy and Define your Power* (Yale University Press 2017), 114.
6 Turow (n 5) 80-81.
7 See, Reuben Binns, Zhao Jun, Max Van Kleek and Nigel Shadbolt, 'Measuring third party tracker power across web and mobile' (2010) 9 *ACM Computer Entertainment* 39; Paul Bernal, *Internet Privacy Rights* (Cambridge University Press 2014); Lilian Edwards and Geraint Howells, 'Anonymity, Consumers and the Internet: Where Everyone Knows You're a Dog' in JEJ Prins and MJM van Dellen C Nicoll (eds), *Digital Anonymity and the Law* (Asser Press 2003).
8 Leigh Gallagher, 'Ad tech has a problem. Fixing it isn't easy' (*Fortune*, 14 July 2015) https://fortune.com/2015/07/14/ad-tech-problems/ (accessed 4 March 2019).
9 On contextual advertising generally see, Kaifu Zhang and Zsolt Katona, 'Contextual Advertising' (2012) 31 *Marketing Science* 873.
10 Turow (n 5) 116.
11 On the history and development of cookies see, Rajiv C Shah and Jay P Kesan, 'Deconstructing Code' (2004) *Yale Journal of Law and Technology* 278.
12 See, Bernal (n 7) 144.
13 Lilian Edwards, 'Data Protection and e-Privacy: From Spam and Cookies to Big Data, Machine Learning and Profiling' in Lilian Edwards (ed), *Law, Policy and the Internet* (Hart 2018) 119, 126-7.
14 Turow (n 5) 92.
15 DoubleClick is the market leader in third party advertising. See, Edwards and Howells (n 7).
16 Joseph Turow, *The Daily You: How the new Advertising Industry is Defining your Identity and your Worth* (Yale University Press 2011), 34-64.
17 Matthew Ingram, 'How Google and Facebook Have Taken Over the Digital Ad Industry' (*Fortune*, 4 January 2017) https://fortune.com/2017/01/04/google-facebook-ad-industry/ (accessed 4 March 2019).
18 Lucas Graves and Rasmus Kleis Nielsen Tim Libert, Changes in Third-Party Content on European News Websites after GDPR, 2018, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-08/Changes%20in%20Third-Party%20Content%20on%20European%20News%20Websites%20after%20GDPR_0_0.pdf (last accessed 10 April 2020).
19 Julian Angwin, 'Google has Quietly Dropped Ban on Personally Identifiable Web Tracking' (*ProPublica*, 21 Aoct. 2016) https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking (accessed 25 February 2019).
20 John D McKinnon and Douglas MacMillan, 'Google Says It Continues to Allow Apps to Scan Data From Gmail Accounts' (*The Wall Street Journal*, 20 Sept 2018) https://www.wsj.com/articles/google-says-it-continues-to-allow-apps-to-scan-data-from-gmail-accounts-1537459989 (accessed 4 March 2019).

The data gathered by Google in relation to its own service offerings, is collected on the basis of user consent when consumers accede to the terms of service and privacy policy attached to the relevant offering. The aggregation of both the data which Google collects through its analytics services and through its own services permits the company to build a detailed data sets for a broad swathe of online users. From this the individual characteristics and preferences of consumers can be analysed or inferred – and detailed profiles of individual consumers can be sold through DoubleClick or aggregated with further information obtained through that platform.

Similarly, Facebook's terms of service and privacy policy require users to consent to the collection, recording and potential sale of the data related to their posts, photos, shared items, group and page memberships, location and installed apps. Facebook has, in the past, also granted its advertising customers access exceeding what was contractually permissible under these terms and policies, including accessing the names of Facebook users' friends and the contents of 'private' messages without the consent of users. [21] Like Google, Facebook can combine this information with the information it obtains through its analytics services to build complex and detailed profiles for sale through the AdTech market.

Many websites may incorporate a Facebook Pixel for analytics purposes, a small piece of code which monitors consumer activity[22] even where consumers are not logged on to Facebook or are not Facebook users (a group Facebook has, rather ominously, dubbed 'non-registered users'[23]) across websites and platforms that contain a Facebook pixel or social plugin.[24]

Facebook's contribution to the erosion of consumer privacy is thus enabled not only through these contractually permitted policies (and their breach) but through these analytics services offered by the pixel. It is also enabled by Facebook's embedded social plugins- the buttons which invite visitors to a website to 'like' or 'share' items or pages online. Where these buttons appear, regardless of whether a consumer interacts with them, Facebook is collecting data related to their activity on that site.

In light of their integrated collection and analysis capabilities, Google and Facebook have become 'triple threats' – offering analytics services to other websites, collecting and aggregating large amounts of data through their own platforms and benefitting from the highly targeted profiles of consumers which they can build and auction to advertisers as a result – part of a broader model which Shoshanna Zuboff has called 'surveillance capitalism'[25] and Danielle Citron and Frank Pasquale have referred to as a central part of the 'scored society.'[26] This threat is only amplified by the further integration of these platforms with other services[27] a pattern noted by Binns et al as part of which consumers sign up or interact with other services by authenticating themselves through their Facebook or Google profiles.[28] The consumer profiles on which Facebook and Google as well as other actors in the AdTech marketplace operate are then sold for use in targeted, behavioural advertising through the real time bidding (RTB) system.

2.2 The Real Time Bidding System (RTB)

When a consumer visits a website, they are shown advertising which is targeted to them based on data gathered and aggregated by data brokers (a group which includes actors like Facebook and Google). The process of a consumer's data being broadcast, advertisers bidding for the attention of that consumer based on their data and the advertiser's ad appearing on the website being viewed by the consumer takes places in milliseconds. During this period the consumer's data is broadcast to an undefined number of advertisers who bid for the available advertising space and the consumer's attention.

This auction system is part of the 'real time bidding' (RTB) mechanism which fuels the AdTech market and operates through one of two markets. The first is Open real time bidding (Open RTB), which is used by a majority of online media providers and advertising industry participants.[29] The second, is Google's proprietary RTB "Authorized Buyers" (AB) system.[30]

The information which is sent to bidders in the auctions (using either system) is referred to as bid request data and can include; the content which the consumer is viewing, their location and a description of the device they are using to access the internet, their unique tracking identities (cookies) as well as their IP address. It may also include additional, enhanced data provided by a data broker based on an analysis and aggregation of other data and which may include the consumers income bracket, age and gender, ethnicity, sexual orientation, religion and political persuasions.

More concerningly, and as highlighted in recent complaints filed with the Irish Data Protection Commissioner[31] and UK's Information

21 Michael LaForgia and Gabriel JX Dance Nicholas Confessore, 'Facebook Failed to Police How Its Partners Handled User Data' (*The New York Times*, 12 Nov 2018) https://www.nytimes.com/2018/11/12/technology/facebook-data-privacy-users.html (accessed 4 March 2019).

22 Facebook pixel is similarly to a cookie, a code for websites which allows websites to measure and analyse their audience. See, https://www.facebook.com/business/learn/facebook-ads-pixel.

23 Günes Acar, Brendan Van Alsenoy, Frank Piessens, Claudia Diaz, Bart Preneel, Facebook Tracking Through Social Plug-ins, Technical report prepared for the Belgian Privacy Commission 2015.

24 In Case C-40/17 *FashionID* EU:C:2019:629, [3], [23] that case the CJEU was asked to consider the integration of social plug-ins, and in particular whether the Facebook 'Like' plug-in on an online retailer's website which transferred the user's IP address and browsing string to Facebook regardless of whether user was a Facebook user or had clicked the like button rendered the appellant a joint data controller for the purposes of the GDPR. In his Opinion, Advocate General Bobek found that, having embedded plug-in in its website resulted in FashionID being considered a joint controller of the data collected though its responsibility should be limited to those operations for which it effectively co-decides on the means and purposes of processing legitimate interests and consent a finding with which the subsequent judgment the CJEU concurred. This decision is one of a growing number of a rapidly proliferating set of challenges by European regulatory and judicial bodies to the activity of actors in the AdTech market which is considered in section four.

25 Shoshanna Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs 2019).

26 Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1.

27 In the United States for example, Facebook has sought to integrate financial services offered by Chase, Wells Fargo, Citigroup and US Bancorp with its messenger service, Deepa Seetharaman and Anna Maria Andriotis Emily Glazer, 'Facebook to Banks: Give us your data, we'll give you our users' (*Wall Street Journal*, 6 August 2018) https://www.wsj.com/articles/facebook-to-banks-give-us-your-data-well-give-you-our-users-1533564049 (accessed 9 April 2019).

28 Binns et al (n 7). The authors proceed to note the negative competition impacts of such consolidation capabilities

29 See, 'Open Real Time Bidding' at https://www.iab.com/guidelines/real-time-bidding-rtb-project/ (accessed 4 March 2019).

30 Google Ads, 'How Ad Exchange works with Google Ads' https://support.google.com/google-ads/answer/2472739?hl=en (accessed 29 February 2019).

31 See, Brave 'Grounds of Complaint to the Data Protection Commissioner' https://brave.com/DPC-Complaint-Grounds-12-Sept-2018-RAN2018091217315865.pdf (accessed 4 March 2019).

Commissioner's Office,[32] the RTB mechanism does not permit control over the dissemination of personal information once it has been broadcast. The advertising industry has countered these complaints with arguments that it abides by its own self-regulatory standards which comply with relevant EU law, which the proceeding section now turns to consider.

2.3 How AdTech Affects our Lives

Consumers' offline lives are deeply integrated with, and are in many ways, as diverse as their digital experiences.[33] As a result, the capacity to track consumers' online activity (and by implication a certain amount of their related, offline activity) naturally generates concern about the impacts of surveillance on individual privacy and the manipulation which can result from such privacy reductions. This is perhaps best illustrated by way of comparison to comparable offline surveillance.

An individual enters a shop. On entering the shop, their name and postcode are given to the shop owner. A private detective who has been following them since they last visited the shop then also hands the shop owner a list of their previous purchases and movements – the names and addresses of the locations they have gone since their last visit to the shop, the area where they live, the types and prices of the goods and services they view most frequently. From this the shop owner can build a rough picture of the shopper's age, socio-economic status and perhaps political and religious persuasions.

As the shopper moves around the shop they are tracked by cameras which record the aisles they visit, the products they looked at and how long they considered each product. On leaving the shop they are then followed again by the private detective who records where they go and what they purchase or consider purchasing. The shopper stops into a coffee shop to meet some friends and the detective records what they eat and drink, and sits nearby listening to their conversation, he obtains a list of their other friends and the shops they enter and goods they purchase building a detailed profile of the social network of the shopper. At the end of the day the private detective gives this information to the shop owner. The shop owner now has an extensive list of the shopper's social connections, geographic movements, areas of interest and purchases from which more intimate details such as his age, gender, race, sexuality, political and religious preferences and socio-economic status can be inferred.

The shop owner can use this information himself to target the shopper with ads for his products or services, hoping by the power of suggestion to influence his preferences. But the private detective who conducted much of the data gathering and analysis for the shop owner might also take his detailed profile of the shopper and sell it to other shop owners trying to influence the shopper to purchase their goods or use their services, to political actors seeking to influence the shoppers preferences in an upcoming election, or to any number of other actors who will bid for the data in order to be able to influence the shopper.

In the online environment, the AdTech market operates on a similar basis to the shopper and those who surveil him in this example. The privacy harm is, of course, evident. What also becomes clear is the negative consequences this surveillance may have for the activities or choices the shopper feels able to make (given that he is being watched) or which he is aware he can make (given that his attention is being vied for constantly by actors who have purchased large quantities of his personal data). Further still, the real world comparison draws to the fore the authoritarian undercurrent of such pervasive surveillance and its capacity to be exploited not only by commercial but also by State actors to influence the shopper.

In a real world scenario, the shopper would not merely notice but might reasonably object to such practices and choose to conduct their business in a shop which did not employ such mechanisms. However, the equivalent prompts to the presence of such surveillance, and alternatives which avoid it, are not necessarily present or available in the digital environment. Individuals are required if not by social, then frequently by professional necessity to engage with the digital market in ways which offer them little alternative but to consent to privacy policies and terms of use which permit their data to be gathered, aggregated, broadcast and sold as part of the AdTech market.

As the decisions outlined in sections below indicate, there is a growing awareness of and unease concerning the AdTech landscape which enables this surveillance of, and influence over, consumers[34] while, as section four examines, the regulatory mechanisms which are currently present do not fully address the privacy impacts which AdTech imports.

3 The Legal Landscape in which AdTech Operates

The AdTech market as detailed in the previous section has to date been governed by a mix of self-regulatory efforts in the form of the Interactive Advertising Bureau Europe Framework, which governs the Open RTB system, Google's AB Guidelines which govern that company's proprietary advertising market platform and those European rules governing the contractual permissions which enable Google and Facebook (as the examples used in this article) to collect and sell user data to advertisers.

3.1 The Interactive Advertising Bureau Europe Framework

The Open RTB system in Europe is currently subject to the voluntary Framework established by the European branch of the Interactive Advertising Bureau in its 'Europe Transparency & Consent Framework.' The IAB Framework provides an open-source, industry standard which aims to ensure actors in the digital advertising chain comply with the GDPR and ePrivacy Directive when processing, accessing or storing information on consumer devices including cookies, advertising identifiers, device identifiers and other tracking technologies.

The Framework is predicated on the collection of consent from data subjects for all subsequent data sharing to third parties during the Open RTB process[35] yet the Framework anticipates that this broadcasting of personal data to third parties may occur without consent stating,

> A Vendor[36] may choose not to transmit data to another Vendor

32 See, Brave 'Submission to the Information Commissioner' https://brave.com/ICO-Complaint-.pdf (accessed 4 March 2019).

33 Helen Nissenbaum, 'A contextual approach to privacy online' (2011) 140 *Daedelus* 32.

34 Case C-40/17 *FashionID* EU:C:2018:1039; C-673/17 *Planet49* EU:C:2019:246; Case C- 311/18 *Schrems*.

35 See, IAB Europe, 'Europe Transparency & Consent Framework' http://www.iabeurope.eu/tcfdocuments/documents/legal/currenttcfpolicyFINAL.pdf (accessed 4 March 2019).

36 In this context vendors are data brokers and buyers may be other brokers or parties interested in obtaining by purchase or license access to the data collected and analysed by such vendors.

> for any reason, but a Vendor must not transmit data to another Vendor without a justified basis for relying on that Vendor's having a legal basis for processing the personal data.
>
> If a Vendor has or obtains personal data and has no legal basis for the access to and processing of that data, the Vendor should quickly cease collection and storage of the data and refrain from passing the data on to other parties, even if those parties have a legal basis.[37]

Those broadcasting bid data are thus afforded significant discretion in determining whether those to whom they broadcast their data possess a "justified basis for relying on that Vendor's having a legal basis for processing personal data" effectively circumventing the consent basis on which the Framework purports to rely and conditioning the integrity of the system on the presence, and rigour, of the vendor's assessment rather than consent or indeed the other basis for processing enumerated under the GDPR.

Motivated, no doubt, by such criticisms IAB Europe announced in 2018 it was developing a tool, in collaboration with The Media Trust, to determine whether the "consent management platforms" (CMPs) that facilitate this passing of data under the IAB Europe Framework are compliant with the Framework's policies.[38]

However, as the CNIL decision detailed in *Vectaury* (examined below) illustrates, the more fundamental concern is that it appears that such consent management platforms are themselves non-compliant with the GDPR. It is also unclear whether a reformatting of the Framework announced by IAB Europe in early 2019 to comply with GDPR can ameliorate the subsisting difficulties with the RTB system itself which broadcasts data so widely, regardless the GDPR compliance efforts of the Framework (which it should be emphasised is a voluntary standard).

3.2 Google's Authorised Buyers Guideline

Google has, thus far, declined to integrate the IAB Europe Framework into its proprietary market[39] and has instead operated its own parallel system in the Google Authoring Buyer Guideline. Similarly to the IAB Framework, the AB Guideline shifts responsibly for data protection from the data controller to those third parties to whom the data is broadcast, noting that buyers may store identifiers in order to evaluate impressions and bids based on user-data previously obtained.[40] The Guideline also permits all other callout data (with the exception of location data) to be retained by a Buyer after responding to an ad call for up to 18 months, in order to enable forecasting of the availability of inventory.[41]

The Guideline does impose limitations on how Buyers use data obtained during the bidding process but notes only that it is not permissible to use callout data to create user lists or to profile users and prohibits the association of callout data with third parties.[42] However, this ignores the practical reality that bidders for such data, of which Cambridge Analytica is an example, can and do perform a 'sync' that uses personal data obtained through the bidding process to augment existing consumer profiles.[43]

Moreover, and in a similar vein to the control issues identified with the IAB Framework, the Google Guideline provides that where a

> Buyer accesses, uses, or processes personal information made available by Google that directly or indirectly identifies an individual and that originated in the European Economic Area ("Personal Information"), then Buyer will:
>
> - comply with all privacy, data security, and data protection laws, directives, regulations, and rules in any applicable jurisdiction;
> - use or access Personal Information only for purposes consistent with the consent obtained by the individual to whom the Personal Information relates;
> - implement appropriate organizational and technical measures to protect the Personal Information against loss, misuse, and unauthorized or unlawful access, disclosure, alteration and destruction; and
> - provide the same level of protection as is required by the EU-US Privacy Shield Principles.

Buyers will regularly monitor your compliance with this obligation and immediately notify Google in writing if Buyer can no longer meet (or if there is a significant risk that Buyer can no longer meet) this obligation, and in such cases Buyer will either cease processing Personal Information or immediately take other reasonable and appropriate steps to remedy the failure to provide an adequate level of protection.[44]

This suggests that once personal data is transferred to a Buyer, AB has no effective control over its use. The result, as the proceeding section examines is that, the AdTech market as it is currently constituted, is operating in manner at odds with the data protection standards under the GDPR and, more fundamentally, with individual privacy.

3.3 Consumer Protection Regulation of AdTech

The most evident regulatory mechanism for the AdTech market is consumer protection, an area in which the Union enjoys an explicit competence and an established history of legislative intervention in the market. However, while the European Union has traditionally placed a high value on consumer protection, a fact reflected in the Treaty Articles,[45] and the Charter, as well as in secondary law[46] there is currently no consumer protection standards which are applicable to AdTech.

The Consumer Rights Directive,[47] which replaced the Distance Selling[48] and Doorstop Selling Directives[49] establishes requirements for information to be provided in distance contracts,[50] formal require-

37 See, IAB Europe, 'Europe Transparency & Consent Framework' http://www.iabeurope.eu/tcfdocuments/documents/legal/currenttcfpolicyFINAL.pdf, para 14.4, 14.5 (accessed 4 March 2019).
38 See, Media Trust, 'IAB Europe CMP Validator Helps CMPs Align with Transparency and Consent Framework' https://mediatrust.com/media-center/iab-europe-cmp-validator-helps-cmps-align-transparency-consent-framework (accessed 4 March 2019).
39 See, Robin Kurzer, 'IAB Europe to release updated consent framework later this year, Google to sign on' (*MarTech Today*, 12 Feb 2019) https://martechtoday.com/exclusive-iab-europe-to-release-updated-consent-framework-google-to-sign-on-230704 (accessed 4 March 2019).
40 Google Authorised Buyer Guidelines, https://www.google.com/doubleclick/adxbuyer/guidelines.html (accessed 7 March 2019).
41 Ibid.
42 Ibid.
43 Ibid.
44 Ibid.
45 Articles 39, 107 and 169 TFEU.
46 See, Stephen Weatherill, *EU Consumer Law and Policy* (2nd edn, Edward Elgar 2014).
47 Directive 2011/83/EC.
48 Directive 97/7/EC.
49 Directive 85/577/EC.
50 Directive 2011/83/EC, Article 6.

ments for distance consumer contracts[51] and a right of withdrawal but does not offer any mechanisms for the regulation of the AdTech market proper.[52] Similarly, the Unfair Consumer Contracts Directive[53] while it includes requirements that contractual terms are drafted in clear language, intelligible to the ordinary consumer is not directly relevant to AdTech.[54] New legislative measures announced in January 2019 including the Directive regulating the supply of digital content and services similarly fail to address the AdTech market.[55] Provisions governing advertising do appear in the 2006 Directive on Misleading Advertising,[56] and in the Directive on Unfair Commercial Practices[57] and while there is no reason, in principle, why these provisions could not be extended to cover AdTech, decisions considering the application of the Directives have been limited[58] and would, in any case, be restricted to the advertising facilitated by AdTech rather than the system which enables it.

3.4 Data Protection, Privacy and the Regulation of AdTech

In the absence of applicable consumer protection laws, the primary regulatory mechanisms currently applicable to the AdTech market emanate from the Union's data protection legislation. The right to data protection enjoys constitutional footing within the Union's legal schema through Article 16 TFEU as well as Article 8 CFR. From this foundation the Union has developed a comprehensive schema for the enforcement of the right to data protection, first through the Data Protection and e-Privacy Directives and more recently with the GDPR.

Of these legislative measures the ePrivacy Directive (ePD) frequently referred to, misleadingly, as the Union's 'e-Cookie' law, is the most direct regulatory mechanism applicable to the AdTech industry. The Directive requires Member States to ensure that the use of electronic communications networks to store information or to gain access to information stored in terminal equipment is permitted only where the subscriber or user concerned is provided with clear and comprehensive information regarding the purposes of the processing, and is offered the right to refuse same.[59] The Directive thus imposes informational and consent requirements on the operation and placement of cookies on consumer's devices. However, as section four examines, the capacity of the Directive to provide for substantive privacy protections, rather than threshold operational requirements for the technologies which cause privacy reductions, is limited, and in practice the right to refuse cookies has been ineffective, frequently resulting in access to a site or service being denied.

This discrepancy between data protection rights and substantive privacy protections lies at the heart of the Union's legislative mechanisms as they apply to the regulation of AdTech. Despite the proliferation of ostensibly privacy orientated legislation during the last two decades, the Union's legislative product while seemingly indicative of a strong commitment to privacy is, on closer examination, notable for its emphasis on market-oriented threshold regulations in the form of information and notice requirements rather than substantive interventions to protect consumer privacy writ large.[60]

The separation of the right to data protection from its ostensible root in the right to privacy, and the continuing ambiguity in the jurisprudence of the CJEU as to the relationship between the two rights has hardly helped matters.[61] However, it is also a distinct product of the legislative preference within the Union for market oriented rather than socially oriented rights protections. Indeed, Antoinette Rouvroy and Yves Poullet have criticized the recognition of the right to data protection, distinct to the traditional fundamental right to privacy on this basis arguing that such division obscures the essential relationship between the rights and estranges data protection from the fundamental values of human dignity and individual autonomy which should justify its existence through its derivation from a privacy interest.[62]

This concern is well placed. While the right to data protection is understood as derived from the right to privacy in the Union's law, the Recitals to the GDPR emphasise only the trade and market-orientated functions of the right, neglecting the social and normative roots of data protection in privacy and that right's function in securing individual dignity and the development of individual personality.[63]

The e-Privacy Directive similarly emphasises in its Recitals the market-based functions of its provisions and while the proposed e-Privacy Regulation includes wording in its explanatory memorandum which makes explicit reference to the right to privacy under Article 7 as distinct from the right to data protection, the Recitals to the Regulation refer to data protection and privacy interchangeably. Moreover, the substance of its guarantees relate largely to interoperability, and digital infrastructures as part of the digital single market with little concern for deeper normative impacts. [64] As such, the provisions of the e-Privacy Regulation appear, in fact, to be a mere extension of the GDPR's focus on market oriented data protection in a differentiated context.

This legislative prioritisation of data protection over privacy is particularly problematic in the context of AdTech. Data protection on a close doctrinal analysis could be considered not to be a right as much as a series of mandatory safeguards which must be present in order to legally infringe privacy rights proper. The GDPR and e-Privacy Directive are thus not so much rights standards in themselves but the enabling frameworks for permissible reductions in rights to individual privacy.

While this is not objectionable *per se*, the use of rights language obfuscates the relationship between data protection and privacy while the promotion of data protection over privacy exposes the alienation of data protection from the justificatory basis of privacy and its ideological foundations in individual autonomy. In practice, the result has been that the contractual practices which form a crucial part of the AdTech landscape have proliferated largely unopposed on the basis that they satisfy current data protection requirements without

51 Ibid, Article 8.
52 Ibid, Article 9-16.
53 Directive 93/13 [1993] OJ L095/29.
54 Ibid, Article 5. Ambiguity in relation to the meaning will be resolved in favour of the consumer under this provision.
55 Council of the European Union, Council and Parliament agree on new rules for contracts for the sales of goods and digital content (2019).
56 Directive 2006/114 [2006] OJ L376/21.
57 Directive 2005/29/EC.
58 Case C-281/12 *Trento Sviluppo* EU:C:2013:859; Case C-122/10 *Ving Sverige* EU:C:2011:299; Case C-428/11 *Purely Creative* EU:C:2012:651.
59 Article 5; Acar (n 23).
60 Gloria Gonzalez Fuster, *The Emergence of Personal Data Protection as a Fundamental Rights of the EU*, vol 16 (Law, Governance and Technology Series, Springer 2014), 243-5.
61 On this see, Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 *International Data Privacy Law* 1.
62 Antoinette Rouvroy and Yves Poullet, 'The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy' in Yves Poullet Serge Gutwirth, Paul De Hert, Cécile de Terwangne andSjaak Nouwt (ed), *Reinventing Data Protection?* (Springer 2009).
63 Recitals 2, 3.
64 Recital 1, 20 – 24.

reference to the deeper impacts which they occasion for privacy.

Cumulatively, this reduction of individual privacy leads in turn to the creation of a population whose preferences and proclivities can be exploited to influence not only individual preferences but also political opinions importing negative consequences for democratic participation, and in turn the Rule of Law. However, before these broader impacts are examined, it is necessary to consider how the current AdTech landscape is accommodated within the current Article 8 framework and the specific shortcomings of the Union's data protection legislation in regulating AdTech.

4. How does AdTech fit within the Article 8 Privacy Framework?

In accordance with Article 6 GDPR, processing of personal data is lawful only if and to the extent that at least one of the listed conditions are present, namely that the data subject has given consent for one or more specific purposes or the processing is necessary for; the performance of the contract,[65] compliance with a legal obligation or to protect vital interests of data subject, for performance of a task carried out in the public interest or the purposes of the legitimate interests pursued by the controller or a third party.

Under the Regulation consent is one the primary grounds for lawful processing of personal data, a position emphasised by Article 7 GDPR, which requires data controllers to demonstrate that the data subject has consented. When assessing the legitimacy of consent the Regulation emphasises in Article 4(11) that consent must be a freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she through a statement or by a clear affirmative action signifies agreement to the processing of personal data relating to him or her, a position reaffirmed in Recitals 42 and 43.[66]

That the collection and sale of consumer data by data brokers as part of the RTB process involves the processing of personal data is evident. The question then is whether such collection satisfies the requirements of consent under Article 6(a) GDPR or is permissible under an alternative ground for lawful processing.

4.1 Adequate Consent under Article 6 and Article 4(11) GDPR

The operation of the RTB system, and the voluntary self-regulatory structures which seek to provide a governance structure for it, ostensibly operate on the basis of consent. However, it is not clear that the IAB Framework or Google AB Guidelines satisfy the GDPR's definition of consent, as the *Vectaury*[67] decision of the French Commission Nationale de l'informatique et des libertés (CNIL) demonstrates.

In January 2019, the CNIL found *Vectaury*, a French AdTech firm, had collected data to create consumer profiles subsequently auctioned through the RTB system without consent. The decision was significant because it found that the validity of consent obtained directly through apps that embed *Vectaury*'s consent management platform and the validity of consent collected elsewhere and signaled to *Vectaury* through use of the IAB Europe Consent Framework ultimately failed to meet the consent criteria required by the GDPR.

The CNIL found consent obtained through consent management platforms was insufficient because it was not informed, specific or affirmative as required by Recital 32 and Articles 4 and 6 GDPR. Crucially, the decision found that consent obtained through the IAB Europe Framework is inherently invalid as consumer consent cannot be passed from one controller to another controller through a contractual relationship.[68] This, of course has broader implications for the operation of consent based AdTech models more generally. The decision also specifically queried whether, in light of the opacity of the RTB system, consumers could be considered to have given valid consent to a process they did not understand or are unaware of and explicitly stated that its decision should be read as placing not only *Vectaury* but the AdTech ecosystem as a whole on notice that existing market practices may violate the requirements of the GDPR. The decision noted separately that the collection of geolocation data for advertising purposes, by *Vectaury*, presented particular risks as it revealed the movements and habits of consumers and could be used to imply sensitive categories of data.[69]

The decision cogently illustrates the false narrative of consumer consent on which the AdTech industry relies and has implications beyond the IAB Framework. For example, Google has traditionally required publishers to collect consent on its behalf for advertising profiling in a similar manner to the IAB's Framework.[70] While Google have stated they will audit this collection for compliance with consent requirements[71] it is no longer clear that this will be sufficient.

IAB Europe responded to the CNIL judgment stating it merely provides a technical, voluntary standard in accordance with which its members may choose to be but are not required to be bound and suggesting that *Vectaury* had fallen foul of the regulator as it had not adequately adopted and complied with the Framework rather than the error subsisting with the Framework itself.[72] However, this conveniently ignores the central, contractual criticism on which the CNIL decision rests – that there is no refuge in packaged, contractual passing of consent and that consumers have not consented to the use of their data in a broader AdTech ecosystem when they agree to use a service or app.

The CNIL decision also congrues with recent CJEU jurisprudence in *Wirtschaftsakademie*[73] and *Planet49*. In *Wirtschaftsakademie* a preliminary reference from the German Courts asked whether the failure by

65 See also, Recital 44 and Article 7(4) which provides that when assessing whether consent is freely given utmost account shall be taken of whether the performance of a contract, including the provision of a service is conditional on consent to the processing of personal data that isn't necessary for the performance of that contract.

66 Recital 42 requires that processing based on the data subject's consent should be demonstrable by the data processor and in the context of a written consent, safeguards should be put in place to ensure that the data subject is aware of the fact that and the extent to which consent is being given by them. Recital 43 provides that in assessing whether consent has been freely given, consent should not be considered to have been given where there is a clear imbalance between the subject and controller.

67 Commission Nationale de l'informatique et des libertés, 'Décision n°MED-2018-042 du 30 octobre 2018' at https://www.legifrance.gouv.fr/affichCnil.do?oldAction=rechExpCnil&id=CNILTEXT000037594451&fastReqId=974682228&fastPos=2 (accessed 1 March 2019).

68 Ibid.

69 As defined under Article 9 GDPR.

70 Natasha Lomas, 'Google accused of using GDPR to impose unfair terms on Publishers' (*Tech Crunch*, 1 May 2018) https://techcrunch.com/2018/05/01/google-accused-of-using-gdpr-to-impose-unfair-terms-on-publishers/ (accessed 5 March 2019).

71 Lara O'Reilly, 'Google Wants Publishers to Get Users' Consent on Its Behalf to Comply With EU Privacy Law' (*The Wall Street Journal*, 22 March 2018) https://www.wsj.com/articles/google-wants-publishers-to-get-users-consent-on-its-behalf-to-comply-with-eu-privacy-law-1521749003 (accessed 5 March 2019).

72 Townsend Feehan, The CNIL's *Vectaury* Decision and the IAB Europe Transparency & Consent Framework (2018).

73 Case C-210/16 *Wirtschaftsakademie* EU:C:2018:388.

Facebook and *Wirtschaftsakademie* (the administrator of a fan page on the platform) to inform visitors that cookies were placed on their device by Facebook constituted a breach of the (then) Data Protection Directive. In particular the appellant's asked whether they could be considered a joint controller with Facebook.[74] The Court noted that though Facebook placed the cookies in accordance with its contract with *Wirtschaftsakademie*, the appellant had benefitted from that placement and was involved in the subsequent analysis in as much as it decided the parameters of the information collected based on its interests and was thus a joint controller of the data and required to institute its own system for informing users of the page that cookies were placed on their devices.[75]

The decision in *Planet49* added to this nascent body of precedent. In that case, the CJEU was asked to consider whether online cookie consents with default pre-ticked boxes submitting to the use of cookies was permissible under the GDPR and e-Privacy Directive. In his Opinion in the case, Advocate General Szpunar noted that the requirements of consent under the GDPR include that consent is active, freely given, separate (i.e. not bundled) and informed, requiring the provision of clear and comprehensive information concerning the duration and operation of the cookies and whether third parties have access to the information collected. The AG noted that these conditions were not met where pre-ticked cookie consent boxes were used.[76] The Court agreed noting that the GDPR standard of consent as freely given, specific, informed and unambiguous was not satisfied in such circumstances.

In the context of the AdTech industry the implication of these decisions would seem to be that where a consent management platform, or otherwise delegated or default consent mechanism is used, a third party who benefits from the data collected and analysed is to be considered a data controller and must satisfy the consent thresholds of the GDPR anew. Given the apparent problems posed by a consent-based processing of user data in light of these decisions it is thus necessary to consider whether the legitimate interest ground under Article 5 might offer an alternative means of legitimate processing.

4.2 Legitimate Interests under Article 6 GDPR

As an alternative to consent, under the GDPR personal data may also be processed on the basis of legitimate interests under Article 6(f). Article 6(f) operates in addition to the more general principle of legitimate interests outlined in Article 5 which provides that personal data shall be processed lawfully, fairly and in a transparent manner and collected for specified, explicit and legitimate purposes and not further processed in a manner incompatible with those purposes. Supplementing Article 6(f), Recital 47 (though non-binding) notes that there should be a relationship between the data controller and data subject on which a legitimate interest is based such as where the data subject is a client, or is in the service, of the data controller. The Recital notes, however, that the existence of a legitimate interest requires careful assessment, including an assessment of the reasonable expectations of the data subject at the time and in the context of the collection of the data.

While this might seem, *prima facie*, to offer a readily available alternative to a consent-based processing in the context of AdTech, any reliance on legitimate interests for the operation of the RTB system would be misplaced. RTB data is broadcast to an undefined list of bidders, who, though they are directed and legally required not to retain or further use such data,[77] are not actively policed by the bid broadcaster to ensure this. Once a bidder is not successful, they no longer have a legitimate interest in processing the data but may retain it. Equally, the data may be received by bidders who have no interest in the segment or consumer data being auctioned but nonetheless receive the data through the RTB system.

The CNIL has previously found that that ticking a box labelled *"I agree to the processing of my information as described above and further explained in the Privacy Policy"* did not satisfy the consent requirements under the GDPR because it attempted to require consent for over one hundred processes and set personalise ads as a default setting.[78] That decision, directed against Google[79] also noted that the processing could not be considered a legitimate interest of the company under Article 6(f) such that consent was not required . The CNIL noted that Google's was particularly intrusive due to the number of services offered by the company, and the quantity and nature of the data processed and combined.

This mirrors the opinion expressed by the Article 29 Working Party that the legitimate interest basis does not cover situations where the processing is not genuinely necessary for the performance of a contract but rather relates to the ancillary use of data and is achieved through terms unilaterally imposed on the data subject. [80] In particular, the Opinion noted that the legitimate interest premise is not a suitable legal basis on which to compile a profile of consumer tastes and choices as the controller has not been contracted to carry out profiling, but rather to deliver particular goods or services and the inclusion of such terms in the contract does not make them necessary for it.[81] This critique is echoed by Frederik Borgesius who notes "the fact that a company sees personal data processing as useful or profitable does not make the processing 'necessary'[82] to provide the contracted service to the user.

4.3 Explicit Consent under Article 9 GDPR

Even where it was possible to establish that processing was permitted on the basis of legitimate interest, under Article 9 GDPR, processing of "special categories" of personal data requires explicit consent if that data has not been "manifestly made public" by the data subject and no other exception applies.[83] Special categories of data include;

74 Case C-210/16, [15].
75 Case C-210/16, [40] noting that as non-Facebook users could visit the page in that circumstance the responsibility of the administrator of the page would be even greater.
76 C-673/17 *Planet49*.
77 See, Article 5.
78 Ibid. It is worth noting in this respect that the Article 29 Working Party in its 2012 Report on Cookie Consent noted that by default social plug-ins should not set a third part cookies in pages displayed to non-members, Article 29 Working Party, Opinion 04/2012 on Cookie Consent, 2012).
79 Commission Nationale de l'informatique et des libertés, 'Déliberation SAN-2019-001 du 21 janvier 2019' https://www.cnil.fr/sites/default/files/atoms/files/san-2019-001_21-01-2019.pdf (accessed 5 March 2019).
80 Article 29 Working Party on Data Protection, Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 2014), 16.
81 Ibid.
82 Frederik Zuiderveen Borgesius and Joost Poort, 'Online Price Discrimination and EU Data Privacy Law' (2017) 40 *Journal of Consumer Policy* 347, 360.
83 The exceptions provided in Article 9(2) include (a) explicit consent, (b) necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law as authorised by member state law (c) protect vital interests (d) carried out in the court of its legitimate activities and with appropriate safeguards by a foundation, association or other non-profit body for phi, religious, trade union aim with regard to its current and former members only (e) relates to

racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership, and the processing of genetic data or biometric data for the purpose of uniquely identifying a natural person or data concerning health or an individual's sex life or sexual orientation. In addition, Recital 51 requires that personal data which are by their nature sensitive merit specific protection in a context where their processing could create significant risks to fundamental rights and freedoms.

As the CNIL decision in *Vectaury* noted, and as Sandra Wachter[84] has argued elsewhere, the collection and aggregation model used by AdTech at present effectively allows individual characteristics or preferences which are classified as 'special categories' of data under GDPR to be deduced or inferred through aggregation and analysis. The result should therefore be, on a purposive reading of the Regulation, that the enhanced, explicit consent requirements under Article 9 are triggered even where the initial data collected are non-sensitive but where their combination with other data, or their geographic of temporal record is such as to allow the imputation of sensitive categories of data.

However, both the IAB Framework and the AB Guidelines permit data to be processed with, at most, implicit consent based on the consumer's previous consents or continued use of a service. This is insufficient under the GDPR in accordance with the threshold established for consent but specifically impermissible in the context of sensitive categories of personal data.[85] This does not appear to have deterred Facebook[86] or its companies WhatsApp[87] and Instagram[88] or Google[89] from processing special categories of data under Article 9 GDPR with basic, rather than explicit permission.

Complaints filed by NOYB against all four companies allege their data collections models fail to specify the legal basis on which data is processed, as required under Articles 6 and 9. In particular the complaints note that the contracts used simply list all possible grounds for lawful processing under Article 6 leading to the assumption that processing is based on consent by failing to indicate on what other Article 6 basis the processing is conducted. However, the privacy policies of the companies only note that they process data of their users as necessary "to fulfil our terms" importing an association with Article 6(b) and (f) which is not clarified. Moreover, such companies do not inform their users of the actual uses to which their data may be put, including sensitive data, as required under Articles 12 and 13.

personal data which are manifestly made public by the data subject (f) establishment, exercise or defence of legal claims (g)necessary reasons of substantial public interest (h) necessary for the purposes of preventive or occupational medicine (i) processing in necessary for reasons of public interest in health.

84 Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2019) 35 *Berkeley Technology Law Journal* Forthcoming.

85 Commission Nationale de l'informatique et des libertés, 'The Restricted Committee of the CNIL imposed a sanction of 150,000 € against Facebook Inc and Facebook Ireland' https://www.cnil.fr/en/facebook-sanctioned-several-breaches-french-data-protection-act (accessed 5 March 2019).

86 NOYB, 'GDPR: noyb.eu filed four complaints over "forced consent" against Google, Instagram, WhatsApp and Facebook' 25 May 2018 https://noyb.eu/4complaints/ accessed 5 March 2019. The complaints appear to have been removed - an article based on the complaints is available at https://www.theguardian.com/technology/2018/may/25/facebook-google-gdpr-complaints-eu-consumer-rights

87 Ibid.

88 Ibid.

89 Ibid.

4.4 Failure to Inform Data Subjects under Articles 12 and 13 GDPR

Article 12 GDPR requires the data controller to take appropriate measures to provide any information about how data will be used to be provided in an intelligible and easily accessible form using clear and plain language. In addition, Article 13 GDPR provides that where personal data are collected, the controller shall provide the data subject with a range of information including, but not limited to, the purposes of processing, the recipients or categories of recipients of the data, the period for which the data will be kept (and how such a period is determined) and the existence of automated decision making including profiling which the data may be exposed to, including meaningful information about means used. Recital 39 further notes that any processing of personal data should be lawful and fair, and clarify what personal data are collected, used, consulted or otherwise processed and to what extent are those data processed by others.

In January 2019, the CNIL fined Google for violating Articles 12 and 13 GDPR Article through its use of contractual terms which lacked transparency and provided inadequate information to data subjects – thus failing to satisfy the requirements for valid consent.[90] In particular, the CNIL found that *"essential information"* such as the data processing purposes, storage periods and the categories of personal data gathered were *"disseminated across several documents"* such that users were required to make additional investigations to find how their data is being processed in personalising advertisements.[91] The decisions noted the information which was communicated to users was not sufficiently clear to enable consent and criticised the vague and obfuscatory nature of the description and purposes of processing presented to users.

In the context of AdTech the decision is particularly relevant, highlighting that information must be unified and should not be provided through a design which renders it deliberately challenging to build a picture of how and for what purposes individual data is used.

In similar decisions relevant to the AdTech market both a Belgian Court, and France's CNIL[92] have found that Facebook's terms do not make it sufficiently clear that apps and therefore Facebook itself systematically collect personal data when consumers visit third party websites that contain Facebook social plugins even where they do not have a Facebook account.[93] These decisions should have had a chilling effect on such activities by Facebook, and indeed other AdTech actors, however, this does not appear to have been the case.[94] Indeed, it appears that while Articles 12 and 13 are well intentioned, the requirements for simple, easily understood language, instead of increasing clarity have been used to excuse the deployment of overly simplified terms which offer a false reassurance to consumers and

90 Commission Nationale de l'informatique et des libertés, 'Déliberation SAN-2019-001 du 21 janvier 2019' https://www.cnil.fr/sites/default/files/atoms/files/san-2019-001_21-01-2019.pdf accessed 5 March 2019.

91 Ibid.

92 Commission Nationale de l'informatique et des libertés, 'The Restricted Committee of the CNIL imposed a sanction of 150,000 € against Facebook Inc and Facebook Ireland' https://www.cnil.fr/en/facebook-sanctioned-several-breaches-french-data-protection-act accessed 5 March 2019; The 16th of February, the court of First Instance rendered its judgment in the proceedings on the merits in the case of the Authority v Facebook' https://www.dataprotectionauthority.be/news/victory-privacy-commission-facebook-proceeding accessed 5 March 2019.

93 Ibid; See also Case C-210/16, [28]-[29].

94 See, Valerie Verdoodt Brendan Van Alsenoy, Rob Heyman, Jef Ausloos, Ellen Wauters and Günes Acar, From social media service to advertising network, 2015 https://www.law.kuleuven.be/citip/en/facebook-1/facebooks-revised-policies-and-terms-v1-2.pdf.

obfuscate the true nature and extent of the dissemination of person data through the AdTech ecosystem.

4.5 Automated Decision-making under Article 22 GDPR

According to Article 22 GDPR explicit consent is required where solely automated decisions are made relating to individuals. Specifically Article 22 requires that subjects shall have the right not to be subject to a decision based solely on automated processing including profiling which produces legal effects concerning him or similarly significantly affects him or her though this does not apply under Article 22(2) where same is necessary for entry or performance of contract or based on explicit consent.[95]

Though Article 22 has not been considered by the CJEU, nor was its precursor Article 15 of the Data Protection Directive, the Article 29 Working Party has identified occasions where behavioural advertising within the AdTech market may have significant effects for the purpose of Article 22 of the GDPR, specifically where consumers are targeted with potentially damaging goods or services, such as gambling or high interest loans.[96] More concerning, is the practical reality that individuals are grouped according to imputed characteristics as part of the analysis of data and the online bidding process in a way that may constitute profiling under Article 22. Underlying these concerns, is the fact that it is not clear that actors in the AdTech system obtain the valid, explicit consent necessary for processing under Article 22, in particular in light of decisions such as *Vectaury*.

4.6 The e-Privacy Directive

In addition to the GDPR, the e-Privacy Directive (ePD), as noted in section three above, operates a particular regulatory regime applicable to the technological mechanisms which enable the AdTech market. The Directive requires in Article 5 that cookies can be set only where the consumer has been 'supplied with clear and comprehensive information' concerning the purposes of the processing and is offered the right to refuse such processing by the data controller. In practice however, this 'informed opt out' has provided little additional protection to individuals with many websites actively employing interfaces that are hostile to consumer choice, or simply blocking consumers from accessing the site or service unless the default cookie settings are accepted.[97]

The ePD Recitals were revised in 2009 to provide that users could opt in through default web browser settings.[98] This was not uncontroversial, the Article 29 Working Party noted that in 2010 three of the four major browsers had default settings which permitted cookies and that user failures to alter such settings could not be interpreted as amounting to consent[99] a suggestion which the Working Party reiterated in 2013.[100]

The CNIL's 2019 decision against Google considered above in the context of legitimate interests was also concerned with default permissions, as was the CJEU's rejection of pre-ticked boxes in its decision in *Planet49*. That latter decision is particularly relevant to the AdTech market as a result of the Court's decision that cookie data under Article 5 need not be personal in order to be covered by the Directive but rather acts to protect the users' broader 'private sphere' in the words of the Advocate General. The decision also noted the need for explicit and transparent information for consumers on the duration of cookies and whether the information they collected was available to third parties.

A reformed e-Privacy Regulation (ePR) was due to enter into force alongside the GDPR, however, as of writing the text has not been finalised. In the drafting process, however, several points of necessary reform, and controversy have emerged which would affect the AdTech industry.[101] The first is the concern highlighted by the EDPS at an early stage, that the Regulation should not permit the processing of metadata under the 'legitimate interest' ground.[102] While the understanding of consent adopted in the Regulation will be required to be equivalent to that afforded under the GDPR (a requirement pre-empted by the Court in *Planet49*) there remained concern that to allow such processing of metadata without consent would dilute existing standards of protection by permitting an over-broad opt out from consent requirements.[103] Instead, the EDPS has opined that such data should be processed only with consent or if technically necessary for a service requested by the user and only for the duration necessary for that purpose.[104]

The second concern, also flagged by the EDPS is the strengthening of Article 10 by requiring privacy protective settings by default which genuinely support expressing and withdrawing consent in a simple, binding and enforceable manner against all parties. This would also require the inclusion of Recital 24 as a substantive provision in the form of a legal requirement such that end users would be afforded the opportunity "to change their privacy settings at any time during use and allow the user to make exceptions, to whitelist websites or to specify for which websites (third) party cookies are always or never allowed."[105]

It is unclear from the draft released in November 2019 whether these concerns will be reflected in the final text. In particular, Article 10 which, in previous versions sought to provide notification and reminder requirements regarding the placement of third party cookies has been deleted in its entirety.[106] While Article 8 (and the related

95 See, also Recital 72 GDPR.
96 Article 29 Working Party, Guidelines on Consent under Regulation 2016/679, 2017), 9.
97 As a result of this concern Acquisti has emphasised the need for a contextual understanding of privacy as part of which the default settings for privacy used by companies are tools used to affect information disclosure and attempt to contextualise privacy in a manner which orientates the status quo toward their contractual practices as part of a malicious interface design through which designers and use features that frustrate or confuse users into disclose information is also widely deployed. See, Laura Brandimarte and George Loewenstein Alesssandro Acquisti, 'Privacy and Human Behaviour in the Information Age' in Jules Polonestsky and Omer Tene & Evan Selinger (eds), *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018)., 187; Ralph Gross and Alessandro Acquisti, 'Information revelation and privacy in online social networks' (2005) *WPES Proceedings of the 2005 ACM workshop on Privacy in the electronic society* 71.
98 E-Privacy Directive, Recital 66.
99 Article 29 Working Party, Guidelines on Consent under Regulation 2016/679.
100 Article 29 Working Party, 'Working Document 02/2013 providing guidance on obtaining consent for cookies' (2013). Exceptions to these requirements are provided in accordance with Article 5 and Recital 25 for technical storage and access cookies and cookies which are 'strictly necessary' to provide an information society service explicitly requested by the subscriber.
101 Formal Complaint by Dr Ryan regarding IAB Europe AISBL website, 2nd April 2019 available at https://regmedia.co.uk/2019/04/02/brave_ryan_iab_complaint.pdf accessed 21 April 2019.
102 European Data Protection Supervisor, Opinion 6/2017 on the Proposal for a Regulation on Privacy and Electronic Communications, 2017), 27.
103 Ibid
104 European Data Protection Supervisor, EDPS Recommendations on Specific Aspects of the Proposed ePrivacy Regulation, 2017), 2.
105 Ibid, 2-3.
106 See, Council of the European Union, 'Proposal for a Regulation of the

Recital 20) which considers consent for cookies remains under consideration[107] the most recent draft has deleted the final sentence of Recital 20 which previously read "Access to specific website content may still be made conditional on the consent to the storage of a cookie or similar identifier."[108] The Recital now provides that monitoring of end user devices should be allowed "only with the end-user's consent and or for specific and transparent purposes" because such monitoring may reveal personal data including political and social characteristics which require "enhanced privacy protection." [109]

This view of 'cookie walls' and similar mechanisms as impermissible is in keeping with the current interpretation of the GDPR by academics[110] and more recently by the Dutch data protection regulator. In a recent decision from the Netherlands the Dutch data protection regulator found that refusing users access to websites unless they consent to cookies was impermissible under the GDPR.[111] That decision, and indeed the content of Recital 20, echo the concerns flagged by the decision in *Vectaury* that special categories of data as classified under Article 9 GDPR are discoverable through the aggregation and analysis of the data collected by cookies.

While the language of the proposed e-Privacy Regulation may thus seem strong, in reality it would achieve little more than a reproduction, albeit in explicit language, of the controls already imposed by the ePD and the GDPR.

4.7 Conclusion

It is clear, that at present there are concrete basis under both the GDPR and ePD on which to ground objections to the operation of the AdTech market. However, the impact of these basis, as well as the decisions in cases like *Planet49* and *Vectaury*, is diminished by the realities of the digital market. That such business models have perpetuated online despite these laws is indicative of a lack of effective enforcement. While it now appears that this shortcoming of enforcement is being ameliorated at a national level by more active regulatory engagement, more fundamental concerns remain.

In particular, as a practical matter for consumers there remains no functional choice for consumers to engage with providers of goods and services who *do not* employ surveillance mechanisms which operate as part of the AdTech market. At present the GDPR and ePD can only impose information requirements and consent thresholds. Neither documents, nor the policies of the Union more broadly, acknowledge that absent a market which also offers goods and services whose provision is not attendant on consenting to such collection and sale of personal data, even the most explicit and informed consent is normatively vacuous as it is given in a context in which no meaningful alternative is present.

This failure goes to the heart of the disconnect between the rights to data protection and privacy in Union law. While data protection is currently conceived of as a right, functionally it operates as the condition under which the infringement of a private right is legally permissible. As such, it is the market-oriented manifestation of privacy, imposing the threshold conditions under, and extents to which, privacy can be forfeited by individuals as the condition for market access and participation.

As such, data protection is commercially, and indeed personally, necessary. However, in the current schema of rights protection within the Union it has taken on an outsize importance to this role, effectively dwarfing the right to privacy which it is intended to enable. More fundamentally, if we consider compliance with data protection requirements as the necessary conditions for legally justified infringements of privacy, the analysis above illustrates that such conditions are being systemically violated by the AdTech market at present such that even this minimal understanding of privacy is not satisfied. The result, as the next section examines, is a perpetuation of a legal context in which privacy rights are ailing, importing consequences for individual autonomy, and the Rule of Law.

5. The Impacts of AdTech on Privacy & Autonomy

The most fundamental harm which results from the consumer surveillance on which AdTech relies is the reduction of individual privacy. It would be remiss to insist this is purely a result of the AdTech landscape, the harm has also been facilitated to no small extent by the Union's curtailing of privacy in operational terms as a result of its preference for a shallow, and market-oriented understanding of data protection as a sufficient privacy protecting mechanism and its failure to systemically analyse the compliance of the AdTech market with even those mechanisms.

By allowing the compilation of large data sets from which layered profiles of individuals' actual and inferred preferences, characteristics and activities can be assembled, AdTech allows the revelation of intimate and detailed portraits of individuals. This, in itself, is harmful in as much as the fundamental right to privacy in EU law propounded by both the CJEU and ECtHR emphasises the right as crucially linked to the development of personal identity.[112]

Where privacy is infringed, individuals' capacity for personal identity development is thus jeopardised by forcing conditions in which individuals are unable, or do not feel able to make choices which accurately or meaningfully reflect their preferences in furtherance of their personal development. This threat is compounded in the context of AdTech which actively seeks to utilise coercive and manipulative tactics to influence consumer attention and preferences, in circumstances where the means of avoiding such tactics are not present. In that context individuals experience proportionate reductions in their capacity to choose without external influences but also experience chilling effects to their exercise of uninhibited choice or action resulting in the active diminution of individual autonomy.

Autonomy in this context can be understood as mirroring the concept articulated by Raz, of 'people controlling, to some degree, their own destiny, fashioning it through successive decision throughout their

European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)' (2018).

107 Ibid, 3.

108 Ibid, Recital 20.

109 Ibid, Recital 20.

110 Sanne Kruikemeier Frederik J Zuiderveen Borgesius, Sophie C Boerman and Natali Helberger, 'Tracking Walls, Take it or leave it Choices, the GDPR and the ePrivacy Regulation' (2017) 3 *European Data Protection Law Review* 353.

111 Autoriteit Perssonsgegevens, Websites must remain accessible when users refuse tracking cookies, https://autoriteitpersoonsgegevens.nl/nl/nieuws/websites-moeten-toegankelijk-blijven-bij-weigeren-tracking-cookies, 2019.

112 Case C-208/09 *Sayn Wittgenstein* EU:C:2010:806, [52]; Case C-391/09 *Malgozata Runevic-Vardyn* EU:C:2011:291, [66]. In the ECtHR see, X v Iceland App No. 6825/74 (1976); *Gaskin v UK* App No. 10454/83(1989). The State's refusal to provide the applicant access to records it held regarding his time in care was a violation of Article 8; *Ciubotaru v Moldova* App No. 27138/04 (2010); Odievre App No. 42326/98 (2003); *Karassev v Finland* App No. 31414/96 (1996); *Stjerna v Finland* App No. 18131/91 (1994).

lives.'[113] Autonomy thus requires the presence of an adequate range of morally acceptable options to choose among. In particular, Raz notes that choice between bad options may not constitute choice sufficient to facilitate autonomy at all.[114] In the context of the digital market this functional choice between viable alternatives is not present. AdTech is not only ubiquitous but systemically engrained and thus impossible to avoid, the only alternative to engaging with it being to forfeit online activity entirely. The dilemma for, and risk to, autonomy thus crystallises in this Razian articulation of the conditions for autonomy.

This understanding of autonomy also requires the presence of meaningful choice free from manipulation, coercion or excessive undue influence[115] and understands autonomy as the capacity for socially situated individuals to make choices which result from deliberative action. Once again in the digital market, AdTech obstructs such freedom, actively seeking to influence the attention and choices of consumers through its collection and analysis of data and its deployment of behavioural advertising practices.

Significantly, Raz's conception of autonomy also presupposes a concept of alienation. When Raz defines autonomy as the capacity to be the author of one's own life – to give it a shape and meaning (an articulation which accords with the understanding of privacy in EU law) - he is not only claiming that the autonomous individual must independently and actively shape her life. In addition, she must presuppose that something matters in her life. Determining oneself then must mean determining oneself *as* something.[116]

Where the capacity to exercise autonomy is hampered, individuals are unable to establish a relation to other individuals, to things, to social institutions and thereby to themselves – they are, in other words, unable to establish themselves *as* something. This inability, referred to as alienation, prevents individuals from distilling meaning from their existence.[117] The commodification of goods and domains that were previously not objects of market exchange is a common historical example of this kind of alienation. AdTech, through its obstruction of individual attempts to relate to those goods or areas which individuals use to define their selves and through its attempts to condition the preferences and thus relations of individuals to other actors actively diminishes individual autonomy and alienates individuals, preventing them from engaging in the development of personality which the right to privacy is explicitly understood as seeking to protect.[118]

Alienation understood in this way is a condition attendant on the reduction of autonomy which itself results in a further loss of individual power - alienated individuals are disempowered, not subject to their own, and vulnerable to the imposition of another's, law.[119] Alienation thus negatively impacts individual liberty, on the basis that it is only when individuals experience and are empowered to experience life as their own, governed by their own choices that they are free.[120] Under this conception autonomy, and the reduction or elimination of alienation, is not merely an individual but is also a social good, acting to ensure the individual development of personality and preference through deliberative choice and to create empowered individuals - pre-conditions central to democratic participation and thus to democratic society.[121]

The European Union's understanding of privacy as fundamentally related to the development of personality, and thus to individual autonomy, recognises that the capacity for individual development diminishes as privacy does.122 Where such restriction of individual self-development occurs, the result is that, at a societal level, individuals are impeded from critical engagement with the processes of democratic self-government due their impaired ability to fulfil their roles as active and engaged citizens. Citizenship, in a European context, is thus understood as more than a status, as a set of social practices whose fulfilment includes voting, public debate, and political opposition which are influenced by institutional mores.[123] The protection of privacy and the promotion of autonomy and individual liberty is thus constitutive of a healthy Rule of Law.

6. AdTech & The Rule of Law

The Rule of Law has been repeatedly proffered as a foundational value of the European project as part of a cluster of ideals constitutive of European political morality, the others being human rights, democracy, and the principles of the free market economy.[124] While neither the Rule of Law nor fundamental rights featured in the Treaty of Rome's text, the Union's constitutional framework has subsequently placed increasing emphasis on both, and affords them a position of centrality in its internal and external policies, featuring them not only as foundational values (identified by the Lisbon Treaty, and later manifested through the Charter and its jurisprudence) but also as central pillars of the Union's external relations.

Article 2 TEU, as the culmination of the Union's commitment to the Rule of Law as an orienting value,[125] links the Rule and fundamental rights to each other alongside the achievement and maintenance of democratic government. The implication of this grouping is an understanding of the three values as interdependent and mutually reinforcing. The Charter's preamble takes a similar stance to Article 2, positioning the Rule of Law as a shared value of the peoples of Europe in the context of fundamental rights, while recent cases linked to ongoing concerns surrounding the Rule of Law in Poland have leant further weight to the suggestion implicit in Article 2's grouping that the theory of the Rule of Law endorsed by the Union is a substantive

113 Joseph Raz, 'Autonomy, toleration and the harm principle' in Susan Mendus (ed), *Justifying toleration: Conceptual and historical perspectives* (Cambridge University Press 1988), 369.
114 Raz (n 113) 372.
115 Bernal (n 7) 24-5.
116 Rahel Jaeggi, *Alienation* (Columbia University Press 2014), 204-5.
117 Jürgen Habermas, *Justification and Application: Remarks on Discourse Ethics* (MIT Press 1993), 48.
118 Jaeggi (n 116) 4-5.
119 Jaeggi (n 116) 22-23.
120 Steven Lukes, *Marxism and Morality* (Oxford University Press 1985), 80.
121 Raz (n 113) 314 'the ruling idea behind the ideal of personal autonomy is that people should make their own lives.'
122 See Rouvroy and Poullet (n 62); *X v Iceland* Application No. 6825/74 (1976); Niemietz v Germany App No. 13710/88 (1992); *Dudgeon v United Kingdom* App No. 7525/76 (1981), [41]; *Klass and Others v Germany* App No. 5029/71 (1978); *Big Brother Watch and Others v The United Kingdom* App No. 58170/13, 62322/14 and 24960/15 (2018); Case C-208/09 *Sayn Wittgenstein* EU:C:2010:806; Case C-391/09 *Malgozata Runevic-Vardyn* EU:C:2011:291, [66].
123 Commission Communication, 'The Commission's Contribution to the Period of Reflection and Beyond: Plan D for Democracy, Dialogue and Debate' COM (2005) 494, 2-3; Andrew Williams, *The Ethos of Europe: Values, Law and Justice in the EU* (Cambridge University Press 2010), 154-156. See also, Ireneusz Pawel Karolewski, *Citizenship and Collective Identity in Europe* (Routledge 2010), 108-112 on the shift from a model of caesarean effective citizenship to one based on a deliberative model; John JH Weiler, 'To be a European Citizen: Eros and Civilisation' in John JH Weiler (ed), *The Constitution of Europe* (Cambridge University Press 1999).
124 Jeremy Waldron, 'The Concept of the Rule of Law' (2008) 43 *Georgia Law Review* 1.
125 It is beyond the scope of this work to engage substantively with the possible implications of this change in language for the justiciability of the values.

one primarily oriented toward to the promotion of individual liberty through democratic government.[126] This understanding of the Rule of Law's practical function is particularly in the Union's commitment to a substantively enforced standard for its Member States in respecting the Rule of Law at a national level, in Article 7 TEU.[127] A reading which finds further support in the constitutional and administrative principles which underpin the EU legal order.[128]

The difficulty with any substantive theory of the Rule of Law is, of course, that its boundaries are difficult to draw, not least as a result of the ambiguous standing of fundamental rights within the Union. Ultimately, the Rule of Law has traditionally functioned to ensure individual liberty, and those fundamental rights which are necessary in enforcing and protecting such liberty must necessarily form part of a substantive theory however widely or narrowly drawn such a theory otherwise is. This is particularly so in the Union, which has explicitly grouped the preservation of democratic order, and fundamental rights alongside the Rule of Law as an orienting principle.

Liberty itself is a porous notion,[129] however, Tamanaha's four concepts of liberty provide a utile framework for assessing the coetaneous nature of liberty and the Rule of Law. Tamanaha posits a layered idea of liberty composed of:

- Political liberty, effected through democratic participation and government[130] and which accords with modern understandings of representative democracy as recognised by Article 2 TEU and enforced by Article 7 TEU,
- Legal liberty which provides that the State act only in accordance with law and in accordance with ideas of legal predictability and equality and which finds expression in the requirement that restrictions on fundamental rights be provided 'by law,'[131]
- Individual liberty which subsists where the government is restricted from infringing upon an inviolable realm of personal autonomy and which finds expression to some extent, the Treaties which seek to delimit the bounds of individual rights and the conditions for State intrusion upon the areas or activities which they protect,[132] and
- Institutional liberty, which holds that individual and therefore societal liberty is enhanced when the powers of government are compartmentalised thus preventing an accumulation to power in a single institution.[133]

In examining Tamanaha's systematisation, it is apparent that the European Union's understanding of the Rule of Law maps onto all four of the distinct categories of liberty identified. Moreover, the categories of liberty identified are coetaneous with the Rule of Law in as much as they seek to ensure an adequately restrained government which adheres to democratic principles of institutional balance, and the equal and predictable application of laws. The result is the creation of a layered, constitutional conception of the Rule of Law not only as seeking the ultimate goal of ensuring individual liberty but also of being fundamentally constitutive of liberty in a broader, political, context.

In accordance with this conceptualisation of the Rule of Law the right to privacy must form part of a minimum content of the Union's substantive conception of the Rule of Law given its centrality in securing individual autonomy and thus facilitating the individual liberty necessary for democratic participation and thus, liberty more broadly. As such, the privacy harms which AdTech imports ultimately reduce individual autonomy and alienate individuals resulting in a loss of liberty which ultimately diminishes the health of the Rule of Law within the Union.

Yet this is not the only mechanism by which AdTech impacts the Rule of Law. On a more practical level, the Rule of Law is affected by AdTech's capacity to enable the development of mechanisms of constitutional avoidance, which permit State actors to bypass limitations to or exemptions from the protective remit of fundamental rights protections through the use of private actors as their proxies. A high-profile example of this pattern in practice was the 2017 Cambridge Analytica revelations. The information uncovered during that episode should have been of little surprise to anyone familiar with the functioning of the AdTech market. Nevertheless, it offered a useful example of the manner in which AdTech harms the Rule of Law.

In 2017 it was revealed that a Cambridge academic working as a researcher for a private company, Cambridge Analytica (CA), had obtained, from Facebook, a large data set containing information relating to an unknown quantity of the company's users. This data set was the combined by the staff at CA with information from other commercial sources to build a data rich system which could target voters with personalised political advertisements based on their psychological profile.[134] The targeting system was then sold to interested actors and was bought and used by candidates for the Republican presidential nomination in the United States,[135] parties campaigning in the Brexit referendum[136] and parties running political campaigns in other jurisdictions including Brazil, India, Kenya, Nigeria and Mexico.[137]

Using the highly specific and personal data profiles sold by Cambridge Analytica, these campaign teams targeted voters on an individual level, as well as identifying and targeting voter blocks by creating

126 See, Case C-216/18 LM EU:C:2018:586; Case C216/18 *PPU Minister for Justice and Equality* EU:C:2018:586, [48]. See also, *Minister for Justice v Celmer (No 2)* [2018] IEHC 153, (Donnelly J). Subsequent to the reference in LM, Donnelly J in *Minister for Justice v Celmer* (No. 5) [2018] IEHC 639 found that the deficiencies in the independence of the Polish judiciary did not meet the threshold for refusal of surrender.

127 European Commission 'Rule of Law: European Commission acts to defend judicial independence in Poland' 20 December 2017, at http://europa.eu/rapid/press-release_IP-17-5367_en.htm (accessed 27 February 2018).

128 Theodore Konstadinides, *The Rule of Law in the European Union* (Hart 2017), 84 et seq.

129 Isaiah Berlin, *Four Essays on Liberty* (Oxford University Press 1969), 121.

130 Jean Jacques Rousseau, 'The Social Contract' in Sir Ernest Baker (ed), *Social Contract: Essays by Locke, Hume and Rousseau* (Oxford University Press 1960), Book II, 6; Brian Z Tamanaha, *On the Rule of Law: History Politics and Theory* (Cambridge University Press 2004), 34.

131 Sharon R Krause, 'Two Concepts of Liberty in Montesquieu' (2005) 34 *Perspectives on Political Science* 88; Tamanaha (n 130) 34-35.

132 Tamanaha (n 130) 35.

133 Tamanaha (n 130).

134 Carole Cadwalladr, ''I made Steve Bannon's psychological warfare tool': meet the data war whistleblower' (*The Observer*, 17 March 2018) https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump (accessed 16 August 2019).

135 Cadwalladr (n 134).

136 Alex Hearn, 'Cambridge Analytica did work for Leave.EU, emails confirm: Parliamentary committee told work went beyond exploring potential future collaboration' The Guardian 30 July 2019 https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm (accessed 16 August 2019).

137 Tactical Tech Report forthcoming: The Influence Industry: The Global Business of Using Your Data in Elections quoted in Julianne Kerr Morrison and Ravi Naik Stephanie Hankey, *Data and Democracy in the Digital Age*, 2018).

tailored messaging and content based on the personal information harvested through the AdTech market.[138] The analytics and aggregation practices used are standard industry practice in online advertising, and rely on the contractual mechanisms and current regulatory approaches which this article has examined.

While this in itself is harmful in a commercial setting, a particular threat to the Rule of Law occurs when public actors capitalise on the AdTech market's capacity to influence individuals to covertly leverage public opinion and influence political choice in a manner they would be constitutionally restrained from doing should they seek to collect and analyse data in a similar way themselves. In this context, privacy rights contribute to the Rule of Law, and seek to ensure a democratic governance, by limiting state intervention with and surveillance of citizens through private proxies.

Privacy thus reinforces the barriers between the individual and the State within the contours of civil society and on that basis is one of the strengths of the democratic model - functioning, in Westin's account, to ensure the 'strong citadels' of autonomous action and personal development which are a prerequisite for liberal democratic society.[139]

The role of online data gathering in facilitating democratic harms was, belatedly, acknowledged following the Cambridge Analytica investigation by the UK Parliament, which revealed that company and indeed Facebook itself, had targeted individuals[140] in a manner which interfered with democratic elections. However, the contributory role of privacy in militating against such data gathering within the AdTech market, and thus against democratic undercutting has yet to be explicitly recognised.

7. Conclusion

The right to privacy in the European Union is premised on an understanding of privacy as enabling the development of individual personality, and as fundamentally linked to the achievement of individual autonomy and liberty. However, this foundation has been obscured by the lack of operational force enjoyed by the right, and the legislative elevation of data protection to the exclusion of more fundamental privacy concerns.

In this context the operation of the AdTech has operated with a significant degree of freedom. While decisions such as *Vectaury* and Planet49 indicate a hardening of attitudes towards the notification and consent thresholds necessary for the data collection practices AdTech, there has not been, as yet any consideration of the need for stricter regulation of the AdTech market or the practices it operates in light of the privacy harms which its operation facilitates.

Most concerningly, there seems little awareness of the crucial nature of such reform given the right to privacy's function in securing democratic participation and as part of a substantive conception of the Rule of Law. In this respect, AdTech is more systemically problematic than is currently acknowledged, importing layered harms at an individual, and societal level. Acknowledging these impacts is the first step toward creating a sustainable online ecosystem which contributes to rather than conflicting with the attainment of individual autonomy and social goods.

138 Information Commissioner's Office, *Investigation into the use of data analytics in political campaigns*, 2018).

139 Alan Westin, *Privacy and Freedom* (Atheneum 1967), 24.

140 ICO (n 138); In March 2019 the EU adopted new rules to "prevent misuse of personal data by European political parties." The move came ahead of the European Parliament elections, which took place across the continent in May 2019, Council of the European Union, EU set to adopt new rules to prevent misuse of personal data in EP elections (2019).

03

Special issue: Governing Data as a Resource

Technology and Regulation

Paving the Way Forward for Data Governance: a Story of Checks and Balances

Editorial

Inge Graef*

data access, data portability, data protection, competition, innovation, intellectual property

i.graef@tilburguniversity.edu

Data governance is a phenomenon that brings many interests and considerations together. This editorial argues that active involvement of various stakeholders is vital to advance discussions about how to create value from data as a means to stimulate societal progress. Without adequate checks and balances, each stakeholder group on its own will not have sufficient incentives to do its utmost to achieve this common goal. Policymakers and regulators need to be stimulated to look beyond short-term results to ensure that the design of their initiatives is fit for purpose. Industry players have to be transparent about their practices to prevent strategic behaviour that may harm society. And researchers must inform their findings with real-world evidence and proper terminology.

1. Setting the scene

It is striking how little is known about effective governance structures for data considering the intensity of discussions about the importance of data as a currency, input or asset.[1] From a legal perspective, debates are still dominated by data protection law – a regime mostly motivated by the need to offer protection against the risks that the processing of personal data entails for the privacy of individuals. Apart from risks, the use of data can provide enormous benefits to consumers, businesses as well as society at large. Data forms a basis for innovative products and services, enables businesses to make their production processes more efficient and can boost economic growth as well as serve societal interests such as through personalized healthcare and improved energy efficiency.[2] Although data protection (at least in the context of EU law) also pursues an objective of market integration and thereby stimulates the free flow of personal data, other considerations beyond data protection need to be taken into account as well in order to design adequate governance models for data. This is not a straightforward exercise, because there is a myriad of legal, economic, technical, and social interests to be reconciled.

Questions about how to regulate data become increasingly complex as datasets typically consist of several types of information (personal, non-personal, machine-generated, organizational, public sector information) over which multiple parties hold overlapping entitlements (data protection and consumer rights of individuals, intellectual property rights of firms as well as confidentiality obligations between parties). The coexistence of such entitlements raises conceptual questions about how various forms of control over data (legal, contractual, technical) can be exercised in parallel and what governance structures should be designed to fully exploit the potential of data across the economy.[3] Because legislators are now preparing to take concrete measures to stimulate data-driven innovation,[4] this is a crucial moment to inform policymaking in the area.

Such a discussion connecting different strands of thought surrounding data governance was the objective of a workshop held at Tilburg University in November 2019 that I co-organized with Martin

1 Beyond the many policy documents published by EU institutions as referenced throughout this editorial, see for instance World Economic Forum, 'Data-Driven Development: Pathways for Progress', January 2015, available at http://reports.weforum.org/data-driven-development/ and OECD, 'Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies', November 2019, available at https://www.oecd-ilibrary.org/science-and-technology/enhancing-access-to-and-sharing-of-data_276aaca8-en

2 The European Commission has argued that "Data-driven innovation will bring enormous benefits for citizens, for example through improved personalised medicine, new mobility and through its contribution to the European Green Deal" and has expressed its intention of creating an attractive policy environment for data "so that, by 2030, the EU's share of the data economy – data stored, processed and put to valuable use in Europe – at least corresponds to its economic weight". See Commission Communication 'A European strategy for data', COM(2020) 66 final, 19 February 2020, 1, 4.

3 The question of how to deal with parallel entitlements to the same data formed the core of the research project 'Conceptualizing Shared Control Over Data' as funded by Microsoft. Next to the workshop on which this editorial reports, the project also involved a call for Microsoft fellows in 2019. Selected candidates obtained funding to visit TILT and join the project for a number of months.

4 In its February 2020 Communication 'A European strategy for data', the European Commission announced its intention to adopt a proposal for a 'Data Act' by 2021, as discussed below in section 4.

* Inge Graef is assistant professor at Tilburg University, with affiliations to the Tilburg Institute for Law, Technology, and Society (TILT) and the Tilburg Law and Economics Center (TILEC). The workshop on which this editorial reports took place in the framework of the research project 'Conceptualizing Shared Control Over Data' that received funding from Microsoft. I would like to thank Martin Husovec for his comments on an earlier version of this piece.

Husovec. The workshop was sponsored by Microsoft within the framework of the research project 'Conceptualizing Shared Control over Data' and brought together scholars from across the globe to reflect on the governance of data from their own expertise in areas such as intellectual property, open data, data protection, data ethics and competition. This special issue entitled '**Governing Data as a Resource**' is the result of that workshop and collects four of the papers presented.

This editorial brings together some of the insights of the workshop and sets out ideas to move the debate around data governance forward.

Data governance is understood broadly here as referring not only to how to set up practical tools or mechanisms for using data, but also including legislative and regulatory actions to enhance the creation of value from data, for instance by facilitating data access and data portability. As key message, this editorial puts forward the claim that active involvement of all stakeholders is needed as a system of checks and balances in order to achieve outcomes that strike a proper balance between the various interests. In what follows, a number of lessons for data governance is discussed from the perspective of the checks and balances relevant for three groups of stakeholders, namely policymakers and regulators, industry players, and researchers. All four contributions to this special issue relate to one of these three angles, so that each of the papers is introduced in the relevant part of this editorial.

2. Policymakers and regulators

Policymakers and regulators are in the front seat of steering the development of data governance in directions that meet societal needs. Because of their commercial interests, the incentives of market players are typically not fully aligned with achieving broader policy goals. Some market players may not want to share data, even when this is societally desirable, due to fear of losing a competitive advantage. Others may be afraid of liability for sharing data in violation of, for instance, data protection rules and be reluctant to share data in the absence of clearer guidance. Policymakers and regulators thus have a key role in facilitating the creation of adequate governance structures for data. The European Commission launched its European data economy initiative in 2017,[5] which in February 2020 culminated in the publication of a Commission Communication 'A European strategy for data' containing a range of specific actions "to enable the EU to become the most attractive, most secure and most dynamic data-agile economy in the world".[6] Such policy and legislative initiatives have to be implementable in practice and be fit for purpose. Input from industry and academia (but also from other actors such as consumer organizations)[7] is therefore vital to guide regulatory actions to prevent that, for instance due to political pressure and a focus on short-term results, suboptimal approaches are taken to achieve a particular policy objective.

How to establish community-based governance of a shared resource is central to the framework of knowledge commons, which **Michael Madison** applies to data in this special issue's opening paper **"Tools for Data Governance"**.[8] His analysis is informed by a distinction between data-as-form, treating data as a fixed object, and data-as-flow, looking at data's fluid attributes and numerous applications. When applying the framework of knowledge commons to data, the author provides two essential tools to develop governance strategies for data focusing on the concepts of groups and things. The first one consists of the identification of relevant social groups in which governance frameworks may be embedded and the second one concerns the identification of relevant resources or things that will contribute to the welfare effects of the data governance system.

The multi-faceted nature of data is also reflected in the various regulatory actions being considered at the EU level to govern data. The European Commission emphasizes the need for sector-specific approaches because of the differences across industries and at the same time aims to create a 'single' or 'common' European data space where data can flow across sectors.[9] There does not seem to be one overarching policy objective behind the Commission's European data strategy. The Commission's February 2020 Communication refers to the existence of market failures as a trigger to adopt data access rights that would make the sharing of data compulsory in specific circumstances.[10] Data-driven innovation is mentioned various times as a notion the Commission wants to support.[11] Reference is made as well to the need for empowering individuals and to more sector-specific goals such as better healthcare, competitiveness in agriculture and tackling climate change.[12] Identifying the underlying objective of policy action is vital, because the objective acts as the benchmark against which to assess the costs and benefits of additional measures and forms the determining factor for how to design new regulatory interventions.

An example explored in my own co-authored work[13] where one can doubt whether legislative design choices are capable of achieving the goal of a single market for data is the artificial distinction between personal and non-personal data made by the EU legislator in the Regulation on the free flow of non-personal data[14] and by the European Commission as policymaker in the context of the European data strategy.[15] Current initiatives to stimulate the European data economy focus on non-personal data in order to complement data protection rules that regulate the processing of personal data. However, because datasets are often mixed, it seems almost practically impossible to maintain two separate legal frameworks.[16]

An underlying assumption of this regulatory choice seems to be that non-personal data is more essential as innovation input than personal data. Statements in the Commission Communication 'A European Strategy for Data' from February 2020 give the impression

5 Commission Communication 'Building a European Data Economy', COM(2017) 9 final, 10 January 2017.
6 European Commission (n 2) 25.
7 See the public consultation that the Commission started in February 2020 to allow stakeholders to comment on the European strategy for data.
8 Michael Madison, 'Tools for Data Governance' (2020) *Technology & Regulation* 29-43.
9 Commission Communication 'Towards a common European data space', COM(2018) 232 final, 25 April 2018 and Commission (n 2) 4-5, 26-34.
10 European Commission (n 2) 13 and footnote 39.
11 European Commission (n 2) 1, 8 15, 16.
12 European Commission (n 2) 10, 20, 22.
13 Inge Graef, Raphaël Gellert & Martin Husovec, 'Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation' (2019) 44 *European Law Review* 605. For a law and computer science perspective, see Michèle Finck & Frank Pallas, 'They who must not be Identified – Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10 *International Data Privacy Law* 11-35.
14 Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L303/59.
15 European Commission (n 2). The notion of non-personal data already came up in the Commission Communication 'Building a European Data Economy', COM(2017) 9 final, 10 January 2017.
16 Graef, Gellert & Husovec (n 13) 608-610.

that initiatives to stimulate data innovation will focus on the sharing of non-personal data. With regard to personal data, the Commission emphasizes the importance of complying with data protection law by stating: "Citizens will trust and embrace data-driven innovations only if they are confident that any personal data sharing in the EU will be subject to full compliance with the EU's strict data protection rules".[17] For non-personal data, the Commission instead stresses its role as "potential source of growth and innovation" by arguing that making non-personal data "available to all – whether public or private, big or small, start-up or giant [...] will help society to get the most out of innovation and competition and ensure that everyone benefits from a digital dividend".[18] However, there is no evidence that non-personal data is more valuable as innovation input than personal data. The two categories of data can hardly be separated in practice and personal data may sometimes even have more value due to its potential to predict new overall trends as well as individual preferences.[19]

When the design of legislative measures or policy initiatives is not properly aligned with their overall objective, there is room for market players to engage in strategic behaviour when deciding how to comply with the law by favouring the interpretation that fits their interests. This concern is further discussed in the next section.

3. Industry players

Industry players play an important role in developing adequate governance structures for data. They will often have more knowledge and insights about the market and available approaches than the other two stakeholder groups discussed in this editorial, namely policymakers and regulators as well as researchers. At the same time, industry players have commercial motives. This implies that they normally have limited incentives to contribute to achieving societal goals on their own initiative, especially when this would go at the expense of their own interests. Pressure to meet the demands of customers and consumers restrains their ability to engage in problematic conduct, as do existing legal regimes ranging from competition, data protection and consumer law to contract, labour and environmental law (and many others). There is a role for researchers as well as policymakers and regulators to keep industry players accountable and to ensure the transparency of industry practices.

An interesting example illustrating the issues data governance can bring about in situations involving multiple competing interests in data is provided by **Teresa Scassa** in her paper **"Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto"**.[20] The paper analyzes the data governance scheme proposed by Sidewalk Labs, a subsidiary of Alphabet (which also owns Google), to develop a smart city project as commissioned by Waterfront Toronto, a Canadian non-profit corporation. In explaining how the chosen governance model failed in the situation at hand, the author draws valuable lessons for the future of data governance more generally. Through the lens of the concept of knowledge commons, the paper concludes that it is vital to address data governance issues at the stage of the project design and to involve a diverse range of stakeholders in the conceptualization and implementation of the data governance model representing different interests that can be both public and private.

As a purely industry-led initiative, the Data Transfer project set up by Facebook, Google, Microsoft and Twitter in 2018 and joined by Apple in 2019 is also worthwhile to discuss here. The project aims to create an open-source platform to enable the transfer or portability of data between online services as initiated by users.[21] In December 2019, Facebook announced the release of a tool developed within the Data Transfer project that allows Facebook users to move Facebook photos and videos directly to Google Photos, with the expectation for other services to be connected to the tool later on.[22] While the tool is presented as a gesture to users at the courtesy of Facebook, Article 20 of the General Data Protection Regulation (GDPR)[23] already requires data controllers to provide data subjects with a right to receive and transmit personal data to another provider. However, Article 20 GDPR only entitles data subjects to a right to have personal data directly transferred between controllers (without having to export and import data themselves) "where technically feasible". Facebook's efforts and those of the Data Transfer project more generally are thus to be welcomed as a way to increase the number of situations in which data portability can be technically implemented.

At the same time, the Article 29 Working Party has interpreted the scope of the right to data portability broadly in its guidelines on data portability from April 2017 – which are not legally binding but do have an authoritative status. According to the Article 29 Working Party, personal data for which data portability can be requested does not only include personal data knowingly and actively provided by data subjects, such as a user name, email address or one's age, but also data observed from the activities of users, including activity logs or history of website usage.[24] Photos are uploaded by users and thus certainly fall within the scope of application of the right to data portability, but also observed data such as one's search history and location data would need to be made portable under the interpretation of the Article 29 Working Party. There is thus a need to remain vigilant as to the efforts made by industry players to comply with the law and to keep developing tools to push for new technical possibilities. Again, this requires involvement of different stakeholders to ensure adequate checks and balances.

Interestingly, when announcing the photo transfer tool, Facebook called upon regulators to step in and balance the benefits and risks of enhancing data portability. If a social network user ports his or her data to another provider, that user does not only reveal information about herself but also about her friends and contacts due to the interactive nature of social networking. According to Facebook, the transfer of data through data portability thereby increases the risks of leaks and raises questions about liability. Facebook argues that it is for regulators to make these trade-offs between the desirability of data portability and the greater risks for privacy, and that such decisions cannot be left to private companies.[25] These are indeed valid concerns requiring proactive approaches by regulators to ensure that such trade-offs are made transparent but also to prevent that industry players use risks for data protection or privacy strategically as an excuse to limit data portability.

17 European Commission (n 2) 1.
18 European Commission (n 2) 1.
19 Graef, Gellert & Husovec (n 13) 617.
20 Teresa Scassa, 'Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto', *Technology & Regulation*, (2020) 44-56.
21 See https://datatransferproject.dev/.
22 Steve Satterfield, 'Driving Innovation in Data Portability With a New Photo Transfer Tool', Facebook Newsroom, 2 December 2019, available at https://about.fb.com/news/2019/12/data-portability-photo-transfer-tool/.
23 Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (GDPR) [2016] OJ L119/1.
24 Article 29 Working Party, Guidelines on the right to data portability, 5 April 2017, WP 242 rev.01, 9-10.
25 Matthew Newman, 'Facebook wants EU lawmakers to weigh up data portability's risks and rewards, Clegg says', MLex, 2 December 2019.

Data portability is an important tool to empower individuals to have more control over how their personal data is used. To make sure this objective of empowerment is achieved, adequate implementation by industry players as well as effective enforcement by regulators is key. Advocacy is also important to make individuals aware of their rights, of which the right to data portability is only one. Researchers, to which the attention turns in the next section, can play a role here to make sure such rights do not merely exist in the books but are actually used in practice.[26]

4. Researchers

Researchers play an important role in commenting on industry initiatives, as well as on enforcement actions and legislative proposals from policymakers and regulators. In their turn, researchers have a responsibility to put checks and balances in place in order to inform their findings with adequate evidence, to be transparent about research funding,[27] and to be clear and consistent with regard to terminology. Scholarship in the area of data governance will often bring different disciplines together. For instance, in order to make findings about how to best implement and enforce the GDPR's right to data portability from a legal perspective, it is necessary to have an understanding of the technical requirements of data portability. To study how to apply or develop the law, legal scholars need to make themselves acquainted with industry initiatives as well as the way products and services work from a more technical perspective. A reality-check with the 'law in practice' as well as with insights from other disciplines is necessary to make a proper analysis.

Data governance indeed increases the need for collaboration between disciplines, ranging from computer science, law, economics and other social sciences such as philosophy and ethics. As it is a topic where so many different interests come together, interdisciplinary research will play an important role in moving discussions about data governance forward. To advance scholarship relating to data governance within a discipline, it is also worthwhile to explore what lessons can be drawn from earlier regulatory experiences. Two of the contributions in this special issue take this approach.

By discussing the governance of electricity data and in-vehicle data, **Charlotte Ducuing** draws lessons regarding the limitations of a so-called 'data flow paradigm' in her paper **"Beyond the Data Flow Paradigm: Governing Data Requires to Look Beyond Data"**[28]. She warns that promoting data exchange as a regulatory aim in itself can lead to imprecise and short-sighted policymaking when there is a lack of consideration for sectoral objectives and constraints. One of the observations relevant for further regulatory initiatives is how the emergence of independent data platforms can help to structure data markets by coordinating supply and demand for data. The creation of such an extra layer in the vertical value chain can be compared with the creation of physical infrastructure managers in some liberalized industries.

The final paper of this special issue **"Defining Data Intermediaries"**[29] creates terminological clarity in an effort to move research and policymaking about data sharing forward in a more systematic way. By categorizing different data governance models, **Alina Wernick, Christopher Olk and Max von Grafenstein** analyze the possibilities data intermediaries can offer depending on the needs of market players and individuals. Drawing an analogy with intellectual property, they argue that the concepts of clearinghouses and patent pools are particularly useful to understand the opportunities and limits of data governance but that there is a need to adapt these governance mechanisms to the peculiarities of data.

With regard to adequate approaches towards regulating data more generally, an analogy can be made with the influential paper published by Easterbrook in 1996 on "Cyberspace and the Law of the Horse". According to Easterbrook, there is no need for specialized legal rules to regulate cyberspace just as it would make no sense to create a separate body of law for regulating all activities relating to horses. In his view, "the best way to learn the law applicable to specialized endeavors is to study general rules" and any effort to collect separate sets of rules into a 'Law of the Horse' "is doomed to be shallow and to miss unifying principles".[30] Easterbrook's views still have impact in the field of technology regulation up to the present day in determining how to regulate new technologies that do not neatly fit within the categories of existing legal frameworks.[31] Data is no exception to this. As the contributions in this special issue will show, many different legal regimes apply simultaneously to data. Not all of them pursue similar objectives so that inconsistencies and tensions are inevitable. Such clashes do not only occur at the level of specific rules but also at the level of general principles Easterbrook referred to.

How can one for instance reconcile the need for data protection and the protection of property with the potential of data sharing for innovation purposes? The GDPR requires data controllers to limit the processing of personal data to what is strictly necessary through principles such as purpose limitation and data minimization.[32] Intellectual property law entitles right holders to exclude third parties from using the protected subject matter, which can include data when it qualifies for protection under copyright, sui generis database protection or as a trade secret.[33] While data protection and intellectual property law thus have mechanisms in place to limit the exchange of data, policymakers are at the same time adopting new measures to stimulate reuse and sharing of data (which will inevitably include personal data and intellectual property protected data) in an effort to create more competition and innovation.[34] Such tensions between policy objectives will

26 In its Communication 'A European Strategy for Data' from February 2020, the Commission also emphasizes the need to further support individuals in enforcing their data subject rights and mentions initiatives to enhance the right to data portability by promoting the use of personal data apps and novel data intermediaries such as personal data spaces. See European Commission (n 2) 20.

27 See for instance the Transparency and Disclosure Declaration that the Academic Society for Competition Law (ASCOLA) developed for competition law scholars: https://ascola.org/content/ascola-declaration-ethics. And note the disclosure of funding from Microsoft in the first footnote of this editorial.

28 Charlotte Ducuing, 'Beyond the data flow paradigm: governing data requires to look beyond data' (2020) *Technology and Regulation 57-64.*

29 Alina Wernick, Christopher Olk and Max von Grafenstein, 'Defining Data Intermediaries' (2020) *Technology and Regulation 65-77.*

30 Frank H. Easterbrook, 'Cyberspace and the Law of the Horse' (1996) *University of Chicago Legal Forum 207.*

31 For a discussion, see Ronald Leenes, 'Of Horses and Other Animals of Cyberspace' (2019) *Technology and Regulation 1, 2-3.*

32 Article 5(1)(b) and (c) GDPR

33 For a discussion of intellectual property protection for data, see Josef Drexl, 'Designing Competitive Markets for Industrial Data: Between Propertisation and Access' (2017) 8 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 257, 267-269.

34 Examples of such measures can be found in the payment and energy sectors as well as in the context of the provision of digital content. See respectively, Directive (EU) 2015/2366 of 25 November 2015 on payment services in the internal market [2015] OJ L337/35; Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity [2019] OJ L158/125; Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L136/1. For a discussion of such sector-specific data

need to be reconciled in practice. Trade-offs as to how to comply with different legal regimes that would lead to diverging outcomes are now mainly left to industry players, the risks of which have been discussed in the previous section.

To prevent that this leads to undesirable strategic behaviour, a question worthy of consideration is whether there is a need to overcome the criticism of Easterbrook regarding the 'Law of Horse' and create some sort of 'Law of Data'. Its purpose would be to set out more concretely how the general principles underlying existing regimes like data protection, intellectual property and competition law should be applied to questions of data governance, and in particular to situations where tensions occur between requirements of separate legal regimes. Additional (sector-specific) measures creating new rights or duties for data access and data portability risk fragmenting the legal landscape even more because of the increasing uncertainty as to how new regimes should be interpreted in light of rules coming from existing frameworks.

An example is how the GDPR's right to data portability of data subjects interacts with the intellectual property rights held by data controllers. Are data controllers obliged to facilitate portability requests for personal data over which they hold intellectual property claims? And if yes, does this also imply that new controllers should be able to reuse this data free of charge without having to obtain a license from the original intellectual property rights holder?[35] The Article 29 Working Party clarified that intellectual property and trade secrets should be considered before answering a data portability request but that "the result of those considerations should not be a refusal to provide all information to the data subject".[36] The Article 29 Working Party suggests data controllers to see if they can transmit the personal data provided by data subjects in a form that does not release information covered by trade secrets or intellectual property rights but does not specify what should happen if this is not possible.[37]

The answers to such questions are too important to be left to ad-hoc solutions. The 'Data Act' that the European Commission announced it intends to adopt in 2021 may provide a necessary overarching framework for the regulation of data by creating clarity about how new(er) mechanisms to promote data access and data portability interact with the existing regimes of general application.[38] As data is affecting all sectors of activity[39] and is becoming relevant for so many different areas of law, there may indeed be a need to set out at a more general level how to prioritize different interests and considerations within an ever-more complex society driven by data.

5. Future steps

There are many unanswered questions about what are the most effective approaches to govern data. To determine the way forward, this editorial has illustrated that continuous interactions between the three groups of stakeholders are necessary to create new insights and learn what mode of governance works best in a given set of circumstances. Policymakers and regulators, industry players as well as researchers each carry their own responsibility in advancing our current knowledge but also have to keep checks and balances in place towards actions of their counterparts. Only with active involvement of all stakeholders will outcomes be achieved that are as optimal as possible.

The four peer-reviewed papers in this special issue aim to contribute to discussions about data governance from a mainly legal perspective by mapping the current thinking around adequate governance approaches for data and by setting out directions to be explored in future work. We hope that this special issue will stimulate further debates, initiatives and research to help move the debate forward.

access regimes, see Inge Graef, Martin Husovec & Jasper van den Boom, 'Spill-overs in data governance: Uncovering the uneasy relationship between the GDPR's right to data portability and EU sector-specific data access regimes' (2020) 9 *Journal of European Consumer and Market Law 3.*

35 For an analysis, see Inge Graef, Martin Husovec & Nadezhda Purtova, 'Data portability and data control: Lessons for an emerging concept in EU law' (2018) 19 *German Law Journal 1359, 1375-1386.*

36 Article 29 Working Party (n 24) 12.

37 Article 29 Working Party (n 24) 12.

38 The Commission intends to cover many different issues in its proposal for a Data Act, such as business-to-government data sharing, business-to-business data sharing, an evaluation of the intellectual property framework to further promote data access and use, a clarification on the compliance of data sharing arrangements with competition law, enhancing the right to data portability for individuals and the creation of usage rights on co-generated industrial data. See European Commission (n 2) 13, 14, 15, 20, 21, 26.

39 In the context of data protection, Purtova has claimed that that with advances in data analytics any information is becoming personal data thereby turning data protection into the 'law of everything'. See Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection' (2018) 10 *Law, Innovation and Technology* 40.

04

Special issue: Governing Data as a Resource

Technology and Regulation

Tools for Data Governance

Michael J. Madison*

governance, institutions, data, commons, knowledge commons, groups, things, objects, resources, communities, collectives

madison@pitt.edu

This article describes the challenges of data governance in terms of the broader framework of knowledge commons governance, an institutional approach to governing shared knowledge, information, and data resources. Knowledge commons governance highlights the potential for effective community- and collective-based governance of knowledge resources. The article focuses on key concepts within the knowledge commons framework rather than on specific law and public policy questions, directing the attention of researchers and policymakers to critical inquiry regarding relevant social groups and relevant data "things." Both concepts are key tools for effective data governance.

1. Introduction

Law offers no single or simple answer to the problems and opportunities afforded by data. For data scientists, commercial entities, and policymakers which may ask, "how should data be generated, or stored, or transferred, or used?," this article offers a short set of basic tools to use in developing suitable possibilities for governance and ethical practice. This is neither a detailed list of prescriptions nor an inventory or checklist of remedies for current controversies. Instead, the article offers two essential tools for imagining how to advance effective data governance. One consists of identifying and describing relevant social groups in which governance frameworks may be embedded. Two consists of identifying and describing relevant resources, or things, whose form and flow will contribute substantially to the welfare effects of the relevant data governance systems.

In part the aim of the article is to broaden relevant perspectives. Preparing the article followed a prompt to consider governing and regulating "data markets" relative to innovation, growth, and societal progress. That premise risks cutting off the inquiry prematurely. Markets, including regulated markets, are often too simplistic as descriptions of relevant problems or solutions, given what is almost self-evidently a complex challenge. State or government control or supply, as the usual alternatives to market regulation of problematic social phenomena, are likewise often too simplistic. Understanding data requires a broader view, adding the concept of commons governance to these two, in which "commons" embraces data sharing in some collectively managed or governed context. Data are almost always significant or valuable because they are shared.

In part the aim of the article is to provide a basic toolkit that is not tethered to immediate needs and that is adaptable and evolutionary in appropriate ways, as data governance questions challenge us to extend our imaginations. Some of this challenge is old. Along with researchers and industry, regulators and ethicists long ago began to confront the speed, breadth, and scale of the raw computing power now available at comparatively modest expense, so-called Big Data, and the rise of disciplines combined under the title "data science." Law and regulation have grappled with widely-deployed artificial intelligence (AI) systems, which feed on massive supplies of data.

What is new, and what calls for newly-flexible modes of thinking and practicing, is the apparent demise of human comprehensibility at the center of technology design and deployment. Computing speed, scale, and autonomous execution of networked computer systems today operate in ways that effectively embody *the absence of meaningful limits on the humans' capacity to discern patterns in data and to draw inferences from them.*

That concern is linked to virtually every area of human endeavor and more. Data undergirds both the "Internet of Things," material objects and environmental contexts in which networked sensors and actuators are embedded, and the "Internet of Bodies," in which connected devices are attached to or ingested by human beings.[1] The influences of data are seen in a growing number of techno-social systems, from manufacturing to health to politics.[2] One can imagine our data-saturated environment as a three-sided blend of the conceptual contri-

* Michael J. Madison is Professor of Law at the University of Pittsburgh School of Law. This paper was presented at the workshop 'Governing Data as a Resource' organized at Tilburg University in November 2019.

[1] Andrea M Matwyshyn, 'The Internet of Bodies' (2019) 61 *William and Mary Law Review* 77.

[2] Julie E Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019); Brett M Frischmann and Evan Selinger, *Re-Engineering Humanity* (Cambridge University Press 2018).

butions of Claude Shannon as to information theory,[3] Alan Turing as to computability,[4] and Manuel Castells as to flows of power in the network society.[5]

To render these broader issues in more tractable terms, data governance asks more mundane questions: How do we get more data? Better data? More useful data? How do we control or limit data generation, or data distribution? How do we prevent or limit harms associated with data acquisition or retention? How do we increase, improve, or optimize social or economic value associated with data? How do we ensure that data are preserved appropriately, or made available for access appropriately?

Lurking close by are related questions about data governance in the context of specific sectors, industries, and fields. What is the role of data governance relative to personal privacy, employment, finance, national security, public administration, public safety, health and medicine, education, transportation, arts and entertainment, and more?

All the while, in almost all settings, understanding that we are sharing data, almost all of the time.

In sum, data governance must be able to accommodate both the broadest data-related questions asked above and also their context-specific applications. Because of that breadth, this article teases out arguments related to foundational questions of data sharing, rather than responding to the litany of questions just identified as "mundane" or sector-specific.

The article begins with a distinction between data form and data flow. This point is primarily descriptive. It has to do with what we focus on rather than simply on what we find. Both as a technical construct and as a social one, data appear to have a quantum character, in loosely metaphorical terms, meaning that data exhibit multiple and seemingly contradictory attributes. In any governance context, a critical and basic problem is: which attributes matter?

At times, data seem thing-like, a fixed object or objects capable of exclusive ownership and control and subject to regulation as if it were an artifact. That characterization of data-as-form seems most apt when data and datasets are subject to commodification and commercialization efforts.

At other times, and sometimes even at the same time, data seem wave-like, fluid, continuously evolving, even moving, aggregations of information that have power or effect by virtue of their scale or density on an ongoing basis rather than at a single moment. That characterization of data-as-flow seems most apt when data and datasets are parts of research programs and are put to other public uses.

In one sense data appear to be "private goods" and in another sense data appear to be "public goods,"[6] but that distinction can be overstated. Data are not always or necessarily "goods" of any sort.

The initial point is that the aims of data governance and regulation begin with exploring and describing both what data "is" and what data "ought to be," not in ontological terms, but in social terms, framed by data-as-form or data-as-flow. Section 2 expands on this.

That apparently simple distinction is fraught with complexity. Breaking down that complexity is the function of the rest of this article, in Sections 3, 4, and 5 below, describing a governance toolkit.

To render the toolkit comprehensible beyond the corridors and conference rooms of regulators and lawyers, the tools are conceptual rather than doctrinal. A conceptual approach avoids entanglement in disciplinary debates. In both descriptive and prescriptive senses, law has wrestled with the character of its basic approach to questions posed by knowledge and information, including data. One might start with issues of trade and commerce; or intellectual property and monopoly. One might focus instead on equity, autonomy, and dignity. A more integrative view would begin at a higher level of generality, asking whether regulatory challenges pose questions that are fundamentally private, including questions of contract (obligation) and tort, or fundamentally public, including questions of constitutional order and administrative law.

The article steers clear of such classification questions. It does not explore the details of specific legal systems or questions of legal rights and stakeholder interests. Instead, it situates questions of legal rule and governance strategy in the context of two distinctive concerns: what about groups, and what about things? Data-as-form and data-as-flow state two responses to a basic problem that data governance should address. It should address them, as an initial matter, by examining data governance as a species of institutional governance, and specifically knowledge commons governance. Section 3 below addresses that topic in greater detail.

The article likewise avoids undue reliance on the usual "either/or" questions that arise when law meets technology and when law meets information, such as individual rights vs. institutions and organizations, and/or the state. Security and stability vs. innovation and opportunity. Exclusivity vs. openness. And so forth. Those are proper governance concerns, and critically exploring groups and things helps us see how to advance them in specific and systematic ways.

But groups and things do something more. They open pathways into emerging research, scholarship, and (critically) experience that teach about a middle ground, in between markets and states, which is broad, useful, and too often overlooked, though it cannot be a panacea or a perfect solution. That middle ground is *knowledge commons*, which means social groups operating in structured ways relative to shared data.

Care must be taken with the language of commons and with what the language signifies. This is an argument for nuanced *governance* of data as a *shared resource* rather than for any hasty or wholesale abandonment of private interests, markets, or even the state. This is also an argument for an *ecological* and *evolutionary* perspective on data and data governance, a perspective that includes accounts of the roles of different actors, agents, and resources in producing both productive and unproductive outcomes of data-related systems. The word "commons" evokes precisely such a system-level perspective.[7]

The discussion of knowledge commons leads, in Sections 5 and 6, into the article's focus on two critical concepts: social groups, and things. These are high-level but nonetheless fundamental topics when investigating effective institutional governance of shared resources, such as data. And with those concepts, the article offers an introductory guide to fundamental data governance questions for the benefit of policymakers; institution and organization designers, builders,

[3] Claude Elwood Shannon and Warren Weaver, *The Mathematical Theory of Communication* (University of Illinois Press 1998).

[4] Charles Petzold, *The Annotated Turing: A Guided Tour through Alan Turing's Historic Paper on Computability and the Turing Machine* (Wiley Pub 2008).

[5] Manuel Castells, *The Rise of the Network Society* (2nd ed. Wiley-Blackwell 2010).

[6] Sabina Leonelli, 'Data — from Objects to Assets' (2019) 574 *Nature* 317.

[7] Donella H Meadows, *Thinking in Systems: A Primer* (Diana Wright ed, Chelsea Green Publishing 2008).

and managers; and researchers and others who wish to find an initial hand-hold in this complex area.

2. The Foundations of Data Pluralism

2.1 Data as Form, Data as Flow

It is a fiction that data "just is" (or "just are"), despite the fact that the word "data" itself derives from Latin for "given." Data are mined, produced, constructed, collected, prepared, cleaned, scrubbed, processed, analyzed, combined, sold, stored, and shared, all with explicit or implicit reliance on interpretive theories and models.[8]

Many metaphors appear in that sentence, some more helpful, some of them less so. All of them, in one way or another, suggest the static character of data. In that sense, data are things; or objects; or commodities. Data are fixed items and collections of information, documenting observations about the world. By implication data are scarce (metaphorically speaking) and valuable. Data-as-form captures the metaphorical instinct to treat data as things, or as a thing.

Metaphors are as inescapable in law as they are elsewhere in social life. By allowing us to describe one (less familiar) phenomenon in terms of another (more familiar) phenomenon, metaphors both describe our thinking processes and promote understanding. If we want to solve a problem, we must capture the problem in its full scope and character. At their best, metaphors are tools for doing that.

Yet metaphors are heuristics, and like all heuristics, they have their limitations and capacities to mislead. Data-as-form is, in this sense, incomplete.

One of the most popular umbrella metaphors for data is "the new oil." *The Economist*, a magazine, invoked that metaphor with the headline, "The world's most valuable resource is no longer oil, but data," alluding to the ubiquity of data, the quantity of data, its value as both commodity and as social and technical lubricant, and the associated economic value and market power of firms that deal in data.[9] The scholarly literature tends to join in the allusion.[10]

"Data as [the new] oil" can be misleading. Oil is tangible, and oil reserves are depletable. In most senses, data are intangible, and pools or collections of data are not depletable. More recently, *The Economist* has invoked a competing metaphor, "data as sunlight," signifying the fundamentally open character that data have, or should have.[11]

But some of the implications of the "oil" metaphor may be helpful. Oil is important and valuable partly because of its commodity character (oil in barrels rather than in untapped pools), but also partly because of its "infrastructural" qualities, in that it can be directed to numerous applications, with diverse value and values. Oil moves and flows, literally. Data are "flow" in related senses. Like oil, it is produced via complex technical processes. It can be "pooled" or dis-aggregated. It can be a commodity itself. It lubricates social and technical process. It can be a vital component of numerous other technical and commercial applications. Data-as-flow captures the metaphorical instinct to look at data's fluid attributes.

Related tensions between data as form and data as flow are suggested by recent efforts by industry to clarify the meanings of metaphors such as "data lake" and "data warehouse" in describing modes of aggregating and managing data resources.[12] A "data lake" may combine data from multiple sources, suggesting flows of data; a "data warehouse" may organize data from a single source, suggesting a well-structured form.

These are not rigid characterizations. One should not be misled by the description of data in metaphorical terms. The key point, illustrated by the necessity of metaphor, is that data are *simultaneously* form and flow. No one, single, correct description of data exists on which we may ground some correct regulatory system. The present, massive moment in computing history, exposing the gap between human cognitive capabilities and computing capabilities, calls for intellectual and pragmatic humility and pluralism.

In that respect, it is important to amend the suggestion in the Introduction that data governance should build on systems perspectives on the origins and functions of data. Systems theory typically teaches a distinction between a resource system, sometimes referred to as a stock, and resource units, sometimes referred to as flows. The political scientist Elinor Ostrom, introducing her research on commons for natural resources, distinguished between fisheries and fish.[13] That distinction is most sustainable where biophysical attributes determine the identities and boundaries of the stock and the units. For data, biophysical attributes typically must give way to characterization and interpretation by humans, including differente modes of technology implementation . A systems perspective is still appropriate, even critical, as this article argues below. But identifying the relevant attributes of the system must be part of governance processes, rather than a lead-in to a governance processes.[14]

Three concrete contexts offer illustrations, before the article moves ahead to discussions of governance and resources more broadly, how current law, public policy, and practice rely on data-as-form and data-as-flow as fundamental framing devices. The illustrations are chosen because of the different respects in which they expose fundamental attributes of data in context. Here as elsewhere in this article, attention is drawn to concepts rather than to debates of the moment.

2.2 Copyright and Data

The first is copyright law. Both in the US and in Europe, data and databases as such are subject either to no copyright protection (data lie in the public domain) or to minimal or thin copyright protection. In the US, the Supreme Court opinion that holds that copyrightable works must reflect at least a modicum of "creativity."[15] Logically-structured collections of facts and data almost always do not. European copyright recognizes copyright in works that reflect the author's own

[8] Sabina Leonelli, 'Data Governance Is Key to Interpretation: Reconceptualizing Data in Data Science' (2019) *Harvard Data Science Review* https://hdsr.mitpress.mit.edu/pub/4ovhpe3v accessed 7 February 2020.

[9] *The Economist*, 'The World's Most Valuable Resource Is No Longer Oil, but Data,' 6 May 2017 https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

[10] Dawn E Holmes, *Big Data: A Very Short Introduction* (Oxford University Press 2017); Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt 2013).

[11] *The Economist*, 'Digital Plurality: Are Data More Like Oil or Sunlight?,' 20 February 2020 https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight.

[12] Daniel E O'Leary, 'Embedding AI and Crowdsourcing in the Big Data Lake,' (2014) 29 *IEEE Intelligent Systems* 70.

[13] Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press 1990), p. 30.

[14] Christiaan Hogendorn and Brett Frischmann, 'Infrastructure and General Purpose Technologies: A Technology Flow Framework' (2020) *European Journal of Law and Economics* http://link.springer.com/10.1007/s10657-020-09642-w accessed 29 April 2020.

[15] Feist Publications, Inc. v. Rural Telephone Service, Inc., 499 U.S. 340 (1991).

intellectual creation.[16] In practice that question is usually the "originality" of the work rather than the author's skill and labor. In cases involving collections of facts and data, originality has often been lacking.[17] The point is that modern copyright tends to advance a doctrinal judgment that data are best conceived in terms of data-as-form (is the work, as "thing," sufficiently original?) and that such a thing-like character is often absent. Although data often are human-created, data are and should be difficult to capture, control, and own, because of their obvious social value. Data might be form, but are not. In practice, as a consequence, data are flow.

Both the data-as-form and data-as-flow constructs can be modified by rule and by practice. Data producers and data controllers often have recourse to alternative legal strategies, both in commercial contexts and in research and government setting. Data-as-form approaches are observed in access controls imposed via contract and/or via technology limitations, as well as via legislative efforts to secure forms of exclusivity in databases that do not sound in copyright. The European Parliament, recognizing the poor fit between copyright and databases that is illustrated in the US by the *Feist* standard, adopted the so-called Database Directive in 1996. The Directive created a *sui generis* right to protect databases from appropriation, so long as the database in question represents a "substantial investment" of resources.[18] The inadequacies of that Directive have, in part, prompted the European Commission recently to propose a new "producer's right' in machine-generated data.[19] Data-as-flow approaches are evident in contract, technology, and commercial considerations combined in "Data as a Service," or "DAAS" arrangements. The categories are not rigid. The key is to see how they provide a conceptual foundation for the simultaneity of the conditions of day-to-day practice.

2.3 Public Health and Data

The second is law and public policy concerning public health and medical research. Data about individual health conditions and treatments is collected, abstracted, and generalized both in order to build predictive models of disease and contagion used for population-level interventions and to build diagnostic heuristics and predictive models used for individual-level interventions. In both settings, where models are built and interventions applied, data-as-flow defines the practice.

Where data are obtained or generated at the level of the individual patient or research subject, data-as-form may dominate. Data-as-form permit researchers and clinicians to describe the individual. Data-as-form permit them to document a collection of attributes about the individual. Data-as-form support policymakers and advocates, in contexts that highlight privacy considerations and human rights, who assert that, intuitively, the data "belong" to the individual because in some respects the data originated with or in that person. Commercial interests (and some research interests) claiming "ownership" of health-related data likewise invoke data-as-form arguments.

Legally, states have developed regulatory regimes to try to manage these conflicts, to protect the interests of researchers, the public, and commercial interests in generating better and more effective public health and clinical medical strategies, and also to protect the interests of individuals in avoidable harm to interests in autonomy, privacy, and bodily integrity.

The US has done this via the Common Rule, a formal regulatory standard that governs ethical practice in biomedical and behavioral research involving human subjects, when that research is conducted (as almost all such research in the US is) with the support of federal funding or in federally-supported institutions. It provides that identifiable individual research subjects must give consent both to their participation in research and also to uses of associated individual data. In effect the Common Rule interposes strong initial data-as-form-based regulation on research programs animated by data-as-flow considerations.

Blends of data-as-form and data-as-flow may change. The Common Rule has now been changed. As of January 2019,[20] the Revised Common Rule substantially lowers the threshold for what amounts to "informed" content, meaning that research subjects no longer need to be provided with detailed and comprehensive information regarding uses to which "their" data may be put (quotation marks are included because, given the earlier discussion of copyright, the law may not support proprietary claims). It may be sufficient for researchers to disclose the simple fact that individual data may be shared. Data-as-form considerations are de-emphasized. Data-as-flow considerations are more prominent.

The illustration suggests both that neither data-as-form nor data-as-flow is necessarily superior in normative terms and also that the two framings may be combined, as in the copyright illustration earlier, in complex ways. Adoption of the Revised Common Rule was prompted by the power and potential of medical and public health research grounded in Big Data techniques, where sharing and combining data from multiple sources is increasingly the norm.[21] Critics point to alternative legal constructions, such as the European Union's General Data Protection Regulation (GDPR),[22] which blend individual patient interests and commercial interests differently.[23] The GDPR imposes significantly higher informed consent requirements with respect to storing and re-using individual health data. Normative assessment is complicated by additional data-as-form and data-as-flow attributes of US medical research systems. Authors of medical and public health research may be required by US law to share their research data by depositing data in public archives, a policy judgment based principally on data-as-flow.[24]

[16] Article 3(1) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

[17] Lionel Bently and Brad Sherman, *Intellectual Property Law* (4th ed. Oxford University Press 2014); C-604/10 Football Dataco v. Yahoo! UK and Others [2012] EU:C:2012:115.

[18] Article 7 of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

[19] Peter K Yu, 'Data Producer's Right and the Protection of Machine-Generated Data' (2019) 93 *Tulane Law Review* 859.

[20] Dept. of Homeland Security et al., Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149, 7150/1 (Jan. 19, 2017).

[21] Willem G van Panhuis, Anne Cross and Donald S Burke, 'Project Tycho 2.0: A Repository to Improve the Integration and Reuse of Data for Global Population Health' (2018) 25 *Journal of the American Medical Informatics Association* 1608.

[22] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L [2016] 119/1.

[23] A Michael Froomkin, 'Big Data: Destroyer of Informed Consent' (2019) 21 *Yale Journal of Law & Technology Special Issue* 27; Lara Cartwright-Smith, Elizabeth Gray and Jane Hyatt Thorpe, 'Health Information Ownership: Legal Theories and Policy Implications' (2017) 19 *Vanderbilt Journal of Entertainment and Technology Law* 207.

[24] Deborah Mascalzoni and others, 'Are Requirements to Deposit Data in Research Repositories Compatible with the European Union's General Data Protection Regulation?' (2019) 170 *Annals of Internal Medicine* 332.

2.4 Biobanks and Data

A third illustration of diverse and changing uses of data-as-form and data-as-flow are biobanks, collections of biospecimens and related data derived from and used for research in both biomedicine and agriculture. Thousands of biobanks operate around the world storing tissue samples, genetic sequence information, and seeds, among other things. Their organizational structures are correspondingly diverse. Many are state-sponsored or supported. Some are private. Some are philanthropic. In cases where these enterprises collect and store data and physical specimens, as many do, their organizational design and governance and relevant legal regulation address data-as-form and data-as-flow perspectives in two layers.

One layer is biospecimens themselves, to which ethical, privacy, contractual, and tangible property interests may attach. They are data-as-form, in the sense of each biospecimen being a "thing," and a collection of biospecimens being a distinct "thing." Biospecimens are also data-as-flow, in that they are data as well as objects, and they have been collected and stored precisely because of their infrastructural, informational value to future researchers. A second layer is the informational data associated with the biospecimens, to which independent ethical, privacy, contractual, and *intangible* property interests may attach and which may have independent infrastructural importance for future research. The informational data are likewise data-as-form (the information associated with each specimen, and with a collection), and data-as-flow.[25]

3. About Governance

Data-as-flow and data-as-form are rhetorical and propositional statements, but they are not pre-theoretical. They are not ontological statements about the true state of data. They are, by virtue of their metaphorical origins, judgments about the world, offered for their utility. They set out the initial conceptual vocabulary of this article. This Section provides the beginnings of its syntactical structure, which animates the analysis. If the challenge of data governance is identifying and advancing respects in which data-as-form should dominate data-as-flow, or the reverse, or neither, then how should that challenge be addressed?

This Section provides the first elements of a toolkit for analyzing situations and possibly recommending courses of action. It is a framework, which describes governance, institutions of governance, and the knowledge commons framework as an instrument for researching governance. Knowledge commons gets particular attention here because it provides as systematic framework for examining governance of shared knowledge resources, and because data governance is above all else, perhaps, a complex and sustained challenge in managing *shared* resources in institutional contexts.

Like a useful theory, a useful framework teaches us what conditions matter and what to look for, and why. As a device for assembling evidence, a framework should not be overly or prematurely precise and should initially accommodate multiple possible theories.[26]

3.1 Governance

The concept of governance is used here in the sense of collective or coordinated decisionmaking by individuals working together, about decisions on matters of collective interest. The emphasis on governance, rather than on law, regulation or public policy specifically or on coordination in the abstract, is based on and justified with respect to a fundamental anthropological instinct rather than a formal or positive legal one. Governance means individuals working together to form groups to solve their own problems.[27] A major thesis of this paper is that with respect to data, we should be asking about governance, not asking simply about law. Starting with governance opens the door to broader and more effective questioning about potential problems and solutions associated with data. Starting with markets or the state, per the Introduction, may pre-judge the character of both.

3.2 Institutions

Governance is best understood via its expression in institutions, rather than via the thoughts and behaviors of individuals. Individuals and their opportunities, thoughts, choices, and behaviors matter and, in a utilitarian sense, often matter most in final welfare judgments. But in practice, individual cognition and motivation are diverse. Efforts to understand governance primarily via references to an imaginary "model" human, responding to commands of the law, are destined to be unsatisfactory to the extent that the models do not match reality. This article foregrounds a framework that is grounded in empirics and pragmatics of institutions, meaning collections of individuals.

Governance is not limited, however, to formal institutions of the state, such as legislatures, courts, and administrative bodies. The reference to "institution" implies a broader view.

For a working definition of "institution," the article adopts the definition given by the economist Douglass North: the rules of the game of a society, devised by humans and shaping human behavior.[28] Also relevant, to similar if not identical effect, is the concept of the institution developed in modern sociology: institutions are stable behavioral patterns that reflect the coordinated behavior of individuals and organizations, where the relations define the actors rather than the other way around.[29]

The difference between the two perspectives, the former focusing more on rules that guide or determine patterned behavior, and the latter focusing on rules that reflect patterned behavior, is not determinative here. What matters is that institutions in either sense (or both senses) simultaneously produce and rely on well-understood sets of human-created norms to determine outcomes among a group of people who significantly self-identify with the enterprise in its own time. Groups may constitute and be denominated "communities" or "collectives" or firms or other enterprises. Membership or participation may be small or large. Group identity may be formally circumscribed or informal, dynamic, and fluid. Groups may exist in specific places and times, as firms or as cities, for example. They may combine mate-

[25] Michael J Madison, 'Biobanks as Knowledge Institutions' in Timo Minssen, Janne Rothmar and Jens Schovsbo (eds), *Global Genes, Local Concerns: Legal, Ethical and Scientific Challenges in International Biobanking* (Edward Elgar Publishing 2019).

[26] Elinor Ostrom and Michael Cox, 'Moving beyond Panaceas: A Multi-Tiered Diagnostic Approach for Social-Ecological Analysis' (2010) 37 *Environmental Conservation* 451.

[27] Donald E Brown, *Human Universals* (McGraw-Hill 1991); Stuart P Green, 'The Universal Grammar of Criminal Law' (2000) 98 *Michigan Law Review* 2104.

[28] Douglass C North, *Institutions, Institutional Change, and Economic Performance* (Cambridge University Press 1990).

[29] Walter W Powell, 'Neither Market nor Hierarchy: Network Forms of Organization' (1990) 12 *Research in Organizational Behavior* 295; John Frederick Padgett and Walter W Powell, *The Emergence of Organizations and Markets* (Princeton University Press 2012).

rial and immaterial forms, transcending place and time in "imagined" communities of the sort described by Benedict Anderson.[30] Groups, loosely specified, are critical loci of governance in institutions.[31]

3.3 Institutional Governance of Resources

The rest of this Section offers a framework for investigating and understanding institutional governance of resources, including institutional governance relative to data, in ways that supplement the two usual sources of legitimate governance, states and markets.

Perhaps the most enduring and influential justification for the roles of markets and states in regulating resources, particularly relative to shared resources, is the story of the tragedy of the commons.[32] Modern researchers have come to identify the story closely with a well-known paper by the ecologist Garrett Hardin from 1968, but the story pre-dates Hardin's work.

The tragic commons offers a powerfully simplistic metaphor. As a result the story has been simultaneously a diagnostic tool, an explanation for historical developments, and a prescription. If resources are shared, they are likely to be over-exploited and ruined. To prevent the expected destruction, regulation should specify an actor or actors responsible for a defined set of resources, accountable either via the marketplace or via state mechanisms, and expect better results.

Legal scholars often have assimilated the tragic commons metaphor to problems in the creation and circulation of information and knowledge, such as production of inventions, new cultural works, management of data, personal information, and interests in privacy. The stereotypical implication is state supply of legal exclusivities in relevant intangibles, to be traded in private markets. Alternatively, the state may simply supply the resource itself, directly (by building and controlling it) or indirectly (by underwriting it). The expected solutions are intended to ensure that the resource exists in the first place, rather than over-exploited.

The tragic commons model works well in some settings. Positive law itself may at times be a resource that would not be adequately supplied absent state direction.[33] Various jurisdictions act differently on that institutional premise. US federal law is committed to the public domain. Other jurisdictions assert proprietary claims over the content of the law, in the name of the state. At best, in short, the tragic commons metaphor offers a helpful beginning. But its shortcomings are more significant. The inadequacies of the metaphor have been critiqued elsewhere at length. Only the briefest review is needed here.

In part, the tragic commons metaphor may mis-describe the resources themselves, particularly as to knowledge and information resources, such as data. The tragic commons metaphor typically posits a depletable resource. Even for tangible resources, that assumption may not hold. Material resources, even biophysical resources such as grazing pastures, may be regenerated or resupplied. For intangible and immaterial resources, such as data, consumption may affect their value but not their existence. They suffer from no depletability problem. One significant problem is creating data resources in the first place, with the right attributes. Further resource-related questions are deferred to Section 5, below.

In part, and as relevant here, the tragic commons metaphor may mis-describe the actors involved. The tragic commons metaphor posits self-regarding, selfish decisionmaking actors with no means or motivation to acquire information about their neighbors' activities, no ability to plan for the future, no practice of coordinating their actions with their neighbors', and no capability for adaptation and innovation in the face of complexity.[34] The metaphor assumes no governance. Instead, it assumes a sort of pre-governmental, pre-political state of nature, with no background customs or rules regarding collective identity or appropriate behaviors, and primitive, one-dimensional individuals.

Obviously, the tragic commons metaphor is not intended generally to describe any actual world. But it may be taken as doing so, and when that happens, the metaphor may become something of a self-fulfilling prophecy. The failure of collective action that the metaphor predicts may provide a premise rather than a conclusion.

One may treat the production, consumption, and preservation of a shared resource as a challenge for collective action, rather than a failure of collective action. Can forms of collective action solve those challenges? Can those forms do so, particularly with respect to shared knowledge and information resources, in ways that are as welfare-enhancing as one supposes state production, distribution, and access?

3.4 Commons Governance

The path to a pluralistic modern understanding of institutional governance and the potential strengths of resource sharing institutions arose initially via the research of Elinor Ostrom. First collected in the 1990 book *Governing the Commons*,[35] the work of Ostrom and her colleagues, collaborators, and students carefully established, via an abundance of fieldwork and comparative analysis, that self-directed collaboration and collective action to solve resource management problems was possible – in practice, if not always in theory. Ostrom's adaptation of the "commons" framing not only enlarged policymakers' and scholars' fields of vision relative to shared resource challenge. This work re-introduced the idea of "commons" in an explicitly ecological sense, referring to actors, institutions, and resources interacting in systems in multiple interdependent ways.[36]

In *Governing the Commons* and later work, Ostrom added to economists' standard taxonomy of types of goods. Beginning with private goods (which area excludable and rivalrous), public goods (which are nonexcludable and nonrivalrous), and club goods (which are excludable but nonrivalrous, and sometimes referred to as toll goods), she added and focused on "common-pool resource systems," or "CPRs." CPRs are resources, rather than goods, a definition that expands their utility and functions to include uses beyond tradeability and consumption. CPRs are nonexcludable and shared but *depletable*, and subject to risks of overconsumption.

For common-pool resources, Ostrom described a series of considerations, or guidelines, indicating when informal systems of collective, community management of the resource was both feasible – contrary

[30] Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (Verso 1983).

[31] Michael J Madison, 'Social Software, Groups, and Governance' (2006) 2006 *Michigan State Law Review* 153.

[32] Madelyn Sanfilippo, Brett Frischmann and Katherine Strandburg, 'Privacy as Commons: Case Evaluation through the Governing Knowledge Commons Framework' (2018) 8 *Journal of Information Policy* 116. A 'shared' resource is one that is produced, used, and/or consumed by multiple actors, either concurrently or sequentially.

[33] Brigham Daniels, 'Legispedia' in Brett M Frischmann, Michael J Madison and Katherine J Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014).

[34] Carol M Rose, 'Commons and Cognition' (2018) 19 *Theoretical Inquiries in Law* 587.

[35] Ostrom, *Governing the Commons* (n 13).

[36] Brett M Frischmann, 'Cultural Environmentalism and "The Wealth of Networks"' (2007) 74 *University of Chicago Law Review* 1083.

to the prediction of the tragic commons metaphor – and likely to generate sustained, welfare-promoting provision of that resource over time.[37] The word "commons" comes forward in this article from Ostrom's work. "Commons" means *not* fully open, unmanaged access to a resource, but instead collective institutional governance of a resource, embodying a set of strategies that solve coordination problems, known as social dilemmas. That mouthful of a phrase can be distilled into something shorter: commons means groups that engage in managed resource sharing. Institutional governance via groups may take the place of or exist in tandem with governance via exclusive rights and markets, (on the one hand) and governance via state provision or determination (on the other hand).

In highlighting the possible virtues of commons-based institutional governance of resources, Ostrom's work is important here in three respects.

One, Ostrom's guidelines for successful commons management have no direct or obvious utility in domains related to knowledge, information, and data. Virtually all of the research conducted for *Governing the Commons* and follow-on research focused on natural (i.e., biophysical) resources, such as water systems, forests, fisheries, and pasturage, which easily fit Ostrom's definition of a CPR. Though late in her career Ostrom and her colleague Charlotte Hess undertook some preliminary explorations of commons governance related to knowledge resources,[38] those efforts should be regarded more as encouraging further investigation rather than as definitive applications of Ostrom's work in new domains. Despite some preliminary efforts to apply Ostrom's work to data governance,[39] shareable knowledge, information, and data resources do not meet the definition of CPRs. In intangible, immaterial forms, knowledge resources are neither excludable nor depletable. Ostrom's commons guidelines should be set aside with respect to data governance. Whether and how collective- or community-based governance of data should function is a matter to be investigated afresh, via examining conditions in the field.[40] Ostrom's body of work exhibits a strong sympathy for collective self-determination and a strong skepticism of the role of the state, via formal property rights systems or otherwise. Those intuitions deserve empirical exploration in contexts related to data.

Two, Ostrom showed that understanding and developing effective institutional governance requires a strong dedication to empiricism and to comparative, contextual analysis.[41] Ostrom and her colleagues were motivated in part by specific resistance to the simplistic conceptual reasoning that is often associated with casual adoption of the tragic commons metaphor. In that spirit, Ostrom formalized her style of research in a strategy labeled the "Institutional Analysis and Development" framework (IAD) in order to support additional research.[42]

That style of analysis, if not that framework itself, is a critical step forward in understanding data governance.

Three, Ostrom highlighted the broad domain of successful resource governance strategies that rely neither on "market exclusivity" nor "state provision of a shared resource" (a strategy that would include a public policy declaring that a resource ought to be unowned and fully "open," as a part of a "public domain"). She titled the address she delivered in association with receiving the Nobel Prize *Beyond Market and States*.[43]

3.5 Knowledge Commons

The proposition that shared knowledge and information resources, such as data, ought to be subject to analysis and possible regulation via commons governance institutions of the sort just described, has been distilled into the knowledge commons research framework. That framework, described sometimes via the shorthand "GKC framework" after *Governing Knowledge Commons*, the title of the first volume of published knowledge commons research,[44] is an analytic tool motivated both by frustration with the tragic commons metaphor, as applied to information, and also by the strengths and style of Ostrom's research on commons. The GKC framework brings the ecological and systems spirit of that research into examinations of knowledge and information governance.

In contemporary research and policymaking, information production problems are simplistically modeled as overconsumption and free riding by multiple actors with access to a shared knowledge resource, leading to depletion and eventually to underproduction. Stereotypical solutions follow, modeled either as exclusive property rights transacted in markets (patents, copyrights), or as public goods provisioned by or underwritten by state authorities (such as scientific research). Problems of information privacy may be subject to equivalent stereotypical treatment, leading to proposals to vest strong exclusive privacy rights in individuals or to empower states to define privacy interests – to the exclusion of collectively self-directed privacy governance, in context.[45]

The GKC framework animates a research program intended to capture and inventory the domain of governance problems and solutions for knowledge and shared information resources. The GKC framework borrows its empiricism, its emphasis on context and setting, and its methodological pluralism from Ostrom's IAD framework. Similarly, the GKC framework anticipates the later development of one more theories or models of institutional design, individual motivation, and normative assessment. While the GKC framework is styled in the manner of Ostrom's IAD framework, it is not simply a special case of Ostrom's thinking or the IAD framework as such. Other scholars of information policy have similarly called for the development of governance strategies based on commons concepts: structured sharing.[46]

Clarifying the terminology helps to introduce the details of commons governance as a system by which some community or collective establishes and enforces principles of managed access to a shared resource. The underlying resource may be "purely" intangible and immaterial or a blend of material and immaterial attributes. The

37 Ostrom, *Governing the Commons* (n 13).

38 Charlotte Hess and Elinor Ostrom (eds), *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007); Charlotte Hess and Elinor Ostrom, 'Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource' (2003) 66 *Law & Contemporary Problems* 111.

39 Joshua B Fisher and Louise Fortmann, 'Governing the Data Commons: Policy, Practice, and the Advancement of Science' (2010) 47 *Information & Management* 237.

40 For the argument that Ostrom's instincts regarding governance, but not the details of Ostrom's program, should be applied to information privacy regulation, see Jane K Winn, 'The Governance Turn in Information Privacy Law' (2019) *SSRN Scholarly Paper ID 3418286* https://papers.ssrn.com/abstract=3418286 accessed 7 February 2020.

41 Brett M Frischmann, 'Two Enduring Lessons from Elinor Ostrom' (2013) 9 *Journal of Institutional Economics* 387.

42 Elinor Ostrom, *Understanding Institutional Diversity* (Princeton University Press 2005).

43 Elinor Ostrom, 'Beyond Markets and States: Polycentric Governance of Complex Economic Systems' (2010) 100 *American Economic Review* 641.

44 Brett M Frischmann, Michael J Madison and Katherine Jo Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014).

45 Sanfilippo, Frischmann and Strandburg (n 32).

46 Jorge L Contreras and JH Reichman, 'Sharing by Design: Data and Decentralized Commons' (2015) 350 *Science* 1312.

resource may be characterized by intellectual property rights and/or other exclusivity interests. The resource may originate in information that is characterized by no IP rights (public domain status). A *patent pool* and a *data pool* are both forms of knowledge commons, as the term commons is used here. A newsgathering and distribution collective, such as the Associated Press "wire" service, is a form of knowledge commons, although individual news stories are typically subject to few if any formal IP rights and in some countries, notably the US, are treated as presumptively open by virtue of constitutional requirements.[47] The relationship between the legal status of the underlying resource and the character of the resource management system is a question to be explored, not declared. Neither ownership nor openness is the end of the matter.

Commons governance includes a range of institutional governance practices under the "commons" umbrella. Because knowledge and information resources may be defined and regulated by positive law, commons governance systems and market-based systems and formal state regulation may be linked and overlap in specific contexts. Further, no bright line exists to divide *knowledge* commons, which are directed primarily to information resources, from other sorts of commons, such as natural resource and environmental commons studied by Ostrom and her colleagues, and urban commons, which refer to governance of urban planning and design.[48] The acronym CPR, which in social science research refers to "common-pool resource," also appears in property law theory as "common property regime," a commons-like governance system anchored in analyses of infrastructural resources such as roads. Common property regimes highlight increasing returns to scale as more and more people consume a resource of a given size.[49] Infrastructural resources, because of their shared character, are often governed as commons.[50] The practice of "commoning" usually refers to politically or ideologically-motivated practices combining local resource governance institutions and self-directed community governance.[51]

The details of the GKC framework as a research instrument are described elsewhere.[52] The key insight of the framework is not whether the institution "is" or "is not" a commons. Rather, the question answered by the framework is whether and how some knowledge or information resource is governed as a *shared resource* via some community or collective, as an alternative to knowledge governance in markets, founded on claims of exclusivity of right, such as patents or copyrights or to knowledge governance via state intervention, provision, or subsidy. Commons governance systems may play important roles with respect to market-based and government-supplied resources. The question is whether some knowledge or information resource presents, in substantial part, hallmarks of structured sharing.[53]

Knowledge commons governance is neither rare nor novel, nor is it limited to specific economic or cultural niches, such as small communities. The GKC framework supplies a means of describing the breadth of the field in a systematic way. The functionality of durable knowledge commons governance – broadly across technical and cultural domains, at different scales, and in specific cases – has been demonstrated in cases across a diverse range of contemporary and historical settings, including both technology development and cultural creation.[54] Janis Geary and Tania Bubela provide an exemplary case study of knowledge commons in a specific and focused case of contemporary life sciences research.[55] Knowledge commons has been used to analyze the field of microbial biology.[56] The GKC framework is consistent with research on patent pools, open source software development, and clearinghouses[57] and other institutions for collective governance of shared resources, including data and datasets. These have been documented in historical settings,[58] in less developed countries,[59] in large-scale, critical scientific and health related research networks,[60] in large scale commercial settings,[61] and in Big Data-enabled scientific research communities.[62]

3.6 Rules and Norms

The GKC framework is primarily descriptive, rather than normative. It aims to surface attributes of institutions via examination of specific cases for potential comparative assessment, using tools borrowed in part from social science, in part from the humanities, and in part from law. (The framework is intended to be accessible to and usable by researchers from each of these domains.) Users of the framework and students of knowledge commons research often focus on the systems of formal and informal rules, norms, customs, and practices by which communities and collectives govern themselves and govern relevant resources. In GKC research as in much of Ostrom's work, these are "rules in use," signifying their empirical rather than normative status. For purposes of comparative institutional analysis, these rules in use may be productively compared with rules and norms in evidence in market-based governance systems and those prescribed

47 Michael J Madison, Brett M Frischmann and Katherine J Strandburg, 'Constructing Commons in the Cultural Environment' (2010) 95 *Cornell Law Review* 657.

48 Sheila R Foster and Christian Ianone, 'Ostrom in the City: Design Principles and Practices for the Urban Commons' in Blake Hudson, Jonathan Rosenbloom and Dan Cole (eds), *Routledge Handbook of the Study of the Commons* (Routledge 2019).

49 Carol M Rose, 'The Comedy of the Commons: Commerce, Custom, and Inherently Public Property' (1986) 53 *University of Chicago Law Review* 711.

50 Brett M Frischmann, *Infrastructure: The Social Value of Shared Resources* (Oxford University Press 2012).

51 David Bollier and Silke Helfrich (eds), *Patterns of Commoning* (Common Strategies Group 2015).

52 Michael J Madison, Brett M Frischmann and Katherine J Strandburg, 'Knowledge Commons' in Blake Hudson, Jonathan Rosenbloom and Dan Cole (eds), *Routledge Handbook of the Study of the Commons* (Routledge 2019).

53 Frischmann, Madison and Strandburg (n 44).

54 Frischmann, Madison and Strandburg (n 44); Katherine J Strandburg, Brett M Frischmann and Michael J Madison (eds), *Governing Medical Knowledge Commons* (Cambridge University Press 2017).

55 Janis Geary and Tania Bubela, 'Governance of a Global Genetic Resource Commons for Non-Commercial Research: A Case-Study of the DNA Barcode Commons' (2019) 13 *International Journal of the Commons* 205.

56 JH Reichman, PF Uhlir and Tom Dedeurwaerdere, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press 2016).

57 Geertrui van Overwalle (ed), *Gene Patents and Collaborative Licensing Models: Patent Pools, Clearinghouses, Open Source Models, and Liability Regimes* (Cambridge University Press 2009).

58 Tine de Moor, 'From Historical Institution to Pars Pro Toto: The Commons and Their Revival in Historical Perspective' in Blake Hudson, Jonathan Rosenbloom and Dan Cole (eds), *Routledge Handbook of the Study of the Commons* (Routledge 2019).

59 Jeremy De Beer and others (eds), *Innovation & Intellectual Property: Collaborative Dynamics in Africa* (Published by UCT Press in association with the IP Unit, Faculty of Law, University of Cape Town and Deutsche Gesellschaft für Internationale Zusammenarbeit 2014).

60 Amy Kapcynski, 'Order without Intellectual Property Law: Open Science in Influenza' (2017) 106 *Cornell Law Review* 1593.

61 Henry Chesbrough, *Open Innovation: The New Imperative for Creating and Profiting from Technology* (Harvard Business School Press 2003).

62 Michael J Madison, 'Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo' in Brett M Frischmann, Michael J Madison and Katherine J Strandburg (eds), *Governing knowledge commons* (Oxford University Press 2014).

in state-based regulatory settings. For students and practitioners of data governance, the intuitive answer to "how should we regulate?" takes the form of "these are the appropriate rules."

In the context of knowledge governance, the temptation to prioritize examination of the rules, empirically or normatively, may be resisted. It risks putting the proverbial cart before the horse. The review of commons governance shows why: commons governance is collective management of a shared resource by or in a group. The role of the collective is largely to define its own governance system relative to dilemmas associated with specified resources, producing a form of institutional governance in context. This article has described the fundamental problem of governing data sharing in terms of two conceptions, data-as-form and data-as-flow. It argues next that understanding data governance should begin not with the rules, but instead with two key phenomena: groups and things.

4. About Groups

"Groups" means formal and informal collections of people, who identify themselves with the group (perhaps closely, perhaps loosely, and perhaps in variable numbers over time) and who adopt and enact practices that are aligned with the interests and identities of the group. When knowledge commons governance research refers to institutional governance of shared resources by self-directed communities and collectivities, it refers to groups solving social dilemmas regarding those resources. *Beyond Markets and States*, the title of Ostrom's Nobel Prize address, is read fairly to claim that governance by groups is an empirically valid mode of resource management.

In practice, that summary opens at least three key lines of inquiry as conceptual matters.

The first is the most pragmatic: In a resource governance context, does one or more groups exist that might serve as governance vehicles? How might such a group be identified, defined, and organized? Should law or regulation be invoked to motivate or to discourage group formation as part of an institutional governance strategy?

The second concerns the possible governance contributions by groups. Groups might generate relevant rules, norms, and practices on their own, such as a voluntary association, or might serve as agents for administering and enforcing rules and norms generated elsewhere, such as employees of a for-profit firm. Groups might serve as collective institutions in a cognitive sense, so that the collective is able to identify and act on information that is not equally accessible or useful to individuals acting alone. Groups might act as loci for interpretive practices by which society gives shape and meaning to places and resources, as suggested in multiple traditions of Science and Technology Studies. In each of these respects, where present, groups may participate in resource governance practices.[63]

The third and most important here concerns ways in which groups may be anchors for two especially critical conceptual foundations for institutional governance of shared resources: *trust* and *polycentricity*. Data governance strategies should explore both.

4.1 Trust

Trust represents the sense that trust mechanisms are critical to cooperative arrangements.[64] It also represents the sense that actual human beings have greater capabilities for understanding and adapting to complex social and environmental challenges, and for doing so cooperatively rather than being coerced to do, than is predicted by tragic commons metaphors and presumptions of selfish behavior following the pursuit of rational self-interest.[65] Trust may operate bilaterally, between individuals or between an individual and an institution. Trust also operates critically among populations of individuals and an institution. For governance by groups, social trust mechanisms must operate at some level among the members of the group, relative to one another and relative to the purposes of the group. It has been argued that trust generally consists of means by which individuals cope with the fact that others may exercise their own freedom.[66] But no single, optimal definition of trust exists.

Likewise, no single social or policy mechanism works universally to promote trust and promote group formation, identity, durability, or adaptability, or to undermine trust or to prevent it from forming. Group-based resource governance may be unhelpful or harmful, or may create unmanageable conflict with other governance institutions. The research literature on trust and cooperation is vast, and it covers sociological, anthropological, economic, political science, and philosophical domains.[67] Emphasizing reciprocal relations between community members, for example, is sometimes suggested as a critical ingredient in effective cooperative settings, an idea that may be traced back to early work on gift economies. But the details matter. "Pay it forward" reciprocity strategies may be as important to trust formation as "pay it back" strategies, or more so.[68] Trust creation and reinforcement may depend on relationships among group decision-making rules (such as enforcement norms, or exit/entry criteria) and the development of shared collective identity (such as "who we are" questions).

This makes trust an ecological and structural question as well as a matter of individual cognition.[69] The research and policy challenge is to design and support institutions where the benefits of individuals' cooperative capabilities can be put to good use, where shared resources can be governed effectively, and where the weaknesses of a trust-based model are minimized. Cooperative capabilities are unevenly distributed, for example, and trust mechanisms may be riddled with harmful power and influence dynamics. Trust is itself, significantly, a shared resource, and governance of that resource is likely necessary as part of broader resource governance strategy.

4.2 Polycentricity

That trust is a shared resource subject to governance, as part of governance of a shared knowledge resource such as data, points to the idea that governing groups may overlap and intersect. *Polycentricity* captures that concept, in the sense that any institutional design for governance is likely to be most effective when it is characterized and implemented in a decentered way, with multiple loci of authority and responsibility, rather than a single center of regulatory agency, intersecting with one another at different scales[70] and relying on individuals' diverse motivations for participating.[71]

[63] Madison, 'Social Software, Groups, and Governance' (n 31).
[64] Kenneth Joseph Arrow, *The Limits of Organization* (Norton 1974).
[65] Bo Rothstein, *Social Traps and the Problem of Trust* (Cambridge University Press 2005).
[66] Niklas Luhmann, *Trust and Power* (English edition, Polity 1979).
[67] Diego Gambetta (ed), *Trust: Making and Breaking Cooperative Relations* (B Blackwell 1990).
[68] Toshio Yamagishi and Karen S Cook, 'Generalized Exchange and Social Dilemmas' (1993) 56 *Social Psychology Quarterly* 235.
[69] Kenneth W Abbott, Jessica F Green and Robert O Keohane, 'Organizational Ecology and Institutional Change in Global Governance' (2016) 70 *International Organization* 247.
[70] Julia Black, 'Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes' (2008) 2 *Regulation & Governance* 137.
[71] Yochai Benkler, 'Law, Innovation, and Collaboration in Networked Econo-

Those multiple centers may be informal or formal or blends of the two. Groups may be organized hierarchically. Smaller groups may be "nested" within a larger group. Groups may be linked to on another in a network of distinct and/or overlapping nodes of different scales. Polycentric systems can be flexible and adaptable across time, scale, and community form. They can support enforcement and accountability mechanisms at different scales, enhancing legitimacy, accountability, and administrability of governance systems as a whole.

So, just as trust is a key governance variable to be explored, polycentricity does not solve all problems. One must still carefully consider the scope of authority and its mechanisms of accountability and legitimacy. Like all governance systems, and like trust, polycentric systems are subject to appropriation and abuse via dynamics of power, wealth, and status. Polycentricity is not a cure-all. It is an analytic strategy, and polycentric systems can be made stronger and weaker.[72]

4.3 Groups and Data

Group-based perspectives, including polycentric governance and emphasis on structures that both generate and rely on social trust, are consistent with but perhaps more nuanced and potentially effective than other norm-based approaches that are not so explicitly pluralistic. Governance of shared data resources with reference to groups helps us organize possible strategies distinguished as data-as-form and data-as-flow. The absence of relevant groups relative to those data resources suggests a different range of strategies distinguished along those lines. For example, certain approaches to "open" data governance (a species of data-as-flow) may be better appreciated and have greater impact if described as parts of polycentric governance, including "best practices" recommendations; "fair practices" approaches, such as the Fair Information Practice Principles (FIPP) for personal data collection and the FAIR Data Principles for scientific data management; suggestions that all of data or all of knowledge constitutes a single, global shared resource;[73] and advocacy under labels such as Open Science and Open Data. In these contexts, "openness" and "fair" practices are achieved by paying careful attention to institutional attributes of groups and fields.[74]

Historical data governance practices are similarly illuminated by prioritizing questions about groups, trust, and polycentricity. The historian Will Slauter argues persuasively that seventeenth century English publishers strategized ways to obtain exclusivity in shipping and price information (data-as-form).[75] Modern copyright and its near-total exclusion of data from legal ownership is in many respects a product of those strategies, their modern analogs, and resistance by other groups in UK and American legal systems (data-as-flow). The political scientist James Scott suggests, provocatively, that central state authority exists not only to enhance the well-being of citizens but to render information about them "legible," as data (data-as-form).[76] He argues that alternatives to the modern state, in localized, collective self-governance, may be equally effective at promoting well-being and offers the benefit of maintaining critical distances between the state and its subjects (data-as-flow).

5. About Things

"Things" captures a broad range of related phenomenon: items, units, commodities, embodiments, objects, artifacts, and stuff, both material and immaterial, analog and digital. With such a broad beginning, semantics and ontologies can get tricky, and interpretive techniques must be developed to sort out relevant distinctions.[77] One object may embody more than one thing, and one thing may be embodied in more than one object. A "work of art" such as a novel may embody a distinct "work of authorship" or "copyright work"; that copyright work may be embodied in numerous copies of the novel. One thing, such as a the novel, may be part of another thing, such as a library, and may itself consist of other things, such as literary elements, and chapters. Identity is another concern. In the larger collection, the smaller unit may be separable. But not always. A gallon of water poured into a river mixes inseparably with the rest of the river. A gallon of water can be extracted from the river, but that gallon is not the same gallon as the water previously poured in. Origins, possession, and authenticity also shape the definitions, meanings, and purposes of things. Things are often associated with specific individuals. They are also often associated with social groups.

The word "thing" is a broad and inclusive way to refer to "resource," as that word and concept have contributed earlier to discussions of governance. When knowledge commons governance research refers to institutional governance of shared resources by self-directed communities and collectivities, it refers to groups solving social dilemmas regarding the creation, use, and preservation of things, treating things as a flexible category that allows researchers and analysts to explore widely.

In practice, that summary opens at least two key lines of inquiry as conceptual matters.

The first returns to the prompt with which the article began: the essential distinction between data-as-form and data-as-flow. That distinction suggests asking, foundationally, what is a thing, and how do we know? Whereas the last Section built conceptually on the contributions of Elinor Ostrom, to a sizable degree this Section moves beyond Ostrom. Ostrom's work on institutional governance and commons typically relied heavily on analysis of natural resources, which come to us with given and mostly unmodifiable biophysical attributes. Ostrom's later work, on knowledge, tended to treat "knowledge" as a single, undifferentiated resource. Neither approach suits the GKC framework. Neither approach suits data.

The second concerns relationships between groups and things. Those relationships are often fundamentally ecological and systemic. The social groups that construct and manage resources may be produced, reinforced, and reproduced by the identity of the resource and by the group's governance practices relative to the resource, both as to the internal dynamics of social groups and as to relationships between social groups.[78] How should those relationships be explained?

my and Society' (2017) 13 *Annual Review of Law and Social Science* 231.

72 Black (n 70). The next Section offers a parallel point regarding resource systems themselves, which can be designed flexibly to operate at greater or lesser scales. The cognitive scientist Herbert Simon characterized organisms with this character as "nearly decomposable." He argued that the "decomposability" strategy for managing adaptation in complex environments rendered such organisms particularly fit in evolutionary terms. HA Simon, 'Near Decomposability and the Speed of Evolution' (2002) 11 *Industrial and Corporate Change* 587.

73 Hess and Ostrom, 'Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource' (n 38); Jennifer Shkabatur, 'The Global Commons of Data' (2019) 22 *Stanford Technology Law Review* 354.

74 Liz Lyon, Wei Jeng and Eleanor Mattern, 'Research Transparency: A Preliminary Study of Disciplinary Conceptualisation, Drivers, Tools and Support Services' (2017) 12 *International Journal of Digital Curation* 46.

75 Will Slauter, *Who Owns the News? A History of Copyright* (Stanford University Press 2019).

76 James C Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed* (Yale University Press 2008).

77 Michael J Madison, 'IP Things as Boundary Objects: The Case of the Copyright Work' (2017) 6 *Laws* 13.

78 John Seely Brown and Paul Duguid, *The Social Life of Information* (Harvard

5.1 What is a Thing

Identifying and defining things are problems in epistemology that go back to Aristotle. The question here is not so broad. The question is governance: what are the things that form parts of governance systems? What are resource systems, and what are resource unit? Borrowing the concept of polycentricity, how do multiple resource systems interact, overlap, and align? Where do relevant resources come from; how are relevant resources used, consumed, and applied; and how, if at all, are relevant resources preserved over time? Data-as-form and data-as-flow are then both inputs into governance analysis and outcomes of governance analysis.

For biophysical resources, answers to most of these questions may be relatively straightforward; resources are the objects of governance. For knowledge and information resources, including data, resources are both subject and objects of governance. Governance often creates (produces, consumes, preserves) the things to which governance applies. Prioritizing things in governance is a way of prioritizing a key set of critical questions. Pragmatically, a critical perspective on governance means that little turns on classifications of things resources as inherently private goods, public goods, club or toll goods, or common-pool resources. The tools of law and policy as well as the experiences of social life teach that boundaries and classifications among these categories can be modified in many settings, disrupting what otherwise might be standard prescriptions based on the logic that gives priority attention to commons tragedies. A functional approach, based on an empirical approach to ecologies in practice, is preferred.

Data depend on their reference and relationships to underlying phenomena. In that sense, data are evidence of something else.[79] They are, almost by definition, both things in themselves and also versions of something else. Data signify a problem long recognized in mathematics, computer science, geography, and literature: to be useful, a model or map must stand in for the whole but not be identical to it.[80] Data are sometimes characterized as "raw" or "cooked," a metaphorical framing that suggests the degree to which data directly (raw, unprocessed) or indirectly (cooked, processed and analyzed) relate to their source. The metaphor departs from its partial origins in the anthropological literature, as a reference to the construction of conceptual oppositions.[81] But the allusion gets at something equally fundamental. Both the identity and the attributes of data, databases, and datasets, including attributes implicating exclusivity and shareability, are matters of design as well as physics or economics.

5.2 Things and Groups

Significantly, social groups are among the most fundamental "designers," even with respect to such traditional resources as property in land. The legal historian Molly Brady, for example, has carefully documented that the historical meaning of the phrase "metes and bounds" in the law of real property refers to boundaries identified by local social and community practices, rather than to fixed boundaries specified by surveyors.[82]

In domains of knowledge and information, including data, the absence of a standard or uniform material reference (unlike land) means that the role of social relationships in constituting things and resources, both in social life and in legal processes, is both broader and deeper.[83] The argument draws on research in Technology Studies and Information Science, rather than legal scholarship.[84] Scholars have researched access to immaterial goods;[85] have explored governance of resources that generate additional resources (so-called "generative" phenomena);[86] and explored modern technologies such as open source computer programs, in which the group and the object constitute each other.[87]

Observing that things may be constructed socially, particularly for purposes of governance, does not imply that those processes of construction are simple or straightforward. (Nor does it imply that material objects do not have a physical reality.) The variability and complexity of those processes; the possibilities that they may or may not be linear and/or purposeful; the fact that they likely involve multiple social systems, including law; and the reality that individual actors in those systems, even within social groups, may have conflicting motivations, are precisely what give rise to the need to examine those processes critically.[88]

In commercial law settings, for example, two actors may agree by contract to treat a dataset as a tradeable commodity even while formal IP law considers that same information to be unowned and unownable. Customary practices in many fields construct domains of things for disciplinary purposes, such as the "copy" that has been the unit of text for both publishers and journalists. For public policy reasons, legal institutions may declare an absence of thing-like character, in order to deprive others of the power to claim property-like exclusivity in them. Patent law resists granting exclusive rights in laws of nature and abstract ideas. Property scholars who are committed to the central role of "things" in property law have begun to explore the legal "toolkit" of doctrines and arguments needed to construct property resources at different scales, producing an architecture of property things.[89]

As noted earlier, the GKC framework for researching knowledge

Business School Press 2000); Madison, 'Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo' (n 73); Thomas C Schelling, *The Strategy of Conflict* (Harvard Univ Press 1960); Susan Leigh Star and James R Griesemer, 'Institutional Ecology, `translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39' (1989) 19 *Social Studies of Science* 387.

79 Christine L Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (MIT Press 2015).

80 Brian Cantwell Smith, 'The Limits of Correctness' (1985) 14,15 *ACM SIGCAS Computers and Society* 18.

81 Claude Lévi-Strauss, *The Raw and the Cooked: Introduction to a Science of Mythology* (Penguin Books 1992).

82 Maureen E Brady, 'The Forgotten History of Metes and Bounds' (2019) 128 *Yale Law Journal* 872.

83 Michael J Madison, 'Law as Design: Objects, Concepts, and Digital Things' (2005) 56 *Case Western Reserve Law Review* 381.

84 Geoffrey C Bowker and Susan Leigh Star, *Sorting Things out: Classification and Its Consequences* (MIT Press 1999); Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford University Press 2007); Henry Petroski, *The Pencil: A History of Design and Circumstance* (Knopf 2006).

85 Jessica C Lai and Antoinette Maget (eds), *Intellectual Property and Access to Im/Material Goods* (Edward Elgar Publishing 2016).

86 Jonathan Zittrain, *The Future of the Internet and How to Stop It* (Yale University Press 2008).

87 Christopher M Kelty, *Two Bits: The Cultural Significance of Free Software* (Duke University Press 2008); Charles M Schweik and Robert C English, *Internet Success: A Study of Open-Source Software Commons* (MIT Press 2012).

88 Ellen P Goodman (ed), *The Atomic Age of Data: Policies for the Internet of Things* (Annual Aspen Institute Conference on Communications Policy 2015).

89 Thomas W Merrill and Henry E Smith, 'The Architecture of Property' in Hanoch Dagan and Benjamin Zipursky (eds), *Research Handbook on Private Law Theories* (Edward Elgar Publishing Forthcoming) <https://papers.ssrn.com/abstract=3462643> accessed 7 February 2020; Lee Anne Fennell, *Slices and Lumps: Division and Aggregation in Law and Life* (University of Chicago Press 2019).

commons emphasizes how social groups develop governance to address social dilemmas, or problems in cooperation.[90] Analysis of social dilemmas in complex settings may be simplified somewhat by techniques of "decomposing" large systems into small components.[91] Larger things may contain small things.

In sum, as to the identity of relevant resources, the possible absence of linearity and the importance of context should be emphasized, and over-reliance on *ex ante* categorization should be avoided. That point has particular significance with respect to data. Modern research demonstrates how scientific research consists of reciprocating processes rather than a progression from "basic knowledge" to "applied knowledge," including technology development and commercial application.[92] Likewise, research data production and management is now likewise often expressed in cyclical terms.[93] Data are sometimes characterized entirely as an infrastructural resource.[94] That focus highlights the many ways in which data use creates spillovers in multiple fields, in both expected and unexpected ways. But that infrastructural designation should be taken only as the beginning of an examination of appropriate governance, because infrastructure is a designed and socially constructed resource much as any other knowledge or information resource is.[95]

5.3 Things and Data

One strength of the word "resource" is that it properly evokes relationships between resources in resource systems or ecologies. Awareness of data ecologies for governance analysis aligns specifically with the emphasis that the GKC framework places on governance in broad context. An ecological perspective requires examining interdependencies between those resources and related resources, as systems, involving both immaterial and material attributes and evolution and variations across scales.

Understanding ecologies of data "things" should take account of the data collection and management practices associated with Big Data, with special attention given to the sources of the now-standard "three v's" of Big Data (volume, variety, and velocity), all the way down to hand-curated data collections. Different settings, resources, and resource systems may call for different governance judgments as to relevant social groups and as data-as-form and data-as-flow considerations.

Those settings and resources may include the following. The classification below is crude. Many overlaps exist among tools, products, services, and research outputs, and multiple opportunities exist to deploy characterizations of data-as-form and data-as-flow.

- Techniques and technologies for observation, experimentation, data collection, association, and construction of databases and datasets. These may include physical devices (the Internet of Things and the Internet of Bodies) as well as digital protocols, including computer programs, data formats, and other digital standards) for sensing and observing, for data transmission and communication, and for creating and managing the resulting data collections. MapReduce is an example of a digital computing paradigm for managing super-large datasets in a distributed computing environment.[96]
- Processes and systems of data stewardship, which emphasize cleaning, scrubbing, normalizing, manipulating, classifying, and maintaining data for storage, analysis, use and application.[97] Data ontologies, data schema, and data storage techniques and models are critical to ensure both technical synthesis and interoperability where data from multiple sources are brought together for use as shared resources, as in data repositories or other data infrastructures.
- Analytics, interpretations, and applications. These occupy an enormous analytic space in their own right, because "data" as governance subjects overlap with "algorithms," "AI," and "platforms" as technologies and institutions for data mining strategies; pattern analysis; and services, products, and new knowledge forms built on those patterns, as governance subjects. As machine learning technologies enable the automatic adjustment of data collection practices via embedded sensors, boundaries blur between data and AI. So-called smart machines learn from old data and collect new data differently. Data visualization tools are critical here, as are conceptual maps and models.[98]

The worlds of data may be changing and expanding so quickly, and this three-part division of data-related resources may be so imprecise, that it may seem unwise to advance the concept of things as a key governance concept. Yet two brief examples illustrate how focusing on things in governance, and particularly in commons governance of shared data, can illuminate specific data-related challenges.

A first example comes from outside the law, in coordination challenges among social groups within a given broad field. Academic researchers know this as the problem of coordinating across research disciplines. Because so much scholarly research now centers on data along with disciplinary knowledge, researchers confront new governance challenges even within institutions long associated with openness and sharing, such as scientific communities and research universities. The knowledge sharing norms of medical researchers overlap with but are also distinct from knowledge sharing norms of engineering researchers and social work researchers, for example. Data-as-form and data-as-flow have no consistent meanings, in practice, across different research traditions. In part, those differences are due to different histories of those fields. In part, those differences reflect different experiences with ethical frameworks, such as the Common Rule mentioned earlier. With respect to making productive uses of data, some of these differences and complexities can be bridged via computational techniques.[99] Others can addressed by research strategies that implement "de-composability" ideas, by building research products that interoperate in modular ways with research products from other fields, like Lego bricks.[100] But commons

[90] Strandburg, Frischmann and Madison (n 54).
[91] Simon (n 72).
[92] Donald E Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation* (Brookings Institution Press 2011).
[93] C Jung and others, 'Optimization of Data Life Cycles' (2014) 513 *Journal of Physics: Conference Series* 032047.
[94] Goodman (n 88).
[95] Frischmann, *Infrastructure* (n 50).
[96] A McKenna and others, 'The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data' (2010) 20 *Genome Research* 1297.
[97] Marcel Boumans and Sabina Leonelli, 'From Dirty Data to Tidy Facts: Clustering Practices in Plant Phenomics and Business Cycle Analysis' in Sabina Leonelli and Niccolo Tempini (eds), *Data Journeys in the Sciences* (Springer 2020) https://ore.exeter.ac.uk/repository/handle/10871/40283 accessed 7 February 2020.
[98] Tony Hey, Stewart Tansley and Kristin Tolle (eds), *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research 2009).
[99] Paul R Cohen, 'DARPA's Big Mechanism Program' (2015) 12 *Physical Biology* 045008.
[100] David Singh Grewal, 'Before Peer Production: Infrastructure Gaps and the Architecture of Openness in Synthetic Biology' (2017) 20 *Stanford Technolo-*

governance strategies based on flexible understandings of the natures of research "things" provide an important set of tools, bringing these approaches together via a systems perspective.[101]

A second example comes from within the law, from intellectual property law and its treatment of data. Here, the problem is that treating a data resource as data-as-form or as data-as-flow in one IP system may push actors to change their characterization of resources with respect to a different system. Recently, the US Supreme Court invalidated patents on genetic sequences isolated from human genes, in *Association for Molecular Pathology v. Myriad Genetics*, Inc.[102] That ruling undercut the power of the patentee, Myriad Genetics, to build a commercial business around genetic testing based on identifying those sequences in individuals. Those who supported invalidation and advocated for eliminating patent coverage of genetic sequences cheered. This appeared to be a win for research science, for the concept of knowledge flow, and to many, for better clinical health outcomes and public health. Yet it appears that Myriad has adjusted its business strategy, applying non-patent strategies to enhance the exclusivity of the pools of research data that were used to develop the patented inventions.[103] What law seems to provide in one legal domain (data-as-flow), it seems to take away in another, at least in part (data-as-form). Similar conflicts now exists with respect to public sector uses of DNA data in criminal proceedings, on the one hand, and trade secrecy law, on the other hand,[104] and between public health objectives and efforts to protect patient privacy by granting property rights in personal data to individual patients.[105] An ecological or systems approach may not solve these specific problems, but it would allow policymakers to anticipate them more clearly.[106]

It should be emphasized that thing-ness or resource forms, whether given, designed, or constructed by law or otherwise, should not be viewed as necessarily hostile to efforts to promote data openness and data sharing. So long as the character and attributes of a knowledge resource are matters of design, including legal reinforcement or disruption of thing-ness, then the design of resources can be tailored appropriately to relevant governance goals. Building a data repository of shared scientific data, for example, typically requires coordination and collaboration as to technical matters (can one dataset be combined or coordinated with another dataset as matters of code?), as to legal matters (are enabling or disabling contracts, licenses, covenants, and/or laws present?), and as to social, cultural, and economic matters (do libraries and archivists and research scientists and institutional administrators each understand, appreciate, and respect how field-specific expertise and other resources are needed to ensure the utility and stability of the repository?).[107]

In conversations that embody those challenges, data-as-flow can be a virtue (because giant data repositories can support streams of new, fantastic research), but it can also be a vice (because contributions of different fields and different resources may be difficult to separately identify and manage, in practice). With things as with groups, no panacea exists, that is, no "one size fits all" solution. Data governance counsels taking an adaptable stance on data-as-form and data-as-flow questions, rather than a rigid or ontological one.

6. Looking Ahead

This article offers a conceptual toolkit for data governance that centers on two big themes: groups and things. Those can be combined in various ways as part of developing approaches to governance data collection, production, storage, stewardship, and use. Knowledge commons is proposed as a significant overarching framework for using these tools in developing data sharing strategies, but the tools are also relevant to understanding market-based or state-based institutional governance. As a conceptual approach, the pair of tools comes with few necessary payoffs or implications. For example, stereotypical lessons such as "define resources with clarity" or "determine boundaries regarding access and use with specificity" may have grounding in research on natural resources by Ostrom and others, but perspectives on knowledge and information resources teaches that different guidance may apply in those contexts, or some of them.[108] The path forward lies as much in imaginative use of the concepts described here as in specific rules for specific problems. Four possible imaginative uses follow.

6.1 Examine Social Groups and Resources in Systems

Neither data governance nor knowledge commons should be implemented in a single way across all fields and domains. Large-scale initiatives to promote openness in research science, AI systems, urban planning, public administration and law, environmental regulation, and public health face the difficult but critical challenge of inventorying, understanding, and analyzing the technical, social and cultural, and legal attributes of polycentric ecologies. Data governance implies that collaboration strategies should be built out of those details.

That implication applies to private collaboratives such as the Open Data Initiative supported by Microsoft and other technology companies,[109] and to individuals and enterprises advancing the Panton Principles, calling for open data in science.[110] It applies to global NGOs focused on forward-looking uses of data such as AI for Good,[111] and private counterparts such as AI Commons[112] and Open AI.[113] It applies to governments. It applies to individual firms, to universities and research organizations, and even to individual policymakers, researchers, data scientists, and archivists.

Relatedly, too much emphasis in developing effective and appropriate data governance may be put on traditional distinctions between public and private enterprises and public and private goods. Similarly, too much emphasis may be put on identifying and reinforcing distinctions between data and algorithms. Last, too much emphasis

gy *Law Review* 143.

[101] Robert Cook-Deegan and Tom Dedeurwaerdere, 'The Science Commons in Life Science Research: Structure, Function, and Value of Access to Genetic Diversity' (2006) 58 *International Social Science Journal* 299.

[102] 569 U.S. 576 (2013).

[103] John M Conley, Robert Cook-Deegan and Gabriel Lázaro-Muñoz, 'Myriad after Myriad: The Proprietary Data Dilemma' (2014) 15 *North Carolina Journal of Law & Technology* 597.

[104] Sonia Katyal, 'The Paradox of Source Code Secrecy' (2019) 104 *Cornell Law Review* 1183.

[105] Jorge L Contreras, 'The False Promise of Health Data Ownership' (2019) 94 *New York University Law Review* 624.

[106] Helen Fay Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford Law Books 2010).

[107] Max von Grafenstein, Alina Wernick and Christopher Olk, 'Data Governance: Enhancing Innovation and Protecting Against Its Risks' (2019) 54 Intereconomics 228.

[108] Madison, 'IP Things as Boundary Objects' (n 77).

[109] https://www.microsoft.com/en-us/open-data-initiative.

[110] https://pantonprinciples.org/index.html.

[111] AI for Good is a United Nations platform for dialogue on future uses of artificial intelligence https://aiforgood.itu.int/.

[112] AI Commons is a non-profit organization collecting diverse contributions to ensure that the benefits of AI systems are broadly distributed. https://aicommons.com/.

[113] OpenAI is a private enterprise whose mission is to ensure that AI systems benefit all of humanity. https://openai.com/.

may be placed on the idea of data as an infrastructural resource and on data infrastructures. None of those distinctions are unimportant. How they are advanced, or modified, are questions for governance discussions.

6.2 Build Pragmatic Models of Policy Problems

Data are sometimes viewed optimistically, as enabling spillover individual and social benefits, and sometimes skeptically, as constraining individuals or imposing harms. An institutional governance framework supplies a useful method of integrating these different and sometimes disparate perspectives into a pragmatic, systems-based matrix.

Efforts to "regulate" data production and use via public/private matrixes or on a field-by-field basis have often proved to be inadequate or inflexible, because regulators, policymakers, and scholars have too often tried to squeeze something that "looks and feels" like an intellectual resource into the IP categories that were constructed over the course of the twentieth century for other intellectual resources: copyright, patent, trade secrets and confidential information, and related fields such as antitrust and unfair competition, and privacy.

Positive law is thus seen in part as providing ways of solving social dilemmas regarding shared resources such as data, by encouraging collaboration via supplying state subsidies for infrastructure; creating safe harbors for commercial collaboration and exemptions from unfair competition and antitrust charges; exempting information from exclusionary IP regimes; offering convening and facilitation services; and in other ways.[114] Positive law is also sometimes seen as impeding collaboration, creating social dilemmas rather than solving them. The idea of the anti-commons, in which a social space is characterized by too many separate property claims recognized by law, is one suggestive example.[115] An approach that organizes data regulation by traditional legal field struggles to reconcile those perspectives.

A promising model for integrating them and others, using a pragmatic approach based on a governance rubric, is the work of the political scientist Martha Finnemore and the legal scholar Duncan Hollis on constructing "cybernorms" for global cybersecurity governance.[116] They argue that managing global cybersecurity data is a systemic and ecological problem; that it does not fit standard policy-specific boxes for diagnoses or solutions; and that polycentric, group-based strategies are most likely to be effective on grounds of legitimacy and adaptability.

6.3 Expect Change, and Borrow From Experience

A pragmatic approach to data governance makes explicit that governance mechanisms must be adaptable, and they must be adaptable at different scales (small to large, slow to fast, local to global, existing to novel) and relative to different resources (human capabilities, social and institutional capabilities, and technological capabilities).

That emphasis on adaptability brings out a possibly surprising feature of governance, and in particular data governance, that focuses on social groups and on things: its receptivity to established governance mechanisms, even those that long pre-date the rise of Big Data, the internet era, or even twentieth century technology. Contemporary IP researchers have acquired a recent interest in informal, norm-governed innovation communities,[117] where formal systems of IP rights as such seem to contribute little or not at all to developing bodies of novel and creative work.

That interest in collective creativity can be traced back to the earliest days of research science, in the Republic of Letters and the early Enlightenment in England, Scotland, and continental Europe. Communities of scientific researchers formed face to face and correspondence networks, eventually becoming formalized in salons, scientific societies, and journals. This was not the practice of formal peer review. It was, instead, a polycentric network of social groups, regulating itself and the contents of their contributions via a complex system of social norms.[118] That centuries-old style of knowledge commons governance has been durable, adaptable, and effective. It may be relevant today.

6.4 Build Assessment Techniques

Perhaps the most difficult challenge to confront in designing and analyzing data governance is the question of assessment. Institutional design is significantly a question of comparative analysis. By what measure is one governance institution preferred to another?

Political theory, economic theory, and social theory have no shortage of answers. Social welfare analysis gives us attention to outputs (utility, including spillovers) and to inputs (human capabilities and capacities). Social choice theory asks us to assess the character of processes of collective choice regarding institutional arrangements. Should institutions aggregate or otherwise accurate reflect the preferences of their participants? Political philosophy directs us to ask questions about legitimacy, transparency, accountability, and protection of primary values of individual human autonomy, including powers of self-determination regarding participation in the polity.[119]

For example, a data governance community that sustains itself in coordination with the state differs from a nominally open community that proceeds only by relying on state-sanctioned legal instruments. Modern scientific research has the former character, given the abundant direct support and tax benefits offered to scientific research institutions and researchers themselves. Users of the Creative Commons licensing tool likely have the second character; mere use of a Creative Commons license, taken alone, does not enroll the user in a collective or community of any sort, and the license instrument itself is a salient and near cousin of proprietary licenses.[120] A group that manages an "open" resource, such as data, entirely via legal instruments, is apt to encounter incompatibility problems. Not every open data license defines "open" the same way.[121]

[114] Jorge L Contreras, 'Leviathan in the Commons: Biomedical Data and the State' in Katherine J Strandburg, Brett M Frischmann and Michael J Madison (eds), *Governing Medical Knowledge Commons* (Cambridge University Press 2017).

[115] Michael A Heller, 'The Tragedy of the Anticommons: Property in the Transition from Marx to Markets' (1998) 111 *Harvard Law Review* 621.

[116] Martha Finnemore and Duncan B Hollis, 'Constructing Norms for Global Cybersecurity' (2016) 110 *American Journal of International Law* 425.

[117] Kate Darling and Aaron Perzanowski (eds), *Creativity without Law: Challenging the Assumptions of Intellectual Property* (NYU Press 2017).

[118] Michael J Madison, 'The Republic of Letters and the Origins of Scientific Knowledge Commons' in Madelyn Sanfilippo, Katherine J Strandburg and Brett M Frischmann (eds), *Governing Privacy Commons* (Cambridge University Press forthcoming).

[119] Hanoch Dagan and Michael A Heller, 'The Liberal Commons' (2001) 110 *Yale Law Journal* 549.

[120] Niva Elkin-Koren, 'Creative Commons: A Skeptical View of a Worthy Pursuit' in Lucie Guibault and P Bernt Hugenholtz (eds), *Future of the Public Domain* (Kluwer Law International 2006).

[121] Alexandra Giannopoulou, 'Understanding Open Data Regulation: An Analysis of the Licensing Landscape' in Bastiaan van Loenen, Glenn Vancauwenberghe and Joep Crompvoets (eds), *Open Data Exposed*, vol 30 (TMC Asser Press 2018).

Measures of experience on the ground matter. Does knowledge commons governance work? Is governance durable and sustainable across time (generations) and space (relevant state and other organizational boundaries and borders)? Does practice align with relevant ideology, including relevant rhetorics, enhancing not only its descriptive legitimacy (acceptability to the community and to society) but also its normative claims?

The adaptability, flexibility, and even fuzziness of commons governance in information and data settings makes assessment even trickier. Stipulating that data-as-form and data-as-flow are key governance attributes, that data may exist in multiple interpreted forms and flows simultaneously, and that resources and groups are often engaged in projects of producing and re-producing one another, complicates classic governance distinctions between individuals and collectives, people and things, and subjects and objects.

7. Conclusion

In almost all contexts of interest for data governance purposes, data are likely to be shared. When, how, and why to share data are governance topics. This article has argued that the fundamental yet nonetheless pragmatic governance question for data is understanding different implications of seeing data-as-form and data-as-flow.

This is a conceptual argument. It is undoubtedly true that where law meets technology, whether on economic grounds or social and cultural terms, rules matter. Positive law matters, along with systems of social norms, customs, and conventions. Rights and interests matter, and their integration into regulatory frameworks matters, too. Nonetheless, the article recommends beginning not with the rules but with questions of institutional design, motivated by key concepts. A well-grounded domain of research exists focusing on shared knowledge, information, and data as objects and subjects of institutional governance. That domain is knowledge commons. Knowledge commons analysis argues for identifying and describing relevant social groups in which governance frameworks may be embedded, and for identifying and describing relevant resources, or things, whose form and flow will contribute substantially to the welfare effects of the relevant data governance systems. Those are tools for data governance.

This perspective takes an ecological or systems approach to regulatory questions, an approach in which market exclusivities and state mandates do not provide the standard two-part regulatory framing. Knowledge commons governance, in which data and information resources are shared according to governance rules tied to identified social and institutional collectives, provides a substantial third storehouse of data governance solutions.

05

Special issue: Governing Data as a Resource

Technology and Regulation

Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto

Teresa Scassa*

data governance, smart cities, data trust, open data, data ownership

Teresa.Scassa@uottawa.ca

Data governance for data sharing is becoming an important issue in the rapidly evolving data economy and society. In the smart cities' context, data sharing may be particularly important, but is also complicated by a diverse array of interests in data collected, as well as significant privacy and public interest considerations. This paper examines the data governance body proposed by Sidewalk Labs as part of its Master Innovation Development Plan for a smart city development on port lands in Toronto, Canada. Using Sidewalk Lab's Urban Data Trust as a use case, this paper identifies some of the challenges in designing an effective and appropriate data governance structure for data sharing, and analyzes the normative issues underlying these challenges. In this example, issues of data ownership and control are contested from the outset. The proposed model also raises interesting issues about the role and relevance of the public sector in managing the public interest; and the need to design data governance from the ground up. While the paper focuses on a particular use case, the goal is to distil useful knowledge about the design and implementation of data governance structures.

1. Introduction

Data is prized as a resource for technological innovation and economic development. Private and public sector companies, researchers, and civil society actors all seek to control and to access data, and governments increasingly seek to facilitate access. Data sharing may be between a few self-selected actors, or on a broad scale, with much in between.[1] Sharing is more complex where data sets include personal information or human behavioral data. These categories of data can significantly impact individuals and communities, raising important questions about privacy, dignity, autonomy, discrimination, expression and association.[2] The increased demand for data sharing creates a concurrent demand for data governance that can address competing claims to rights and interests in the governed data. These are not so much 'ownership' claims, although some may be framed in those terms. Rather, they are claims by groups and individuals to benefit from, or to not be harmed by, data in which they have an interest.

This paper considers data governance for data sharing through the lens of the data governance scheme proposed by Sidewalk Labs as part of its Master Innovation Development Plan (MIDP)[3] for a 'smart city' development on the waterfront of Toronto, Canada. Recognizing the diverse interests in the data that might be collected in the development, the MIDP called for the creation of an Urban Data Trust (UDT) as a data governance body to address both the collection and the sharing of the novel category of 'urban data'. Data governance bodies, whether labelled data trusts or otherwise, have generated considerable interest as a means of facilitating data sharing while accommodating different interests in data. This paper uses the example of urban data and the UDT to illustrate some of the challenges that are central to data governance for data sharing.

This paper begins with a discussion of the concept of data governance for data sharing. Recognizing both the importance of and the unique characteristics of data in a digital society, it draws upon Frischmann, Madison and Strandburg's proposal for a framework for governance that seems well suited to the smart cities context. They define the 'knowledge commons' as "the institutionalized community governance of the sharing, and, in some cases, creation, of information, science, knowledge, data, and other types of intellectual and cultural resources".[4] Their framework recognizes that the goal is not just to store data securely, but rather to share it to build new knowledge and tools in service of a common goal or in accordance with shared values. The first part of this paper explores the 'knowledge commons' as an organizing framework for data governance and links it to the

[1] A diversity of sharing arrangements can fit within the concept of a 'data trust'. See: Jack Hardinges, 'What is a Data Trust?' (*Open Data Institute*, 10 July 2018) https://theodi.org/article/what-is-a-data-trust accessed 27 April 2020.

[2] Although the widespread collection of personal data is most often associated with individual privacy, some scholars raise concerns about other privacy harms to both individuals and communities. See, e.g., Linnet Taylor, Luciano Floridi and Bart van der Sloot, 'Introduction: A New Perspective on Privacy' in Linnet Taylor, Luciano Floridi, and Bart van der Sloot (eds) *Group Privacy* (Springer 2017).

Teresa Scassa is the Canada Research Chair in Information Law and Policy at the University of Ottawa. An earlier version of this paper was presented at the workshop 'Governing Data as a Resource' organized at Tilburg University in November 2019. I am grateful to the thoughtful feedback received from many participants, and particularly to the comments of my paper's discussant, Michael Madison, and the anonymous peer reviewers. Thanks also to Pamela Robinson for her comments on an earlier draft and to my research assistant Tunca Bolca. Although I am a member of Waterfront Toronto's Digital Strategy Advisory Panel, my comments in this paper are my own and do not represent the views of that panel or of Waterfront Toronto.

[3] Sidewalk Labs, 'Master Innovation Development Plan' (*Sidewalk Labs*, June 2019) https://quaysideto.ca/sidewalk-labs-proposal-master-innovation-and-development-plan accessed 27 April 2020 [MIDP].

[4] Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg, *Governing Knowledge Commons* (OUP 2014) 3.

particular context of Sidewalk Toronto. Section 2 introduces the Sidewalk Toronto project and considers Sidewalk Labs' initial data governance proposal and how it shifted over time and in response to public reaction. The knowledge commons framework requires a consideration of the background issues that shape the data sharing context, the resources to be shared, and the key governance elements. Each of these issues is dealt with respectively in sections 3, 4 and 5 of this paper. The conclusion identifies the issues that led to the failure of the UDT, and extracts key lessons.

2. Data Governance

The growing importance of data in the information society and economy, and the rise of data-dependent technologies such as Artificial Intelligence (AI) have created a demand for data sharing on an unprecedented scale. While data governance has always been a part of the operational reality of governments and organizations that collect and use data, data governance for data sharing is something quite different. In the first place, it is no longer about the managing of data to meet the needs and legal obligations of a specific organization. Rather, it is about governing data so as to enables its sharing with other entities or organizations to meet polycentric objectives. Data sharing can be broad and indiscriminate (as with government open data regimes), or it can be limited to one or two consenting organizations – or anything in between.

Typically, governance obligations fall on those who 'own' or 'control' data. Conventional forms of data governance – usually within a single organization (whether public or private sector) are premised on some notion of control, whether it is expressed as 'ownership' or as custody over the data. Rights to and interests in data are rooted in law, shaped by policy and practice, and negotiated in private agreements.[5] In some jurisdictions such as Canada and the UK, public sector interests in data are framed as a kind of ownership right.[6] In other jurisdictions there is no specific legal construct, other than a general custodial duty with respect to state information and data. Europe's Directive on Open Data and Public Sector Information,[7] for example, is not framed in terms of public ownership of data or information. Nevertheless, it clearly recognizes obligations of each member state to manage its information and data in the public interest.[8] Public sector right-to-know legislation establishes government as an information steward; it holds it and provides (or denies) public access in the public interest.[9] Open data policies also guide how and in what circumstances public sector data is shared with the public.[10]

A private sector organization may base its rights to control access to and use of its data through a combination of intellectual property law (copyright law[11] and the law of confidential information[12] in particular) as well as physical barriers and the laws that support them (such as trespass, technological protection measures, and criminal law).[13] Access to and use of data is governed by contracts and licences. The organization's data governance practices may also be shaped by data protection laws, as well as evolving standards regarding cybersecurity.

Law also shapes the different interests of individuals and organizations in data. Individuals have interests in their own personal data, notwithstanding any proprietary claims to the same data that might be asserted by public or private sector actors. Public and private sector data protection laws provide a framework for the recognition and exercise of individual rights and interests in personal data. These interests confer a degree of control, including rights of access, erasure (in some circumstances), and portability (in some contexts).[14] As the number and nature of the rights of individuals to their personal data expands, these rights are increasingly labeled 'ownership' rights.[15]

Normally any plan to collect data will include data governance. Where private sector companies collect data, data protection laws establish parameters for managing the data, and these must be integrated into an organization's overall data governance scheme. Data protection laws also establish the data subjects' interest in their data in the

5 Joshua B. Fisher & Louise Fortmann, 'Governing the data commons: Policy, practice, and the advancement of science' (2020) 47 *Information & Management* 237, 237.

6 In Canada, Crown copyright is provided for in s. 12 of the Copyright Act, RSC 1985 c. C-42. In the UK, Crown copyright is found in s. 163 of the Copyright, Designs and Patents Act 1988. See also: Copyright Act 1968, No. 63, 1968 (Australia), Part VII. In the United States, the Copyright Act, 17 USC §107 declares that there is no copyright in works of the federal government. However, this does not prevent state governments from asserting copyright in their works. Different states take different approaches. See: Marketa Trimble, 'U.S. State Copyright Laws: Challenge and Potential' (2017) *Scholarly Works* 1019, 84-85 https://scholars.law.unlv.edu/facpub/1019 accessed 27 April 2020.

7 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information [2019] OJ L 172/56.

8 It also identifies the different public interests in access to and re-use of public sector information. See, e.g. Ibid., Recitals 16 and 31.

9 Ontario's Information and Privacy Commissioner emphasizes the importance of the role of the public sector as data steward in smart cities under public sector data governance legislative frameworks. (Brian Beamish, 'Open Letter to Stephen Diamond, Chairman of the Board of Directors, Waterfront Toronto' (Information and Privacy Commissioner of Ontario, 24 September 2019) https://www.ipc.on.ca/wp-content/uploads/2019/09/2019-09-24-ltr-stephen-diamond-waterfront_toronto-residewalk-proposal.pdf accessed 27 April 2020.

10 See, e.g., EU Directive on Open Data and Public Sector Information (n 7); Simpler, Faster, Better Services Act, 2019, SO 2016, c 7, Sch 56; Treasury Board Secretariat, 'Directive on Open Government' (Canada, 9 October 2014) https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=28108 accessed 27 April 2020.

11 Although copyright law places facts in the public domain, compilations of data can be protected as 'works'. Article 10(2) of the TRIPS Agreement provides: "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself." (Agreement on Trade-Related Aspects of Intellectual Property Rights, 15 April 1994, [1994] 1869 U.N.T.S. 299, 33 I.L.M. 1197.

12 E.g., art. 39 of the TRIPS Agreement, ibid, establishes criteria for the protection of confidential information. What is rewarded is not just the investment in the collection of commercially important information, but the efforts made to control that information and to maintain its confidentiality. Any 'proprietary' dimensions are rooted in physical and legal control, as opposed to 'authorship'.

13 See, e.g., Teresa Scassa, 'Data Ownership', (2018) *CIGI Papers* No. 187 https://www.cigionline.org/publications/data-ownership accessed 27 April 2020.

14 Rights of erasure and of data portability are features of the EU GDPR in respectively , articles 17 and 20. General Data Protection Regulation, persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (GDPR) [2016] OJ L119/1. The Australian Consumer Data Right also includes a data portability element. See: Treasury Laws Amendment (Consumer Data Right) Bill 2019, Bills Digest No. 68, 2018–19 https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6370 accessed 27 April 2020.

15 Although inaccurate, the term 'ownership' recognizes the steady expansion of these interests. For an example of such usage, see: British Academy, 'Data ownership, rights and controls: Reaching a common understanding: Discussions at a British Academy, Royal Society and techUK seminar on 3 October 2018', (British Academy 2018), 3-4 https://royalsociety.org/-/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf accessed 27 April 2020.

hands of the organization. While, as discussed below, an organization might assert rights over its data, individual interests in personal data must also be respected.[16] The public sector is similarly also responsible for the governance of the data it collects and is bound by laws, including those relating to access to information and data protection, as well as internal policies and directives. In these cases, however, the collecting organization is either private or public sector in nature, and its data governance is shaped by existing legal frameworks.

The smart cities context implicates multiple parties with interests in data. This can include different private sector actors, one or more levels of government, and a range of other stakeholders that include urban residents individually and collectively. In some cases, the nature and/or volume of the data to be collected, the obvious demand for access to the data, the individual or group interests in the data, or the need for compromise between public and private sector partners, may call out for the creation of a new data governance framework to facilitate data sharing according to articulated values. This is particularly the case where there is a more systematic collection of greater volumes of data, along with plans for more extensive data sharing – particularly personal data or human behavioural data.[17] Data governance for data sharing in this context goes well beyond bilateral data sharing agreements and requires a novel approach. Such approaches have come to be labelled almost colloquially as 'data trusts'.

Beneath the label of 'data trust' is a concept of pooled or shared resources subject to a collective understanding around access or use. This evokes the concept of a 'knowledge commons'. This term invokes both pooled resources and collective governance,[18] reflecting a collective decision-making process.[19] Governance is of this sort incorporates collective action and approaches. Michael Madison invokes Elinor Ostrom's concept of governing the commons, where she states that "a core goal of public policy should be to facilitate the development of institutions that bring out the best in humans."[20] According to Madison, commons as governance involves communally or collectively determined principles that shape and enforce managed access to a shared resource.[21]

Frischman et al created a set of questions organized around key issues for analyzing and understanding a knowledge commons. Their framework recognizes four key elements: 1) the background environment or context in which the commons arises; 2) the attributes of the commons, including what resources are to be pooled, who the relevant community members are, and what goals and objectives it is meant to serve; 3) the governance framework for the commons including governance mechanisms, decision-makers, and relevant institutions and infrastructures. Relevant governance issues also include the applicable norms and laws and the ways in which members interact; and 4) what patterns and outcomes are relevant, including the benefits, costs and risks.[22] These questions shape the discussion below, with a particular emphasis on the background and context, the attributes of the commons, and the governance framework. Concerns about patterns and outcomes are integrated into the discussion of the background environment and context. Not only were these not well articulated in the governance proposal, its failure renders them moot. Nevertheless, benefits, costs and risks are part of the public discussion that shaped the development of the governance framework.

This knowledge commons framework identifies and organizes the issues at the core of data governance design. The relevance of these questions is evident in the case of Sidewalk Toronto. Yet as will be seen in the discussion below, there were, in a sense, two parallel processes for developing a governance framework. One came from Sidewalk Labs itself in the form of the Urban Data Trust (UDT) proposed in the MIDP. The other was a kind of public discussion occurring on many fronts that articulated different visions of a commons based in part on other known models and in part on critiques of the UDT. Not only did the public discussion shape the UDT, it likely also informed Waterfront Toronto's rejection of the proposal. In the governance vacuum created by the demise of the UDT, the knowledge commons framework remains a useful tool to shape a new approach to governance.

3. Background Environment and Context: The Origins of the Sidewalk Toronto Smart Cities Project

The knowledge commons framework identifies the background environment and context in which a commons arises as a primary consideration. In the case of Sidewalk Toronto, this context was particular and unusual and clearly played a significant role in shaping both the governance solution proposed and indeed the entire conversation around governance.

The Sidewalk Toronto smart city development originated in a Request for Proposals (RFP) issued by Waterfront Toronto for the development of a portion of port lands in the City of Toronto, Canada.[23] The Sidewalk Toronto[24] development had unique features that sharply distinguished it from other smart city projects. Most importantly, it was not led by Toronto City Council, nor was it part of a broader smart city initiative.[25] This distinguished it from cities such as Barcelona, which reflect a concerted, overall strategy driven by an elected municipal

[16] In Canada, this includes a right to access one's personal data in the hands of an organization. More extensive rights, such as rights to data portability or the right to erasure are features of the GDPR, (n 14).

[17] Diverse data governance frameworks are emerging to address data sharing in a range of contexts. See, e.g., Teresa Scassa and Merlynda Vilain, 'Governing Smart Data in the Public Interest: Lessons From Ontario's Smart Metering Entity' *CIGI Paper* #221 (CIGI 2019) https://www.cigionline.org/publications/governing-smart-data-public-interest-lessons-ontarios-smart-metering-entity accessed 27 April 2020; Open Data Institute, 'Data Trusts: Lessons from Three Pilots' (*ODI* 2019) https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT30ObnYJGibEhspr2v87jg/edit accessed 27 April 2020.

[18] Frischmann et al, (n 4).

[19] Michael J. Madison, 'Tools for Data Governance' (2020) *Technology and Regulation* 29.

[20] Elinor Ostrom, 'Beyond Markets and States: Polycentric Governance of Complex Economic Systems' (2010) 100:3 *American Economic Review* 641, 665.

[21] Madison (n 19).

[22] Frischman et al, (n 4), 20-12.

[23] The focus of this paper is on data governance. For an examination of the broader project, see: Ellen P. Goodman and Julia Powles, 'Urbanism Under Google: Lessons From Sidewalk Toronto' (2019) *Fordham LR* 457.

[24] I use the name 'Sidewalk Toronto' to refer to the proposed development. 'Quayside' is the name of the parcel of land to be developed under the agreement between Waterfront Toronto and Sidewalk Labs. Waterfront Toronto has often referred to the project as the Quayside development; Sidewalk Labs uses 'Sidewalk Toronto'.

[25] Note that Sidewalk Labs largely avoids the 'smart city' label. See, e.g., Sidewalk Labs, 'Sidewalk Labs is reimagining cities to improve quality of life', n.d. https://www.sidewalklabs.com/ accessed 27 April 2020. In the MIDP, Sidewalk Labs writes: "This effort defines urban innovation as going beyond the mere pursuit of urban efficiencies associated with the 'smart cities' movement, towards a broader set of digital, physical, and policy advances that enable government agencies, academics, civic institutions, and entrepreneurs both local and global to address large urban challenges." (*MIDP*, 'Overview' Volume 0 (n 3), 138.

government.[26] It was also different from developments in other Canadian cities such as Montreal[27] or Edmonton.[28] Unlike a smart city in which city council and city officials take the lead, the primary 'public' actor in the Sidewalk Toronto project was not a public body at all. Rather, Waterfront Toronto is a non-profit corporation created by three levels of government – federal, provincial and municipal – to oversee development of Toronto's port lands, [29] in which, due to the vagaries of geography and the Canadian constitution, all three levels of government have an interest.

According to Waterfront Toronto, its mandate is "to deliver a connected waterfront that belongs to everyone, serving as a leading example of innovation and excellence in urban design, a magnet for investment and job creation, and a source of pride and inspiration for Canadians."[30] Created in 2001 (as the Toronto Waterfront Revitalization Corporation), Waterfront Toronto had already coordinated several development projects along Toronto's waterfront. On March 17, 2017 Waterfront Toronto issued an RFP for the development of Quayside, a 12-acre parcel of port land.[31] In May 2017, Waterfront selected a proposal by Sidewalk Labs, an Alphabet company.[32] The parties entered into a Framework Agreement on October 16, 2017.[33] On July 31, 2018, Sidewalk presented its Plan Development Agreement (PDA) to Waterfront Toronto,[34] in which it set out its preliminary proposal for the Quayside project. The proposal was for a 'smart city' to be developed from the ground up, embedded with sensors as part of a vision of the "potential for technology to improve urban life and to create people-centered communities that are more livable, connected, prosperous and resilient."[35] The PDA also referred to the creation of "a destination for people, companies, start-ups and local organizations to advance solutions to the challenges facing cities . . . and make Toronto the global hub of a rising new industry focused on urban innovation."[36] After receiving feedback on the PDA, Sidewalk began working on its Master Innovation Development Plan (MIDP), which was submitted in June 2019 and made public on June 24, 2019.[37]

While the initial press coverage of the Sidewalk Toronto project showed interest in and openness to its futuristic promise,[38] the project quickly sparked a vocal public reaction. Critics raised multiple concerns, including lack of transparency,[39] experimentation on Toronto's citizenry,[40] lack of long-term viability,[41] and insufficient civic participation.[42] Local start-ups were concerned that they would be shut out of the development, and that proprietary standards might create a kind of vendor lock-in.[43] There was considerable public

[26] See, e.g., Tuba Backici, Esteve Almirall, and Jonathan Wareham, 'A Smart City Initiative: the Case of Barcelona' (2013) 4 *J Knowl Econ* 135; Josep-Ramon Ferrer, 'Barcelona's Smart City vision: an opportunity for transformation' (2017) 16 *Field Actions Science Reports* 70; Mila Gasco, 'What Makes a city smart? Lessons from Barcelona' (2016) *49th Hawaii International Conference on System Sciences* 2983. See also the discussion of the concept of an 'open smart city' in : Tracey Lauriault, Rachel Bloom, and Jean-Noé Landry, 'Open Smart Cities Guide V1.0' (Open North 2018) https://docs.google.com/document/d/1528rqTjzKWwk4s2xKuPf7ZJg-tLlRK8WcMZQbi-coGTM/edit accessed 27 April 2020.

[27] See, e.g., Montreal Urban Innovation Lab https://laburbain.montreal.ca/en accessed 27 April 2020.

[28] See City of Edmonton, 'Edmonton: Smart City' n.d. https://www.edmonton.ca/city_government/initiatives_innovation/smart-cities.aspx accessed 27 April 2020.

[29] For a description and a map of the Quayside area, see: Waterfront Toronto, 'Quayside', n.d. https://www.waterfrontoronto.ca/nbe/portal/waterfront/Home/waterfronthome/projects/quayside accessed 27 April 2020.

[30] Waterfront Toronto, 'Note to Reader: Waterfront Toronto's Guide to reading the draft Master Innovation and Development Plan proposal submitted by Sidewalk Labs' (28 June 2019), 3 https://quaysideto.ca/wp-content/uploads/2019/06/Note-to-Reader_June-28-2019_Waterfront-Toronto.pdf accessed 27 April 2020.

[31] Waterfront Toronto, 'Request for Proposals: Innovation and Funding Partner for the Quayside Development Opportunity' (17 March 2017) https://quaysideto.ca/wp-content/uploads/2019/04/Waterfront-Toronto-Request-for-Proposals-March-17-2017.pdf accessed 27 April 2020 [RFP]. The RFP called for a "globally significant demonstration project that advances a new market model for climate-positive urban developments" (at 9).

[32] Sidewalk Labs, 'About Sidewalk' n.d. https://www.sidewalklabs.com/ accessed 27 April 2020. In its 2018 report, the Auditor General of Ontario criticized the very short time period allowed for responses to the RFP. See: Office of the Auditor General of Ontario, Annual Report 2018, Vol. 1, 31 http://www.auditor.on.ca/en/content/annualreports/arreports/en18/2018AR_v1_en_web.pdf accessed 27 April 2020.

[33] Waterfront Toronto, 'Framework Agreement among Toronto, Waterfront Revitalization Corp., Sidewalk Labs LLC and Sidewalk Toronto LP' (16 October 2017) https://www.waterfrontoronto.ca/nbe/wcm/connect/waterfront/035e8ad1-6ba2-46f6-8915-707176baa40f/Framework+Agreement_Executed_SUPERSEDED.pdf?MOD=AJPERES accessed 27 April 2020.

[34] Waterfront Toronto, 'Plan Development Agreement between Toronto Waterfront Revitalization Corporation and Sidewalk Labs LLC', as amended, July 31, 2018 https://waterfrontoronto.ca/nbe/wcm/connect/waterfront/73ac1c93-665b-4fb8-b19b-6bfa23c2a427/PDA+July+31+Fully+Executed+%28002%29.pdf?MOD=AJPERES accessed 27 April 2020.

[35] Plan Development Agreement (n 34) 2-3.

[36] Plan Development Agreement (n 34) 3.

[37] MIDP (n 3). After review and feedback, the MIDP was followed by a Digital Innovation Appendix, released. See: Sidewalk Labs, 'Master Innovation Development Plan: Digital Innovation Appendix' (14 November 2019) https://quaysideto.ca/wp-content/uploads/2019/11/Sidewalk-Labs-Digital-Innovation-Appendix.pdf accessed 27 April 2020.

[38] Kate McGillivray, 'Inside Quayside, the hyper-modern, tech-friendly development coming to Toronto's waterfront' (*CBC News*, 10 May 2017) https://www.cbc.ca/news/canada/toronto/quayside-waterfront-toronto-1.4108717 accessed 27 April 2020; Patrick Lynch, 'Sidewalk Labs Announces Plans to Create Model Smart City on Toronto's Waterfront,' (*Arch Daily*, 17 October 2017) https://www.archdaily.com/881824/sidewalk-labs-announces-plans-to-create-model-smart-city-on-torontos-waterfront accessed 27 April 2020; Andrea Hopkins & Alistair Sharp, 'Toronto to be home to Google parent's biggest smart city project yet' (*Financial Post*, 17 October 2017) https://business.financialpost.com/technology/google-to-be-anchor-tenant-at-toronto-innovation-hub-government-source accessed 27 April 2020; David George-Kosh, 'Alphabet's Sidewalk Labs to Create 'Smart' Neighborhood on Toronto Waterfront' (*Wall St Journal*, 17 October 2017) https://www.wsj.com/articles/alphabets-sidewalk-labs-to-create-smart-neighborhood-on-toronto-waterfront-1508266001 accessed 27 April 2020.

[39] Alanna Rizza, 'Critics call for more transparency for Sidewalk labs neighbourhood in Toronto' (*CTV News*, 8 December 2018) https://www.ctvnews.ca/sci-tech/critics-call-for-more-transparency-for-sidewalk-labs-neighbourhood-in-toronto-1.4210519 accessed 27 April 2020; Brian Barth, 'The fight against Google's smart city' (*The Washington Post*, 8 August 2019) https://www.washingtonpost.com/news/theworldpost/wp/2018/08/08/sidewalk-labs, accessed 27 April 2020; Mariana Valverde, 'The controversy over Google's futuristic plans for Toronto' (*The Conversation*, 30 January 2018) http://theconversation.com/the-controversy-over-googles-futuristic-plans-for-toronto-90611 accessed 27 April 2020 .

[40] See, e.g., Molly Sauter, 'Google's Guinea-Pig City', (*The Atlantic*, 13 February 2018) https://www.theatlantic.com/technology/archive/2018/02/googles-guinea-pig-city/552932/ accessed 27 April 2020; Star Editorial Board, 'Sidewalk Labs community can't be just a techno-experiment' (*Toronto Star*, 10 October 2018) https://www.thestar.com/opinion/editorials/2018/10/10/sidewalk-labs-community-cant-be-just-a-techno-experiment.html accessed 27 April 2020.

[41] This was one of the many concerns on a published list of questions for Sidewalk Labs. See 'Key (Mostly Unanswered) Questions Regarding Sidewalk Toronto Project' n.d. https://cfe.ryerson.ca/key-resources/guidesadvice/key-mostly-unanswered-questions-regarding-sidewalk-toronto-project accessed 27 April 2020. See also: John Lorinc, 'A Mess on the Sidewalk', (The Baffler, March 2019) https://thebaffler.com/salvos/a-mess-on-the-sidewalk-lorinc accessed 27 April 2020.

[42] See, e.g., Jathan Sadowski, 'Google wants to run cities without being elected. Don't let it' (*The Guardian*, 24 October 2017) https://www.theguardian.com/commentisfree/2017/oct/24/google-alphabet-sidewalk-labs-toronto accessed 27 April 2020.

[43] Kurtis McBride, 'Monetizing Smart Cities' (*Building*, 24 August 2018) https://building.ca/feature/monetizing-smart-cities/ accessed 27 April

outcry over issues of privacy, surveillance, and data sovereignty.[44] As the project evolved, some critics questioned the business plan underlying the deal, voicing concerns that it might be a 'real-estate grab' orchestrated by Sidewalk Labs.[45] The opposition culminated in a #BlockSidewalk movement,[46] and the Canadian Civil Liberties Association launched a lawsuit against Waterfront Toronto and the three levels of government in April of 2019, alleging that the project breached residents' constitutional rights.[47] While some have criticized opponents for their resistance to the benefits that the proposal might have for Toronto,[48] the lack of democratic/civic participation in the high-technology development was, for many, a fundamental defect.[49] Ultimately, the project involved private development with significant consequences for more than just land, creating new governance challenges. The project fell outside of traditional public sector participatory governance frameworks and outside of traditional land development paradigms.[50]

Although the original RFP called for plans to develop the Quayside district, the MIDP submitted by Sidewalk Labs distinguished between Quayside and the Innovative Design and Economic Acceleration (IDEA) District, and made proposals for both. The IDEA district included Quayside, but was much larger.[51] While Quayside represented a 4.9 hectare or 12-acre area, the IDEA district extended over 77 hectares or 190 acres. Sidewalk Labs suggested that issues of scale were behind this geographic extension. The data governance scheme proposed in the MIDP was for the larger IDEA district.[52]

Sidewalk Labs and Waterfront Toronto both sought to backfill the perceived democratic deficit with extensive public consultations leading up to the release of the MIDP and continuing afterwards.[53] The MIDP itself sought to allay many of the concerns raised by opponents of the project. Sidewalk Labs stepped back from the 'smart cities' label, recasting the project as one focusing on urban innovation.[54] Its data governance scheme (which changed shape from the PDA to the MIDP) was designed to address multiple concerns relating to the collection and sharing of data within the proposed development.

The initial proposal for Quayside framed it as a high-tech smart city development from the ground up, with a digital layer fully integrated from the design stage.[55] However, the proposal contained no clear plan for data beyond assurances that privacy would be protected through deidentification at source and the adoption of Privacy by Design (PbD) principles.[56] Data governance was an awkward issue for this project. This might have been in part because Waterfront Toronto is not the agent of any particular government and is itself not a party that would assert 'ownership' in generated data. The process by which the MIDP came about was therefore different from normal government procurement. In addition, the proposed development was not clearly either public or private in character. The project had

2020.

44 Bianca Wylie, 'Sidewalk Toronto and the Manufacturing of Consent — Thoughts Heading into Public Meeting 2 of 4' (*Medium*, 18 April 2018) https://medium.com/@biancawylie/sidewalk-toronto-and-the-manufacturing-of-consent-thoughts-heading-into-public-meeting-2-of-4-9acd289e9fa8 accessed 27 April 2020; Laura Bliss, 'How Smart Should a City Be? Toronto Is Finding Out' (*Citylab*, 7 September 2018) https://www.citylab.com/design/2018/09/how-smart-should-a-city-be-toronto-is-finding-out/569116/ accessed 27 April 2020; John Lorinc, 'A Mess on the Sidewalk' (n 43). On the issue of data localization, Sidewalk Labs was initially resistant to the concept. See, e.g., Alyssa Harvey-Dawson, 'An Update on Data Governance for Sidewalk Toronto' (Sidewalk Labs, 15 October 2018) https://www.sidewalklabs.com/blog/an-update-on-data-governance-for-sidewalk-toronto/ accessed 27 April 2020.

45 See, e.g., David Skok, 'Cracks in the Sidewalk' (*Macleans*, 15 February 2019) https://www.macleans.ca/opinion/cracks-in-the-sidewalk/ accessed 27 April 2020; Bianca Wylie, 'Sidewalk Toronto: Here's the Business Model Framework' (*Medium*, 7 June 2018) https://medium.com/@biancawylie/sidewalk-toronto-waterfront-toronto-digital-strategy-advisory-panel-meeting-1-before-6a158971eb65 accessed 27 April 2020.

46 #BlockSidewalk, n.d. https://www.blocksidewalk.ca/ accessed 27 April 2020.

47 Canadian Civil Liberties Association, 'CCLA Commences Proceedings Against Waterfront Toronto' (16 April 2019) https://ccla.org/ccla-commences-proceedings-waterfront-toronto/ accessed 27 April 2020.

48 Stephanie Marotta, 'Business leaders push for Sidewalk Labs smart-city development to be built on Toronto's waterfront' (*Globe and Mail*, 4 July 2019) https://www.theglobeandmail.com/business/article-business-leaders-push-for-sidewalk-labs-smart-city-development-to-be/ accessed 27 April 2020. This article references an open letter published by local business leaders. See Toronto Region Board of Trade, 'Open Letter from Civic Leaders', 4 July 2019 https://www.bot.com/Portals/0/NewsDocuments/742019Civic%20Leaders%20Open%20Letter%20final.pdf accessed 27 April 2020.

49 Bianca Wylie, 'Democracy or Sidewalk Toronto. You Can Have One But You Can't Have Both' (*Medium*, 14 May 2019) https://medium.com/@biancawylie/democracy-or-sidewalk-toronto-you-can-have-one-but-you-cant-have-both-a40e4d1d8daa accessed 27 April 2020; Michael Oliviera, 'Critics decry lack of 'democratic participation' over Sidewalk Labs' proposed neighbourhood' (*Toronto Star*, 2 May 2018) https://www.thestar.com/news/gta/2018/05/02/critics-decry-lack-of-democratic-participation-over-sidewalk-labs-proposed-neighbourhood.html accessed 27 April 2020. See also Goodman & Powles (n 23).

50 An interesting analogy might be made with projects that have significant environmental impacts. These projects necessarily combine economic and development priorities with complex public interest and environmental concerns. In the environmental regulation context, there are complex frameworks for the assessment and approval of such projects. It is also worth noting that the concept of 'social licence' has its roots in the environmental context. See: Kristen van de Biezenbos, 'The Rebirth of Social Licence' (2019) 14 *McGill J. Sust. Dev. L.* 149.

51 See the map of the areas under discussion in: Swerhun, Inc., Waterfront Toronto's Public Consultation on the draft MIDP: Round One Feedback Report (Toronto, 19 September 2019) 5, https://quaysideto.ca/wp-content/uploads/2019/09/Round-One-Public-Consultation-Feedback-Report-September-19-2019.pdf accessed 27 April 2020.

52 Waterfront Toronto, Note to Reader (n 30) 5. The expansion of the area to form part of the proposal was not approved by Waterfront Toronto, and the project has since been scaled back: Stephen Diamond, 'Open Letter from Waterfront Toronto Board Chair' (31 October 2019) https://waterfrontoronto.ca/nbe/wcm/connect/waterfront/waterfront_content_library/waterfront+home/news+room/news+archive/news/2019/october/open+letter+-from+waterfront+toronto+board+chair+-+october+31%2C+2019 accessed 27 April 2020.

53 See, e.g., Waterfront Toronto, 'Quayside: Participate in a Public Consultation' n.d. https://quaysideto.ca/get-involved/public-consultation/ accessed 27 April 2020. Sidewalk Labs' public outreach is described in the MIDP (n 3), Volume 0, 67. A summary of Waterfront Toronto's public consultation, carried out after the release of the MIDP, was published in September 2019. See: Swerhun, Inc (n 52) 5.

54 In the introductory Volume to its MIDP, Sidewalk Labs writes: "This effort turned Sidewalk Labs' initial ideas, as expressed in the RFP response, into a development plan with the potential to serve as a demonstration for an inclusive community that puts urban innovation to work for better quality of life." (See MIDP (n 3), Volume 0, 86).

55 PDA (n 34) 49. See also Dan Doctoroff, 'Reimagining cities from the Internet up' (*Medium*, 30 November 2016) https://medium.com/sidewalk-talk/reimagining-cities-from-the-internet-up-5923d6be63ba accessed 28 April 2020.

56 Brian Jackson, 'Sidewalk Toronto commits to Privacy by Design principles amid citizen concerns' (*IT World Canada*, 4 May 2018) https://www.itworldcanada.com/article/sidewalk-toronto-commits-to-privacy-by-design-principles-amid-citizen-concerns/404887 accessed 28 April 2020; Ann Cavoukian, 'De-identifying data at the source is the only way Sidewalk can work' (*Toronto Life*, 4 September 2019) https://torontolife.com/city/de-identifying-data-at-the-source-is-the-only-way-sidewalk-can-work/ accessed 28 April 2020. Privacy by design principles focus on embedding privacy into the design of technology. See: Ann Cavoukian, 'Privacy by design: The 7 Foundational Principles' (Information and Privacy Commissioner of Ontario, January 2011) https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf accessed 28 April 2020.

definite public dimensions: it involved publicly owned lands, was originally labelled a 'smart city', and it implicated traditional, municipal services. At the same time, it was also a real estate development and a technology innovation hub.[57] The knowledge commons framework demands consideration of the background and cultural context for the knowledge commons. In the case of Sidewalk Toronto, the relationship between a private sector company and a non-profit corporation around the digital integration of public and private sectors within a real estate development/technology innovation lab created a particular challenge for data governance.

The deadline to finalize an agreement based on the MIDP was extended from September 2019 to March 31, 2020,[58] with a possibility for the parties to terminate the PDA by October 31, 2019 if no agreement could be reached on key issues.[59] The project survived the October 31, 2019 cut-off date after Sidewalk Labs agreed to a number of conditions set by Waterfront Toronto. These included abandoning the UDT and avoiding the novel category of 'urban data' both of which are the focus of this paper.[60] Sidewalk Labs subsequently produced a lengthy Digital Innovation Appendix[61], which provided greater detail about its plans and a more cautious approach to data governance, which recognizes that Waterfront Toronto must play a central role.[62] The project came to an abrupt end on May 7, 2020. In a statement released by Sidewalk Labs' Dan Doctorow, the "unprecedented economic uncertainty [that] has set in around the world and in the Toronto real estate market" was cited as the reason for its termination.[63]

In spite of the demise of the project, the UDT and 'urban data' remain of interest and importance both to understand their origin and concept as a novel form of data governance for data sharing, as well as the reasons for their rejection.

4. The Emergence of Key Governance Issues

The second category of considerations in the knowledge commons framework relate to key attributes of the emerging commons, including the nature of the resources to be governed, the members of the relevant governance community, and the goals and objectives of the commons. In part because of the way in which this project evolved, there was considerable pushback around these issues once the plans for the project became public. As a result, at the same time as a governance framework was being developed, there was a parallel set of conversations that raised particular concerns and preoccupations around many of the core attribute issues. This section considers issues that emerged in public reactions and how the shaped the development of what ultimately became the UDT.

Elements of public pushback can be organized into four broad challenges that Sidewalk Labs subsequently sought to address in the data governance scheme that it proposed[64] and later refined in the MIDP.[65] As a result, these four publicly expressed data-related concerns played an important role in shaping the evolution of the governance scheme in the MIDP. The sheer breadth of the concerns made governance increasingly complex, perhaps overburdening the proposed framework.

The first set of issues related to **data sharing and access**. The initial announcement of the project raised concerns among local technology developers who felt that it might exclude them from opportunities to participate in the development of smart city technology in Toronto, with a large US corporation instead being invited to both shape and occupy the market.[66] Although Sidewalk Labs talked of making data from the project open, the extent of this commitment was unclear.[67] Developers' inclusion issues extended beyond data;[68] nevertheless, there was a desire that smart city data be made available in real-time and under open licences so that developers could use it to generate innovative and competing applications for the city.[69] The data sharing and access concerns were ones that suggested a need for some form of knowledge commons.

Developers also wanted to be able to **participate in the data collection** that would take place within the development zone. In other words, they resisted a vision in which Sidewalk Labs had a monopoly on the applications that would be used to collect smart city data. Sidewalk Labs responded with assurances that it would not monopolize innovation within the district. However, permitting more developers to innovate also meant that there would be new data governance challenges. While Sidewalk Labs could make commitments about data sharing, deidentification, or privacy by design with respect to its

57 See Steve McLean, 'Sidewalk Labs' Sirefman updates Toronto development plans' (*Real Estate News Exchange*, 18 September 2019) https://renx.ca/sidewalk-labs-sirefman-toronto-waterfront-development/ accessed 28 April 2020; James McLeod, 'Did Sidewalk Labs overstep with their masterplan? It certainly raised concerns at Waterfront Toronto' (*Financial Post*, 24 June 2019) https://business.financialpost.com/technology/sidewalk-labs-long-awaited-smart-city-masterplan-raises-concerns-at-waterfront-toronto accessed 28 April 2020. In the MIDP (n 3), Vol 1, 17, Sidewalk Labs describes its 'Innovative Design and Economic Acceleration(IDEA) District that represents an innovative new development model for how the private sector can support the public sector in tackling the toughest growth challenges.'

58 Waterfront Toronto and Sidewalk Labs, Amending Agreement (31 July 31 2019), 1 https://quaysideto.ca/wp-content/uploads/2019/04/Plan-Development-Agreement-July-31-2018-and-Amendment-July-31-2019.pdf accessed 28 April 2020. This deadline was subsequently extended to take into account the COVID-19 crisis.

59 Ibid.

60 Diamond (n 55).

61 Sidewalk Labs, (n 38).

62 Ibid. These timelines have been further extended as a result of the COVID-19 pandemic.

63 Daniel L. Doctorow, "Why we're no longer pursuing the Quayside project — and what's next for Sidewalk Labs" (*Medium*, 7 May 2020) https://medium.com/sidewalk-talk/why-were-no-longer-pursuing-the-quayside-project-and-what-s-next-for-sidewalk-labs-9a61de3fee3a accessed 8 May 2020.

64 Sidewalk Labs, 'Digital Governance Proposals for DSAP Consultation' (October 2018) https://waterfrontoronto.ca/nbe/wcm/connect/waterfront/41979265-8044-442a-9351-e28ef6c76d70/18.10.15_SWT_Draft+Proposals+Regarding+Data+Use+and+Governance.pdf?MOD=AJPERES accessed 28 April 2020.

65 MIDP (n 3) Vol II, Ch 5.

66 See, e.g., Aeman Ansari, 'Toronto doesn't need Google to build a smart city, says open data expert' (*betakit*, 20 November 2017) https://betakit.com/toronto-doesnt-need-google-to-build-a-smart-city-says-open-data-expert/ accessed 28 April 2020; Bill Bean, 'The world is watching as data drives Toronto's Smart City experiment' (*Communitech News*, October 30, 2017) http://news.communitech.ca/the-world-is-watching-as-data-drives-torontos-smart-city-experiment/ accessed 28 April 2020; Jim Balsillie, 'Sidewalk Toronto has only one beneficiary, and it is not Toronto' (*Globe and Mail*, 15 October 2018) https://www.theglobeandmail.com/opinion/article-sidewalk-toronto-is-not-a-smart-city, accessed 28 April 2020.

67 See, e.g., Bianca Wylie, 'Civic Tech: On Google, Sidewalk Labs, and Smart Cities' (*Torontoist*, 24 October 2017) https://torontoist.com/2017/10/civic-tech-google-sidewalk-labs-smart-cities/ accessed 28 April 2020.

68 Some even expressed the concern that discussions around data distracted from issues of ownership/control of the underlying source code. See: Terry Pender, 'Miovision CEO sees great value in Sidewalk Labs data' (*The Record.com*, 3 November 2018) https://www.therecord.com/news-story/9004728-miovision-ceo-sees-great-value-in-sidewalk-labs-data accessed 28 April 2020.

69 Donovan Vincent, 'Who will reap the benefits of Quayside's smart city data?' (*Toronto Star*, 16 December 2018) https://www.thestar.com/news/gta/2018/12/16/who-will-reap-the-benefits-of-quaysides-smart-city-data.html accessed 28 April 2020.

own technologies, it could not do the same for other actors.[70] Instead, it decided to make compliance with the data governance scheme a precondition for participation in the data ecosystem that was being developed.[71] Those seeking to collect data within the IDEA District, or those seeking to use certain types of 'urban data' that were not otherwise available as open data, would have to request permission and comply with requirements established as part of the data governance framework. Not only did this undermine the potential for the design of the kind of consensual data governance framework required for a knowledge commons, the potential scale and cost of managing this more complex data sharing framework, would also have implications for 'openness'. In the MIDP, Sidewalk Labs indicated that there might be fees for approvals of plans to collect or use data submitted to the Urban Data Trust.[72]

A third wave of opposition related to data came from those who were concerned that the ubiquitous collection of data within the smart city posed a **risk to privacy and other values**. Privacy issues had already been anticipated by Sidewalk Labs, which had retained former Ontario Information and Privacy Commissioner Ann Cavoukian as a consultant. Based on principles of Privacy by Design (PbD)[73] Sidewalk Labs had promised that all data it collected would be de-identified at source.[74] However, critics found this unsatisfactory for two main reasons. The first was a growing lack of confidence in deidentification as a means of protecting privacy.[75] In a context in which vast quantities of different types of data are collected and analyzed using big data analytics and AI, reidentification risks are high.[76] A second concern was that even deidentified human behavioural data posed risks of harm both to individuals and to communities. These harms could flow from the use of the data to profile individuals or communities/groups in ways that might impact their access to resources or benefits, or that might incorporate or contribute to bias and oppression.[77] If PbD and deidentification were not complete solutions to the problem, then something more was needed. That something would have to include a mechanism to ensure that the data collected would be used in an appropriate, ethical and responsible manner. This is suggestive of the need for some form of framework for governing the 'knowledge commons'. The UDT proposed in the MIDP was therefore designed to oversee the collection and use of data, under a Responsible Data Use Agreement (RDUA)[78] similar to a privacy impact assessment.

A fourth issue around **data localization** arose from the considerable opposition to the idea that data collected in the smart city environment might end up stored on servers located outside Canada. On one level this was a privacy issue – Canadians have long been wary about the impact of the U.S. PATRIOT Act[79] on data about Canadians stored in the United States.[80] On another level, it is a data sovereignty issue.[81] Because the data was collected within and about a Canadian city, many saw it has having a public quality and that it should there-

[70] See Gabrielle Cannon, 'City of Surveillance: Privacy Expert Quits Toronto's Smart City Project' (*The Guardian*, 23 October 2018) https://www.theguardian.com/world/2018/oct/23/toronto-smart-city-surveillance-ann-cavoukian-resigns-privacy accessed 28 April 2020; John Buntin, 'Technopolis: Google's Sister Company Wants to Build the City of the Future on Toronto's Waterfront. Should a private tech giant be designing smart cities?' (*Governing*, July 2019) https://www.governing.com/topics/urban/gov-google-toronto.html accessed 28 April 2020.

[71] The MIDP (n 3), is clear that meeting the requirements of the Responsible Data Use Framework (RDUF) is independent of meeting all legal obligations. In other words, developers would not only have to meet the requirements of applicable laws, they would also have to meet what might be additional requirements imposed by the UDT. Adding another layer of compliance– and one for which fees might be charged -- would increase the burden for participation of SMEs.

[72] MIDP (n 3) Vol II, Ch 5, 422 and 434-435.

[73] Cavoukian, 'Privacy by Design' (n 58).

[74] Cavoukian, 'De-identifying data' (n 58).

[75] Concerns over reidentification risk have existed for some time (see, e.g., Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 *UCLA LR* 1701. These are exacerbated with the advance of technology. A recent article found that the reidentification risk was so high even for anonymized medical data that anonymization techniques in use today were unlikely to meet the rigorous norms of the GDPR. See: Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye, 'Estimating the success of re-identifications in incomplete datasets using generative models' [2019] 10 *Nature Communications* Article #3069 https://www.nature.com/articles/s41467-019-10933-3 accessed 28 April 2020. By contrast, Yakowitz argues that re-identification risks are exaggerated. See: Jane Yakowitz, 'Tragedy of the Data Commons' (2011) 25 *Harv. J L & Tech* 1.

[76] See Rocher et al, ibid.

[77] Concerns over the adverse impacts of data profiling on individuals and groups are longstanding. See, e.g., David Lyon, *Surveillance as Social Sorting* (Routledge 2002); Oscar H. Gandy, Jr., *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (Routledge 2009); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (MacMillan 2018).

[78] MIDP (n 3) Vol II, Ch 5, 424-440.

[79] Uniting (and) Strengthening America (by) Providing Appropriate Tools Required (to) Intercept (and) Obstruct Terrorism Act of 2001, Pub L 107-56 [The U.S.A. PATRIOT Act].

[80] For example, the governments of British Columba and Nova Scotia each passed laws that prohibited the storage of certain public data outside of Canada. See: Personal Information International Disclosure Protection Act, SNS 2006, c 3; Freedom of Information and Protection of Privacy Act, RSBC 1996, c 165, s. 30.1. See also: David Loukidelis, 'Privacy and the USA PATRIOT Act: Implications for British Columbia Public Sector Outsourcing' (Office of the Information and Privacy Commissioner of British Columbia, October 2004) https://www.oipc.bc.ca/special-reports/1271 accessed 28 April 2020.

[81] Banks characterizes it as a situation where "vast troves of data in a public-private partnership would be exfiltrated from Canada." He asks, "Once the data is outside of Canada, could Canadian governmental bodies ever reclaim control of that data should future voters decide that this is appropriate for security or other reasons?" (Timothy Banks, 'Will Sidewalk Labs' civic data trust hush critics of Waterfront Toronto?', (*IT and Data Governance*, 23 October 2018) https://timothy-banks.com/2018/10/23/will-sidewalk-labs-civic-data-trust-hush-critics-of-waterfront-toronto/ accessed 28 April 2020. Sean McDonald notes: "Framing data localization around the Canadian Government's enforcement of privacy law narrows the potential benefits of localization, and ignores the threats emanating from internationalizing the processing and storage of public data." (Affidavit of Sean McDonald in Canadian Civil Liberties Assn and Lester Brown v. Toronto Waterfront Revitalization Corporation, et al, Court File No. 211/19, 16 https://ccla.org/cclanewsite/wp-content/uploads/2019/06/Affidavit-of-Sean-McDonald-2019-05-28.pdf accessed 28 April 2020. The term "data sovereignty" is sometimes confused with other concepts such as data residency or data localization. Data localization typically involves legal requirements to store data within a specified jurisdiction. Data residency involves ensuring that enough of a company's data processing activities are 'located' in a legal sense within a country's borders in order to take advantage of certain beneficial laws or policies. Data sovereignty, in its narrowest sense refers to data being subject to the laws of a particular jurisdiction. However, data sovereignty can have a broader meaning, as it does in the context of the Indigenous Data Sovereignty movement. In that context, data sovereignty involves not only claims to self-governance with respect to the storage and management of data about the self-governing community. See, e.g., Tahu Kukutai and John Taylor, eds., *Indigenous Data Sovereignty: Toward an Agenda*, (ANU Press 2016) https://www.oapen.org/download?type=document&docid=624262#page=25 accessed 28 April 2020. Note that the term "data sovereignty" is now also used in relation to person control over personal data. See, e.g., the statement that "sovereign data subjects are those who are in a position to articulate and enforce claims to power about their data." Patrik Hummel et al, 'Sovereignty and Data Sharing' (2018) *ITU Journal: ICT Discoveries*, Special Issue No. 2, 2 https://www.itu.int/en/journal/002/Documents/ITU2018-11.pdf accessed 28 April 2020.

fore be located in Canada.[82] Although Sidewalk Labs initially resisted data localization arguments,[83] by the time the MIDP was published, this commitment had softened somewhat – storage in Canada would take place if adequate facilities existed.[84] The discussion over data localization suggests that the proposed UDT was meant to house the data it governed, rather than simply managing access to the data stored on the servers of the actors that generate it – although this was not entirely clear.[85]

Because this was not a public or city-led project, public concerns could only be raised after the announcement of the project. This led to the development of a governance framework in rather unique circumstances that ultimately proved problematic. Not only did the timing and context prevent the collaborative development of the data governance framework by all stakeholders, the project went ahead before the question of who was to have custody or control over what data was resolved. It is fair to say that many considered that smart city data would, by default, be municipal data under the custody and control of the City of Toronto – at least so far as the data was collected in relation to the infrastructure, streets, and other public spaces of the development.[86] This view was evidently not shared by Sidewalk Labs, although Sidewalk Labs remained cagey on the issue.[87] The issue is important. Contributors of data to a knowledge commons are stakeholders entitled to participate in the shaping of the governance framework. By moving ahead without addressing who was contributing what data to the commons, there could be no consensual governance model.

Sidewalk Labs ultimately proposed an independent data governance body to oversee its data-sharing framework. In doing so, it also attempted to hive off a category of data suitable for governance in this way. Rather than identifying particular data sets, whether controlled by public or private sector actors, that should be pooled and governed collectively in the public interest, it chose to create a whole new category of data – "urban data". Any data falling within the definition of 'urban data' was subject to governance by the Urban Data Trust.

5. Attributes: Urban Data

The second category of questions in the knowledge governance framework asks what resources are to be pooled, who the relevant stakeholders are, and what the goals and objectives are. The category of 'urban data' in the MIDP was, in many ways designed to answer these sorts of questions. As will be seen below, it defined a category of data for governance (urban data), characterized it as a kind of communally shared resource, and identified a fairly general concept of public interest. But, as will be seen below, this category of data was inherently problematic, creating fundamental problems for the data governance scheme. In this sense, the category also interacts with the next set of questions over knowledge commons governance, as the novel category of 'urban data' made it difficult to identify how existing legal frameworks would apply. 'Urban data' was defined as either unowned or communally owned. The data was conceived of as existing independently of its collectors, who would have to seek permission and follow rules regarding its collection. The category of urban data therefore defined the commons in terms of data in a geographic context, rather than data sets collectively pooled by stakeholders to serve common ends.

One reason why the UDT might have been built around 'urban data' could be to avoid the legal barriers to the contribution of public sector data to a communal governance regime. Under the laws of Ontario at the time of the MIDP,[88] the management of data collected by a public sector entity could not simply be delegated to a third party with its own governance rules. The public body was legally required to manage that data according to public sector laws and policies.[89] There was therefore a jarring and unresolved relationship between public ownership as represented by the public sector, and the notion of 'public' or 'communal' ownership of urban data in the UDT. These challenges were not insurmountable, but they might have required some legislative change.

The MIDP defined "urban data" as "information gathered in the city's physical environment, including the public realm, publicly accessible spaces, and even some private buildings".[90] The category 'urban data' was largely based on geography and concepts of public versus private space. Urban data could be personal or non-personal data, and could

[82] Lauriault et al (n 80) 24, state that "Data residency is a critically important consideration for Open Smart Cities because many firms that provide cloud computing for smart cities (Google, Microsoft, etc.) store their data in servers outside of Canada."

[83] See, e.g., Alyssa Harvey-Dawson, 'An Update on Data Governance for Sidewalk Toronto' (Sidewalk Labs, 15 October 2018) https://www.sidewalklabs.com/blog/an-update-on-data-governance-for-sidewalk-toronto/ accessed 28 April 2020.

[84] Specifically, Sidewalk Labs committed "to using its best efforts at data localization, as long as there are Canadian-based providers who offer appropriate levels of security, redundancy, and reliability.", MIDP (n 3), Vol II, Ch 5, 460. In an Open Letter dated October 31, 2019, the Chair of Waterfront Toronto confirmed that the parties had agreed that all personal information would be stored in Canada (Diamond (n 55)) For a critique of Sidewalk Labs' initial approach to data localization, see Affidavit of Sean McDonald (n 82).

[85] For example, when discussing access to data collected under the supervision of the Urban Data Trust, the MIDP states: "Facilitating access could be accomplished in a variety of ways, from having the Urban Data Trust actually hold the data to having it set rules that require collectors to publish de-identified, aggregate, or non-personal data in real time." MIDP (n 3), Vol II, Ch 5, 434.

[86] See, e.g., the letter of Ontario's Information and Privacy Commissioner that criticizes how the MIDP negates the role of the public sector in governing Sidewalk Toronto data (Beamish (n 9)).

[87] The MIDP makes reference to the need to comply with "all applicable laws" (See, e.g., MIDP (n 3) Vol II, Ch 5, 421.) and identifies both public sector (FIPPA and MFIPPA) and private sector (PIPEDA), data protection laws without specifying which would apply in what contexts (MIDP (n 3) Vol II, Ch 5, 421.).

[88] Ontario has since amended its Freedom of Information and Protection of Privacy Act, RSO 1990, c F.31, to allow for the creation of entities outside government that can engage in the governance of data from multiple sources.

[89] For municipal governments in Ontario, this includes the Municipal Freedom of Information and Protection of Privacy Act, RSO 1990, c M.56. As the Information and Privacy Commissioner for Ontario notes, it would have been possible for the UDT to take a position on privacy different from that of the provincial regulator. See Beamish (n 9), 6. It is perhaps no surprise that the federal government is also contemplating legislative change to facilitate collective data governance in a manner consistent with data protection obligations. See ISED, 'Strengthening Privacy for the Digital Age: Proposals to modernize the Personal Information Protection and Electronic Documents Act' 21 May 2019, https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html accessed on 28 April 2020.

[90] MIDP (n 3), Vol II, Ch 5, 377. 'Publicly accessible spaces' is a complex category. It appears to mean that owners or lessors of publicly accessible private properties– such as retail spaces (MIDP (n 3), Vol II, Ch 5, 426) – would, to the extent that they collect data in these spaces, be collecting 'urban data' and would therefore be subject to the RDUA and the UDT. In the MIDP, Sidewalk Labs uses the example of a parking garage lessor who would need to go through the RDUA process in order to install security cameras in its garage (MIDP (n 3), Vol II, Ch 5, 439-440.) The 'public realm' includes public spaces such as streets or parks. It apparently also includes atmospheric or environmental data. MIDP (n 3), Vol II, Ch. 5, 379, 417

include aggregate or de-identified data.[91] The definition neatly avoided the traditional dichotomy of public and private sector data; the identity of the party collecting the data was irrelevant to its characterization as 'urban data'. The creation of a new category evaded both ownership and control issues, as well as the collaborative approach to governance that different 'ownership' interests would entail. Yet the recognition and reconciliation of diverse interests is both an important process and an outcome of commons governance.

As defined in the MIDP, Sidewalk Labs' 'urban data' has the following characteristics:

1. It is defined based upon where it is collected (i.e. location is a key element in the definition of urban data);[92]
2. The "where" is linked to some concept of shared or communal space;
3. Shared or communal space can cut across the boundaries of publicly and privately-owned spaces;[93]
4. Urban data may include personal information and/or human behavioural data, as well as other types of non-personal data;[94]
5. Urban data is not defined by who is collecting it (i.e. it can be collected by public or private sector actors and possibly even by individuals).[95]

Geography or location was therefore a core component of the definition.

'Urban data' was defined both in terms of what it was and what it was not. For example, 'urban data' is distinct from what Sidewalk Labs labeled as "transaction data". Transaction data was data relating to any specific transactions carried out by individuals with the providers of particular services (such as ride-sharing, utilities, etc.).[96] The distinction between urban data and transaction data was explained by Sidewalk Labs' Alyssa Harvey-Dawson:

> For clarity, we call the original information collected in a physical place in the city "urban data." Urban data is different from data created when individuals agree to provide information through a website, mobile phone, or paper document. It presents unique challenges, including that it could reasonably be considered a public asset, and that it raises potential concerns around surveillance and privacy.[97]

In the MIDP, transaction data was also explained as not fitting within the category of urban data because it would be "unworkable given the lack of a relationship between this kind of data and a specific geography."[98] Yet 'transaction data' would not always be easy to separate from information collected in a physical space. For example, contracting for municipal water services will generate transaction data such as the amounts billed to a particular customer. However, it is unclear whether data about the volume, frequency and timing of water consumption (on which billing is based) is solely transaction data or also 'urban data', since it is linked to a particular geographic location (the point of consumption). Perhaps the answer is that some data would be transaction data when linked to a particular individual, but could become urban data in aggregate or anonymized form. As another example, the MIDP distinguished between data from sensors such as cameras on ride sharing vehicles (urban data for which permission to collect in the IDEA district is required) and consumer trip and payment data, which would be transaction data.[99] Yet arguably, data about the movement of a person from point A to point B within the IDEA District (which is data relevant to the transaction) has links to physical space and could be construed as urban data, particularly if it were useful data for understanding traffic patterns or transit demands. These questions about where transaction data ended and urban data began illustrates how challenging the definition of a novel category of data can be.

A major reason why transaction data was separated from urban data was because it is seen as specific to a contractual relationship between an individual and a service provider, and would be governed by terms of service and a separate privacy policy. In other words, this data was not collectively owned because it was seen as proprietary to the party that collected it from an individual under the terms of a contract. Sean McDonald criticized this distinction between transaction data and urban data, stating that the result is that "the more sensitive the data the more proprietary it would be."[100] Yet this seems precisely part of the rationale. For Sidewalk Labs, urban data was suitable for collective governance because it was 'owned' by no one. The relationship between the individual and the provider both makes the data proprietary and enhances its sensitivity. By contrast, urban data involves no specific relationships. Sidewalk Labs' insistence on geography as a core characteristic of urban data nevertheless created a tension with transaction data because the two categories – urban data and transaction data – depended on different characteristics that were not mutually exclusive.[101] Urban data relied upon collection in shared geographical spaces, while transaction data was defined in terms of specific relations between an individual and an organization. The fact that specific relationships can arise with respect to data that – in aggregate – can provide information about shared public space creates conceptual problems. These are only augmented by the ambiguity around the notion of public versus private spaces. It raises the question of why aggregate transaction data with the appropriate geographical dimensions is not also communally owned urban data.

The definition of urban data is also interesting because it relied upon concepts of 'public' and 'private' tied to geography and in particular to concepts of public and private spaces defined not necessarily in terms of land ownership but in terms of access and usage. This

91 MIDP (n 3) Vol II, Ch 5, 417.

92 "The term 'urban data' nods to the fact that it is collected in a physical space in the city and may be associated with practical challenges in obtaining meaningful consent." MIDP (n 3) Vol II, Ch 5, 416.

93 In the MIDP (n 3) this seems to include privately owned or controlled spaces with public dimensions, such as retail stores, the lobbies of apartment buildings, or public spaces within publicly owned buildings.

94 For example, in the MIDP (n 3) Vol II, Ch 5, 416, it states that urban data "includes both personal information and information that is not connected to a particular individual." It goes on to say that "Urban data would be broader than the definition of personal information and includes personal, non-personal, aggregate, or de-identified data. . .".

95 It is not clear whether "There is no obvious means for individuals to consent to its collection" should be a sixth factor in this list, or whether this statement is simply a conclusion that can be drawn from the listed features of urban data. In other words, it is not clear if it is an 'and', or if the problem of consent is considered inherent to data within this category.

96 MIDP (n 3), Vol II, Ch 5, 416: "urban data would be distinct from more traditional forms of data, termed here "transaction data", in which individuals affirmatively – albeit with varying levels of understanding – provide information about themselves through websites, mobile phones, or paper documents."

97 Harvey-Dawson (n 45).

98 MIDP (n 3) Vol II, Ch 5, 427.

99 MIDP (n 3) Vol II, Ch 5, 427.

100 Sean McDonald, 'Toronto, Civic Data, and Trust', (*Medium*, 17 October 2018) https://medium.com/@McDapper/toronto-civic-data-and-trust-ee-7ab928fb68 accessed 28 April 2020.

101 This is also noted in the Ontario Information and Privacy Commissioner's open letter to the Chair of Waterfront Toronto; Beamish (n 9).

jarred with established understandings of data ownership that turn on in who collects or controls the data.[102] Data could be 'urban data' regardless of whether it was collected by public or private sector actors

The linking of urban data to location was probably at least partly driven by concerns over the collection of human behavioural data. At one point in the MIDP, Sidewalk Labs noted that the location elements "may be associated with practical challenges in obtaining meaningful consent."[103] In other words, the data governance scheme was designed to address the privacy problem of the requirement of consent for collection of personal information in a context in which consent would be impractical to request or obtain – such as urban public spaces. Yet since technology might evolve to enable consent in a broader range of contexts, this added another layer of uncertainty about what would constitute 'urban data'.

The consent requirement for the collection of personal data is different in Canada depending on whether the collector is a public or private sector actor. Consent is not required for personal data collection by public sector actors, although notice is.[104] This recognizes the imbalance of power between governments and citizens as making true consent impossible.[105] Instead, data collection by government is legitimized by democratic processes that enable the government's action and the public policy considerations that motivate the collection. Where the collector is a private sector actor, consent is required.[106] The UDT was intended to provide a substitute process to legitimize collection without consent in public spaces by private sector actors.[107] It did so by establishing an independent governance framework that would set the rules for both collection and for subsequent uses of this data. Yet this shifted the role of the UDT from data steward to a kind of data protection authority or even a mini-municipal government. For example, if Toronto's municipal government decided to collect a certain type of data from light standards or from other municipal infrastructure throughout the city, it would have to apply to the UDT for permission to deploy these sensors in the IDEA district – and such permission could be refused, or conditions could be attached.[108] This created the possibility that the UDT could deny permission to the population's democratically elected municipal government to implement a city-wide policy decision.[109]

The problems were not just with new governance for public sector data. The MIDP offered an example of a parking garage operator in the development area who decides to install security cameras. Although use by patrons of the garage is consent-based and transactional, Sidewalk Labs considered the camera data to be 'urban data' subject to the governance regime. Thus, the garage operator, who would already be subject to private sector data protection legislation, would have to go through the RDUA process. It seems problematic to suggest that security camera footage should be contemplated as shareable through the UDT, even in deidentified form. There is no compelling case for public or communal 'ownership' of such data. Any governance process beyond data protection law seems unnecessary. Other problematic "publicly accessible spaces" might include the lobbies of apartment buildings or condominiums, retail stores, shopping malls, or restaurants. In all of these cases, there are already data protection laws that would govern collection of personal information, and in many instances, collection would be for fairly specific purposes such as security. In most cases, governance through privacy legislation would suffice to place strict limits on what could be collected, how it might be used, how notice would have to be provided, and how long the data could be retained. It is unclear what added value would be provided by a further layer of data governance. Adding such data to a data governance regime for data sharing would only raise additional privacy and ethical concerns.

By making geography (particularly 'public space') the primary characteristic of 'urban data', the definition also became dangerously over-inclusive. Sidewalk Labs provided at least two examples of urban data collection in which the problems of over-inclusivity are evident. The first involved the use of an app to collect non-personal data about park usage by a civil society group.[110] The data collection was an automated version of what might otherwise be recorded by volunteers equipped with pens and paper. Sidewalk Labs offered this as an example of data collection that would have to go through the RDUA process and that would require approval by the UDT prior to collection because the data is 'urban data' collected in public space. Another example from the MIDP is the collection of air quality data

[102] In copyright law, for example, authorship of compilations of data is determined based upon who is responsible for the selection or arrangement of the data within the compilation. See, e.g., Geophysical Service Inc v Encana Corp, 2016 ABQB 230, 38 Alta LR (6th) 48, aff'd 2017 ABCA 125, leave to appeal denied 2017 CanLII 80435 (SCC). Under the EC, European Database Directive 96/9/EC of the European Parliament and of the Council of the European Union of 11 March 1996 on the legal protection of databases, [1996] O.J, L 77/20, article 4, ownership is determined based upon who created the database.

[103] MIDP (n 3) Vol II, Ch 5, 416. This is contrasted with "transaction data" for which, according to Sidewalk Labs, consent can be directly obtained from the individual. (See: MIDP (n 3), 426).

[104] See, e.g., Beamish (n 9); Department of Justice, Canada, 'Privacy Principles and Modernized Rules for a Digital Age' (Canada, 21 August 2019), 12-13 https://www.justice.gc.ca/eng/csj-sjc/pa-lprp/dp-dd/modern_1.html accessed 28 April 2020.

[105] For example, a consultation document from Canada's Department of Justice states: "some individuals might fear adverse consequences and feel compelled to consent to the collection of personal information". Since true consent is not possible, collection is instead based on the link to a legal activity by government. See: Department of Justice (n106) 12.

[106] See, e.g., the critique by David Young, 'Sidewalk Labs – Public or Private Data' (*David Young Law*, 2019) http://davidyounglaw.ca/compliance-bulletins/sidewalk-labs-public-or-private-data/ accessed 28 April 2020.

[107] See, e.g., Natasha Tusikov, "'Urban Data' and 'Civic Data Trusts' in the Smart City", (Centre for Free Expression, 6 August 2019) https://cfe.ryerson.ca/blog/2019/08/%E2%80%9Curban-data%E2%80%9D-%E2%80%9Ccivic-data-trusts%E2%80%9D-smart-city accessed 29 April 2020; Keri Grieman, 'Pedestrian Curiosity: A Brief Examination of Consent and Privacy in Swath Section Smart City Spaces' (2019) 7(5) *Spatial Knowledge and Information Canada* 1 http://ceur-ws.org/Vol-2323/SKI-Canada-2019-7-5-1.pdf accessed 28 April 2020. Ontario's Office of the Information and Privacy Commissioner raises concerns that unclear and overlapping roles for private regulators and the UDT create a confusing compliance context. See: Beamish (n 9) 6.

[108] The Ontario Information and Privacy Commissioner (Beamish (n 9), 8) commented on how problematic this would be. In his view, to "expect the City to apply to a non-profit Trust, go through the evaluation process, and commit to contractual undertakings would be inappropriate given the experience, mandate and statutory authority of the City."

[109] In his letter to the Chair of Waterfront Toronto, Commissioner Beamish notes that it is "problematic that, as proposed, the City and other public sector organizations would be expected to apply to the Trust in order to collect or use any Urban Data in the geographical area of the project." (Beamish (n 9) 8). He observes that where the city is required by law to collect data: "To then expect the City to apply to a non-profit Trust, go through the evaluation process, and commit to contractual undertakings would be inappropriate given the experience, mandate and statutory authority of the City."

[110] MIDP (n 3) Vol II, Ch 2, 185. This project was publicized by Sidewalk Labs prior to the development of the MIDP, and was part of the discussion around their RDUA. See: Farrah Merali, 'Sidewalk Labs partners with Toronto groups to collect data for public life study' (*CBC News*, 16 December 2018) https://www.cbc.ca/news/canada/toronto/sidewalk-labs-thorncliffe-park-womens-committee-1.4946336 accessed 28 April 2020.

by an environmentalist.[111] These examples reveal the tension between data in the public domain – free for all to gather and use – and "urban data" in which the UDT would assert some form of control over who collects the data and why. Under this approach, public domain data becomes collective data, subject to control over both collection and use.[112]

This confusion between public domain and collectively 'owned' data was evident in the MIDP. Writing about data governance for Sidewalk Toronto, the company's Alyssa Harvey-Dawson stated "No one has a right to own information collected from Quayside's physical environment — including Sidewalk Labs."[113] At the same time, Sidewalk Labs characterized urban data as a "community or collective asset",[114] suggesting a kind of communal ownership distinct from public sector data.[115] Harvey-Dawson acknowledges the governance gap created by this novel concept when she states: "If no one owns urban data, the question remains: Who manages it in the public interest?"[116] Sidewalk Labs' answer, was of course, the Urban Data Trust, which is discussed in more detail in the following part.

6. Governance: The Urban Data Trust

The third set of questions in the knowledge commons framework addresses governance. This includes a consideration of governance mechanisms and decision-makers, infrastructures and institutions, as well as informal norms and legal structures. In the case of Sidewalk Toronto, the UDT was presented as the governance body for the pool of 'urban data'.

In the PDA, Sidewalk Labs proposed that it would explore the creation of a "data trust" to govern data collected in the Quayside development. This mention of the data trust was short on detail; it was referred to as a "novel form of data governance".[117] This concept evolved into a "civic data trust",[118] which is described by McDonald and Porcaro as "an organizational and legal model that protects the public's interest" in data.[119] Both proposals generated debate and uncertainty about what they meant in terms of governance, with some raising concerns that they could not be 'trusts' in a legal sense.[120] In addition, some critics challenged the appropriateness of using the 'civic data trust' label for the scheme proposed by Sidewalk Labs, which was ultimately a top-down arrangement.[121] In any event, perhaps in response to both sets of criticism, the MIDP, dropped 'civic data trust' and proposed instead an Urban Data Trust, with the qualification that it was not using the word "trust" in its trust law sense.[122] Sidewalk Labs also indicated in the MIDP a reluctance to adopt any solution that depended upon new legal infrastructure (i.e. legislative amendment or new legislation).[123] This reluctance might have been due to a concern about delays and uncertainty that could arise from any solution that would be subject to the vagaries of a political process. Nevertheless, Sidewalk Labs left open the possibility that the new governance body might at some point evolve into a public body, although how this might occur was unclear.[124]

The MIDP contemplated that the final development agreement with Waterfront Toronto would provide for the establishment of the UDT.[125] Once created, it would be a non-profit organization independent of both Sidewalk Labs and Waterfront Toronto. It would have the mandate "to address the digital governance challenges related to urban data while also promoting data driven innovations that benefit

[111] MIDP (n 3), Vol II, Ch 2, 183.

[112] It is difficult to see how a communal data ownership argument could be used to prevent anyone from collecting non-personal data in the public realm. Even in the case of personal data, data protection laws do not prevent the collection of personal data by individuals for purely private reasons, nor do they apply to the collection, use or disclosure of personal information when it is for journalistic, artistic, or literary purposes. (See: Personal Information Protection and Electronic Documents Act, SC 2000, c 5, s. 4(2)(b) and (c); Personal Information Protection Act, SA 2003, c P-6.5, s. 4(3), and Personal Information Protection Act, SBC 2003, c 63, s. 3.) This is due to freedom of expression concerns (Alberta (Information and Privacy Commissioner) v. United Food and Commercial Workers, Local 401, [2013] 3 SCR 733, 2013 SCC 62.).

[113] Harvey-Dawson (n 45). Data in the public domain is not owned. By contrast, a data commons is a pool of data that, although shared, is nonetheless controlled. Observing that data held in a commons is often for specific purposes, Yakowitz describes a data commons as consisting of "public-use research datasets" (Yakowitz (n 76) 6).

[114] MIDP (n 3) Vol II, Ch 5, 418.

[115] Sidewalk Labs gives the example of traffic data, stating: "Since that data originates on public streets paid for by the taxpayers and since the use of that data could have an impact on how those streets operate in the future, that data should become a public resource." MIDP (n 3) Vol II, Ch 5, 418.

[116] Harvey-Dawson (n 45). Goodman and Powles (n 23), 18, argue that "creating a term unrecognized in law, would effectively negate any default privacy setting: everything done within the bounds of the Sidewalk Toronto project would be potentially up for grabs."

[117] Plan Development Agreement (n 34), Schedule 1, 47. The concept of a data trust is quite fluid and open-ended. See, e.g., Hardinges (n 1); Element AI/Nesta, 'Data Trusts: A New Tool for Governance' (*ElementAI*, 2019) https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf accessed on 28 April 2020; 'A Primer on Civic Digital Trusts' (*MaRS*, December 2018) https://marsdd.gitbook.io/data-trust/ accessed on 28 April 2020; Sylvie Delacroix and Neil D. Lawrence, 'Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance' (2019) *International Data Privacy Law* https://doi.org/10.1093/idpl/ipz014 accessed on 28 April 2020; Sean McDonald, 'Reclaiming Data Trusts' (*CIGI*, 5 March 2019) https://www.cigionline.org/articles/reclaiming-data-trusts accessed on 28 April 2020.

[118] See: Harvey-Dawson (n 45). Sidewalk Labs defined the 'Civic Data Trust' as "an independent third party that ensures that value from data goes to the people, communities, government, industry and society from which it was collected and that data privacy and security are protected." (Sidewalk Labs, Digital Governance Proposals (n 65), 12.)

[119] Sean McDonald and Keith Porcaro, 'The Civic Trust' (*Medium*, 4 August 2015) https://medium.com/@McDapper/the-civic-trust-e674f9aeab43 accessed on 28 April 2020. McDonald acknowledges that as the concept is still in evolution, there may be different understandings of what constitutes a civic data trust. Affidavit of Sean McDonald (n 82) 5.

[120] According to the trust model, ownership in data is transferred to the trust which then manages it according to the specified terms. Some argued that data was incapable of this kind of transfer and ownership. A civic trust must also act in the interest of the broader population, and some argued that a data trust would not easily fit within the concept of 'charitable trusts' developed in Canadian law for public benefit trusts. See: Goodman and Powles (n 23), at 19; Mariana Valverde, 'What is a data trust and why are we even talking about it? Sidewalk Labs' magic tricks' (*Centre for Free Expression*, 14 January 2019) https://cfe.ryerson.ca/blog/2019/01/what-data-trust-and-why-are-we-even-talking-about-it-sidewalk-labs%E2%80%99-magic-tricks accessed on 28 April 2020.

[121] McDonald and Porcaro (n 123), state that a civic data trust and its private sector data contributors have a fiduciary duty "to develop participatory governance processes that keep each other in check." This lack of process in the proposed UDT was seen as a key failing.

[122] MIDP (n 3) Vol II, Ch 5, 423.

[123] According to Sidewalk Labs, "housing the Urban Data Trust in a public-sector entity would require new or amended legislation, and the passage of legislation can take time and would need to account for emerging technologies." MIDP (n 3) Vol II, Ch 5, 422. Ontario's Information and Privacy Commissioner, by contrast, urges law reform to provide the necessary legal infrastructure: Beamish (n 9), 8. An Australian report on smart cities notes that "It is those cities that actually enact legislation around their data ecosystem and the panoply of smart cities initiatives that are best placed to shape and control their urban digital futures." ('Governance and the Smart City' (*Energy of Things*, December 2016), 10 https://www.fishermansbend.vic.gov.au/__data/assets/pdf_file/0015/33243/Governance-and-the-Smart-City_EoT_December-2016.pdf accessed on 28 April 2020.

[124] MIDP (n 3) Vol II, Ch 5, 422.

[125] MIDP (n 3) Vol II, Ch 5, 420.

individuals and society."[126]

The UDT as proposed in the MIDP would have consisted of five members (at least initially). The nature and composition of the UDT was dictated by the concept of 'urban data' as being neither public nor private sector data, and subject to some form of 'public' or communal ownership. Thus, the five proposed members were meant to represent different interested parties in this data. One would be chosen for his or her expertise in data governance and legal issues. The other four would represent different 'interest groups': academic, public, private and community.[127] This suggested a commonality in interests within each of these categories – something that could not be safely said about any of them.[128] Beyond this, although the data was seen as being collective or communal data and while it was clearly expected that much of this data might be human behavioural data, the "community" received only one seat on a board of five.[129]

The UDT was meant to govern urban data by controlling who was entitled to collect and use this data, and by setting the terms and conditions. This was to be carried out through the RDUA – a combination of application form and ethics approval request to be filed prior to commencing the collection of data in the designated area. Parties seeking to use urban data collected by someone else would file RDUAs explaining the nature and purpose of their proposed use. The RDUA would require the incorporation of privacy-by-design principles, and would specify that data must be used for a "beneficial purpose" which "must incorporate Canadian values of diversity, inclusion, and privacy as a fundamental human right."[130] The purposes for collection and use would have to be clear and transparent. Data would be deidentified by default, stored securely, and collection would be minimized. Data must not be sold or used for advertising without explicit consent of the data subjects. Those who wish to use data for the development of AI must also conform to responsible AI use principles.[131] The actual RDUA process would be similar to a privacy impact assessment.[132] While Sidewalk Labs acknowledged that the UDT could establish its own guidelines, it proposed the RDUA for at least the initial start up period.[133] Any sensors would have to be mapped and registered with the UDT in a public registry to enhance transparency.[134] Although collected data would be publicly accessible by default, the UDT would have the ability to impose access conditions where this was warranted to protect the public interest. The UDT would also oversee data sharing agreements, access terms and fees. A "data collection and use administration fee" would be part of each data collection/use agreement and would be payable to the UDT to offset its operating costs.[135] The UDT would have the authority (presumably under the terms of the agreements with individual data collectors or users) to audit an organization's practices, to remove sensors in cases of non-compliance, and to seek legal remedies for breaches of conditions.[136] However, as it would not be a public body, nor would it be created by statute, it was unlikely to have any special enforcement powers.[137]

An alternative to the UDT might have been to turn to the public sector for a governance framework. For example, the OIPC suggested that:

> Rather than relying on Sidewalk Labs to develop an appropriate solution, this is an opportunity for the provincial government to take the lead and modernize the laws to address the legislative shortcomings. Amendments could include mandatory requirements for data minimization, additional protections for individual and group privacy, ethical safeguards, and greater enforcement tools for my office, including additional investigation, order making and audit powers.[138]

Public sector governance was specifically rejected by Sidewalk Labs. The Toronto Board of Trade, in a separate proposal, suggested that the Toronto Public Library should operate as a trusted data steward.[139]

In an article on the Sidewalk Labs proposal, Alyssa Harvey-Dawson suggested that the UDT would fill a void because: "Existing laws on urban data do not address ownership."[140] It was thus a concept of governance premised on the idea that the captured data were a communal asset. Yet data exist because someone has captured them, and this act of capture reflects specific choices made by the data collector. In addition, some data, such as personal data, reflect layers of interests. The idea of urban data as a kind of 'terra nullius' masked the existing interests in the data, and it was these interests that needed to be reflected in the design and implementation of a governance framework.

Ultimately, in proposing the UDT, Sidewalk Labs chose a governance model developed unilaterally, and not as part of a collective process involving data stakeholders. It was driven by a sense of urgency that allowed neither collaboration nor even legislative change that might have provided some institutional legal infrastructure. It is perhaps not surprising, therefore, that after its review of the MIDP, Waterfront Toronto rejected both the concept of 'urban data' and the UDT, and

[126] MIDP (n 3) Vol II, Ch 5, 420. Note that Commissioner Beamish (n 9), 7, expresses concerns over limited oversight of the UDT and the fact that it would not be subject to data protection and transparency laws.
[127] MIDP (n 3) Vol II, Ch 5, 420
[128] For example, having a single community representative mistakenly presumes a homogeneous community. It is also not clear whether academia is represented in their research capacity or as a substitute for civil society, which is unrepresented. As for the public sector, three levels of government have an interest in the port lands that are the subject of development, and their interests are not necessarily common. The "business industry representative" presupposes common interests across large, medium and small enterprises.
[129] In Data Trusts (Element AI) (n 123) 21, the authors observe that the UDT "failed to address the types of power imbalances at the core of the issues being discussed, and further exemplified the disenfranchisement of citizens in the decision-making process as to how their personal data is to be used, as the terms of the trust were chosen by Sidewalk Labs in the first place." Commissioner Beamish (n 9) 6, argued that it would be more appropriate to focus on the expertise required by the work of the UDT rather than on representation by sector.
[130] MIDP (n 3) Vol II, Ch 5, 424.
[131] MIDP (n 3) Vol II, Ch 5, 425. Ontario's Information and Privacy Commissioner, in an open letter to Waterfront Toronto, criticized the extent to which the UDT duplicated existing governance regimes for what the Commissioner clearly considers public sector data: Beamish (n 9).
[132] MIDP (n 3) Vol II, Ch 5, 428-429.
[133] MIDP (n 3) Vol II, Ch 5, 424.
[134] MIDP (n 3) Vol II, Ch 5, 433.
[135] MIDP (n 3) Vol II, Ch 5, 422.
[136] MIDP (n 3) Vol II, Ch 5, 435
[137] Enforcement is challenging. The Office of the Information and Privacy Commissioner notes that the UDT would have limited powers of oversight and redress (Beamish (n 9) 7). Commenting on an earlier iteration of the trust, McDonald, supra note 99, expressed concerns that if the UDT were to have the enforcement powers it needed, "we would have to substantially devolve and privatize limited forms of regulatory investigation and punishment authority."
[138] Beamish (n 87) 8.
[139] Donovan Vincent, 'Toronto Public Library should control data collected at Quayside, Board of Trade says' (*Toronto Star*, 9 January 2019), https://www.thestar.com/news/gta/2019/01/09/toronto-public-library-should-control-data-collected-at-quayside-board-of-trade-says.html accessed 3 May 2020.
[140] Harvey-Dawson (n 45).

the parties agreed that the project would move to the next phase without these elements.[141]

While the UDT and the concept of 'urban data' were problematic, their abandonment did not resolve the project's data governance issues. It returned the development to the status quo ante, leaving the private and public sector actors each to manage their data according to existing frameworks. The termination of the project in May 2020 made the immediate issue of data governance moot, although Waterfront Toronto remains committed to developing the Quayside area and any new partner or project may well have to design some form of data governance framework. Recent legislative amendments may have since created more room to innovate in the creation of a knowledge commons in which both public and private sector data can be shared. It remains to be seen whether there will be a willingness among new partners to invest in the design of an appropriate knowledge commons framework.

7. Conclusions

The preceding discussion of the data governance model proposed by Sidewalk Labs for the Sidewalk Toronto development offers an example of a failed governance scheme from which useful lessons may be drawn.

One problem with the UDT as a governance model was that it developed, in part, in response to a diverse range of public criticisms and concerns that were raised following the announcement of the Sidewalk Toronto project. A first problem was the reactive nature of the design of the data governance regime. The knowledge commons and its governance are ideally part of project design from the outset. The concerns were brought forward by many different urban stakeholders, from developers to residents. They included the ability to participate in innovation within the district, concerns over undue surveillance, ethics and human rights, and data localization arguments that combined privacy and sovereignty considerations. An attempt to build governance in response to these diverse concerns led to a data governance framework that tried to do too much and for many different reasons. While Frischman et al observe "Commons governance confronts various obstacles to sustainable sharing and co-operation",[142] not all of the obstacles sought to be overcome by the UDT were about the pooling or sharing of information assets. Rather, some related to the very nature of the development itself. In many ways, the UDT was designed to do too much and to satisfy too many disparate concerns.

A second flaw in the proposal was the decision to base the framework on the novel category of 'urban data'. This category was meant to capture a kind of data in which there might be a multiplicity of stakeholder interests. Yet by basing the definition on a combination of physical geography and uncertain notions of public and private space, the category was both unwieldy and uncertain. Rather than create governance for a pool of data shared by collaborating partners, the MIDP defined a category of data in which no one could claim ownership and subjected it to governance by the UDT. Quite apart from the problems with identifying data as independent of its collectors, this approach distanced the data to be governed from those who would have a clear stake in its governance.

A third flaw was that the governance model proposed was a top-down model originating from a single stakeholder in a complex environment with multiple participants and diverse interests in the data. The lack of an organic process with broad stakeholder engagement was a serious defect. Such a process should have identified who the stakeholders were and then involved them in considering what the data sharing model should look like, what data it should govern, according to what principles, and for whose benefit. It is clear that Sidewalk Labs saw some urgency in the task of designing data governance, found existing legal frameworks lacking, and felt legal change could not happen with sufficient speed or flexibility. Yet all of these factors undermined the legitimacy of what was proposed. "Urban data" was a profoundly problematic category of data, and the "trust" was not a trust in any real sense of the word. The result was a data governance scheme doomed to failure.

The failure of the UDT illustrates the importance of addressing data governance issues at the project design stage; these issues are often intricately intertwined with questions about what data to collect and for what purposes, which in turn are both project design and data governance issues. Further, data sharing necessarily implicates multiple interests, which may be both public and private. The diverse stakeholders need to be able to participate in the conceptualization and design of the data governance model and need representation in its implementation. In this respect, the concept of the 'knowledge commons' is useful and instructive. A knowledge commons does not depend upon the existence of a new type of data. Rather, it is premised upon different data 'owners' choosing to pool or share their data to achieve common goals within carefully set parameters. A knowledge commons shifts the focus from ownership/control over data to governance for sharing, but it does not deny or undermine the rights and interests of those who contribute to the commons. Rather, these form the basis for the interests of the contributors to participate in the governance of the commons.

[141] Diamond (n 55).
[142] Frischman et al (n 4) 23.

06

Special issue: Governing Data as a Resource

Technology and Regulation

Beyond the data flow paradigm: governing data requires to look beyond data

Charlotte Ducuing*

data governance, data ownership, data sharing, data flow paradigm, data commons

charlotte.ducuing@kuleuven.be

The paper aims to contribute to the discussion on how to regulate and govern data as an economic asset. It critically discusses the 'data flow paradigm', defined here as the regulatory focus on data (transactions) with the purpose to enhance data exchange by establishing data markets. Based on the examples of the electricity and the automotive sectors with respect to data governance, the paper finds that the data flow paradigm alone is too narrow. This paradigm seems to bear the idea that there should be well-operating data markets, possibly by the operation of the law, and that such markets alone would deliver the grand policy expectations, such as 'AI' or 'data-driven innovations'. Yet, fostering data exchange is not an end in itself and should be regarded with respect to the sectoral objectives and constraints. As the study of the examples shows, the quest for appropriate mechanisms to govern data often leads to rediscovering old concepts, such as (data) commons or (data) platform. Finally, the paper discusses future possible regulatory intervention.

1. Introduction

The paper submits that the 'data flow paradigm', defined here as the regulatory focus on data (transactions) with the purpose to enhance data exchange by establishing data markets, is too narrow to govern data as an economic resource. The data flow paradigm is particularly exemplified, at the European Union ('EU') level, by the regulatory attempts to create ownership(-like) rights on data or, conversely, to impose data sharing obligations, as considered in the Communication from the European Commission 'Building the European data economy' of 2017.

Based on two sectoral examples in the electricity and automotive industries, the paper discusses *limitations* suffered by the data flow paradigm. As a matter of fact, the regulatory options discussed to govern data as an economic resource in both sectors are already much broader in scope and diversified. Although sometimes implicitly and/or disguised in technical considerations, the governance of data in both cases is discussed in terms of institutional arrangements between the stakeholders. They resemble well-known governance mechanisms, such as 'commoning' practices on the one hand and the creation of a monopolist (platform) operator on the other hand. One can observe a growing interest in scholarship and amongst policy makers to adapt these older governance mechanisms by apprehending data as a resource. This can be seen in the recently published 'European Data Strategy' from the European Commission.

The paper starts with a characterisation of the 'data flow paradigm'. Then the two following sections outline, in turn, the data governance mechanisms discussed in the electricity and in the automotive sectors, in order to fairly govern data as a resource. Against this background, the fourth section draws critical lessons with respect to the data flow paradigm. While this paradigm can be characterised as horizontal, in the sense of being general and context-agnostic, the determination of the fitness of data governance mechanisms appears to be highly contextual, both in terms of objectives and constraints. This being said, the fifth and last section concludes by opening avenues for further research. Although essentially contextual, many lessons can indeed be drawn from the analysis of data governance mechanisms in specific sectors, in order to better understand the factors influencing positively or negatively their fitness. The data flow paradigm is mainly a regulatory one. By showing its limitations, the paper also aims to contribute to opening avenues for further regulatory initiatives to regulate data as a resource.

2 Owning or sharing: the data flow paradigm

In order to define what is called here the 'data flow paradigm', the section presents, in turn, two of its sides, namely the creation of an ownership(-like) rights on data and, second, the enactment of data access or data sharing obligations. The data flow paradigm may obviously also encompass other regulatory measures.

The creation of ownership(-like) rights on data has been contemplated, in the Communication from the European Commission 'Building a European data economy', with the purposes to bring legal certainty as for entitlements on data and to empower parties providing or, respectively, producing data.[1] The aim was to "improve[...] the operation of data markets by transforming data into merchandisable private goods in much the same way as do intellectual property rights

* Charlotte Ducuing is a doctoral researcher at the Centre for IT and IP Law (CiTiP) at Katholieke Universiteit Leuven (KU Leuven), Belgium

[1] European Commission, Communication 'Building a European data economy', COM/2017/09 final, 10.1.2017 and the accompanying Commission Staff Working Document 'On the free flow of data and emerging issues of the European data economy', SWD/2017/02 final.

in regard of their subject matter".[2] In other words, the basic implicit rationale is that the law should endorse - and adapt to - the economic reality where data are being commodified. The creation of an ownership(-like) right on data has been discussed in the scholarship and mostly opposed by lawyers, based on a wealth of both conceptual and practical arguments.[3] This option was not retained in the ensuing proposal from the European Commission for a Regulation on the free flow of non-personal data,[4] which led to the adoption of the Regulation 2018/1807.[5] Yet, the discussion on data ownership is still on-going, somehow further developed around the newly-coined expression "data sovereignty" (or '*Datensouveränität*', as the expression arose in Germany).[6]

In contrast, the adoption of data sharing obligations boomed while the European Commission contemplated them as a means to foster data exchange in its Communication 'Building a European Data Economy'. Data sharing legal regimes have been imposed on public sector bodies for a long time with the Open Data and PSI Directive, which was recast and reinforced in 2019.[7] Not only does the recast Directive impose stricter rules on public sector bodies, but the scope *rationae personae* was also extended to i.a. public undertakings. A new category as "high-value datasets" was created: subject to delegated and implementing acts of the European Commission, these high-value data sets shall be made available for re-use by third parties, with limited (in any)conditions.[8] Additionally, *lex specialis* data sharing legal regimes are increasingly being adopted in many brick-and-mortar industry sectors, such as banking, farming, electricity, automotive industry and road vehicles.

Outside the world of online platforms, data sharing obligations have mostly targeted (public and private) entities in their quality of 'monopolist data holders'.[9] In the Open Data and PSI Directive for example, data are created in the course of public service activities operated by regulated entities outside market conditions (in particular "public sector bodies") in an exclusive manner. Similarly, the European Commission contemplates data sharing obligations to be imposed on vehicle manufacturers (or Original Equipment Manufacturers, 'OEMs') described as "exclusive [in-vehicle] data gatekeepers",[10] in addition to existing legislation on access to vehicle repair and maintenance information.[11] In the electricity sector, the recast of the Electricity Directive in 2019 includes new obligations to share electricity data.[12] The data holder, namely the entity in the electricity value chain which collects the data from the (smart) energy meter (usually the Distribution System Operators, 'DSOs', or the

[2] Hanns Ullrich, 'Technology Protection and Competition Policy for the Information Economy. From Property Rights for Competition to Competition Without Proper Rights?', *SSRN Scholarly Paper* (Rochester, NY: Social Science Research Network, 12 August 2019), https://papers.ssrn.com/abstract=3437177."

[3] Alain Strowel, 'Les Données : Des Ressources En Quête de Propriété - Regards Sur Quelques Développements Récents En Droit Européen', in Elise Degrave, Cécile de Terwangne, Séverine Dusollier, Robert Queck (eds) *Law, Norms and Freedoms in Cyberspace / Droit, Normes et Libertés Dans Le Cybermonde - Liber Amicorum Yves Poullet*, Collection Du CRIDS (Larcier, 2018), 251–68; Serge Gutwirth and Gloria Gonzalez Fuster, 'L' éternel retour de la propriété des données: de l'insistance d'un mot d'ordre', in *Law, norms and freedoms in cyberspace. Droit, normes et libertés dans le cybermonde. Liber amicorum Yves Poullet*, Collection du CRIDS (Larcier, 2018), 1717–140; Andreas Wiebe, 'Protection of Industrial Data – a New Property Right for the Digital Economy?', *Journal of Intellectual Property Law & Practice* 12, no. 1 (1 January 2017): 62–71, https://doi.org/10.1093/jiplp/jpw175; Josef Drexl, 'Designing Competitive Markets for Industrial Data – Between Propertisation and Access', *JIPITEC* 8, no. 4 (2017). Gutwirth and Gonzalez Fuster mostly emphasise the public good nature of information and knowledge, based on the principle of freedom of expression and fear that ownership(-like) right on data would amount to a privatization of information. Drexl, for his part, looks at how value is created in the data economy and warns against the possibility of anti-competitive effects of data ownership. He mainly opposes the creation of an ownership(-like) right on data, as the conceptual rationales for such a creation (e.g., to incentivize the generation and collection of data) are not met. For legal scholars in favor of the creation of an ownership(-like) right on data, see Eric Tjong Tjin Tai, 'Data Ownership and Consumer Protection', *Journal of European Consumer and Market Law*, no. 4 (2018): 136–140, https://doi.org/10.2139/ssrn.3172725; Herbert Zech, 'Data as a Tradeable Commodity – Implications for Contract Law', in *Josef Drexl (Ed.), Proceedings of the 18th EIPIN Congress* (The New Data Economy between Data Ownership, Privacy and Safeguarding Competition, Rochester, NY: Social Science Research Network, 2017), https://papers.ssrn.com/abstract=3063153. Finally, some have argued in favor of qualified forms of ownership on data, such as 'defensive' or 'non-exclusive ownership', which may eventually amount to unbundling the bundle of ownership rights, see Benoit Van Asbroeck, Julien Debussche, and Jasmien César, 'Building the European Data Economy Data Ownership', White Paper, 2017. See also the on-going project of the American Law Institute (ALI) and the European Law Institute (ELI) with the purpose to propose a 'data law', PRINCIPLES FOR A DATA ECONOMY, https://www.europeanlawinstitute.eu/projects-publications/current-projects-feasibility-studies-and-other-activities/current-projects/data-economy accessed 9 February 2020. Besides, the creation of an ownership(-like) right on data has been discussed also in the economic scholarship, see for instance Nestor Duch-Brown, Bertin Martens, and Frank Mueller-Langer, *The Economics of Ownership, Access and Trade in Digital Data*, 2017.

[4] Proposal for a regulation of the European Parliament and of the Council on a framework for the free-flow of non-personal data in the European Union, COM/2017/0495 final - 2017/0228 (COD).

[5] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ 2018 L 303/59.

[6] The novelty of 'data sovereignty', compared to data ownership, consist in its attempt to enforce data 'right' of the data 'owner' to keep control over 'his' data technically, based on a 'reference architecture'. See for instance the recently created International Data Space Association ('IDSA'), supported by the German government, IDSA https://www.internationaldataspaces.org/the-principles accessed 10 February 2020. The expression "digital sovereignty" is also particularly discussed in Germany, which the prospect of a Data Law, see Jeffrey Ritter and Anna Mayer, 'Regulating Data as Property: A New Construct for Moving Forward', *Duke Law & Technology Review* 16, no. 1 (6 March 2018): 229–32.

[7] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, OJ 2019 L 172/56 ('Open Data and PSI Directive').

[8] Open Data and PSI Directive, Art. 2 (10), Chapter V and Annex I.

[9] Björn Lundqvist, 'Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet-of-Things World: The Issue of Accessing Data', in *Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach?*, ed. Mor Bakhoum et al., MPI Studies on Intellectual Property and Competition Law (Berlin, Heidelberg: Springer Berlin Heidelberg, 2018), 191–214, https://doi.org/10.1007/978-3-662-57646-5_8; Charlotte Ducuing, 'Data as Infrastructure? A Study of Data Sharing Legal Regimes', *Competition and Regulation in Network Industries*, 23 December 2019, https://doi.org/10.1177/1783591719895390.

[10] European Commission, Communication to the European Parliament, the Council, The European Economic and Social Committee, the Committee of the Regions, On the road to automated mobility: an EU strategy for mobility of the future, COM/2018/283 final, and the Proposal from the European Commission for a Regulation of the European Parliament and of the Council amending Regulation (EC) No 715/2007 on type approval of motor vehicles [...] and on access to vehicle repair and maintenance information, COM(2019) 208 final. See also, Bertin Martens and Frank Mueller-Langer, 'Access to Digital Car Data and Competition in Aftersales Services', Working Paper, JRC Digital Economy Working Paper (Brussels, Belgium: JRC, European Commission, 2018).

[11] Regulation (EC) N° 715/2007 of the European Parliament and of the Council of 20 June 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information, OJ 2007 L 171/1.

[12] Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU, OJ 2019 L 158/125 (the Electricity Directive), see Art. 23 and 24.

Transmission System Operators, 'TSOs'), could easily reserve an exclusive access and use of such data. In this paradigm, every (smart) energy meter – just like every (smart) car – constitutes a market with respect to the data that it produces. The data sharing legal regime thereby confirms – or even establishes, such as in the case of the PSI Directive – the regulated entities in a role as (raw) data providers in the data economy. The market for data is conceived of as a parallel market, beside the original market on which the regulated entities are active (or aside public service activities, in the case of public sector bodies), such as the manufacturing and sale of road vehicles or the distribution of electricity.

Data sharing obligations depart from their competition law inspiration, regarding their purpose and also possibly the range of beneficiaries. In the name of ensuring a 'fair data level-playing field' or 'fair competition for data', they were often found to pursue at least two different objectives: First, the objective of *preventing* potential abuses from being caused by the exclusive (raw) data holder to its competitors or to companies active in related markets (*ex ante* approach as opposed to the *ex post* effect of competition law). Second, data sharing obligations are also increasingly ascribed a *proactive* objective, that is to feed the data economy and data-driven innovation by benefiting a broader range of parties, without harm or abuse to be necessarily involved. Data are then considered as a *purposive infrastructure* for the data economy, in the sense that data sharing obligations are expected to turn them into infrastructural resource feeding *yet-to-be-created* downstream activities.[13] Such an approach is visible, for instance, in the automotive industry, where the Commission observed, in a 2018 Communication, that in-vehicle data "have an enormous potential to create new and personalized services and products, revolutionize existing business models [...] or lead to the development of new ones".[14] It is this general purpose that the European institutions, businesses and scholars[15] have attempted to achieve by fostering or even imposing[16] data sharing. Interestingly enough, the regulatory focus no longer seems to target only public sector bodies and public undertakings, but also private actors, based on their consideration as 'raw data exclusive holder'.[17]

The creation of ownership(-like) rights on data on the one hand and data sharing obligations on the other seem, at first glance, to be at odds with one another. The latter makes it mandatory for the data holder to grant access and re-use to (some) third parties while, on the contrary, the former grants *control* on data to the data holder. Yet, both regulatory options appear to have in common to treat (raw) *data as the regulatory subject-matter*, and more specifically the (raw) data transaction or market for (raw) data. The implicit aim is to support or even create data markets, deemed instrumental to data exchange, in turn viewed as a desirable objective. This was well captured by Zech: "The task of the law is to ensure that data markets exist (since the exchange and use of data are desirable)".[18] This is essentially, in our view, the 'data flow paradigm', characterised thereby both a regulatory objective and a regulatory subject-matter. The regulatory objective is to foster the flow of data with the aim to feed the data economy and to let data-driven innovation develop, based on data markets. It should finally be noted that the data flow paradigm is not limited to the two types of regulatory options outlined in this section. In this respect, the Regulation on the free flow of non-personal data laid down a general prohibition of national data localisation requirements.[19]

The two following sections outline the data governance mechanisms discussed in two sectors, namely the electricity and the automotive ones, in order to critically analyse the data flow paradigm. Although broadly used, the expression 'data governance' is not consensually defined. It is sometimes simply equated with "data management". From an information security or quality perspective, it may broadly refer to the control of - or alternatively to decision-making and -maker(s) with respect to - data management,[20] which may include intra-organisational division of tasks. 'Governance' generally refers to the high level management of organisations or countries, as well as the decision-making system and institutions for doing it.[21] From a policy and regulatory perspective, data governance can be defined as a system of rights and responsibilities that determine who can take what actions with what data. To be clear, the purpose is not to engage into a normative discussion on which data governance mechanisms would best serve the objectives and constraints in these sectors.

3. Electricity data governance

After having experienced liberalisation and vertical unbundling, the electricity sector is now undergoing major transformations along two trends. First, the integration of renewable electricity supply resulted in a decentralisation of the electricity supply. Second, the electricity sector is undergoing digitisation or the application of information and communication technology to the electricity system, particularly with the deployment of smart meters delivering near-real time consumption data.[22] As a result, distribution networks are expected to turn into "smart (distribution) grids",[23] in the sense that they allow for a better adjustment of electricity capacity demand and offer. Data are also expected to make existing markets more contestable, given the existence of information asymmetries between market operators and to allow for the creation of new data-driven personalised products and services, with the entry of new players on the market and possibly new markets. Data are thereby considered a required resource for concurrent purposes. "Information and data management [becomes] the interface between network and commercial side" and has become

[13] Ducuing (n 9) 7–8.
[14] European Commission, Communication 'On the road to automated mobility: an EU strategy for mobility of the future', COM/2018/283 final.
[15] Wolfgang Kerber, 'Data Governance in Connected Cars: The Problem of Access to In-Vehicle Data', *JIPITEC* 9, no. 3 (2018).
[16] See Report of 23.2.2018 on a European Strategy on Cooperative Intelligent Transport Systems (2017/2067(INI)) of the Committee on Transport and Tourism, point 41.
[17] Ducuing (n 9)
[18] Herbert Zech, 'A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data', *Journal of Intellectual Property Law & Practice* 11, no. 6 (1 June 2016): 462, https://doi.org/10.1093/jiplp/jpw049.
[19] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ 2018 L 303/59, Art. 4.
[20] Rene Abraham, Johannes Schneider, and Jan vom Brocke, 'Data Governance: A Conceptual Framework, Structured Review, and Research Agenda', *International Journal of Information Management* 49 (1 December 2019): 424–38, https://doi.org/10.1016/j.ijinfomgt.2019.07.008.
[21] See the definition of 'governance' in the Cambridge online Dictionary, GOVERNANCE, https://dictionary.cambridge.org/dictionary/english/governance accessed 11 February 2020, and in the Oxford online Dictionary: GOVERNANCE https://www.oxfordlearnersdictionaries.com/definition/english/governance?q=governance accessed 11 February 2020.
[22] Marius Buchmann, 'The Need for Competition between Decentralized Governance Approaches for Data Exchange in Smart Electricity Grids—Fiscal Federalism vs. Polycentric Governance', *Journal of Economic Behavior & Organization* 139 (1 July 2017): 106–17, https://doi.org/10.1016/j.jebo.2017.05.011.
[23] Christine Brandstätt et al., 'Balancing between Competition and Coordination in Smart Grids - a Common Information Platform (CIP)', *Economics of Energy & Environmental Policy* 6, no. 1 (2017), http://dx.doi.org.kuleuven.ezproxy.kuleuven.be/10.5547/2160-5890.6.1.cbra

"a new task in the electricity supply chain".[24] For this reason, data but also information and communication technology more generally were described as "the key infrastructure [...] in smart grids". These transformations triggered new questions on the role of data and on the institutional and organisational aspects thereto.[25]

According to the European Commission's Impact Assessment for the adoption of the Electricity Directive in 2019,[26] electricity data management constitutes a market entry barrier. Electricity data are data of the final electricity customer and include (smart and conventional) metering and consumption data as well as data required for customer switching, demand response and other services.[27] Data management is described in the Impact Assessment as comprising the processes by which data are sourced, validated, stored, protected and processed and by which they can be accessed by suppliers or customers. With the purposes to make existing markets more contestable and to enable the creation of new products and services, the Electricity Directive adopted in 2019 regulates the conditions in which a range of third parties ("eligible parties") can access and use electricity data stemming from data holders ('DSOs' or 'TSOs'). Data holders shall provide electricity data under transparent, fair, reasonable and non-discriminatory conditions ('FRAND') to eligible parties. Further interoperability requirements shall also be adopted by the European Commission, as facilitating technical measures. The Electricity Directive, in its final version, refrains from regulating the "data management model". It remains therefore within the jurisdiction of the Member States to "organise the management of data in order to ensure efficient data access and exchange". As a matter of fact, a study of the Council of European Energy Regulator ('CEER') issued in 2016 showed a clear trend towards centralisation of electricity data management amongst Member States.[28] Yet, many options exist.

First, and notwithstanding the competence of Member States to regulate data management models, the Electricity Directive goes beyond mere data sharing obligations and lays down requirements applying to the data management operator, who shall either be supervised by the competent authority or "authorised and certified".[29] When the data manager is a vertically integrated DSO dealing with smart meters data, additional 'compliance program' obligations apply to the internal processing of the company. They are copied from the independence requirements applying to electricity distribution activities, with a view to ensure that discriminatory conduct is excluded, that impartiality is ensured and that observance with such obligations is adequately monitored within the company.[30] Such an option is inspired by the model of the 'DSO as neutral market facilitator'.[31]

Secondly, the Impact Assessment considered a further-reaching option, where data management would be operated by an 'independent central data hub', namely a third party as a market facilitator interacting with different smart grid stakeholders and aggregating data from them.[32] Such an independent platform would ensure impartiality vis-à-vis new entrants and thereby ensure the existence of a level playing field for the access to data, subject to regulatory oversight. The European Commission further notes that the existence of a central player would ease legal enforcement, while also reckoning that its creation is likely to be costly and time-consuming, especially for TSOs and DSOs.

Thirdly, the so-called market-based approach builds on standardised interfaces installed with each consumer, that allow storing and accessing the data locally ('Data Access Point Manager' option).[33] Such a commercial role is played by companies acting as data gatekeepers, providing data access to stakeholders. The Data Access Point is close to the relevant device (eg. the smart meters), so that this option is a decentralised one. In contrast, there is no central handling of data in such option.[34] This option easily enables consumers to make choices on their preferences as for the (re)use of data relating to them.

Fourthly, Brandstätt and al. suggest yet another governance option, the 'Common Information Platform ('CIP')'. The CIP constitutes a collaborative governance of data management activities by interested stakeholders, including, horizontally, network operators to prevent fragmentation. The authors hold that such a collaborative governance of data management activities would best allow to balance between competition and coordination objectives that are ascribed to data. Taking into account the history of network industries regulation, a CIP-based approach would allow to avoid discrimination in the access to data by third parties, by including them as stakeholders in the governance mechanisms. On the other hand, the CIP would not stumble over weak coordination challenges faced by unbundled or independent operators since it would *not unbundle the smart systems itself*, but merely the decision-making process, to which stakeholders would be associated. Subject to reliable decision-making mechanisms in place, a collaborative governance approach could mitigate the risk of anticompetitive behaviours of monopolies. With respect to consumer protection and personal data protection, the representativeness of consumers and data subjects in the CIP could be a means to collectively empower them.

4. (In-)vehicle data governance

Road vehicles are increasingly becoming connected devices. They produce a wealth of data, expected to feed the creation of new and personalised services and products and to optimise existing business models in the whole automotive value chain. On the flip side, some of the data could constitute an essential facility for some actors in the automotive sector, such as independent repairers, in the sense that denial of access would prevent them from operating in the maintenance markets. An interest in in-vehicle data and resources has indeed been expressed by repairers and maintainers, parts producers, distributors, but also insurers, entertainment service providers, navigation providers, road authorities and others.[35] OEMs are

[24] Marius Buchmann, 'Governance of Data and Information Management in Smart Distribution Grids: Increase Efficiency by Balancing Coordination and Competition', *Utilities Policy* 44 (1 February 2017): 63–72, https://doi.org/10.1016/j.jup.2017.01.003.

[25] Tijs van den Broek and Anne Fleur van Veenstra, 'Governance of Big Data Collaborations: How to Balance Regulatory Compliance and Disruptive Innovation', *Technological Forecasting and Social Change* 129 (1 April 2018): 330–38, https://doi.org/10.1016/j.techfore.2017.09.040; Buchmann (n 22).

[26] Commission Staff Working Document Impact Assessment Accompanying the document Proposal for a Directive of the European Parliament and of the Council on common rules for the internal market in electricity (recast) [...], SWD/2016/0410 final – 2016/0379 (COD).

[27] Electricity Directive, Art. 23 (1).

[28] Council of European Energy Regulation (CEER), Review of Current and Future Data Management Models CEER report Ref: C16-RMF-89-03, 2016, available here: https://www.ceer.eu/documents/104400/-/-/1fbc8e21-2502-c6c8-7017-a6df5652d20b (last visited 27th April 2020).

[29] Electricity Directive, Art. 23 (4).

[30] Electricity Directive, Art. 34.

[31] Smart Grid Task Force – EG3 Report: EG3 First Year Report: Options on handling Smart Grids Data, 2013, <https://ec.europa.eu/energy/sites/ener/files/documents/xpert_group3_first_year_report.pdf> accessed 30 April 2020, 8-9.

[32] Ibid, 10-11.

[33] Brandstätt et al., (n 23).

[34] Smart Grid Task Force (n 31) 12-13.

[35] M McCarthy et al., 'Access to In-Vehicle Data and Resources', Publications

tempted to secure the centralisation of vehicle data by implementing the so-called 'extended vehicle' model, in which data from all vehicles of the same brand are directly transmitted to a proprietary back-end server of theirs, where they could possibly be made available to third parties.[36] OEMs argue that such a closed system would be necessary to ensure safety and (cyber)security of data and vehicles, by preventing third parties' applications to enter the vehicle system directly. As a result, vehicle data are *de facto* held by OEMs, who enjoy an exclusive access and control over such resources[37] - some even talk about a form of technological 'ownership'.[38] A broad consensus was therefore formed around the idea that vehicle data shall be shared to a range of actors in the automotive industry or even possibly to actors outside the sector (e.g. to feed infotainment services operators),[39] in order to prevent anti-competitive behaviours from OEMs and to boost innovation.

In order to ensure fair and undistorted competition between independent operators and authorised dealers and repairers, EU Regulation 715/2017 does already provide OEMs with data sharing obligations to the benefit of independent operators with respect to vehicle repair and maintenance information. Data sharing obligations are based on 'FRAND' conditions, and especially non-discrimination between authorised dealers and repairers on the one hand, and independent operators on the other. They are accompanied by requirements regarding data format as well as the channel by which data shall be made available for reuse (through websites[40] and the 'On-Board Diagnostic' ('OBD') system amounting to a quasi-open technical standard for access and data interoperability).[41] Adopted prior to the arrival of digital and real-time car data,[42] Regulation 715/2017 is however limited in scope, both in terms of data categories and resources and in terms of beneficiaries, and has been outpaced by technological progress. OEMs are now, again, in a position to foreclose adjacent markets and prevent data from being broadly reused, which would call for further anticipatory regulatory initiatives.[43]

It has been clear from the outset that extending data sharing obligations falling on OEMs shall be balanced with other – possibly contradictory - parameters, such as safety and cybersecurity of vehicles, the risk of extending the liability exposure of the OEMs, the need to secure return on investment made by OEMs and the need for consumer protection.[44] Importantly, individuals (drivers and/or holders of nomadic devices) shall be protected with respect to the processing of their personal data. A consensus emerged that they should be given the opportunity to consent prior to any re-use of personal data. Although no major regulatory action has been taken so far, there have been significant discussions. In this context, the "extended vehicle" data model has been contrasted with alternative ones, accompanied by a large number of options and sub-categories. The names of the models are not uniformly used and keep evolving, which adds another layer of complexity.

To simplify and give a taste of the discussion, the "on-board application platform" model provides access to vehicle data and the execution of (third parties') applications inside the vehicle environment, either based on vehicle embedded systems or not. In turn, the "In-vehicle interface" model consists in an upgraded OBD interface inside the vehicle. While data would be directly accessible via the OBD interface, applications would remain outside the vehicle.[45] Both options are criticised (especially by OEMs) for not providing sufficient security assurance. They would also lack operational maturity, when it comes to real-time data provision.[46] Save the implementation of specific (regulatory) safeguards, the in-vehicle interface model could also prevent OEMs from exploiting their control of the data to reward their investment, which could remove their incentive to keep developing the necessary technical solutions.[47]

While the "extended vehicle" model put for by OEMs has been criticised for allowing them to retain exclusive control over data stored and processed in their back-end server, other technical solutions propose to retain the back-end server option but to have it controlled by other entities. In the "shared server" model, the back-end server would be controlled by a consortium of stakeholders, beyond the sole OEMs, with equivalent link to the vehicle. In turn, the "B2B marketplace" model (also called "commercial platform provider" or "neutral server provider")[48] would consist in creating an additional layer between the vehicle and the service providers, fed by the OEMs back-end servers but maintained by a service provider who would facilitate access by the market (such as Google or IBM).[49] As evaluated by Martens and Mueller-Langer, such a model could generate efficiency gains from economies of scale and scope in data collection across car brands, and by incurring the high fixed cost of setting up a data platform. This model would also facilitate the adoption of standards across brands. On the flip side, they also highlight that such platform would turn into monopolies and may be prone to new anti-competitive behaviours. Whether they would have sufficient room of manoeuver to negotiate with large OEMs as exclusive data providers remains an open question therein.[50]

5. Governing data: Learning from the electricity and automotive sectors

In both cases and with sectoral differences, much of the discussion focusses essentially on the determination of which governance mechanisms shall be established to best regulate data as a resource. This section draws critical lessons from these two cases with respect to the data flow paradigm.

While the existence of governance mechanisms is necessary, there can be a great array of them. In both sectoral cases, the data market - as data governance mechanism underpinning the data flow paradigm – appears to constitute (only) one of the available options, whose respective benefits and drawbacks are assessed against the context-specific objectives and constraints at stake. The above sections provide neither an exhaustive overview of all objectives and constraints nor their impact on the assessment of the various data governance mechanisms. Yet, several of them come to light, such as the data protection law, reliability, safety and (cyber)security considerations, 'time to market' of technical tools, the need to ensure a

Office of the European Union, 2017, 29.

36 Kerber (n 15) 311.

37 Ibid.

38 Cynthia Delronge and Alain Strowel, 'Data Sharing For a Smarter Mobility and For Connected Vehicles: How the Design of the Data Flows Contributes (or Not) to Transport Policy and Innovation', in *Des Véhicules Autonomes à l'intelligence Artificielle - Droit, Politique et Éthique*, Christophe Lazaro, Alain Strowel (Bruxelles: Larcier, 2020), 200–201.

39 European Commission (n 14).

40 Regulation (EC) No. 715/2007 (n 11).

41 Martens and Mueller-Langer (n 10) 11.

42 Ibid.

43 Delronge and Strowel (n 38).

44 McCarthy et al., (n 35) 9.

45 Ibid 43–45.

46 Ibid 43–45.

47 Ibid (n 35) 27.

48 Ibid (n 35) 47.

49 Ibid (n 35) 6.

50 Martens and Mueller-Langer(n 10).

return on the investment made by the incumbent data holder, fair and undistorted markets at all levels, innovation as an objective, the need to ensure that the making available of data does not affect the original business of the incumbent data holder, the need for coordination of stakeholders, etc.

Although with obvious differences, the on-going discussions in both the electricity and the automotive sectors display striking similarities with respect to the considered governance mechanisms. For example, the CIP proposal in the electricity sector and the 'shared server model' in the automotive one seem to have in common that a range of pre-determined stakeholders, although not the exact same categories, would jointly make decisions about the resources at stake. In both cases, this option is justified by the need to empower (deemed) weaker parties, particularly independent operators and new entrants in both cases. The CIP adds to that representatives of individuals in their quality as customers and data subjects, which does not seem to be present in the shared server model, mostly viewed as an industrial consortium. Both the CIP and the shared server model are expected to preserve against monopole and/or monopolisation of the resources, by bringing together both big and small players.

In the same vein, the 'independent central data hub' option in the electricity sector and the 'B2B marketplace' in the automotive one do share similarities. They both consist in the deliberate (regulatory?) creation of a new data platform layer in the value chain operated by an independent player. Such an operator would assume a new monopolist role, with a view to facilitate the relationships between data providers on the one hand (mainly DSOs and TSOs in the electricity sector and OEMs in the automotive one) and data customers on the other. In both cases, the creation of such a central player is motivated by the need to ensure non-discrimination and impartiality vis-à-vis the activities conducted by the data provider (electricity distribution in the electricity sector and vehicle manufacturing in the automotive one), and therefore fair and undistorted markets. The creation of such a new data platform is expected to be resource- and time-consuming. By creating a new layer between data providers and data customers, it could also create transaction costs, following observations on the creation of monopolist physical infrastructure managers in some liberalised industries. Such governance option could take advantage of economies of scale and scope and could facilitate the adoption of standards, as a result of the monopolisation of the activity. Ironically, the creation of a central player to fight anticompetitive behaviours of incumbent data holders may, in turn, raise competition law issues.

Several legal and regulatory conclusions can be drawn from the analysis of the two sectoral examples. First, the study of the selected data governance mechanisms in both sectors shows a clear concern for the economic environment of data, *beyond the sole data market and data transaction phase*, to the sector value chain more broadly. The data governance mechanisms are evaluated, *inter alia*, against their ability to empower deemed weaker parties, such as new entrants, consumers and data subjects. The specific risks of 'platformisation' of data intermediation is somehow accounted for, namely, in the parlance of Montero and Finger, the restructuring around the business model of online platforms, which can imply substitution and commoditisation of traditional activities demoted to a mere side of the platform.[51] This is especially so as data-driven online markets were found to nearly always tip, moving "towards monopoly" based on data network effects.[52] It is all the more so that, in both cases, substantial *advantages* can also simultaneously be derived from the aggregation of data across brand (of TSOs and DSOs in the electricity sector and of OEMs in the automotive one), or at least from the possibility to have a comprehensive governance of them across brands, such as better coordination. The independent data platform option would anticipate and 'embrace' such platformisation. For its part, the CIP or shared server model would aim at preventing monopolist platformisation from happening. This meets a more general statement made by Lundqvist. Taking into account the network effects of data, as already observed in data-driven online markets, collaborative governance mechanisms (such as 'data pools') could mitigate the risk of monopolisation since *all relevant stakeholders* participate in the same arrangement.[53]

Second, while, in the data flow paradigm, the regulatory focus is mainly placed on data as a subject-matter, it is never solely about data as a resource. To some extent, it is also about data management as a set of data activities and, specifically in the automotive sector, about the underlying technologies, whether servers, platforms or interfaces. As a technological asset, data are indeed not standalone but remain highly reliant on their technological environment. As underlined by Delronge and Strowel, it is the control over the technology which enables some well-placed stakeholders to retain a form of *de facto* ownership, exemplified by the use of brand-specific back-end servers by OEMs in the automotive sector.[54] Or, in the parlance of Lessig, "code is law".[55] In turn, regulating the access and control over the server *as a means to arrange access and control over data* is an illustration of the phenomenon, described by Lessig, where the law is designed to have an indirect effect, by leveraging "code" or the technological architecture as another "modality of regulation".[56] While striking, it should therefore not surprise us that, in the automotive sector, much of the discussion on the governance of data consists in a technical discussion on the supporting technologies thereto.

Besides, many of the other governance mechanisms would require to partly shift - or *extend* – the regulatory focus to the *stakeholders*, whether an independent data platform or the decision-making rules for a consortium of stakeholders in the above examples, to establish them and/or to *regulate their operation*. The independent data platform model comes with obvious risks of anti-competitive behaviour, which could require *ex ante* regulatory intervention beyond the sole operation of competition law. Much of the regulatory 'pressure' would similarly shift to the CIP or shared server model. In order to ensure that all relevant stakeholders take part, they could for instance be mandated by law to participate, inspired by the "open data pool" model described by Lundqvist.[57] It would remain to be seen whether the legislator should further intervene with respect to the decision-making process, for example to re-balance power asymmetries between stakeholders or to prohibit certain data processing activities (eg. to protect data subjects and consumers).

Finally and to wrap up, the study of data governance mechanisms in the electricity and automotive sectors challenges the implicit assump-

[51] Juan J Montero and Matthias Finger, 'Platformed! Network Industries and the New Digital Paradigm' [2018] *Competition and Regulation in Network Industries* 1783591718782310.

[52] Jens Prüfer and Christoph Schottmüller, 'Competing with Big Data', Discussion Paper, Tilburg Law and Economics Center (TILEC) *Law and Economics Research Paper Series* (Tilburg: Tilburg University, 2017) 1.

[53] Bjorn Lundqvist, 'Competition and Data Pools', *Journal of European Consumer and Market Law* 7, no. 4 (14 August 2018): 146–54.

[54] Delronge and Strowel (n 38) 198.

[55] Lawrence Lessig, 'The Law of the Horse: What Cyberlaw Might Teach', *Harvard Law Review* 113, no. 2 (1999): 501–549.

[56] Ibid.

[57] Lundqvist (n 53).

tion of a naturalness of the data flow paradigm. It also illustrates the importance of having a regulatory purpose. Viewing data and the market for such data as the regulatory target in the data flow paradigm seems to bear the implicit idea that there *should* be well-operating such markets, possibly by the operation of the law, and that *they alone* would deliver the grand policy expectations, such as 'AI' and 'data-driven innovations'. The data flow paradigm seems to detach the data transaction phase from the technological and economic environment of data. As a result, the policy objectives linked to the data flow paradigm seem both imprecise, short-sighted, and not context-specific enough. Fostering data exchange is not an end in itself and should thus be regarded with respect to the sectoral objectives and constraints, sometimes contradictory to each other. To be clear, this conclusion should not be interpreted as pleading against any form of horizontal 'data law' which could particularly be necessary to democratically determine *who* has legitimate entitlements on data (or 'data rights').[58]

6. Conclusion: brand new, same old song, or somewhere in between?

While it is contended here that the (sectoral) context, in terms of both objectives and constraints, shall be taken into account when regulating data as a resource, this should not be interpreted as an obstacle to knowledge, action and improvement of how data could be governed. Based on the study of the electricity and automotive sectors, this concluding section opens avenues for further research and regulatory intervention.

However new they may be, it is striking that the quest for appropriate mechanisms to govern data often leads to rediscovering old concepts, as can be observed in the electricity and automotive sectors. The independent data platform, as data intermediary, coordinates data demand and offer. A quick look back at recent history shows that online platforms have emerged in environments characterized by fragmentation, where they offer new types of data-driven aggregation and intermediation. Such scenarios have for example been observed in the context of network industries characterised by large number of actors in freshly liberalised environments.[59] Subject to both vertical unbundling and decentralisation of supply, the electricity sector is obviously a prominent illustration thereof. The scholarship also began to observe the emergence of data platform intermediaries in the data sharing economy or data marketplaces.[60] The independent data platform model does not take away the markets for data, but it structures them by adding a layer in the vertical value chain. The creation of this new layer can be compared to the creation of independent managers of physical infrastructure as a result of the liberalisation of network industries, such as in the railways or the aviation sectors. Taking the vertical unbundling mechanism to the extreme, it results in the creation of both a new market and a new product, namely train paths and airport slots in the railways and in aviation. The independent data platform model goes however a step further. As a platform, it brings coordination in both the data demand side *and the data offer side*, by bringing together various brands of data producers. There can of course be a variety of options for the independent data platform. For instance, whether the data platform would pool the data or whether it would merely facilitate the data transaction constitutes a crucial question. Whether such independent data platforms would be established by law or not, the regulation of their operation remains an important question, which could be informed by on-going discussions on the regulation of online platforms. Particularly in sectors (such as electricity) characterised by public service activities, it cannot be excluded that the independent data platform could be viewed – and regulated - as a novel form of (data) utility.

For their part, the CIP and the shared server model could be akin to commons, as defined by Ostrom in 1990, in the sense that they amount to "institutionalised arrangements of community management or governance of shared resources".[61] A reservation should however be made regarding the CIP, which is portrayed as a decision-making body without actual sharing of the resources at stake, namely the data. In any case, a commons-like model is viewed, in both situations, as a means to accommodate the competing – and sometimes contradictory – needs of the various stakeholders, subject to decision-making arrangements between them. As a matter of fact, 'data commons' have recently gained traction as a form of collaborative governance mechanism to govern data in many instances. Just like in the electricity and automotive examples, they are often advocated for as a means to counterbalance power asymmetries in data environments, such as with online platforms like Facebook[62] or in the Smart Farming industry.[63] Much can therefore be learned from other 'commoning' experiences and, from a regulatory perspective, on how the law can support the establishment or even the operation of such governance mechanisms.

This calls for a empiricist and pragmatic perspective, following the work of Ostrom with the institutional analysis and development framework, and then by Frischmann, Madison and Strandburg with their Governing Knowledge Commons framework.[64] Data governance concerns have (re)surfaced with technological developments, which have multiplied the value of – and thus the greed for – data and have prompted governments to enhance data (re)use, in expectation of innovation and growth benefits. Many factors are found to have an influence on the respective fitness of governance mechanisms in a given context, as outlined in the electricity and automotive sectors. Gathering and analysing these factors can certainly inform the governance of data in other situations. Such an exercise is beginning to be carried out in the scholarship. For instance, Van den Broek and Van Veestra showed that compliance with data protection is of great concern for participants of what they call "big data inter-organisation collaborations" (or data pools). Their empirical research finds that the presence of personal data has an impact on the design of the governance mechanisms, and results in more hierarchical control in the collaboration.[65] To begin with, the European Commission could be well advised to launch an observatory, just like for other topics.[66]

[58] On this question, see also the on-going work of ALI-ELI on the "Principles for a Data Economy" (n 3).

[59] Montero and Finger (n 51).

[60] Heiko Richter and Peter R. Slowinski, 'The Data Sharing Economy: On the Emergence of New Intermediaries', *IIC - International Review of Intellectual Property and Competition Law* 50, no. 1 (1 January 2019): 4–29, https://doi.org/10.1007/s40319-018-00777-7.

[61] Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, The Political Economy of Institutions and Decisions (Cambridge ; New York: Cambridge University Press, 1990); Jennifer Shkabatur, 'The Global Commons of Data', *Stanford Technology Law Review* 22, no. 1 (2019): 354–411.

[62] Shkabatur (n 61).

[63] Jeremiah Baarbé, Meghan B, and Jeremy de Beer, 'A Data Commons for Food Security', *Proceedings of the 2017 IASC Conference* ; Open AIR Working Paper No. 7/17 (Rochester, NY: Social Science Research Network, 2017).

[64] Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg, eds., *Governing Knowledge Commons* (Oxford, New York: Oxford University Press, 2014).

[65] van den Broek and van Veenstra (n 25).

[66] To remain in the digital economy, the European Commission launched for instance the EU Blockchain Observatory and Forum in 2018 https://www.

The observatory could map existing or considered data governance mechanisms[67] and analyse the contextual factors for their success or failure. The outcome would be valuable for researchers, players in the field as well as policy- and law-makers alike.

Telling from its Communication 'A European Strategy for Data', the European Commission appears to be embracing data governance mechanisms, beyond the sole data flow paradigm, as measures to share and govern data, account being had to their (sectoral) environment.[68] The Communication significantly refers to 'data cooperatives', 'data pools', 'data trusts' as data governance mechanisms. The Communication reckons the need for "organisational approaches and structures (both public and private)". It is based on a seeming attempt to balance between horizontal and context-specific regulation of data that the European Commission commits to regulate the governance of 'common European data spaces' in the coming months. The Communication includes an Appendix listing the common European data spaces in "strategic sectors and domains of public interest" where the EU shall therefore be specifically involved. The automotive industry is indicated as part of the 'Common European mobility data space'. The Communication does not expressly anticipate regulation of data governance, but refers to the on-going review of the current EU type-approval legislation for motor vehicles, in order to "open it up to more car data based services" by early 2021. According to the Communication, the review shall look at "how data is made accessible by the car manufacturer, what procedures are necessary to obtain it in full compliance with data protection rules and the role and rights of the car owner".[69] The electricity sector makes part of the 'common European energy data space'. While the Communication confirms that "the specific governance frameworks" shall be defined at national level, the European Commission will further regulate interoperability requirements, as laid down in the Electricity Directive. The concern of the European Commission for contextual data governance mechanisms can be analysed as a move beyond the data flow paradigm and shall be welcomed positively.

eublockchainforum.eu/about accessed 11 May 2020, and a group of experts for the Observatory on the Online Platform Economy, launched also in 2018 https://platformobservatory.eu/about-observatory/introduction accessed 11 May 2020.

[67] Data governance classifications are already being elaborated, based on concrete illustrations, although with different angles and scope and with no uniform taxonomy being used. See for instance Gov Lab with respect to DATA COLLABORATIVES https://datacollaboratives.org accessed 11 May 2020, or the OPEN DATA INSTITUTE https://theodi.org accessed 11 May 2020.

[68] European Commission, Communication 'A European strategy for data', COM(2020) 66 final, 19.2.2020.

[69] Ibid, 27-28.

07

Special issue: Governing Data as a Resource

Technology and Regulation

Defining Data Intermediaries

A Clearer View through the Lens of Intellectual Property Governance

Alina Wernick*, Christopher Olk** and Max von Grafenstein***

Data governance, data ownership, data protection, intellectual property, data intermediaries

alina.wernick@hiig.de
chrstphr.olk@gmail.com
max.grafenstein@hiig.de

Data intermediaries may foster data reuse, thus facilitating efficiency and innovation. However, research on the subject suffers from terminological inconsistency and vagueness, making it difficult to convey to policymakers when data governance succeeds and when data sharing requires regulatory intervention. The paper describes what distinguishes data intermediaries from other data governance models. Building on research on intellectual property governance, we identify two distinct types of data intermediaries, data clearinghouses and data pools. We also discover several governance models that are specific to data and not present in the context of intellectual property. We conclude that the use of more refined terminology to describe data intermediaries will facilitate more accurate research and informed policy-making on data reuse.

1. Introduction

1.1 Siloed data

Data is becoming increasingly important for innovation in contemporary industries. Despite its status as an intermediate, non-rival good with the ability to create strong spillover effects,[1] it is often siloed. The insufficient reuse of data is likely to adversely impact economic efficiency and innovation, and it may lead to wasteful, duplicative investments into the reproduction of data.[2]

In some contexts, the obstacles to data sharing are legal (and frequently justified). These constraints may arise from data protection law or the protection of intellectual property rights and trade secrets. However, while data is not subject to property rights, the data holder may still exclude others from using it. In most cases, the constraints to data reuse stem from the factual control of data and the data holders' incentives to share it with others. There may be a number of reasons why motivation to share data is lacking.

These motivations may be divided into two categories: a) interest to maintain competitive advantage in the market and b) obstacles arising from operating in a particular context, such as transaction costs. As an example of the former, economic agents may be reluctant to share data with others out of fear of losing a competitive advantage derived from the data.[3] Risk aversion with respect to breaching relevant legislation, such as data protection and intellectual property law,[4] as well as imperfect information on whether the reuse could pose a competitive threat may hence discourage sharing. The data holder may also overestimate the data's value due to an endowment effect.[5] In terms of business strategy, if the appropriability and criticality of a resource are perceived as too high and its substitutability as low, firms tend not to cooperate with other players even if the potential benefit from cooperation is very large.[6] Furthermore, there are several

[1] OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being* (OECD Publishing 2015) 38, 177, 180.

[2] Josef Drexl, 'Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy' (2018) *Max Planck Institute for Innovation & Competition Research Paper* No 18-23,17 https://ssrn.com/abstract=3274519 accessed 28 June 2019; Nestor Duch-Brown, Bertin Martens and Frank Mueller-Langer, 'The economics of ownership, access and trade in digital data' (2017) *JRC Digital Economy Working Paper* 2017-01, 46-47 https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf accessed 13 February 2020.

* Alina Wernick is researcher at the Alexander von Humboldt Institute for Internet and Society in Berlin and doctoral candidate at the Ludwig-Maximilians-Universität in Munich, Germany.
** Christopher Olk is student assistant in the Data Governance project at the Humboldt Institute for Internet and Society in Berlin, Germany.
*** Max von Grafenstein is co-head of the research program „Governance of Data-Driven Innovation" at the Alexander von Humboldt Institute for Internet and Society as well as professor for „Digital Self-determination" at Einstein Center Digital Future appointed to Berlin University of the Arts.

[3] In some situations, data may also qualify as a trade secret. See Josef Drexl, Reto Hilty, Luc Desaunettes, Franziska Greiner, Daria Kim, Heiko Richter, Gintare Surblyte, and Klaus Wiedemann, 'Data Ownership and Access to Data-Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate' (2016) *Max Planck Institute for Innovation and Competition Research Paper* No 16-10, 6 https://ssrn.com/abstract=2833165 accessed 28 June 2019.

[4] Max von Grafenstein, Alina Wernick and Christopher Olk, 'Data Governance: Enhancing Innovation and Protecting Against Its Risks' (2019) 54 *Intereconomics* 228, 228–232; Heiko Richter and Peter R Slowinski, 'The Data Sharing Economy: On the Emergence of New Intermediaries' (2019) 50 *IIC* 4, 7 fn 15.

[5] Daniel Kahneman, Jack N Knetsch and Richard Thaler, 'Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias' (1991) 5 *J Econ Perspectives* 193, 195; Angela G Winegar and Cass R Sunstein, 'How Much Is Data Privacy Worth? A Preliminary Investigation' (2019) 42 *Journal of Consumer Policy* 425, 425–440.

[6] Anne-Sophie Fernandez and Paul Chiambaretto, 'Managing tensions related to information in coopetition' (2016) *Indust Mar Mgmt* 53.

characteristics that differentiate a data economy from other economies, including the non-linear returns from the scope of data, the intense concentration observed in many data markets,[7] or the 'growth before profit' strategies of many data holders,[8] which create additional incentives against sharing data in cases of strategic uncertainty and imperfect information.[9]

Other obstacles to data sharing are more context dependent and do not directly reflect the strategy of an individual data holder. They are related to more common market failures, namely the discrepancies between social and private interests. These discrepancies create inefficient market outcomes even under conditions of perfect information. Even if all parties involved can assess the risks adequately and see that the benefits of sharing are greater than the risks, the collective action problem remains: each party may have insufficient incentives for participating in data sharing and in creating an infrastructure for sharing if they can each expect a sufficiently large number of the other parties to share data and invest in the infrastructure.[10] However, other market failures arise due to excessive transaction costs,[11] which may hinder data holders and the potential users of data from finding each other. The costs of identifying and devising a method for sharing data which complies with data protection, trade secret, intellectual property, or competition law may also limit its reuse, even if such a method exists. Furthermore, transaction costs can also arise due to insufficient interoperability between data sets, data formats semantics, application programming interfaces (APIs), and other structures. Excessive transaction costs may lead to a situation akin to a 'tragedy of the anti-commons'[12] where data transactions are so costly that data sets end up not being shared and combined, even if they are highly complementary.

1.2 Data governance models as a potential solution

How should the disincentives to sharing data be addressed by policymakers or legislators? One approach is to explore to what extent different governance models can foster forms of data sharing that are both efficient and legally compliant. We use the term data governance models (DGMs) to refer to institutions, i.e. assemblages of legal and social norms, and organizational and technical designs that interact and determine the conditions for the interorganizational sharing of data. DGMs may be particularly helpful for addressing the more context-dependent obstacles to data sharing. According to the approach outlined above, a legislator or policymaker should only interfere when the market fails (or in this case, when private ordering[13] through data governance fails).[14] For example, this may occur when a data holder's incentives to permit the reuse of data are insufficient and the lack of access to data proves detrimental to social welfare. Data governance may also fail due to other obstacles, such as excessive implementation costs. From the legal perspective, data sharing market failures may be resolved by enacting an access right.[15] However, other policy measures may also be employed, such as financial incentives to sharing data or found intermediaries.

However, making policy recommendations in favour of data reuse is difficult at this moment because we lack a systematic review of existing or potential DGMs in different sectors and their effectiveness in fostering data reuse. Furthermore, the vague and heterogeneous terminology applied to data intermediaries both in practice and in research[16] makes it difficult to learn from existing practice and studies on DGMs for the purposes of policymaking.

The existing research on sharing intellectual property (IP) may be relevant for fostering understanding of opportunities and limits of data governance. In particular, the research on IP clearinghouses and patent pools is helpful for categorizing DGMs and for enhancing the terminology applied to data intermediaries. However, one should exercise caution when applying findings from IP to data because, both from an economic and a legal perspective, they represent different types of goods. While both IP and data are inputs for innovation, unlike patented inventions and copyright protected works, data is not subject to exclusive rights which would give rise to a right to exclude others from using this knowledge resource.[17] As a result, any agreements to transfer, share, and maintain the data within a specific circle of recipients would only have inter partes effects[18] and would require additional organizational and technical measures to maintain de facto control of the data.[19] Furthermore, the General Data Protection Regulation (GDPR) sets out specific conditions for processing personal data. It impacts data governance, for example, by mandating the implementation of appropriate technical and organizational measures by means of pseudonymization technologies in the data intermediary's infrastructure.[20]

We present our categorization of DGMs in Section 3 in order to illustrate the role of data intermediaries among other data governance solutions and to clarify the terminology that is used to refer to diverse DGMs for future research. We identify two main categories of data intermediaries: data clearinghouses and data pools. Sections 4 and 5 discuss these DGMs in detail and review to what extent these data intermediaries differ from their counterparts in IP governance. Drawing on this analysis, we identify several DGMs specific to data.

7 Vikas Kathuria, 'Greed for data and exclusionary conduct in data-driven markets' (2019) 35 *CLS Rev* 89; Jacques Crémer, Yves-Alexandre de Montjoye and Heike Schweitzer, 'Competition policy for the digital era' (Report for the European Commission 2019), 2, 4-5, 99. https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf.

8 Nick Srnicek, *Platform Capitalism* (Polity 2017), 75-76.

9 There are also obstacles to the sharing of data which are attributable to intra-organizational dynamics. Although we draw partly from management literature, these aspects are beyond the scope of our research.

10 Mancur Olson, 'Collective action' in John Eatwell, Murray Milgate, Peter Newman, (eds.) *The Invisible Hand* (Palgrave Macmillan 1989) 61.

11 Ronald H Coase, 'The Problem of Social Cost' (1960) In C Gopalakrishnan (ed.) *Classic Papers in Natural Resource Economics* (Palgrave Macmillan)

12 Michael A Heller and Rebecca S Eisenberg, 'Can patents deter innovation? The anticommons in biomedical research' (1998) 280 *Science* 698; von Grafenstein, Wernick and Olk (n 4) 229 fn 16.

13 We understand private ordering in the meaning of 'self-regulation voluntarily undertaken by private parties', Niva Elkin-Koren, 'What contracts cannot do: The limits of private ordering in facilitating a creative commons' (2005) 74 *Fordham L Rev* 375, 376.

14 See Drexl (n 2) 8.

15 Drexl (n 2) 8

16 von Grafenstein, Wernick and Olk (n 4) 232.

17 Furthermore, similar to trade secrets, data may be subject to Arrow's information paradox as it is difficult to assess the value of data without getting access to it and once the prospective buyer sees the data, she may no longer be interested in paying the price for it. Kenneth J Arrow, *The economics of information* (Vol. 4 Harvard UP 1984)). By contrast, the information on patented technology is by definition public, making it easier to assess the value of a patent.

18 see Josef Drexl et al (n 3) 3.

19 In the same vein, even though data transfer agreements are often referred to as 'licensing agreements', their conditions apply only to the contracting parties.

20 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR), [2016] OJ L119/1, art 25 s 1.

2 Research approach

2.1 Method

We explored data sharing practices in DGMs from an interdisciplinary perspective, focusing especially on data intermediaries. We relied on the legal, economic, and policy literature analysing the governance of intellectual property[21] and data,[22] the economics of privacy,[23] competition in data-driven industries,[24] open (and user) innovation,[25] and co-opetition.[26] We also studied the literature and online resources on data governance in the advertising, automotive, and e-health sectors and conducted interviews with experts in these fields in order to map the possible constellations of stakeholders, conflicts of interest, and sharing practices in different legal, economic, and technological contexts. We determined which DGMs to discuss in the paper by means of iterative comparison between concepts and practices identified in the literature and those present in the reviewed sectors.

2.2 Terminology

Acknowledging that different legal norms apply to personal data and non-personal data[27] unless stated otherwise, we use the term 'data' to refer to both of its legal subcategories. In the description of the DGMs, we employ the concept of a 'data holder' to refer to the natural and legal persons who have actual control over non-personal data or over personal data of which they themselves are not the subject. 'Data users' are natural or legal persons interested in data for the purposes of reuse and to whom data is transferred in the particular DGM. In alignment with Article 4 section 1 of the GDPR, we use the term 'data subject' to refer to a natural person who is the 'source' of personal data, especially when she is an active subject in the context of a specific DGM. Whenever we discuss DGMs specific to personal data, we employ the GDPR's terms 'controller' and 'processor' to specify the roles and responsibilities of data holders.[28]

The term 'platform' is often used to refer to a number of different DGMs.[29] For this reason, we consciously refrained from referring to DGMs as platforms. Instead, we look at the degree of platformization in an individual DGM, i.e. the extent to which it employs a platform-type business model. We further define platforms as intermediaries that leverage the data being transacted via their infrastructure and that capture part of the value created through them.[30] For example, picture two intermediaries that facilitate the exchange of data. If one of them accesses the exchanged data and uses it to train an algorithm while the other does not, then it exhibits a higher degree of platformization than the other.[31]

3 Data Governance Models – a typology

In our research, we focused on three different governance layers (i.e. the normative / legal layer, the organisational layer, and the technical layer)[32] and we identified five categories of DGM's based on their defining features: closed DGMs, single source DGMs, clearinghouses, data pools, and distributed DGMs (Figure 1). These DGMs represent abstract solutions for governing interorganizational data exchange which are not specific to any sector. In essence, one could think of them as "ideal types," a concept introduced by Max Weber and fruitfully applied in earlier research on governance structures.[33] In essence, the DGMs introduced in this paper represent boundary objects - abstract "concepts [that are] plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites."[34]

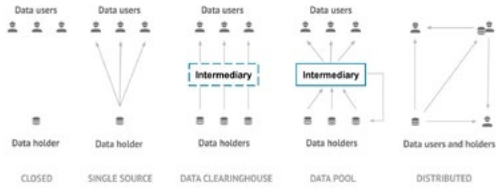


Fig. 1: Data Governance Models

[21] See Michael A Heller, 'The tragedy of the anticommons: property in the transition from Marx to markets' (1998) 111 *Harv L Rev* 621; Heller and Eisenberg (n 12); Robert P Merges, 'Contracting into liability rules: Intellectual property rights and collective rights organizations' (1996) 84 *Cal L Rev* 1293; Geertrui Van Overwalle, Esther van Zimmeren, Birgit Verbeure, and Gert Matthijs 'Models for Facilitating access to patents on genetic inventions' (2006) 7 *Nature Reviews Genetics* 143.

[22] See Michael Mattioli, 'The data-pooling problem' (2017) 32 *Berkeley Tech LJ* 179; Björn Lundqvist, 'Competition and data pools' (2018) 77 *J Europ Consumer and Market L* 146; Richter and Slowinski (n 4); Stefaan G Verhulst, Andrew Young, Michelle Winowatan, Andrew J Zahuranec 'Data Collaboratives: Leveraging Private Data for Public Good. A descriptive analysis and typology of Existing Practices (GovLab Report 2019) https://datacollaboratives.org/static/files/existing-practices-report.pdf accessed 14 February 2020; OECD, Enhancing Access to and Sharing of Data : Reconciling Risks and Benefits for Data Re-use across Societies (OECD iLibrary 2019)

[23] Alessandro Acquisti, Curtis Taylor and Liad Wagman, 'The economics of privacy' (2016) 54 *J Econ Literature* 442.

[24] Maurice E Stucke and Allen P Grunes, Big data and competition policy (OUP 2016); Srnicek, (n 8); Kathuria (n 7); Crémer, de Montjoye and Schweitzer (n 7).

[25] Henry Chesbrough *Open innovation: The new imperative for creating and profiting from technology* (2006 Harvard UP); Eric von Hippel Democratizing innovation (2005 MIT Press); Eric von Hippel and Georg von Krogh 'Open source software and the "private-collective" innovation model: Issues for organization science' (2003) 14 *Org Sci* 209.

[26] Ricarda B Bouncken, Johanna Gast, Sascha Kraus and Marcel Bogers, 'Coopetition: a systematic review, synthesis, and future research directions' (2015) 9 *Rev Managerial Science* 577; Fernandez and Chiambaretto (n 6); Bruno Carballa Smichowski, 'Determinants of coopetition through data sharing in MaaS' (Hal Archieves Ouvertes 2018) https://hal.archives-ouvertes.fr/hal-01872063/document accessed 19 June 2019.

[27] GDPR art 1 s 1, art 4 s 1.

[28] GDPR art 1 ss 7-8.

[29] For instance in: Mark de Reuver, Castren Sørensen and Rahul C Basole, 'The digital platform: a research agenda' (2018) 33 *J Info Tech* 33, 124; European Commission, 'Towards a common European data space' SWD (2018) 125 final, 11; European Commission, 'Guidance on sharing private sector data in the European data economy' COM (2018) 232 final 8-11.

[30] Srnicek (n 8).

[31] MindSphere is a 'platform as a service' for intra- and interorganizational data exchange that offers analytics that learn from the data exchanged through the platform; hence, a part of the business model is to leverage the data it transfers. Siemens, 'Mindsphere: Enabling the world's industries to drive their digital transformations' (White paper 2018) https://www.plm.automation.siemens.com/media/global/en/Siemens-MindSphere-Whitepaper-69993_tcm27-29087.pdf?stc=wwiia420000&elqTrackId=e0d6520bc42f4e44952b0a7cf107f372&elq=0859ca3b11b848b7952b9760250a5a6c&elqaid=2984&elqat=1&elqCampaignId= accessed 19 June 2019.

[32] See the previous publication of the authors, von Grafenstein et al (n 4) 231 et seq. Also governance of intellectual property has previously been reviewed from the perspective of three layers of governance. See Elkin-Koren (n 13), 392-397, analyzing creative commons as a social movement from the perspective of law, social norms and technology.

[33] See for example Henrik P Bang, (ed) *Governance as social and political communication* (Manchester UP 2003), 43; Anna Grandori, 'Governance structures, coordination mechanisms and cognitive models' (1997) 1 *J Mgmt & Governance* 29, 31.

[34] Susan L. Star, 'The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving' in Michael Huhs and Lens Gasser (eds): *Distributed Artificial Intelligence, vol 2* (Morgan Kaufmann Publishers Inc 1989) 46, 49.

The categorization is subject to two limitations. First, it focuses on illustrating the governance of reuse of data that is already collected by a data holder - therefore it does not address the governance of the initial data collection, for example via web-scraping or obtaining data from sensors. Second, the described DGMs represent abstractions. In practice, data governance constellations are considerably more complex and may simultaneously display features from several DGMs described below. Furthermore, DGMs, as institutions, may also in practice be nested in one another.[35]

The following subsections briefly introduce the main characteristics of each DGM. We will also briefly discuss practical examples of the three models that do not involve an intermediary: the closed, the single source and the decentralized model. We will review the DGM's that qualify as data intermediaries in more detail in Sections 4 and 5.

3.1 Closed DGM

Closed DGM refers to a situation where data is deliberately not shared with other organizations or people.[36] In the closed DGM for non-personal data, a data holder takes legal, organizational, and/or technological measures to maintain control of her data. Despite the objective to refrain from interorganizational data sharing, an organization adopting a closed DGM may nevertheless feature a sophisticated governance model for intraorganizational sharing of data.[37] In fact, implementing the appropriate policies, processes, and mechanisms for intraorganisational data sharing has been for long the focus of data governance literature.[38] The typical case for this DGM is the so-called data silo. As mentioned in the introduction of this paper, the closed DGM was for a long time the natural state, until data has been "discovered" as the new oil of the digital society. An example for this could be public sector information, before legal regulation enforced public agencies to open their data silos for the public.[39]

The closed DGM cannot effectively be adopted by natural persons with respect to their personal data, as data subject living in a modern society cannot to completely prevent the processing of her personal data,[40] since records containing personal data are kept since birth. From a more relative perspective, home environment has been traditionally perceived a the most private sphere, which an individual valuing privacy can choose to govern his personal data following the closed DGM, i.e keeping the doors shut.[41] However, the adoption of smart phones and smart home technology is currently undermining the individual's control over the processing of personal data derived from the home environment.[42] In turn, a controller of personal data can also not employ a pure closed DGM, due to data subjects' access rights. As a consequence, the closed DGM is most closely associated with non-personal data.[43]

As discussed before, the reasons for not sharing data are heterogeneous. From a normative perspective, employing a closed DGM is undesirable in situations where data sharing would facilitate innovation without undermining the rights of data subjects or creating anticompetitive effects, for example where the withholding of access to data would preclude competition in the downstream market.[44]

3.2 Single-source DGM

In the simplest form of interorganizational governance of non-personal data is the single-source DGM,[45] wherein the access is provided by an individual holder of the data on the terms she decides upon.[46] For example, in the automotive sector, this governance model is represented by the "extended vehicle" proposition, where access to vehicle data is under the control of the original equipment manufacturers (OEMs)[47] and provided to other companies on the basis of bilateral agreements and though an OEM-controlled technical interface.[48] The classical data brokers that sell access to consumer data, such as Acxiom, represent single source DGM's, regardless of whether they had collect the data themselves from heterogeneous sources, or buy from other commercial actors. Typically, data brokers provide data for the purposes of marketing, risk mitigation and the so-called "people search".[49] However, a single-source DGM can be employed also in the context of R&D, and used for the purposes of open innovation. For instance, Astrazeneca offers data from preclinical studies in its

35 See Michael J Madison, Brett M Frischmann and Katherine J Strandburg, 'The University as Constructed Cultural Commons' (2009) 30 *Wash U JL & Pol'y* 365, 385-386.

36 The closed DGM reflects data governance in the spirit of the "closed innovation" paradigm, wherein innovation process is governed strictly within the firm boundaries. Henry Chesbrough, 'Open innovation: a new paradigm for understanding industrial innovation.' in Henry Chesbrough, Wim Vanhaverbeke and Joel West (eds): *Open innovation: Researching a new paradigm* (OUP 2006) 2-3. However, the other DGMs discussed in this chapter are not ranked on the basis of their openness, as each of the models can be used to facilitate sharing only to a limited set of users or to anyone willing to access data. See OECD, Enhancing Access, (n 22) ch 2.

37 Furthermore, the adoption of the closed DGM with respect to data does not preclude the data holder from sharing the results of data analysis more openly. Verhulst et al (n 22), 36.

38 See, for instance, John Ladley, *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program* (Morgan Kaufmann 2012); however, see more recent approaches taking also the sharing between organisations into account, for instance, Barbara Engels, 'Data Governance as the Enabler of the Data Economy' (2019) 54 *Intereconomics* 216, 217, referring to the DEMAND project, online accessible at https://demand-project.de/.

39 See Council Directive 2019/1024 on open data and the re-use of public sector information (Open Data Directive) [2019] OJ L192/27, 56–83 revising the earlier Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. (PSI Directive)

40 GDPR arts 6 and 9.

41 Gabriele Britz, 'Informationelle Selbstbestimmung zwischen Grundsatzkritik und Beharren des Bundesverfassungsgerichts' in Wolfgang Hoffmann-Riem (ed.) *Offene Rechtswissenschaft: ausgewählte Schriften von Wolfgang Hoffmann-Riem mit begleitenden Analysen* (Mohr Siebeck 2010), 588-591.

42 See, for example, Alexa D Rüscher, 'Siri und Google als digitale Spione im Auftrag der Ermittlungsbehörden? Zur Abgrenzung von Quellen-TKÜ, Onlinedurchsuchung und akustischer Wohnraumüberwachung' (2001) 12 *NStZ*, 687 et seq.

43 GDPR arts 15 and 20.

44 Autorité de la concurrence and Bundeskartellamt 'Competition Law and Data (2016), 15-24 https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2 accessed 19 June 2019; Josef Drexl 'Designing Competitive Markets for Industrial Data - Between Propertisation and Access (2016) *Max Planck Institute for Innovation and Competition Research Paper* No 16-13, 42-59 https://ssrn.com/abstract=2862975 accessed 28 June 2019.

45 See Richter and Slowinski (n 4) 21, qualifying "single source data" as data that is difficult to replace.

46 Cf Richter and Slowinski (n 4) 11, who with respect to nonpersonal data, describe such DGM as a company-owned platform.

47 Bertin Martens and Frank Mueller-Langer 'Access to digital car data and competition in aftersales services' (2018) *JRC Digital Economy Working Paper* 2018-06, 6, 8-9 https://ec.europa.eu/jrc/sites/jrcsh/files/jrc112634.pdf accessed 14 February 2020.

48 Verband der Automobilindustrie. 'Access to the vehicle and vehicle generated data' (Position paper 2016), 2. https://www.vda.de/en/topics/innovation-and-technology/network/access-to-the-vehicle.html accessed 14 February 2020.

49 Federal Trade Commission, 'Data Brokers. A Call for Transparency and Accountability' (2014) 8, 14, 23 > https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf accessed 28 April 2020. Indeed, data brokers can create complex networks of data transactions. Ibid. 46.

Data Library service.[50]

By definition, a single-source DGM has one centralized data access point controlled by the data holder. The conditions for accessing the data are typically determined contractually, at the legal level. Indeed, bilateral contract represents the most widespread means of governing data reuse.[51] The breadth and cost of access can vary greatly from case. In principle, data sharing in a single source DGM may also be facilitated by a data holder's pledge to granting access to its data on specific conditions. When the data holder commits to granting access to data to someone, such pledges to provide access to data could be reminiscent of commitments to license patents under fair, reasonable, and nondiscriminatory terms (FRAND).[52] At the organizational level, single source DGM requires practices that facilitate the transfer of data to the prospective user.[53] At the technical level, the data holder needs to execute the standardisation of data and device a method for a secure interorganizational transfer of data. The access may be implemented through an API[54] by downloads or in the context of data sandboxes.[55] On the other end of the spectrum, data transfer may also take place offline and in an unstructured form, such as vie delivery of hand-written documents.

Generally, the access to data in the single source DGM is characterized by supply-side control of the data access points, where access to data is dependent on the incentives of data holders. As a consequence, the data may not be shared at the socially optimal level.[56] For example the 'extended vehicle' proposal has been viewed to feature risks of distorting competition in favour of OEM's controlling the access to data.[57] In the worst case scenario, data may not be shared at all, or it may only be shared in a discriminatory manner if the data qualifies as an essential facility for competing in a specific market and is in the exclusive control of a dominant market player.[58] From the legal perspective, market failures in the sharing of data may be resolved by enacting an access right.[59]

In theory, a data subject could govern her personal data through a single source DGM on conditions that she alone determines. In practice, this is almost impossible. First, at the legal level, the personal data of a data subject may be lawfully collected and processed by another entity without her consent on a number of grounds.[60] Second, in many contexts, it is questionable whether a data subject is fully informed about the content and scope of the consent she gives to allow her personal data to be processed.[61] Also, the power relations between the data subject and a controller are rarely balanced in a manner where the data subject is free to determine the conditions for processing her data.[62] Third, individual data subjects rarely have the means to create and control a technical interface through which all their personal data would be transmitted.

3.3 Data clearinghouse

We employ the concept of a "data clearinghouse" to characterize DGMs that position themselves clearly as an intermediary between data subjects and controllers or data holders and data users.[63] Clearinghouses are either governed by a neutral actor that represents neither the demand nor the supply side of the market for data, or by a collective of actors operating,[64] for example, in the relevant sector/market. When not displaying any features of a platform, data clearinghouses can be described as agencies that explicitly seek to facilitate the sharing of data. Their business model, if it exists, is based on facilitating data exchange, for example in the form of taking commissions.[65] Clearinghouses, as institutions, have been adapted for use in diverse contexts, including the governance of intellectual property. As an example of a data clearinghouse, consider the company Prifina. They develop an infrastructure that enables data subjects to securely store their data and to share it with selected service providers. These providers then process the data under the conditions specified by the data subject.[66] The specific features of data clearinghouses, as opposed to those of clearinghouses for IP, will be reviewed in the Section 4.

3.4 Data pool

The previously described DGMs (i.e., single source DGMs and clearinghouses) focus on providing access to individual data holders' data sets. However, DGMs may also provide access to predetermined combinations of data sets. In the literature, such approaches are often referred to using the term "data pool" based on an analogy with patent pools,[67] wherein "companies and other data holders agree to create a unified presentation of datasets as a collection accessible by multiple parties.[68] To illustrate the concept of a data pool, consider the recent initiative coordinated by Berlin's Charité hospital to aggregate data on COVID-19 patients from all German university hospitals into a comprehensive database to facilitate academic research on the virus.[69] The particularities of pooling data as opposed to patents are

[50] 'Data Library' (Openinnovation 2019) https://openinnovation.astrazeneca.com/data-library.html accessed 4 May 2020.
[51] Duch-Brown, Martens and Mueller-Langer (n 2), 25. https://www.vda.de/en/topics/innovation-and-technology/network/access-to-the-vehicle.html accessed 14 February 2020.
[52] See Richter and Slowinski (n 4) 17-21 and chapter 5. Such a commitment would represent a more open spectrum of single source DGM.
[53] Reflecting the organizational level of this form DGM, Verhulst et al (n 22) 28-29 refer to it as "Data Transfer".
[54] Verhulst et al (n 22) 14.
[55] See OECD Enhancing Access, (n 22) ch 2.
[56] It should be noted that the sharing of data is not in all cases favourable from the perspective of economic welfare. For example, the sharing of sales prices and output data with competitors may enable tacit or explicit collusion. Stucke and Grunes (n 24).
[57] Mike McCarthy, M Seidl, S Mohan, J Hopkin, A Stevens, F Ognissanto, 'Access to In-Vehicle Data and Resources' (European Commission 2017) CPR 2419, 136–138.
[58] See on the applicability of the essential facilities doctrine to data, Autorité de la concurrence & Bundeskartellamt (n 44), 17-18; Drexl (n 44) 42-59; Crémer, de Montjoye and Schweitzer (n 7) 98-107.
[59] See Drexl (n 2).
[60] GDPR arts 6 and 9.
[61] Alessandro Acquisti, Laura Brandimarte and George Loewenstein, 'Privacy and human behavior in the age of information' (2015) 347 *Science* 509.
[62] For an overview over various forms of such power asymmetries and their origins see Shoshana Zuboff, The age of surveillance capitalism: the fight for the future at the new frontier of power (Profile Books 2019).
[63] On this basis, data brokers, ie companies which actively collect data to which they provide access to, are deemed to rely on single-source DGM, since they position themselves as an intermediary on a two-sided market. See OECD Enhancing Access (n 22) ch 2.
[64] Reiko Aoki and Aaron Schiff 'Promoting access to intellectual property: patent pools, copyright collectives, and clearinghouses' (2008) 38 R&D Mgmt 189, 196.
[65] For example, clearinghouses in the automotive sector charge a certain percentage from the price of transferred data. Martens and Mueller-Langer (n 47) 22.
[66] See, for example, 'Core Concept' (Prifina) https://www.prifina.com/core-concept.html accessed 4 May 2020.
[67] Lundqvist 'Competition and data pools' (n 22).
[68] Verhulst et al (n 22) 11.
[69] 'Coronavirus / SARS-CoV-2: Charité Coordinates Network of Academic Medical Research into COVID-19' (Charité-Universitätsmedizin Berlin) https://www.charite.de/en/the_charite/themen/coronavirus_sars_cov_2_charite_coordinates_network_of_academic_medical_research_into_covid_19/ accessed 4 May 2020.

discussed in Section 5.

3.5 Distributed DGMs

Distributed DGMs enable data transfers between data subjects and controllers or between data holders and users without the direct involvement of an intermediary or another centralized entity. Decentralized access to data may be enabled on different governance levels and typically involves efforts to standardize elements of the data sharing process.

At the legal level, decentralized access may be facilitated by model contractual clauses, which are similar to the Creative Commons copyright license model.[70] Several data holders may employ these clauses independently of each other. Open Data Commons, initiated in 2007, was developed to offer multiple license options for data and databases.[71] However, especially in jurisdictions that do not recognize a sui generis right to databases or copyright in the arrangement of a database, the bindingness of such instruments is unclear.[72] Nonetheless, they may still function to reinforce a social norm of providing access. More recently, advocating for a more 'user-centric approach' to data, scholars have proposed a spectrum of six licenses for personal data. In the spirit of Creative Commons Licenses, these licenses range from providing full anonymity to granting permission to sell personal data. These licenses may be accompanied by further qualifications about the duration of access, identification of the accessing person, and a personalized value proposition.[73] Initiatives to standardize licenses for non-personal data are also emerging.[74]

Decentralized access to data may also be enabled by a technical standard. For example, in the automotive sector, in-vehicle data from individual cars is accessible to any repair shop or other service provider via a standardized, on-board diagnostics port (OBD-II).[75] In the automotive sector, this decentralized DGM is deemed more procompetitive than a single-source access model of in-vehicle data.[76] However, standardization is not a panacea for sustaining a distributed DGM. The automotive sector is displaying signs of competition between different standards, with OEMs pushing for the adoption of the 'extended vehicle' solution, which is a single-source DGM.[77] This raises concerns for aftermarket participants about losing access to real-time, in-vehicle data. The current OBD-II standard was set before the surge in the datafication of vehicles and has issues both with respect to bandwidth and cybersecurity. An update of the standard would require an internationally coordinated effort.[78]

Distributed DGMs can also be found in the context of medical research. Actors engaging in research and care can take a modular, networked organizational structure, where a central node is responsible for identity management. This unit facilitates data transfers between other nodes of the network, such as units specializing in clinical care, research, or biobanking. The identity management unit ensures that data concerning individual patients is consistent and pseudonymized when processed for research purposes.[79]

Distributed DGMs may be implemented through emerging technological solutions . For example, edge computing, which takes place on a data holder's device instead of transmitting data to the cloud,[80] may support the adoption of personal use data licenses.[81] Especially the medical sector has explored the use of distributed ledger technology (DLT) for decentralized data sharing .[82] However, due to data protection and security concerns, its use remains mostly experimental.[83] DLT challenges the underlying logic of the GDPR, which presumes the centralized governance of data. However, when DLT is designed to support data protection, it may also uphold data sovereignty.[84] When it is integrated with other technology that ensures adequate data protection, DLT may also be used for the interorganizational sharing of data.[85] As a case in point, in Estonia, DLT is used in the national system for managing electronic health records for ensuring their integrity.[86]

Distributed DGMs often involve two layers of governance to support decentralized access to data. Paradoxically, despite featuring decentralization at one or two levels of data governance, distributed DGMs often require a certain level of centralized coordination at the organizational level. At least a minimal organizational structure is required to draft the standardized license conditions of a distributed DGM, to set a technical standard, or to design the distributed data transfer infrastructure[87] and ensure its technical functioning. It appears that not a single distributed DGM is governed in a purely decentralized manner. Rather, as Contreras and Reichman explain, DGMs can display varying degrees of centralization.[88]

70 European Commission 'Free flow of data and emerging issues in the European data economy' SWD (2017) 2 final, 31.

71 'Licenses' (Open Data Commons 2019) https://opendatacommons.org/licenses/index.html accessed 20 June 2019. 'About' (Open Data Commons, 15 December 2007) https://opendatacommons.org/about/ accessed 4 May 2020.

72 'Licenses FAQ' (Open Data Commons 2019) https://opendatacommons.org/faq/licenses/index.html accessed 20 June 2019.

73 Paul Jurcys, Chris Donewald, Jure Globocnik and Markus Lampinen, 'My Data, My Terms: A Proposal for Personal Data Use Licenses' [2020] *Harv J L & Tech Dig* 8–11 https://jolt.law.harvard.edu/digest/my-data-my-terms.

74 See Misha Benjamin, Paul Gagnon, Negar Rostamzadeh, Chris Pal, Yoshua Bengio and Alex Shee, 'Towards Standardization of Data Licenses: The Montreal Data License' (2019) arXiv:190312262 https://arxiv.org/abs/1903.12262 accessed 28 April 2020; Paul Jurcys et al, (n 73) 13 discussing the Montreal Data License for sharing data in the fields of machine learning and artificial Intelligence 'About MDL' (Montreal Data License) https://www.montrealdatalicense.com/en/about accessed 28 April 2020.

75 Martens and Mueller-Langer (n 47) para 11 and fn 16, 18.

76 Martens and Mueller-Langer (n 47) 18, see also chapter 4.4.

77 See Wolfgang Kerber and Daniel Gill, 'Access to Data in Connected Cars and the Recent Reform of the Motor Vehicle Type Approval Regulation' (2019) 10 JIPITEC 244 para 1, para 11 and fn 27, para 29 and fn 60.

78 McCarthy et al (n 57) 85, 131-132, 151.

79 Klaus Pommerening and T Müller, *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: generische Lösungen der TMF* 2.0 (MWV, Med Wiss Verl-Ges 2014) 3, 106.

80 Paul Miller, 'What Is Edge Computing?' (*The Verge*, 7 May 2018) https://www.theverge.com/circuitbreaker/2018/5/7/17327584/edge-computing-cloud-google-microsoft-apple-amazon accessed 4 May 2020.

81 Paul Jurcys et al, 'My Data, My Terms: A Proposal for Personal Data Use Licenses' [2020] *Harvard Journal of Law & Technology Digest* 4 and fn 10 https://jolt.law.harvard.edu/digest/my-data-my-terms.

82 See Qi Xia et al, 'MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain' (2017) 5 *IEEE Access* 14757, 5; Alevtina Dubovitskaya Petr Novotny Zhigang Xu and Fusheng Wang, 'Applications of Blockchain Technology for Data-Sharing in Oncology: Results from a Systematic Literature Review' [2019] *Oncology* 1.

83 Dubovitskaya et al (n 82) 5.

84 M Finck, 'Blockchains and Data Protection in the European Union' (2018) 4 *EDPL* 17, 17, 35.

85 Dubovitskaya et al (n 82) 1.

86 'E-Health Records' (e-Estonia) https://e-estonia.com/solutions/healthcare/e-health-record/ accessed 4 May 2020.

87 See Jessica Schmeiss, Katharina Hölzle and Robin P Tech. 'Designing governance mechanisms in platform ecosystems. Exploring the potential of blockchain technology' (2019) 62 *Cal Mgmt Rev* 121.

88 Jorge L Contreras and Jerome H. Reichman, 'Sharing by design: Data and decentralized commons' (2015) 350 *Science* 1312.

4. Data Clearinghouse

4.1 Background

In the context of data, the term 'clearinghouse'[89] has been used interchangeably with 'intermediary',[90] 'platform',[91] 'trusted third party',[92] or 'data brokerage'.[93] However, we take a view that the concept of a clearinghouse is sufficiently flexible to be able to accommodate and thus identify a number of legal and technical constellations of varying complexity that facilitate the sharing of data.

Clearinghouses are governance mechanisms that were initially developed in the banking sector.[94] In this account, the concept is understood as 'an intermediary between buyers and sellers of financial instruments. It is an agency or separate corporation of a futures exchange responsible for settling trading accounts, clearing trades, collecting and maintaining margin monies, regulating delivery, and reporting trading data.'[95] They can also be described to 'take the opposite position of each side of a trade. When two investors agree to the terms of a financial transaction, such as the purchase or sale of a security, a clearing house acts as the middle man on behalf of both parties. The purpose of a clearing house is to improve the efficiency of the markets and add stability to the financial system.' [96]However, clearinghouses have been adapted for use in other contexts, and their governance model can be defined as 'a central agency for the collection, classification, and distribution especially of information.'[97] Clearinghouses have also been used to govern intellectual property. A well-known example of this are collective copyright management organizations,[98] such as the GEMA.[99] Scholars have discussed whether or not the model is suitable to facilitating the reuse of patents,[100] and Van Overwalle et al. and van Zimmeren et al. have identified five different subcategories of patent clearinghouses on the basis of the scope of services offered, organizational complexity, and the extent to which they engage in licensing patents (Figure 2).[101] Their findings illustrate how flexible the concept of a clearinghouse can be when one seeks to describe heterogeneous patent intermediaries.

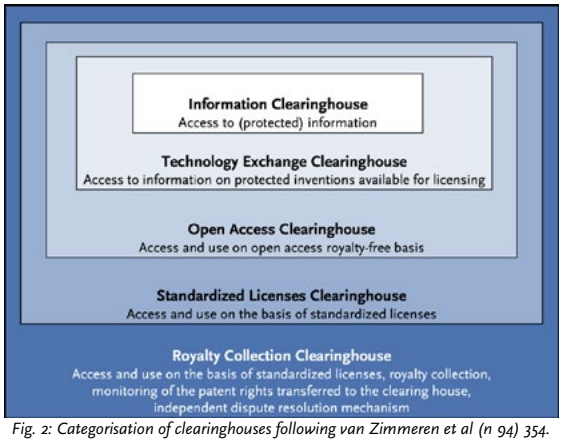


Fig. 2: Categorisation of clearinghouses following van Zimmeren et al (n 94) 354.

4.2. Defining data clearinghouses

Clearinghouses identified in patent law[102] are not directly applicable to data because the economic and legal qualities of data are distinct from intellectual property. For reasons described above, we found it necessary to adapt the known patent clearinghouse models as identified by Van Overwalle et al. and van Zimmeren et al.[103] for use with personal and non-personal data and to conceptualize the subcategories of data clearinghouses.

The IP literature divides clearinghouses into two main categories: 'an informational clearinghouse ... [which] collects and provides information about the existing IP' and 'a licensing clearinghouse...[which] provides information and also sells licenses directly, and may perform royalty collection functions'.[104] An analogous division could also be applied to data clearinghouses. In the context of data, 'information clearinghouses'[105] could be seen to provide information about the location of a data set and its owner and to facilitate the negotiations for obtaining access to the data. However, they do not determine the price for accessing the data or control the flow of data between the two sides of the market. In contrast, the more complex model of 'data

[89] A clearinghouse for geo-spatial data has been defined as 'a service for searching, viewing, transferring, ordering, advertising and disseminating over the internet geo-data stored at many different locations in digital format.' Mathias Lemmens 'Spatial Data Clearinghouses' (*GIM Magazine*, 24 July 2006) https://www.gim-international.com/content/article/spatial-data-clearinghouses accessed 19 June 2019; See also `Regional Transportation Data Clearinghouse' (Regional Transportation Data Clearinghouse 2019) http://rtdc-mwcog.opendata.arcgis.com accessed 19 June 2019.

[90] E.g., Tuukka Lehtiniemi, Yki Kortesniemi,'Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach.' (2017) 4 *Big Data & Society* 3. See also Verhulst et al (n 22) 11, using the term 'Trusted Intermediary'.

[91] E.g., Annabelle Gawer, 'Bridging differing perspectives on technological platforms: Toward an integrative framework.' 43 Research policy (2014) 1239; European Commission 'Free flow of data' (n 70) 17; Verhulst et al (n 22) 20.

[92] E.g., Susan W Van den Braak, Sunil Choenni, Ronald Meijer and Anneke Zuiderwijk, 'Trusted third parties for secure and privacy-preserving data integration and sharing in the public sector' *Proceedings of the 13th Annual International Conference on Digital Government Research 2012.*

[93] E.g., Verhulst et al (n 22) 19-20.

[94] Esther van Zimmeren, Birgit Verbeure, Gert Matthijs and Geertrui Van Overwalle, 'A clearing house for diagnostic testing: the solution to ensure access to and use of patented genetic inventions' (2006) 85 *Bull WHO* 352, 353.

[95] 'Central Clearing Houses' (CFA Institute 2019) https://www.cfainstitute.org/en/advocacy/issues/central-clearing-houses accessed 19 June 2019.

[96] 'Clearinghouse' (Investopedia 2019) https://www.investopedia.com/terms/c/clearinghouse.asp accessed 16 June 2019.

[97] 'Clearinghouse' (Merriam Webster) https://www.merriam-webster.com/dictionary/clearinghouse accessed 28 April 2020.

[98] Van Overwalle et al (n 21) 146.

[99] The GEMA is the centralized organization in Germany responsible for collecting royalties on behalf of musicians for every performance and copy that is made of their works.

[100] van Overwalle et al (n 21), 146; van Zimmeren et al (n 94) 352; Aoki, and Schiff (n 64); Reiko Aoki and Aaron Schiff 'Intellectual property clearinghouses: The effects of reduced transaction costs in licensing' (2010) 22 I *Econ & Pol* 218. Geertrui Van Overwalle, 'Patent pools and clearinghouses in the life sciences: back to the future' in Duncan Matthews & Herbert Zech (eds) *Research Handbook on IP and the Life Sciences* (Edward Elgar 2017), 304. Even if a clearinghouse reduces transaction costs, its overall effect on welfare can be positive or negative depending on the number of patents used in downstream value creation and other factors. Aoki and Schiff (n 100)

[101] van Zimmeren et al (n 94) 354, Figure 1.

[102] Van Overwalle et al (n 21) 146; van Zimmeren et al (n 94), 352-354; Aoki and Schiff (n 64) 195-197.

[103] Van Overwalle et al (n 21) 146; van Zimmeren et al (n 94), 352-354. However, the above-mentioned authors' models for IP clearinghouses are more directly suitable to governing copyrighted collections of data or exclusive rights to databases.

[104] Aoki and Schiff (n 64) 196 and Figure 8. Similar, but they use the term 'information clearinghouse' instead of informational clearinghouse. Van Overwalle et al (n 21), 145-147; van Zimmeren et al (n 94) 352- 353; Van Overwalle (n 100) 304, uses the terms 'information clearinghouses' and 'technology transfer clearinghouses'.

[105] We are using the term employed by Van Overwalle et al (n 21), 145-147; van Zimmeren et al (n 94) 352- 353.

transfer clearinghouses' establishes the conditions for data access and transfer,[106] controls access to the data, and manages the data transfer.[107] The subcategories of information and data transfer clearinghouses (Figure 3) are discussed in the following subsections.

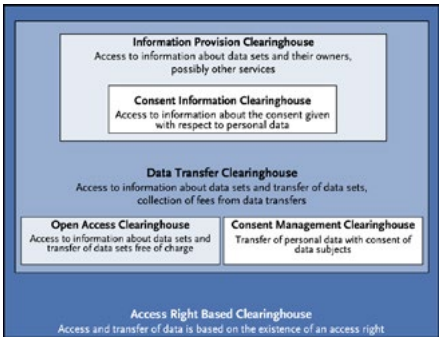


Fig. 3: Data Clearinghouses

4.3 Information clearinghouses

We identify two types of information clearinghouses in the context of data. An 'information provision clearinghouse' offers information about data sets and potentially also their owners.[108] An example of an information provision clearinghouse is a Wikipedia page that lists datasets for machine-learning research according to their type (i.e. image data, text data) and application (i.e. face recognition, action recognition).[109] In principle, information provision clearinghouses may also offer other services, such as facilitating negotiations and allowing data holders and those interested in obtaining access to data to enter into a contractual agreement. However, they do not take part in the transfer of data.[110]

A 'consent information clearinghouse' is an information clearinghouse specific to personal data. Such a clearinghouse provides information regarding the scope of consent given with respect to personal data and whether it is possible to process this data in the way envisioned by its prospective users. For example, netID is an association of advertisers and website owners ('publishers') that seeks to launch a 'consent module' solution. This solution provides publishers with a tool to obtain informed consent from website visitors to share their personal data with other members of the NetID consortium.[111] NetID also enables these members to communicate the fact that this consent has been given to each other through its own servers. Moreover, it also collects some voluntary data (name, address, or birthday) of data subjects from all its members so that it can compare and correct them if necessary. Its position as an intermediary distinguishes it from a purely standard-setting organization. Following the signalling of consent, NetID consortium partners can then negotiate bilateral deals for data sharing in order to aggregate data from multiple websites and target online advertisement space more effectively.[112]

By revealing the consent status of the relevant data subject, consent information clearinghouses can be seen as both reducing the information costs associated with the processing and sharing of personal data and as helping to solve the 'tragedy of anti-commons'.[113] Without the consent information clearinghouse, purpose-specific and context-specific consent would have to be obtained individually from each data subject. This would not only lead to high and possibly prohibitive transaction costs. It may also produce differences between the types of consent given by data subjects and thus to uncertainty among the data controllers and processors on whether they are allowed to reuse all the data under the same conditions. With standardized consent, consent given by users once can be used by all controllers, and controllers and processors can be certain that all the data collected by members of the consortium can be used and exchanged under the same conditions. The standardisation of consent requires to standardise its legal components, in particular, the types of personal data collected and the types of purposes for that the data is used. [114] An example for such data types can be found in the GDPR (e.g. biometric data, data about religious beliefs etc.). An example for purposes that could be standardised are IT security, marketing, etc. Such standards are non-exclusive, which means that the data subjects and the controllers can always fall back to individual purposes that do not match with the standardised purposes. However, in this case, they have to assess it on their own on a case-by-case basis how to specify the data and purposes in a GDPR-compliant way. Just one example for such data and purpose standards can be found at netID, which acts as an intermediary by determining which data type (e.g. email addresses) can be used for which purpose (esp. online advertising), and facilitates the flow this information.

4.4 Data transfer clearinghouses

Generally, data transfer clearinghouses seek to facilitate the actual transfer of data from its source, a data subject or controller, to its user. The accessibility of the data in question is dependent on the incentives of data holders and subjects to share it. Unlike information clearinghouses described above, data transfer clearinghouses also have a certain level of control over the conditions under which the access to data is provided, and they also engage in the actual transfer of data between the data holder and the data user. For example, the Luxembourgian data repository ELIXIR-LU offers a service for storing and archiving transnational medicine data from multiple scientific projects while enabling easy accessibility to the data sets.[115] The 'B2B marketplace solution' is an example of a data transfer clearinghouse in the automotive sector. A neutral intermediary controls the server

[106] The data exchange clearinghouse hence corresponds to the 'technology exchange clearinghouse' for patents as defined by van Zimmeren et al (n 94) 353; Van Overwalle (n 100) 304.

[107] This feature of clearinghouses is not present in clearinghouses for IP but is specific to data clearinghouses.

[108] van Zimmeren et al (n 94) 353-354. In the context of patents, such clearinghouses provide information about patented inventions and possibly of their owners. Examples of these clearinghouses are patent search websites and patent offices' databases. van Zimmeren et al (n 94) 353-354.

[109] 'Lists of datasets for machine learning research' (Wikipedia 2019) https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research accessed 19 June 2019.

[110] Such information provision clearinghouses would thus adopt features of a 'technology exchange clearinghouse' for patents, see: van Zimmeren et al (n 94) 353.

[111] 'NetID' (NetID 2019) https://netid.de accessed 19 June 2019; the information about the netID consent module is drawn from telephone interviews with netID representatives

[112] The NetID initiative does not involve a solution for storing data.

[113] This means that they reduce transaction costs that were previously so high as to prohibit the sharing of data, so that the benefits of sharing can now be realized. See Heller and Eisenberg (n 12).

[114] Max von Grafenstein, *The Principle of Purpose Limitation in Data Protection Laws: The Risk-Based Approach, Principles, and Private Standards as Elements for Regulating Innovation* (1st edition, Nomos 2018) 616 et seq.

[115] 'Sustainability of Data' (ELIXIR-LU 2019) https://elixir-luxembourg.org/sustainability-data accessed 19 June 2019. DAWEX, which offers data monetization and sourcing services for companies in multiple industries, appears to be a data transfer clearinghouse. See 'DAWEX' (DAWEX 2019) https://www.dawex.com/en/ accessed 30 June 2019; European Commission 'Free flow of data' (n 70) 17; Richter and Slowinski (n 4) 11.

on which vehicle data is stored and offers B2B access to data from multiple OEMs while also providing services that facilitate building partnerships and concluding B2B contracts,[116] such as assistance in determining the price for data.[117]

At the legal level, data transfer clearinghouses presuppose a mandate from the data holder or consent from a data subject to provide controllers (third-party users) access to their data. The clearinghouse must also determine the conditions under which it transfers data from one party to another. Departing from the typology of clearinghouses presented by van Zimmeren et al., we regard the degree of standardization offered by data transfer clearinghouses in their terms and conditions as a contingent feature of this type of clearinghouse at their legal data governance level.[118] In our view, the granularity of standardized terms exists on a continuum. There are clearinghouses that offer little flexibility in determining the conditions for sharing or accessing the data or for giving consent to its processing, and there are those that offer full freedom to data subjects and holders to determine the conditions of access and for data users to agree to them. In between these two poles we find all the data transfer clearinghouses that offer a degree of customizability regarding the terms and fees associated with the data transfer. Indeed, the more customizable the terms and fees are, the higher the transaction costs become. This is an unavoidable trade-off.

By definition, data transfer clearinghouses do not merge or recombine data sets from multiple sources, which would lead to the creation of a data pool (see below for more details). Therefore, at the organizational and technical level, the individual data sets and data transactions are kept apart. However, in order to facilitate the reusability of data, the data transfer clearinghouses may engage in the harmonization of data or its conversion into a specific standard accepted by the data users.

Data transfer clearinghouses may be further distinguished into three subcategories: 'open access clearinghouses', 'consent management clearinghouses', and 'access-rights-based clearinghouses'. The latter two types are specific to data and do not have a counterpart among IP clearinghouses.

An 'open access clearinghouse' can be defined as a DGM facilitating data transfers to any willing party for free.[119] This model is also found in the field of IP.[120] The wide accessibility of data is based on its owners' voluntary interest in sharing it. Examples of non-personal data range from curated public open data (e.g. the US federal government's Data.gov)[121] to peer-to-peer exchange sites for open data sets (e.g. Awesome Public Datasets on Github).[122] Interestingly, such open access clearinghouses also exist for personal data: one such example is OpenSNP, where users can make the results of a genotyping test openly available.[123]

'Consent management clearinghouses' are specific to personal data. Here, the DGM concentrates on enabling the transfer of data to users with the consent of the data subject. Upon enabling the transfer of personal data, consent management clearinghouses must comply with the GDPR and hence are subject to stricter conditions for the legal, organizational, and technical levels of governance than data transfer clearinghouses. For example, Vivy[124] is a mobile application that allows data subjects to store their health data in one place and to share it with selected healthcare providers. Such clearinghouses are designed to serve the interests of the data subject but they also benefit the recipients and subsequent processors of data by ascertaining the lawfulness of data processing and by facilitating more efficient data exchange. From an economic perspective, clearinghouses such as Vivy reduce healthcare costs by avoiding the need to duplicate health data such as x-rays and reduce information asymmetries between different actors in the healthcare sector. The solution will most likely benefit social welfare as it fosters competition between healthcare providers while also enabling the data subject to obtain better quality healthcare. It also somewhat reduces data subjects' transaction costs when setting the optimal privacy level, although there are limits to the user-empowering potential of such clearinghouses.[125]

We also observe the emergence of 'access-rights-based clearinghouses' that, at the legal level of governance, rely on the existence of a right to access data, such as the right to access personal data under Article 15 GDPR or the right to data portability under Article 20 GDPR. The access rights on which the clearinghouse relies may also be sector specific.[126] The access-rights-based clearinghouse may, in theory, be formed around an access right to non-personal data.[127] Examples of such access-rights-based clearinghouses include many 'Personal Data Spaces',[128] such as Cozy Cloud,[129] fair&smart,[130] Datafund,[131] or Personium.[132] These services provide data subjects with a tool to retrieve their personal data from one or more controllers and transfer them to another, sometimes for a fee. Although the legal infrastructure of these initiatives is not always transparent, they appear to be based on the use of several access rights to personal data.[133]

'Royalty collection clearinghouses,' which include collective copyright management organizations, are among the most complex clearinghouses in IP law. They function for the purpose of obtaining licenses, collecting and distributing licensing fees, monitoring the fulfilment of

[116] C-ITS Platform 'Final Report' (European Commission 2016), 81-82 https://ec.europa.eu/transport/sites/transport/files/themes/its/doc/c-its-platform-final-report-january-2016.pdf accessed 19.06.2019. In practice, companies such as Caruso and Otonomo offer clearinghouse services in the automotive sector.

[117] 'Marketplace' (Caruso 2019) https://www.caruso-dataplace.com/marketplace/ accessed 19 June 2019.

[118] In the context of clearinghouses for IP, a 'standardized licences clearing house' is identified as a distinct type of clearinghouse that enables the reuse of IP on the basis of standardized licence conditions, van Zimmeren et al (n 94) 354. In our concept of data governance, the fact that certain terms and conditions of data access and use are standardized does not necessarily lead to an autonomous type of clearinghouse.

[119] van Zimmeren et al (n 94) 354 describes this type of clearinghouse to disclose patented or patentable inventions to the public domain.

[120] van Zimmeren et al (n 94) 352, 354.

[121] 'Data.gov' (Data.gov 2019) https://www.data.gov/ accessed 19 June 2019.

[122] 'Awesome Public Datasets' (GitHub 2019) https://github.com/awesome-data/awesome-public-datasets accessed 19 June 2019.

[123] 'OpenSNP' (openSNP 2019) https://opensnp.org/ accessed 19 June 2019.

[124] see 'Vivy' (Vivy 2019) https://www.vivy.com accessed 19 June 2019.

[125] Lehtiniemi and Kortesniemi (n 90).

[126] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) N 1093/2010, and repealing Directive 2007/64/EC (Revised Payment Service Directive) [2015] OJ L337/35, arts 66 and 67.

[127] See Drexl (n 44), 57-58; European Commission 'Free flow of data (n 70), 46-49.

[128] Tuukka Lehtiniemi 'Personal Data Spaces: An Intervention in Surveillance Capitalism?' (2017) 15 *Surveillance & Society*, 626.

[129] 'Cozycloud' (Cozy.io 2019) https://cozy.io/en/ accessed 19 June 2019.

[130] 'Fair and Smart' (Fair & Smart 2019) https://www.fairandsmart.com/en/ accessed 19 June 2019.

[131] 'Datafund' (Datafund 2019) https://datafund.io/ accessed 19 June 2019.

[132] 'Personium' (Personium 2019) https://personium.io/ accessed 19 June 2019.

[133] In particular, GDPR art 15 and 20 but also Revised Payment Service Directive art. 66 and 67.

license obligations, and providing a mechanism for dispute resolution.[134] This model of clearinghouse appears too difficult and bureaucratic to apply to data, especially given the fact that data is not subject to a property right.[135] However, access-rights-based clearinghouses, which also enable the monetization of the use of personal data, may adopt features from the IP-based, royalty collecting clearinghouses, especially if they become very popular and begin to process large numbers of data transactions.

Finally, a clearinghouse operator may gain additional leverage through the extraction and control of shared data between two markets, especially because this expands the portfolio of services that support the exchange of data. In other words, a clearinghouse may come to acquire platform-like features over time by leveraging the data it has collected by facilitating data transactions. We will refer to this phenomenon as the 'clearinghouse platform'.[136] Over time, this clearinghouse may begin to compete with either the controllers or the users data, or it may begin to operate on a third market. The operator may also leverage network effects to coerce the users of its services to disclose data relevant for its business model.[137]

5. Data Pools

5.1 Organizational structure

Data pools aggregate data from multiple sources and provide access to the aggregated data to several users from a single point of access. Some authors have adopted wider definitions of a data pool. For example, Lundqvist describes them as models where 'firms agree to share their digitalised [sic] information regarding a given market, in reference to a given service or generally in an industry, or an e-ecosystem.' [138] Mattioli qualifies this DGM by the performance of data analytics by the pool and provision of access to the results of analysis of the aggregated data.[139] Furthermore, data pools have been defined by whether they offer access to data in a standardized format. 'A data pool is a centralized repository of data where trading partners (e.g., retailers, distributors or suppliers) can obtain, maintain and exchange information about products in a standard format. Suppliers can, for instance, upload data to a data pool that cooperating retailers can then receive through their data pool.' [140] In our view, the locus of analytics and the format of data, are additional, but not definitive qualities of a data pool as a DGM. Yet, the formation of data pools plays an important role in facilitating data analytics and machine learning, as well as other applications of artificial intelligence.[141]

Data pools may either be governed by actors who contribute data to the pool or by a third party.[142] The aggregated data may be processed and structured in various ways depending on the specific function of the pool. For example, the data may be combined into a 'data lake', where data remains unstructured and raw, or it may be processed and curated into a 'data warehouse'.[143] The analysis of the pooled data may either be undertaken by the users accessing the data or by the data pool itself. The entity governing the data pool may also outsource the analysis to another entity.

Technology pools can, according to the European Commission, 'take the form of simple arrangements between a limited number of parties or of elaborate organisational arrangements whereby the organization of the licensing of the pooled technologies is entrusted to a separate entity. In both cases, the pool may allow licensees to operate on the market on the basis of a single licence.'[144] Similar organizational variety may also be found among data pools. In its simplest form, two or more data holders combine their data sets and provide each other access to the pooled data. At the legal level, this requires a multilateral contractual arrangement, but parties typically also need to create a technical infrastructure for combining their data and accessing it. Further relevant organizational measures may also include harmonization of different data types as well as other measures ensuring interoperability of the data and the technological infrastructures of the pooling arrangement. As an example, consider Moovel[145] or Compte Mobilité.[146] Both are providers of 'mobility as a service', which is examined in Carballa's paper on data sharing as co-opetition.[147] Here, several mobility providers share data in a common pool in order to create one tool where customers can book a route that combines all of the providers' services. The mobility app Jelbi is an example of such a provider from Berlin. The municipal transport company runs and governs a data pool that 25 mobility providers can access and contribute to.[148]

Data pools may also be configured in a more open manner, for instance by permitting parties who do not contribute data to still also access the pooled data. BRCA Exchange operate in this manner by pooling information on the BRCA1 and BRCA2 gene variants in a curated and classified form.[149] From an organizational perspective, such pools are usually governed by an intermediary, which, similarly to clearinghouses, operates in a two-sided market between data holders and data subjects or controllers and data users, respectively. The main difference between these two DGMs is that clearinghouses

[134] Van Overwalle et al (n 21) 146; van Zimmeren et al (n 94) 354-355.
[135] Such complex clearinghouses, paired with property rights, would be required to realize the 'radical data markets' proposed in Imanol Arrieta-Ibarra, Leonard Goff, L, Diego Jiménez-Hernández, Jaron Lanier, & E Glen Weyl, 'Should We Treat Data as Labor? Moving beyond "Free"' (2018) 108 *aea Papers and Proceedings* 38 and elaborated in Eric A Posner, E Glen Weyl *Radical markets: Uprooting capitalism and democracy for a just society* (Princeton UP 2018). However, in the absence of an exclusive right to data, a data transfer clearinghouse would not be able to collect licensing fees for third-party usage of data.
[136] Cf Richter and Slowinski (n 4) 10, who understand platforms to 'enable a systematic exchange of data sets and streams on a large scale between many actors.
[137] Richter and Slowinski (n 4) 16.
[138] Lundqvist, (n 22) 146.
[139] See Mattioli (n 22).
[140] Justine Rodian 'The complete A-Z of Master Data Management' (StiboSystems 2018) https://blog.stibosystems.com/the-complete-a-z-of-master-data-management accessed 19 June 2019.
[141] European Commission 'A European Strategy for Data' (Communication) COM 2020 66 final, 5 and fn 13;
European Commission 'On Artificial Intelligence – A European approach to excellence and trust (White Paper) COM(2020) 65 final, 3.
[142] See Lundqvist (n 22), 149.
[143] For a distinction, see: Rodian (n 140); 'Data Lake vs. Data Warehouse' (talend 2019) https://www.talend.com/resources/data-lake-vs-data-warehouse/ accessed 19 June 2019; Sherry Tiao 'What's the difference between a Data Lake, a Data Warehouse and a Database' (Oracle Big Data Blog 2020) https://blogs.oracle.com/bigdata/data-lake-database-data-warehouse-difference accessed 14 February 2020.
[144] Communication from the Commission of 28 March 2014 Guidelines on the Application of Article 101 of the Treaty on the Functioning of the European Union to Technology Transfer Agreements (Technology Transfer Guidelines) [2014] OJ C 89/3, para 244.
[145] 'Moovel' (Moovel 2019) https://www.moovel.com/de/referenzen/moovel-mobility-app accessed 19 June 2019-
[146] 'Compte Mobilité' (Compte Mobilité 2019) https://www.compte-mobilite.fr/ accessed 19 June 2019-
[147] Carballa Smichowski (n 26).
[148] Stefan Krempl 'Jelbi: App von BVG und Trafi vereint Berliner Mobilitäts-Angebote' (Heise Online 2 February 2019) https://www.heise.de/newsticker/meldung/Jelbi-BVG-will-uebergreifende-Mobilitaets-App-fuer-Berlin-im-Sommer-starten-4311779.html accessed 16 June 2019.
[149] 'BRCA Exchange' (BRCA Exchange 2019) https://brcaexchange.org/about/thisSite accessed 20 June 2019

focus on transactions of individual data sets and data pools focus on aggregated data sets that, prior to their aggregation, have been retrieved from different data subjects or data holders. In addition, a person or organization may play a dual role in the data pool, namely both that of a contributor of data to a set and of a user of that same data set. Furthermore, this actor may also take part in the governance of the pool, when the pool is not operated by an independent organization. The actor(s) having the authority to govern the data pool are also in the position to steer it towards or away of platformization.

Personal data may also be pooled, which requires the pool infrastructure to be GDPR compliant.[150] However, given the amount of aggregated data and the easier access to it in data pools compared to clearinghouses, it is a daunting task to design such DGMs in a GDPR-compliant way.[151] Provided that data pools are necessary for training artificial intelligence, this form of DGM calls for further research into the risks of re-identifiability of data subjects following triangulation of anonymized or pseudonymized data sets.

Besides platformization, a data pool may display degrees of decentralization [152]. For example, European Commission's data strategy describes model where data is not physically transferred to a centralized repository, but a number of distributed data sets are analyzed by a centrally governed entity, who provides the results of the analysis to those contributing data to the pool.[153] Such constellation bears closer resemblance to distributed DGMs. We expect to see more further variety in data pooling in the future. However, for the purposes of identifying DGMs that are procompetitive and GDPR compliant, it would be desirable to use more nuanced terminology and explicit description of DGMs such as pools both in research and in policy.

5.2 Data Pools and Competition

Just like clearinghouses, data pools may alleviate the 'tragedy of anti-commons' among data sets and overcome the problem of duplicative investments.[154] They can foster pro-competitive effects like encouraging wider reuse of data for a variety of innovative purposes, including algorithm training and facilitating market entry.[155] They may improve efficiency and foster competition especially in connection with the Internet of Things.[156]

However, data pools may facilitate collusion or give rise to abuse of collective dominance.[157] It is not clear under which conditions data pools may be deemed pro-competitive under Art. 101 and 102 TFEU and whether the rules for ensuring the pro-competitiveness of patent pools, most importantly Technology Transfer Guidelines and Guidelines for Horizontal Co-operation, also apply to data pools.[158] The European Commission's revision of the Horizontal Co-operation Guidelines is expected to clarify the legal framework that regulates the pro-competitiveness of data pools.[159]

According to Lundqvist, existing guidelines are not directly applicable to data pools due to the different nature of patents and data. For example, it is difficult to classify pooled data as either essential or non-substitutable because data subjects may be multi-homing their data.[160] Nonetheless, scholars recognize that data available from only a single source may, under certain circumstances, be deemed essential for participation in a certain market.[161] With regard to patent pools, the European Commission has determined that open licensing for all willing non-members of the pool is one of the affirming factors of compliance with competition law.[162] However, data pools may be established to facilitate access to data for a small market player as well as competitive advantage for tech giants with large data repositories of their own.[163] In such cases, it should not be mandatory to grant data access to a competitor with a larger market share.[164] Instead, the requirement for openness of the pool should correlate with its market power.[165]

Information sharing in a data pool may facilitate collusion,[166] especially in its tacit form. This is particularly true for constellations that combine data pooling with price setting algorithms. This may lead to the emergence of hub-and-spoke cartels where collusion is facilitated through an algorithm.[167] Data pools may also give rise to other types of market manipulation,[168] such as excessive or discriminatory pric-

[150] See Sophie Stalla-Bourdillon, Gefion Thuermer, Johanna Walker, Laura Catherine and Carmichael, 'Data protection by design: building the foundations of trustworthy data sharing' *Proceedings of Data for Policy Conference 2019* 6 doi:10.5281/zenodo.3079895 access date 16 June 2019.

[151] See, for instance, the Data Protection Impact Assessment conducted for a hypothetical Smart City scenario in Berlin, which came to the conclusion, that the data collection for research purposes in the area of smart urban traffic planning based on the legitimate interests-clause under Art. 6 sect. 1 lit. f) GDPR requires a decentralised infrastructure, Max von Grafenstein, 'Innovationsoffener Datenschutz durch Folgenabschätzungen und Technikgestaltung: Ein Anwendungsbeispiel mit Empfehlungen für die Evaluierung der DSGVO sowie Verhandlungen zur ePrivacy-VO' (2020) 44 *Datenschutz und Datensicherheit - DuD* 172.

[152] Contreras and Reichman (n 88) 1312-1313. See also Section 3.5

[153] European Commission 'A European Strategy for Data' (n 141) 5 and fn. 13.

[154] See Verhulst et al (n 22) 25; 'Accelerating Medicines Partnership (AMP)' (*DataCollaboratives.org* 2020) https://datacollaboratives.org/cases/accelerating-medicines-partnership-amp.html accessed 14 February 2020; see also Heller and Eisenberg (n 12).

[155] Crémer, de Montjoye and Schweitzer (n 7) 92, 95.

[156] Bundeskartellamt, 'Big Data Und Wettbewerb' (2017) 9 https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Schriftenreihe_Digitales/Schriftenreihe_Digitales_1.pdf?__blob=publicationFile&v=3 accessed 28 April 2020.

[157] Martina Anzini and Anne-Carine Pierrat, 'Data Pools as Information Exchanges between Competitors: An Antitrust Perspective' (2020) 5 *cepInput* 3 https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepInput_Data_pools/cepInput_Data_Pools_as_Information_Exchanges_between_Competitors_An_Antitrust_Perspective.pdf accessed 28 April 2020.

[158] Technology Transfer Guidelines (n 147), paras 248-273; Communication from the Commission — Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-operation agreements Text with EEA relevance [2011] OJ C11/1, ch 2. For a review of data pools in light of Art. 101 and 102 TFEU, see Crémer, de Montjoye and Schweitzer (n 7) 93-107. For an analysis in light of Art. 101 TFEU, see Anzini and Pierrat (n 158).

[159] European Commission 'A European Strategy for Data' (n 139).

[160] See Lundqvist 'Competition and data pools' (n 22) 149.

[161] Richter and Slowinski, (n 4) 21.

[162] Consolidated version of the treaty of the European Union [2012] OJ C326/13, art 101; Technology Transfer Guidelines (n 147), para 261; Crémer, de Montjoye and Schweitzer (n 7) 97.

[163] Björn Lundqvist, 'Data Collaboration, Pooling and Hoarding under Competition Law' (2018). Faculty of Law, Stockholm University Research Paper No 61 < https://ssrn.com/abstract=3278578 accessed 28 April 2020.

[164] Crémer, de Montjoye and Schweitzer (n 7) 9, 97.

[165] Lundqvist, 'Data Collaboration' (n 162) 26.

[166] See Horizontal Co-Operation Guidelines (n 159) ch 2; Anzini and Pierrat (n 158) 4.

[167] Ariel Ezrachi and Maurice Stucke, 'Algorithmic Collusion: Problems and Counter-Measures. Note' (2017) OECD Doc. DAF/COMP/WD (2017) 25, 10, 25. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/WD%282017%2925&docLanguage=En accessed 28 April 2020. See also Crémer, de Montjoye and Schweitzer (n 7) 96. Usually, hub-and-spoke cartels refer to collusion facilitated by a third party. See Communication from the Commission — Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-operation agreements Text with EEA relevance [2011] OJ C11/1, 1– 72, para 55. See also Case C-74/14 Eturas UAB and Others v Lietuvos Respublikos konkurencijos taryba EU:C:2016:42.

[168] Maurice E Stucke and Ariel Ezrachi. 'Antitrust, algorithmic pricing and tacit collusion'. In Woodrow Barfield and Ugo Pagallo (eds) *Research Handbook of the Law of Artificial Intelligence* (Edward Elgar 2018) 627.

ing.[169] Opening the data pool to third-party membership would reduce the likelihood of collusion. Furthermore, data pools that mainly share technical data for R&D purposes and do not involve direct competitors are less likely to breach Art. 101 TFEU,[170] whereas pools sharing data on consumers may require more nuanced analysis.[171] Besides the risk of collusion and discriminatory practices towards competitors, the question has been raised whether data pooling may disincentivize pool members from refining their data collection and analysis methods.[172]

Lundqvist takes a view that competition norms for R&D collaboration and standard setting may complement or replace the Technology Transfer Guidelines addressing patent pools, especially when the governance model of a data pool deviates substantively from that of patent pool.[173] Emerging legal scholarship also seeks to review to what extent the economic features and legal instruments associated with patents in the context of standard setting, such as FRAND licensing agreements, are applicable to data.[174] It is unclear whether such a pledge to grant access to data could be legally binding for third parties.[175] Nevertheless, even if FRAND commitment is not enforceable, it may have a limited facilitative effect on the sharing of data at the organizational level of data governance among parties with aligned interests. It is also debated whether the FRAND commitment should be introduced into the context of data at all, given the history of litigating FRAND obligations for standard essential patents.[176] Concerns have also been raised about the possibility of market-dominant data holders in a data pool engaging in exploitative behaviour by requesting supra-FRAND licensing fees and violating Article 102 TFEU.[177]

6. Conclusions

IP governance models are also relevant for governing data. In particular, the concepts of data clearinghouses and data pools are helpful for recognizing data intermediaries and allow us to distinguish between different DGMs. In comparison with IP-based governance models, we recognize intermediaries that are specific to data: consent information clearinghouses, consent management clearinghouses, and access-right-based clearinghouses. Due to the legally non-excludable nature of data, DGMs require more contractual, organizational, and technical measures that ensure data integrity and inter-partes control of the transferred data. Similarly, it is unclear whether certain instruments that are familiar from IP, such as patent pledges[178] and FRAND-commitments, are effectively applicable to data. Standardized data licenses merit further research, especially since they may offer the legal foundation for a number of different DGMs.

Whereas patent pools must include complementary and essential technology,[179] how exactly one might qualify pooled data to ensure that the pool has favourable effects on social welfare remains elusive. This is especially true with regard to the role of analytics for the value of a data pool and the risks associated with pooling personal data. Similarly, the procompetitive nature of data clearinghouses and their treatment under competition law has not been researched. In the same vein, it would be relevant to study more closely the quality and locus of analytics in DGMs and their impact on data access, competition, and innovation, especially in AI applications. Furthermore, more research is needed on the relevance of data clearance, homogenization, and standardization for the success and costs of employing data intermediaries. All too often, policy initiatives on data markets presume that the mere existence of and access to data are sufficient for its meaningful reuse.

Besides legal and economic analysis of DGMs, we deem it important to conduct further empirical research both on individual types of DGMs as well as sector-specific analyses. We take the view that qualitative research following 'Governing Knowledge Commons' (GSC)[180] may be an appropriate framework for further research of diverse DGMs, including those which at first glance do not qualify as 'commons'.[181] In doing so, it seems worth to going into the details of the different data governance layers, i.e. the normative (e.g. legal) layer, the organizational layer, and the technological layer.[182] It is likely that such further empirical research will reveal presence hybrid DGMs, which may combine features of data pools and clearinghouses,[183] and will offer a more qualified taxonomy of DGMs that display platformization. Furthermore, technological development may advance the design of distributed DGMs as well as access-rights-based and consent management clearinghouses. Whether such solutions will succeed at truly supporting a data subjects' right to self-determination remains to be seen.

On the one hand, case studies of DGMs may expand our understanding of how legally compliant and effective DGMs should be designed. On the other hand, they will show where they succeed and fail, given the context-dependent interests of diverse stakeholders.[184] This includes the pull of platformization for the DGM's business model as well competitive tensions present in the particular industry. Special attention should also be given to the role of public actors in DGMs, for example in the context of smart cities. Such studies would also be

[169] Lundqvist (n 162) 3.

[170] Bundeskartellamt, 'Big Data Und Wettbewerb' (2017) 9 https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Schriftenreihe_Digitales/Schriftenreihe_Digitales_1.pdf?__blob=publicationFile&v=3; Lundqvist, (n 162) 11, 16.

[171] See Case C-238/05 Asnef/Equifax [2006] ECR I-111125; Case C-7/95 P John Deere [1998] ECR. I-3111 Anzini and Pierrat (n 158), Lundqvist, 'Data Collaboration' (n 162) 13-15; Crémer, de Montjoye and Schweitzer (n 7) 9, 94-95.

[172] Crémer, de Montjoye and Schweitzer (n 7) 9, 96-97.

[173] Lundqvist (n 162) 17; Horizontal Co-Operation Guidelines (n 159) ch 7; Commission Regulation on the application of Article 101(3) of the Treaty on the Functioning of the European Union to certain categories of research and development agreements [2010] OJ L335/36.

[174] For example, Richter and Slowinski (n 4) 17-23 address the possibility of applying FRAND (Fair reasonable and non-discriminatory) licenses as an instrument of private ordering.

[175] With respect to data protection law, consider the idea of 'sticky policies'. Siani Pearson and Marco Casassa-Mont 'Sticky policies: An approach for managing privacy across multiple parties' (2011) 44 *Computer* 60.

[176] Oscar Borgogno and Guiseppe Colangelo 'Data sharing and interoperability: Fostering innovation and competition through APIs' (2019) 35 CLSR 1, 15, 17; cf Mathew Heim and Igor Nikolic 'A FRAND Regime for Dominant Digital Platforms' (2019) 38 *J Intell Prop Info Tech & Elec Com* L 10.

[177] Crémer, de Montjoye and Schweitzer (n 7) 9.

[178] On patent pledges, see Jorge Contreras 'Patent Pledges' (2015) 47 *Ariz St LJ* 543.

[179] Technology Transfer Guidelines (n 147) para 261 (b).

[180] Katherine Strandburg, Brett Frischmann, B. M and Michael Madison, (Eds.) *Governing Medical Knowledge Commons* (2017), 13-17. The GSC framework is based on Ostrom's Institutional Analysis and Development (IAD) Framework. See Elinor Ostrom *Understanding Institutional Diversity* (Princeton UP 2005) 7-31.

[181] The term 'commons' describes systems of governing shared resources that are not subject to property rights, such as information and knowledge. Elinor Ostrom and Charlotte Hess *Understanding knowledge as a commons* (MIT Press 2007), 4-5.

[182] See regarding the three analytical layers of data governance, for instance, von Grafenstein, Wernick and Olk (n 4).

[183] Van Overwalle (n 100) 325, observes the occurrence of hybrid governance models for patents that contain features of clearinghouses and patent pools.

[184] See Mattioli (n 22) 180-181 on motivations for not sharing data in the context of cancer research.

relevant for policy-making on data and its regulation,[185] especially in light of the European strategy of creating 'data spaces' to foster seamless data exchange and innovation in nine strategic sectors, including health and mobility. [186]

[185] Cf Heiko Richter and Reto Hilty 'Die Hydra des Dateneigentums – eine methodische Betrachtung' (2018) *Max Planck Institute for Innovation and Competition Discussion Paper* No 12-2018, 9-10 https://ssrn.com/abstract=3263404 accessed 27 June 2019, on challenges of using empirical methods to inform lawmaking on data.

[186] European Commission 'A European Strategy for Data' (n 141) 5, 21-22. The data spaces aim to provide an infrastructure to support an ecosystem of diverse actors both from the private and public sectors. Ibid. 5.

08

Technology and Regulation

~~Content Not Available~~
Why The United Kingdom's Proposal For A "Package Of Platform Safety Measures" Will Harm Free Speech

Mark Leiser* & Edina Harbinja**

#online harms, duty of care, platform regulation, online safety

m.r.leiser@law.leidenuniv.nl

e.harbinja@aston.ac.uk

This article critiques key proposals of the United Kingdom's "Online Harms" White Paper; in particular, the proposal for new digital regulator and the imposition of a "duty of care" on platforms. While acknowledging that a duty of care, backed up by sanctions works well in some environments, we argue is not appropriate for policing the White Paper's identified harms as it could result in the blocking of legal, subjectively harmful content. Furthermore, the proposed regulator lacks the necessary independence and could be subjected to political interference. We conclude that the imposition of a duty of care will result in an unacceptable chilling effect on free expression, resulting in a draconian regulatory environment for platforms, with users' digital rights adversely affected.

1. Introduction

In April 2019, the UK Government's Department of Digital, Culture, Media and Sport ("DCMS") released its White Paper for "Online Harms" which, if accepted, would impose a new duty of care standard for online platform users to be overseen by an independent regulator.[1] If the White Paper proposals are brought into force, a regulator will be empowered to decide what activities and content are deemed harmful to Internet users.[2] After making this determination, the regulator can mandate intervention by internet providers to protect users from these harms.[3] If the recommendations in the DCMS White Paper are enacted into law, the UK could soon have a new Internet regulator (provisionally referred to as "OfWeb"[4]) that will have the statutory obligation of imposing a duty of care on online services to prevent a series of yet-to-be defined "online harms."[5] It moves enforcement of laws regulating content and free speech from courts and passes the obligation to private actors like Facebook, Google, and Twitter under the threat of penalties for non-compliance (possibly a loss of "license to operate"[6] or attaching personal liability to directors[7]). Rather than using the courts or other legitimate democratic institutions, platforms are obliged to determine and assess the harmfulness of user behavior before-and-after content is generated by users. The "duty of care" and the imposition of liability will change platforms and social media from a safe space for exercising fundamental speech rights to one where the state forces platforms to regulate content and decide what actions could be harmful.

However, the White Paper's "world-leading package of safety measures"[8] leaves important terms undefined, empowering politicians of the day to force platforms to respond to harms where there is little evidence to support its dangers. Based on a statutory duty of care to prevent users from harm, platforms will be forced to monitor, moderate, and remove user-generated content under the threat of "substantial fines".[9] As tight compliance deadlines strongly incentivize online service providers to comply with complaints swiftly, there is ample evidence from takedown regimes that platforms err on the side of caution, regardless of the actual merits of the claims.[10] The proposal also "empowers users" to hold platforms to account for failing to live up to their duty of care.[11] Fortunately, for people who care about the value of public discourse and are willing to resist the moral panic about the dangers of unregulated platforms, the White Paper should unite a disparate crew of civil society groups, desperate for a cause to rally behind since the Digital Economy Act 2010, the

1 Online Harms White Paper, Gov.UK, https://www.gov.uk/government/consultations/online-harms-white-paper (last updated Feb. 12, 2020) [hereinafter Online Harms White Paper].
2 Online Harms White Paper (n 1) ¶ 2.2 and ¶ 5.15.
3 Idem at ¶ 6.5.
4 Akin to Ofcom (The Office of Communications), broadcast and telecoms regulator, https://www.ofcom.org.uk/home, It is imagined that the office of the web' would be a newly created regulator named Ofweb.
5 Online Harms White Paper (n 1) 7 §-
6 Online Harms White Paper (n 1) 60 ¶ 6.5.
7 Idem.
8 Online Harms White Paper (n 1).
9 Online Harms White Paper (n 1) 59 § 6.
10 Hosting Intermediary Services and Illegal Content Online, Inst. for Info. L. (2018), https://www.ivir.nl/publicaties/download/hosting_intermediary_services.pdf.
11 Online Harms White Paper (n 1) 10 §16-18.

* Dr. Mark Leiser is Assistant Professor in Law and Digital Technologies, Leiden Law School, The Netherlands.
** Dr Edina Harbinja, Senior Lecturer in Media/Privacy Law, Aston University, United Kingdom.

Received 30 March 2020, Accepted 8 Aug 2020, Published 8 Sept 2020

SOPA[12]/ACTA[13] protests of 2012[14], and Articles 10 and 17 of the new Copyright Directive.[15] The DCMS intervention might just also help the typical citizen understand why Article 10 of the European Convention of Human Rights[16] is so important to the functioning of our modern society and end the present cycle of everything on the Internet is bad. With free expression at the heart of Western concepts like the marketplace of ideas, democratic deliberation, and the media's role in holding power to account, the challenge of any regulatory intervention online is targeting the effort at the right people, the legitimacy of the intervention, the proportionality of the measure and the effectiveness of steps taken, while ensuring media pluralism and the protection of low-level speech.[17]

This Article provides a brief overview of the events and the numerous hearings and interventions by the UK Parliament and the Government that led to the production of the White Paper. The Article then critiques the "duty of care" proposed in the White Paper, concluding that the imposition of a duty will chill free speech. The next section focuses on the role and independence of the proposed regulator, "OfWeb." This is followed by a critique of the harms identified by the Department of Culture, Media and Sport as justification for the duty of care. The Article concludes with recommendations and next steps.

2 The Bureaucratic Responses

The White Paper mirrors large parts of the output from the House of Lords' Communication Committee Report[18] and a previous report from the House of Commons DCMS Committee titled "Disinformation and 'fake news': Final Report"[19]: an amalgamation, in part, of what special interest groups believe is in the best interests of their members, rather than the wider digital community at large. The White Paper relies heavily on evidence from NGOs and charities like Doteveryone, the Children's Commissioner, Internet Matters, and Girl Guiding.[20] Upon reading the evidence submitted, one could easily conclude that the digital environment remains extremely hazardous and generally unsafe. On the contrary, the public generally believes the Internet is a good thing. In Ofcom's examination of online users, 59% of adults said the benefits outweigh the risks. Only a small percentage said the opposite. Furthermore, 61% of children said the Internet makes their lives better.[21] So where does this nuanced vision come from?

After the Cambridge Analytica scandal,[22] a series of hearings across a wide range of UK government entities were launched into the role of platforms in everything from disrupting democracy to causing long-term harm to children, facilitating abusive content.[23] The general consensus was that "something must be done."[24] Ironically, those same Members of Parliament (MPs) took to platforms to publicize how their plan was going to make the Internet safe again. The White Paper follows the Government's proposals set out in the Internet Safety Strategy Green Paper from October 2017[25] and the Digital Charter from January 2018.[26] The key principles for future regulation are parity ("what is unacceptable offline should be unacceptable online"), *openness*, *transparency*, the *protection of human rights*, and the *protection of children*.[27] In February 2019, the House of Commons DCMS Committee published their report, "Disinformation and 'Fake News',"[28] calling for government action to curtail the dissemination of deceptive content.[29] The White Paper goes even further, including disinformation within the list of harms that platforms will be under a duty of care to prevent.[30]

The Government and both Houses of Parliament agreed that there needs to be extensive regulation of the Internet and, in particular, social media.[31] To justify the need for intervention, they cite everything from issues with political advertising[32] (in particular, the UK's referendum on the Continued Membership of the European Union (Brexit)), to fake news and online manipulation,[33] data breaches by the tech giants,[34] the lack of competition in the Internet's mainstream (social

[12] https://www.congress.gov/bill/112th-congress/house-bill/3261/text

[13] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX-:52010IP0432&from=GA.

[14] The US Congress debated two bills (Stop Online Piracy Act (SOPA) was designed to protect the copyright creative industries. The bills were ultimately rejected after unprecedented protests and complaints to American representatives and Senators. The Anti-Counterfeiting Trade Agreement (ACTA) protests flared in Europe out of the belief that the openness of the Internet would be compromised.

[15] Council Directive (EU) 2019/790, 2019 O.J. (L 130) 92 (EC).

[16] The Convention for the Protection of Human Rights and Fundamental Freedoms was opened for signature in Rome on 4 November 1950 and came into force on 3 September 1953. As of 16 May 2018, it counts 47 States parties.

[17] Jacob Rowbottom, To Rant, Vent and Converse: Protecting Low Level Digital Speech, 71 *Cambridge L.J.* 355, (2 Apr. 2012).

[18] Regulating in a digital world: Final report published, House of Lords Communication Committee (9 Mar. 2019), https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-to-regulate/.

[19] Disinformation and 'fake news': Final Report published, House of Commons Digital, Culture, Media and Sport Committee (14 Feb. 2019), https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.

[20] Online Harms White Paper (n 1).

[21] Online Nation, Ofcom 3 (30 May 2019) https://www.ofcom.org.uk/__data/assets/pdf_file/0025/149146/online-nation-report.pdf 141.

[22] The Cambridge Analytica files, (*The Guardian*) https://www.theguardian.com/news/series/cambridge-analytica-files (last accessed 6 May 2019).

[23] See ICO issues maximum £500,000 fine to Facebook for failing to protect users' personal information, Info. Commissioner's Off. (25 Oct. 2018), https://ico.org.uk/facebook-fine-20181025; House of Commons Digital, Culture, Media and Sport Committee, (n 19); Addressing harmful online content, Ofcom (18 Sept. 2018), https://www.ofcom.org.uk/__data/assets/pdf_file/0022/120991/Addressing-harmful-online-content.pdf; Government response to the Internet Safety Strategy Green Paper, HM Government (May 2018), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf.

[24] Sonia Livingstone, Rethinking the rights of children for the Internet Age, Available at https://blogs.lse.ac.uk/parenting4digitalfuture/2019/04/03/rethinking-the-rights-of-children-for-the-internet-age/ (last accessed 25 April 2019).

[25] HM Government, Internet Safety Strategy – Green paper, October 2017, Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf (last accessed 6 May 2019).

[26] Digital Charter, Gov.UK, https://www.gov.uk/government/publications/digital-charter (last accessed 8 Apr. 2019).

[27] Ibid.

[28] House of Commons Digital, Culture, Media and Sport Committee (n 19).

[29] (n 19) 64.

[30] Online Harms White Paper (n 1) 31

[31] Online Harms White Paper (n 1); House of Lords Communication Committee (n 18) and House of Commons Digital, Culture, Media and Sport Committee (n 19).

[32] Online Harms White Paper (n 1) 28-29. See also Vote Leave's targeted Brexit ads released by Facebook, (*BBC News*, 26 July 26 2018), https://www.bbc.co.uk/news/uk-politics-44966969.

[33] Online Harms White Paper (n 1) 22-23. See Articles on Fake news, The Conversation, https://theconversation.com/uk/topics/fake-news-33438 (last accessed 30 May 2019).

[34] Idem at 31-32; see generally Sam Schechner & Mark Secada, You Give Apps Sensitive Personal Information. Then They Tell Facebook, (*Wall St. J.*, 22 Feb. 2019), https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636.

networking, search engines, advertising),[35] child abuse[36] and harms to children (including self-harm threats),[37] terrorist and extremist content,[38] and even knife crime.[39] The Committee expressed confidence that their proposal will address all of these issues, regardless of how different their causes and consequences.[40]

The White Paper recommends establishing a new independent regulator for the Internet[41] and the adoption of a co-regulatory model similar to broadcast regulation,[42] despite the fact that wireless internet is transmitted in a similar manner to the broadcasting signal, the Internet has almost zero resemblance to broadcasting. The new "OfWeb" will draft codes of conduct that set out principles of online safety and the "duty of care,"[43] backed up by reporting requirements and effective enforcement powers.[44] The regulator will also have responsibilities to promote education and awareness-raising about online safety, take a risk-based approach, 'prioritising action to tackle activity or content where there is the greatest evidence or threat of harm, or where children or other vulnerable users are at risk.'[45] Additionally, the regulator will be tasked to safeguard innovation, and to protect digital rights, 'being particularly mindful to not infringe privacy and freedom of expression.'[46] This regulatory effort actually goes hand-in-hand with what Facebook CEO Mark Zuckerberg recently proposed in his regulatory vision for Facebook.[47] Understandably, the proposal has attracted opposition and warnings from civil society groups, human rights advocates, lawyers, and academics.[48] Given the DCMS's promises and passion for new regulation, it is unlikely that the submissions substantially alter the proposal, no matter how well-evidenced and reasonable.

The White Paper is the latest in a long line of government reports and policy documents emanating from, among others, the controversy surrounding the vote on the United Kingdom's continued membership in the European Union (Brexit) and concerns about Russian interference in democratic discourse. The DCMS White Paper is the latest of these reports and focuses on the identification of "online harms" that are then used to justify the creation of a new regulator for Internet platforms. The harms are linked to and supported by evidence and reports filed by a large number of civil society groups, NGOs, charities, and child protection advocates.[49] The DCMS White Paper argues that these online harms are severe enough to warrant new and Internet-specific regulation. It also claims to reflect the changing tide in the way the government and society think about the Internet.

Despite numerous laws already in place to tackle some of the identified harms and numerous laws regulating content, actions, and behavior,[50] the White Paper attempts to pass the government's own policing responsibilities onto platforms; in other words, "it is *your* platform, *you* have to deal with it". The ethos of the White Paper is simple: platforms are seen as public spaces and are no different from theme parks, offices, and restaurants. Risk-based legal regimes like the UK's Health and Safety Act and the EU's General Data Protection Regulation have successfully deployed a duty of care before; therefore, as Facebook et al. are places where people gather, the imposition of a duty of care will work between platforms and users too. As we discuss in the next section, there are fundamental flaws with this argument.

3 How a Duty of Care Will Chill Free Speech

A duty of care normally carries with it a "three-stage test of foreseeability, proximity, and policy."[51] Foreseeability and proximity involve an examination of the factual circumstances of the parties.[52] Policy considerations usually require the court to deploy foresight into the consequences for other parties, not part of the dispute. The test requires the court to determine whether there is a legal relationship between the parties of the dispute;[53] for example, does an employer have a duty of care to its employees? Does a business have a duty of care to its customers? Does a building site operator have a duty of care to its visitors? A legal requirement to keep a place safe not only makes sense, but also puts prevention at the heart of the legal regime. Secondly, one of the central purposes of a duty of care is to apply similar duties to similar facts.[54] Once a court establishes that a duty of care is owed between x and y in circumstances z, then that decision applies to all future cases of the same kind. Of course, this duty will then be foreseeable and more certain, and not vague as suggested in the White Paper, both in terms of harms and individuals owed to.

35 Online Harms White Paper (n 1) 27-28; Matt Binder, Google hit with $1.7 billion fine for anticompetitive ad practices, (*Mashable,* 20 Mar. 2019), https://mashable.com/article/google-eu-antitrust-fine-ads/.

36 Online Harms White Paper (n 1) 50; Jamie Grierson, Met police 'overwhelmed' by surge in online child sexual abuse, (*The Guardian*, 28 Mar. 2019), https://www.theguardian.com/uk-news/2019/mar/28/london- metropolitan-police-overwhelmed-by-surge-in-online-child-sexual-abuse-watchdog-finds.

37 Online Harms White Paper (n 1) 19; Sarah Marsh & Jim Waterson, Instagram bans 'graphic' self-harm images after Molly Russell's death, (*The Guardian*, 7 Feb. 2019), https://www.theguardian.com/technology/2019/feb/07/instagram-bans-graphic-self-harm-images- after-molly-russells-death.

38 Online Harms White Paper (n 1) 14; Preventing the dissemination of terrorist content online, European Parliament (Sept. 2018), http://www.europarl.europa.eu/legislative- train/theme-area-of-justice-and-fundamental-rights/file-preventing-the-dissemination-of-terrorist-content-online.

39 Online Harms White Paper (n 1) 13. The report implies that online content allegedly glorifies gangs, and has led to an increase in knife crimes (the Report cites ONS (2019). Crime in England and Wales, Year Ending September 2018. Available at: https://www.ons.gov.uk/ peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingseptember2018). This is unsupported and the causation/correlation is vague. See generally Knife Crime, (*The Guardian)*, https://www.theguardian.com/uk/knifecrime (last accessed 31 May 2019).

40 Online Harms White Paper (n 1) 11–41.

41 Idem at 57 (OfWeb is the suggested name. However, the paper suggests Ofcom may initially be given the task.).

42 Idem.

43 William Perrin & Lorna Woods, Reducing harm in social media through a duty of care, CarnegieUK Trust (8 May 2018), https://www.carnegieuktrust.org.uk/blog/reducing-harm-social-media-duty-care/.

44 Online Harms White Paper (n 1) Section 7.42 and 6.

45 Ibid, p. 53.

46 Ibid.

47 Regulating in a Digital World, House of Lords Select Committee on Comm. (9 Mar. 2019), https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/29902.htm.

48 See, e.g., PI's take on the UK government's new proposal to tackle "online harms", Privacy Int'l (8 Apr. 2019), https://privacyinternational.org/news/2779/pis-take-uk-governments-new-proposal-tackle-online-harms; Jim Killock & Amy Shepherd, The DCMS Online Harms Strategy must "design in" fundamental rights, Online Rts. Group (8 Apr. 2019), https://www.openrightsgroup.org/blog/2019/the-dcms-online-harms-strategy-must-design-in-fundamental-rights; UK: More regulation of content online risks right to privacy and free expression, Article 19 (19 June 2018), https://www.article19.org/resources/uk-more-regulation-of-content-online-risks-rights- to-privacy-and-free-expression/.

49 Examples include reports cited in the Online Harms White Paper (n 1) 13 – 14.

50 Abusive and Offensive Online Communications: A Scoping Report, L. Commission at 66-96 (1 Nov. 2018), https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage- 11jsxou24uy7q/uploads/2018/10/6_5039_LC_Online_Comms_Report_FINAL_291018_WEB.pdf.

51 *Caparo Industries v. Dickman* [1990] 2 AC 605 (HL).

52 Ibid per Lord Roskill at 629 - 627

53 Ibid per Lord Oliver of Aylmerton at 632 - 634

54 Ibid at 618 – 619 per Lord Bridge of Harwich; Brennan J. in the High Court of Australia in *Sutherland Shire Council v. Heyman* (1985) 60 A.L.R. 1, 43–44.

Applying a duty of care between platforms and users to speech, however, will require platforms to block entire categories of speech, based on a legal obligation to block specific kinds of harm in the future. Cyber-bullying might cause individualized harms, but another user might not view others' comments as bullying. Trolling and swearing, for example, might be completely unacceptable to one person, but acceptable to another. Trolling is purely subjective speech that may, on occasion, rise to the threshold of criminal speech. An example would be grossly offensive, obscene, indecent or menacing communications regulated by s. 127 of the Communications Act 2003. Trolling that does not pass this threshold, would not be considered criminal now. For this reason, there are no laws regulating this kind of speech or content wherein the legal test of harm or offence is subjective and a recipient of speech/content gets to be the sole determinant of whether something causes harm.[55] In a recent high-profile event, a columnist for the New York Times accused an academic of abuse for referring to him as a "metaphorical bedbug".[56] The DCMS report offers no guidance about how platforms should police metaphors.

In the UK, the test required before criminal charges will be brought against content posted on social media is one of such gross offensiveness that criminal charges should be brought.[57] After several controversial and high-profile prosecutions,[58] the Public Prosecutor issued guidelines for the prosecution of grossly offensive speech.[59] These limit prosecutions under Section 127 of the Communications Act 2003 to cases which go beyond those which are "[o]ffensive, shocking or disturbing; or [s]atirical, iconoclastic or rude; or [t]he expression of unpopular; or unfashionable opinion about serious or trivial matters, or banter or humor, even if distasteful to some or painful to those subjected to it."[60] The threshold for bringing criminal charges is high, yet the DCMS bases their report on broad categories of speech that does not come close to the threshold of criminality.

One of the challenges of regulating content is differentiating between the risk of harm and the *uncertainty* that the harm may or may not occur.[61] Imposing a duty of care on platforms to tackle harms associated with content *ex post* is fundamentally different from the imposition of a duty of care on platforms for the *uncertainty* that users may generate content that causes harm. Risk can be accounted and modelled for and quantified through pricing strategies, while *uncertainty* is a risk than cannot be measured. Its reliance on the evidence of numerous NGOs, charities, consumer protection groups, and other advocates informs us of numerous *types* of harms[62], but none can predict *when* or *how* these harms will take place.

The imposition of a duty of care backed by financial sanctions onto platforms, spins both the *uncertainty* about the frequency and the likelihood and validity of the harms themselves into the *risk* of harms associated with user-generated content. Once this occurs, risk can be modelled and priced. This is a dangerous path leading to a chilling effect on permitted content. First, are the unknowns. As Douglas Adams writes in the Hitchhiker's Guide to the Galaxy, "the major difference between a thing that might go wrong and that thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong, it usually turns out to be impossible to get at or repair."[63] Second, once something is priced, the imposition of a duty of care establishes a transaction cost for content; speech deemed too costly to the platform will be filtered, blocked and/or removed *ex ante* rather than *ex post*, especially when the uncertainty surrounding content is determined to have too high a transaction cost, regardless of its *actual risk*. In other domains, where one sees the imposition of a duty of care (i.e. environmental law[64] or health and safety[65]), the law serves to mitigate the distribution costs of uncertainty through legal conventions like the precautionary principle[66] or the preventative measures rule.[67] Prentice-Dunn & Rogers argue that preventative regimes operate best when it is possible to predict outcomes that are contrary to totally rational decision making.[68] But regulating speech through precaution or prevention is a disproportionate response to the various forms of uncertainty. An online harm may come about from one or more causes, and if it occurs, one or more effects. In isolation, an innocuous comment may cause little harm, but the cumulative effect might cause substantial harm.

The duty of care would require platforms to avoid content that would place them at fault for a list of harms. In practice, this means that platforms would be under a statutory obligation to take reasonable care. If there was no duty of care to prevent a certain harm; in these circumstances, the law would permit platforms to act unreasonably. Yet, cases in English law about duty of care are limited to whether

55 Section 127 of the Communications Act 2003 should be interpreted as an objective test. Would a reasonable person view the communication as 'grossly offensive or of an indecent, obscene or menacing character'.

56 @davekarpf, Twitter (26 Aug. 2019, 2:07 PM), https://twitter.com/davekarpf/status/1166094950024515584. See Allan Smith, A professor labeled Bret Stephens a 'bedbug.' Here's what the NYT columnist did next, (*NBC News*, 27 Aug. 2019), https://www.nbcnews.com/politics/politics-news/professor-labeled-bret-stephens-bedbug-here-s-what-nyt-columnist-n1046736.

57 Communications Act 2003, c. 21, § 127 (UK).

58 *Chambers v Director of Public Prosecutions* [2012] EWHC 2157 (PC). See also Azhar Ahmed Sentencing Remarks, Available at https://www.judiciary.uk/wp-content/uploads/JCO/Documents/Judgments/azhar-ahmed-sentencing-remarks-09102012.pdf (last accessed 12 Sept 2019); "Man who racially abused Stan Collymore on Twitter spared prison", (*The Guardian*, 21 Mar 2012), Available at https://www.theguardian.com/technology/2012/mar/21/man-racially-abused-collymore-twitter-spared-prison (last accessed 11 Sept 2019).

59 Director of Public Prosecutions, "Social Media - Guidelines on prosecuting cases involving communications sent via social media", Revised: 21 August 2018, Available at: https://www.cps.gov.uk/legal-guidance/social-media-guidelines-prosecuting-cases-involving-communications-sent-social-media (last accessed 23 March 2020).

60 Director of Public Prosecutions, "Interim guidelines on prosecuting cases involving communications sent via social media", Available at https://adam1cor.files.wordpress.com/2012/12/117342720-social-media-dpp.pdf, (last accessed 12 September 2019).

61 For a detailed explanation on the difference between risk and uncertainty, see Volz, K. G., & Gigerenzer, G. (2012). Cognitive processes in decisions under risk are not the same as in decisions under uncertainty. *Frontiers in Neuroscience*, 6, 105.

62 Department for Digital, Culture, Media & Sport and UK Council for Internet Safety, Online harms research publications, Available at https://www.gov.uk/government/collections/online-harms-research-publications (last accessed 3 Sept 2019).

63 Douglas Adams (2000), *Mostly Harmless* (New York: Del Rey).

64 At the European level, the precautionary principle was enshrined in the Maastricht Treaty in 1992. It is now included in Article 191 of the Treaty on the Functioning of the European Union among the principles underpinning EU environmental policy.

65 The Health and Safety at Work etc Act 1974; See also Management of Health and Safety at Work Regulations (MHSWR) 1999.

66 M.D. Rogers. Scientific and technological uncertainty, the precautionary principle, scenarios and risk management. *Journal of Risk Research*, 4(1):1–15, 2001; See also doteveryone, "A digital duty of care", Available at https://doteveryone.org.uk/wp-content/uploads/2019/02/Dotevery-one-briefing-a-digital-duty-of-care.pdf (last accessed 3 September 2019).

67 Niskanen, T., Naumanen, P., & Hirvonen, M. L. (2012). An evaluation of EU legislation concerning risk assessment and preventive measures in occupational safety and health. *Applied ergonomics*, 43(5), 829-842.

68 Prentice-Dunn, S., & Rogers, R. W. (1986). Protection motivation theory and preventive health: Beyond the health belief model. *Health Education Research*. 1(3), 153-161.

the duty of care was applied properly and whether the party with the obligation acted reasonably.[69] Most duties of care cases involve an examination of the application of vague concepts like "foreseeability," "proximity," and "fair, just and reasonable."[70] The scope of the duty is unrelated to legal causation: should we limit the extent of the defendant's responsibility for the harm despite the fact that the defendant's fault was a but-for cause of the harm? It is difficult to contemplate that platforms, rather than user-generated content, is the but-for cause of harm, that can only be identified by a new regulator.

4 Online Harm Offensive

The government defines "online harm" as "online content or activity that harms individual users, particularly children, or threatens our way of life in the UK, either by undermining national security or by reducing trust and undermining our shared rights, responsibilities and opportunities to foster integration."[71] The vague definition ironically refers to the UK's "way of life" and "rights," but the list of harms contradicts this proposition. Having analyzed the White Paper,[72] the EDPS report on Online Manipulation,[73] the High Level Working Group's Report on Disinformation,[74] the House of Lords' Communication Committee Report on Regulating the Internet,[75] and any of the hundred other reports into deceptive media online, one could be forgiven for thinking that "harms" are an invention of the World Wide Web. Although it is clear that the Internet's architecture and scale make some harms easier to facilitate,[76] it is also true that the diffusion of harms, especially within the context of communication, has always been a danger for society.

The White Paper's framework aims to regulate harms "based on an assessment of their prevalence and impact on individuals and society."[77] Rather than relying on compelling evidence, this is based on a handful of surveys and media reports, largely provided by a variety of outside groups with their own agendas.[78] Even more troubling is the fact the list of harms is "by design, neither exhaustive nor fixed."[79] This is justified by claiming that a "static list could prevent swift regulatory action to address new forms of online harm, new technologies, content and new online activities."[80] Consequently, in the right political climate, OfWeb could theoretically proclaim *anything* harmful, including academic articles that criticize OfWeb's approach for its chilling effect on free expression.

The White Paper identifies twenty-three harms in total.[81] Some of them are already criminal offenses, others are so vague it would be a regulatory achievement for OfWeb to come up with a definition that sounds good in theory, but also works in practice. The DCMS report categorizes these harms in three groups: harms with a clear definition; harms with a less clear definition and underage exposure to legal content.[82] In the table below, we provide an overview of all the harms included in these three groups, referring to their current legal status, i.e. whether some of them are already a criminal offence, of their status is less clear from the perspective of the current laws.

Table 1 Harms with a Clear Definition[83]

Group 1: **Harms with a Clear Definition**	**Status**	**Criminal Law Provision**
Child sexual exploitation and abuse	Criminal Offense	Sections 47-51 of the Sexual Offences Act 2003.
Terrorist content and activity	Criminal Offense	Section 58 of the Terrorism Act 2000; Section 3 of the Counter-Terrorism and Border Security Act 2019.
Organized immigration crime	Criminal Offense	Modern Slavery Act 2015; Section 1 of the Asylum and Immigration (Treatment of Claimants, etc.) Act 2004; Sections 57 to 59 of the Sexual Offences Act 2003.
Modern slavery	Criminal Offense	Modern Slavery Act 2015.
Extreme pornography	Criminal Offense	Section 63 of the Criminal Justice and Immigration Act 2008.
Revenge pornography	Criminal Offense	Section 33 of the Criminal Justice and Courts Act 2015.
Harassment and cyber-stalking	Criminal Offense	Section 2, 2A, 4, 4A, Protection from Harassment Act 1997; Section 1 Malicious Communications Act 1988.
Hate crime	Criminal Offense	Public Order Act 1986; Racial and Religious Hatred Act 2006; Criminal Justice and Immigration Act 2008; Criminal Justice and Public Order Act 1994; For England, Wales, and Scotland, the Crime and Disorder Act 1998 makes hateful behavior towards a victim based on the victim's membership (or presumed membership) in a racial group an "aggravating factor" for the purpose of sentencing in respect of specified crimes. Sections 2, 2A, 4, 4A of the Protection from Harassment Act 1997 also apply for racially and religiously aggravated offences of harassment and stalking and putting people in fear of violence, and stalking involving fear of violence. Finally, there are communication offence under section 127(1) of the Communications Act 2003, or section 1 of the Malicious Communications Act 1988, with enhanced sentencing due to hostility towards one of the five protected characteristics.
Encouraging or assisting suicide	Criminal Offense	Section 2 and 2A of the Suicide Act 1961.
Incitement of violence	Criminal Offense	Section 44, 45 of the Serious Crime Act 2007.
Sale of illegal goods/ services such as drugs and weapons on the open internet	Criminal Offense	Section 1(1) of the Criminal Law Act 1977; Section 46 of the Serious Crime Act 2007; Fraud Act (2006), Misuse of Drugs Act (1971), or Firearms Act (1968).
Content illegally uploaded from prisons	Criminal Offense	Section 40D(3A) Prison Act 1952.
Sexting of indecent images by under 18s	Criminal Offense	Section 45 of the Sexual Offences Act 2003.

As seen in the table above, these harms are *already* illegal and there is no need to introduce entirely new laws for them. Some of them could benefit from further clarifications (e.g. terrorist content or harass-

69 Key cases: *Donoghue v Stevenson* [1932] AC 562; *Topp v London Country Bus* [1993] 1 WLR 976; *Home Office v Dorset Yacht Co Ltd* [1970] AC 1004.
70 Howarth, D. (2006). Many Duties of Care—Or a Duty of Care? Notes from the Underground. *Oxford Journal of Legal Studies*, 26(3), 449-472.
71 Online Harms White Paper (n 1) 30.
72 Online Harms White Paper (n 1).
73 EDPS Opinion on online manipulation and personal data, Opinion 3/2018, available at: https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf.
74 The final report "A multi-dimensional approach to disinformation" is available at https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation.
75 House of Lords Communication Committee (n 18).
76 Council of Europe, Parliamentary Assembly, Resolution 1970 (2014), Internet and politics: the impact of new information and communication technology on democracy, Available at https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=20447&lang=en, Accessed 30 May 2019.
77 Online Harms White Paper (n 1) 30.
78 Ofcom (2018). Adults' Media Use and Attitudes Report. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0011/113222/Adults-Media-Use-and-Attitudes-Report-2018.pdf Ofcom and ICO (2018). Internet users' experience of harm online 2018. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/120852/Internet-harm-research-2018-report.pdf Accessed 6 May 2019.
79 Online Harms White Paper (n 1) 30.
80 Ibid.
81 Ibid, p. 31.
82 Ibid.
83 Ibid.

ment and cyberstalking).[84] We argue that this should not be done within the overarching bundle of dissimilar harms as suggest in the White Paper.

Table 2 Harms with 'less clear' definition[85]

Group 2: **Harms with 'less clear' definition**	**Status**	**Provision/Comment**
Cyberbullying and trolling	Potentially criminal in some instances, but vaguely defined and with serious implications for free speech.	Potentially subset of communication offences under section 2, 2A, 4, 4A of the Protection from Harassment Act 1997; Section 1 Malicious Communications Act 1988, but vague and depends on the definitions, which are vague and overlapping.[86]
Extremist content and activity	Criminal in many instances, but vaguely defined and difficult to apply uniformly.	Potentially Section 58 of the Terrorism Act 2000; Section 3 of the Counter-Terrorism and Border Security Act 2019, but this is already covered by terrorist content, so it is unclear why the extremist content is necessary as a "new harm."
Coercive behavior	Vaguely formulated.	Potentially Section 2, 2A, 4, 4A, Protection from Harassment Act 1997; Section 1 Malicious Communications Act 1988, but vague and depends on the definition. This harm can also be potentially confused with existing offences, such as harassment. Further offence is found in section 76 of the Serious Crime Act 2015, but it only relates to domestic abuse cases.
Intimidation	Potentially criminal, but also vaguely defined.	Section 4 and 4A, Protection from Harassment Act 1997, already illegal, and serious fear of violence offences in Section 4 and 4a of the Public Order Act 1986, so unnecessary as a vaguely defined and subjective harm here. Its vagueness could mean that the harm may include legitimate free speech.
Disinformation	Vague, regulation of "fake news" is in progress.	Potentially covered by Section 127(2)(a) or (b) of the Communications Act 2003 and Section 1 Malicious Communications Act 1988, ongoing law reform in the area. Section 51 of the Criminal Law Act 1977 covers a bomb hoax; Hoaxes involving noxious substances or things are covered under section 114(2) of the Anti-Terrorism, Crime and Security Act 2001; giving a false alarm of fire exists under section 49(1) of the Fire and Rescue Services Act 2004; impersonating a police officer - section 90 of the Police Act 1996; section 106 of the Representation of the People Act 1983 offence to make or publish any false statement of fact in relation to the personal character of a candidate prior to or during an election.[87]
Violent content	Vaguely defined and problematic - any violent content online, including artistic speech could be harmful.	It is unclear how this is different from harassment, fear of violence, threat, stalking and extreme pornography, and other already existing criminal offences, as noted above. Does it include artistic speech, video games, films and what implication can this vaguely defined harm have on free speech?
Advocacy of self-harm	Dangerous precedent, blurs the lines between free speech and 'advocacy', as well as support self-harm support groups on social media.	Not illegal, but the UK government has threatened to introduce legislation if platforms do not remove content promoting self-harm. The Law Commission notes "[publicizing] or glorifying self-harm is not ostensibly criminal either."[88] However, offence of causing grievous bodily harm with intent, contrary to section 18 of the Offences Against the Person Act 1861, could be used here, provided that the victim caused herself serious harm with intent, so assisting or encouraging such behavior could be guilty of an offence under sections 44 to 46 of the Serious Crime Act 2007.[89]
Promotion of female genital mutilation	Criminal	Female Genital Mutilation Act 2003 makes the Act illegal, but there is no offence relating to its promotion. but see ss. 44 - 46 of the Serious Crime Act, intentionally encouraging or assisting an offence; encouraging or assisting an offence believing it will be committed; and encouraging or assisting offences believing one or more will be committed.[90]

[84] The Law Commission, 'Abusive and Offensive Online Communications: A Scoping Report', Law Com No 381, at: https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2018/10/6_5039_LC_Online_Comms_Report_FINAL_291018_WEB.pdf.

[85] Online Harms White Paper (n 1)31.

[86] The Law Commission (n 84) ¶ 8.8. The Commission also rightly points out the issues with studies that analyze these phenomena, such as: 'it is unclear whether the offending behavior being discussed would constitute a criminal offence under the criminal law, or whether the study is focused on generally unkind and unacceptable behavior, which falls short of an offence but is capable nevertheless of causing harm.' (emphasis by the authors).

[87] But there are no offences of creating or spreading fake news per se in the UK, see Disinformation and 'fake news': Interim Report, Report of the Digital, Culture, Media and Sport Committee (2017-19) HC 363, available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf. There are however civil remedies in defamation and media regulation. The ECHR in *Salov v Ukraine* (2007) 45 EHRR 51 held that Article 10 does not prohibit discussion or dissemination of information received, even if it is strongly suspected that this information might not be truthful. The Court suggested that otherwise it would deprive persons of the right to express their views and opinions about statements made in the mass media, and would therefore place an unreasonable restriction on the freedom of expression. For a broader discussion see B McNair, Fake News Falsehood, Fabrication and Fantasy in Journalism, Routledge, 2018; T McGonagle, "'Fake news': False fears or real concerns?" (2017) 35(4) *Netherlands Quarterly of Human Rights* 203

[88] The Law Commission (n 84) ¶ 12.93, see also A Gillespie (2016), *Cybercrime: Key Issues and Debates*, Routledge, 200. As the Commission also finds, online communications promoting self-harm would need to pass the threshold of "obscene, indecent or grossly offensive" to be prosecuted under section 127 of the CA 2003. If considered obscene, its publication may be prosecuted under section 2 of the Obscene Publications Act 1959. This would need to ensure compatibility with the Human Rights Act 1998, as both the Commission and Gilespie warn. See The Law Commission (n 84) ¶ 12.95, or Gilespie ibid p. 201.

[89] See Law Commission (n 84) ¶ 12.94. The Commission also questions the appropriateness of using criminal law in this context, ¶ 12.99.

[90] The Law Commission (n 84) ¶ 12.64 notes: 'However, a general glorification of certain conduct, without an intention for a specific crime to be committed, would be difficult to fit within the terms of sections 44 to 46 of the SCA 2007'..

Table 3 Underage exposure to Legal Content[91]

Group 3: Underage exposure to Legal Content	Status	Comment
Children accessing pornography	Service providers' liability for making pornographic content available to persons under 18.	Digital Economy Act 2017, s 14 requires providers to prevent children from accessing pornography, under a threat of financial penalties (implementation has been delayed).[92]
Children accessing inappropriate material	Vague, undefined, and problematic - who decides what is 'inappropriate' and who decides whether a child can access? What is the role of parents and education in helping kids understand what is appropriate for them to engage with? Do we really want parents determining what content a child accesses about sexual health is appropriate.	There are existing provisions preventing children from accessing pornographic, obscene and other prohibited materials, as noted above. 'Inappropriate' as a category is extremely vague and open to interpretation, it is not certain whether it includes harmful online advertising, for example. It could also affect free speech of adults, children as well as other rights such as privacy.[93]
Under 13s using social media and under 18s using dating apps	Already the rule; however, rarely enforced; moral panic.	This is a question of adequate enforcement and age verification, as noted above.[94]
Excessive screen time	Moral panic.	Evidence that the risk outweighs benefits have not been conclusive and the real harm is often overestimated by the media and advocacy groups.[95]

Even for "clear" harms, there is some dispute whether the definition is clear enough and to what extent these should even be criminalized (e.g. the definition of what defines "extreme pornography" is limited to the anus, genitals, breasts; necrophilia and bestiality, while excluding other injuries to the body).[96] Some other harms are, of course, indisputably illegal, e.g. content related to child abuse. In the "less clearly" defined harms group,[97] the scope of harm goes far beyond what is permitted by UK law in the offline world.

Some harms are not illegal in the offline world and some are difficult to define without adversely affecting protected free speech (e.g. trolling, violent content, intimidation or disinformation). The legislators have been avoiding to criminalize all trolling for instance, as a legal definition would potentially include legitimate free speech. Furthermore regulating "to achieve equivalent outcomes online and offline" requires a platform to determine the comparable offline offence. For example, Section 127 of the Communications Act 2003 is far more restrictive to online speech than any offline equivalent.[98] Thus, DCMS are suggesting that in order to achieve equivalent outcomes one needs tighter restrictions on online speech.

5 Free Speech

Article 10 of the European Convention on Human Rights reflects a middle ground between unfettered speech under the First Amendment of the US Constitution and the authoritarian approach to direct control of the media within a territory proposed by the Soviets.[99] Beyond the Convention, the European "project" is rooted in concepts of dignity and social justice. The first pillar of the Treaty of Maastricht on European Union refers to and of social protection and equality between men and women.[100] The European Social Model is based on the concept of social cohesion. That individuals should not have to put up with promulgation of views deeply hurtful to themselves or their communities is a basic tenet of this approach to the needs of a pluralistic society.

Article 10, Para 1 ECHR provides that the freedom of expression "shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers." The same wording appears almost verbatim in Article 19, Para 2, of the International Covenant on Civil and Political Rights (ICCPR).[101] Recent judgments from the ECtHR have highlighted State obligations to not only respect freedom of expression by refraining from interference and censorship, but also to ensure a favorable environment for inclusive and pluralistic public debate.[102] Thus, Article 10 requires the adoption of "positive measures" to protect the "passive" element (right to be correctly informed) of free expression.[103] There is a strong link between the two fundamental rights: an election process is "free" when, not only the electorate's choice is determined by access to the widest possible range of proposals and ideas, but also if the election results do not risk being distorted or altered by the dissemination of false information.

Article 10(2) of the European Convention contains a list of the exceptions to the right of free expression contained in 10(1).[104] Any limitations to this must be not only proportional but achieve a legitimate aim for which the restriction is in place, in accordance with the law, and is necessary in a democratic society. European Union activities must respect the proportionality principle and, in areas that did not fall within its exclusive competence, the principle of subsidiarity[105] which encourages regulation at the local level, "as close to the citizen

91 Online Harms White Paper (n 1) 31.

92 The UK Government has struggled to find the appropriate way to implement the age verification system, so the implementation and enforcement of this provision has been delayed a few times already, see J Waterson and A Hern, 'UK age-verification system for porn delayed by six months', (*The Guardian*, 20 Jun 2019), at: https://www.theguardian.com/technology/2019/jun/20/uks-porn-age-verification-system-to-be-delayed-indefinitely

93 Regulation is not the silver bullet for all the risks associated with children using the Internet. The matrix of opportunities and risk associated with this is very complex, and researchers have identified various model to assess and address this risk, see e.g. Livingstone, Sonia, Mascheroni, Giovanna and Staksrud, Elisabeth, 'European research on children's internet use: assessing the past and anticipating the future', 2018 *New Media and Society*, 20 (3). pp. 1103-1122; E Staksrud (2013), *Children in the Online World: Risk, Regulation, Rights*, Aldershot: Ashgate.

94 There are, however, many myths associated with this issue and the picture may not always be as presented in the media. See e.g. Livingstone, Mascheroni and Staksrud, ibid.

95 Livingstone, Sonia and Franklin, Keely (2018) Families with young children and 'screen time' advice. *Journal of Health Visiting*, 6 (9). pp. 434-439.

96 Section 63. Criminal Justice and Immigration Act 2008.

97 Online Harms White Paper (n 1) 31.

98 Section 127 Communications Act 2003.

99 For a good overview of the debates of the foundations of free expression in international law, see Leiser M.R. (2019), 'Regulating Computational Propaganda: Lessons from International Law', *Cambridge International Law Journal* 8(2): 218-240.

100 Article 1 and 2 of the Treat of Maastrich on European Union; See also https://www.europarl.europa.eu/RegData/etudes/fiches_techniques/2013/010103/04A_FT(2013)010103_EN.pdf

101 UN General Assembly, International Covenant on Civil and Political Rights, 16 December 1966

102 See *Lingens v Austria* (App No 9815/82) 8 July 1986; *Janowski v Poland*, Judgment (Merits), (App No 25716/94) 21 January 1999; *Tammer v Estonia*, (App. 41205/98), 6 February 2001; *Janowski v Poland*, Judgment (Merits), (App No 25716/94) 21 January 1999

103 J.F. Akandji-Kombe, Positive obligations under the European Convention on Human Rights. *A guide to the implementation of the European Convention on Human Rights—Human rights handbooks*, No. 7. 2007, Strasbourg: Council of Europe. Available at: https://rm.coe.int/168007ff4d.

104 Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950

105 Article 5 of the EC Treaty; now Article 5(3) of the Treaty on European Union (TEU) and Protocol (No 2) on the application of the principles of subsidiarity and proportionality.

as possible."[106]

While the European Court of Human Rights will generally not interfere with a Member State's margin of appreciation to determine whether a particular measure is "necessary," the ECtHR does not take kindly to measures that are not properly prescribed or satisfy the quality of law test, especially when it comes to expression.[107] However, Strasbourg is only engaged in "...applying the principle of subsidiarity when national authorities have demonstrated in cases before the court that they have taken their obligations to secure Convention rights seriously."[108]

What is it about free speech that irritates UK regulators? The Internet is a communications system without any front-end filter for user-generated content. It exists inside a legal system that has historically regulated different forms of non-digital speech. The framework for free expression as a *fundamental right* had to find a way to slot on top of an existing body of law that restricts speech in certain instances; for example, the longstanding law of copyright restricts the use of intellectual creation[109] and defamation law generally restrains speech that lowers the standing of one's reputation in the eyes of right-minded members of society.[110] Attributing responsibility (and in many cases, liability) was relatively straightforward when systems of information dissemination were limited to one-on-one or mediated communication.

Mass communication systems, like broadcasting and print journalism, have fought to be subject to narrow controls.[111] Media freedom in Europe has come about from hard battles that have established a set of legal principles and rules, alongside general duties and obligations.[112] As a result, the media and the political class have a symbiotic relationship wherein the press might sit as a vital check on political power in one newspaper column and play the role of public relations conduit in another. The roles are enshrined in Convention Law and the European Union's Charter of Fundamental Rights and judgements of the European Court of Justice.[113]

The question that requires a clearer answer is what is an "offense" and when/why people do take offense. This may include some of the following considerations: unwarranted critique of/interference with sense of personal/collective identity; political opportunism; taking offence seriously: questions of principle/practice. Further, this begs three questions: first, where are appropriate limits to be drawn? Second, who should decide where these limits are? Third, is it legitimate ever to restrict freedom of expression to avoid the causing of offense to recipients of message?

Starting point to answering these key questions should be the assertion of the importance of free speech: "Freedom of expression constitutes one of the essential foundations of such a society, one of the basic conditions for its progress and for the development of every man. Subject to Article 10 (2), it is applicable not only to 'information' or 'ideas' that are favorably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no 'democratic society'."[114]

Will users be banned or have their speech curtailed for spreading fake news, just because they are misinformed or not well-educated? Is sharing a review of a violent Korean film or a death metal song with a bloody video 'violent harmful content'? Does a platform need to take down Quentin Tarantino film previews too? The right to offend, shock and disturb is part of free expression. Speech should never be judged on its subjective effects on a user, but carefully weighed against clearly defined public interest and other human rights. The vague nature of harms as a group that is not illegal could be challenged under principles of the rule of law, proportionality and legal certainty.[115] This also contravenes the longstanding principle from Handyside v UK:[116]

> Freedom of expression ... is applicable not only to "information" or "ideas" that are [favorably] received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population.[117]

However, the Internet has brought a wholesale change in not only how information was disseminated, but also in respect of the actors doing the broadcasting. Users now generate most of the Internet's content, with Article 10 engaging everything from low-level speech to news commentary to search engines display results.[118] With a variety of technology available to hide user identities and the web's architecture empowering the user to not only speak without fear of social censure, but also automate and propagate their voice, the search for the *right* actor to regulate has frustrated regulators who have spent the last decade searching for a way to characterize platforms as publishers to attach a regulatory code of content for the "harms" associated with social media platforms.

As platforms have no general obligation to monitor content, the White Paper also fails to address how it intends to comply with the e-Commerce Directive and corresponding case law on platform liability.[119] The Online Harms White Paper claims, without explanation, that a 'duty of care' will somehow increase compliance:

> 'The new regulatory framework will increase the responsibility of

[106] http://www.europarl.europa.eu/factsheets/en/sheet/7/the-principle-of-subsidiarity (last accessed 18 September 2019).
[107] *Sunday Times v United Kingdom* (Application no. 6538/74), 26 April 1979
[108] *Handyside v United Kingdom* (5493/72) at ¶ 49.
[109] UK, Copyright, Designs and Patents Act 1988 c. 48.
[110] UK, Defamation Act 2013, c. 26.
[111] See e.g. F S Siebert (1952), *Freedom of the Press in England, 1476–1776: The Rise and Decline of Government Controls* (Urbana: University of Illinois Press) or J Rowbottom (2015), 'Entick and Carrington, the Propaganda Wars and Liberty of the Press' in A Tomkins and P Scott (eds), *Entick v Carrington: 250 Years of the Rule of Law* (Oxford, Hart).
[112] J Rowbottom (2018), *Media Law* (Oxford, UK: Hart Publishing) 1-7.
[113] Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950; *Sunday Times v United Kingdom* (Application no. 6538/74), 26 April 1979
[114] *Handyside v United Kingdom* (5493/72) at ¶ 49.
[115] UN, Declaration of the High-level Meeting of the General Assembly on the Rule of Law at the National and International Levels (2012), ¶ 8 http://www.unrol.org/article.aspx?article_id=192; EU, Charter of Fundamental Rights of the EU (2009), Article 49 (concerning the principles of legality and proportionality of criminal offences and penalties); European Convention on Human Rights, in particular 6(1), 7, 8(2), 9(2), 10(2) and 11(2).
[116] ECtHR (1976) *Handyside v UK* (5493/72).
[117] Ibid ¶ 49.
[118] See e.g. *Google Spain, SL, Google Inc v Agencia Espanola de Proteccion de Datos* Case C-131/12 for search engine results; *Sunday Times v United Kingdom* (Application no. 6538/74), 26 April 1979 for news or *Mosley v United Kingdom* (Application no. 48009/08) 10 May 2011 for celebrity gossip as 'low level' speech.
[119] EU, Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), article 15; *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL* (SABAM) Case C-70/10; or the ECtHR case *Delfi AS v. Estonia* (Application no. 64569/09).

> online services in a way that is compatible with the EU's e-Commerce Directive, which limits their liability for illegal content until they have knowledge of its existence, and have failed to remove it from their services in good time'.[120]

However, 'intermediary liability' is not merely a creation of the e-commerce Directive. It may have been established by the e-Commerce Directive,[121] but it was implemented into UK law through the e-Commerce Regulations[122] and developed by subsequent case law in both the CJEU and UK Courts.[123]

This will have implications on the status of the law post brexit as well. Under Section 2 of the European Union (Withdrawal) Act 2018, the directives are not 'retained EU law', yet the domestic legislation that gives a directive effect in national law remains:

> 'Whereas other provisions of the e-Commerce Directive were implemented by the Electronic Commerce (EC Directive) Regulations 2002, article 15 was not specifically implemented through UK domestic legislation. Under section 2 of the European Union (Withdrawal) Act 2018 directives are not in themselves "retained EU law", only the domestic legislation made to implement them. However, under section 4 of the Act any prior obligations or restrictions of EU law which are "recognised and available in domestic law" will continue after Brexit. As article 15 has been recognised by domestic courts, including the Supreme Court in Cartier International AG and others v British Telecommunications Plc, it is likely to be considered retained law, **but uncertainty may remain until the matter is tested by the courts**'[124] [Emphasis Added]

The Directive provides a safe harbor for internet "hosts" (most of the companies the Government aims to regulate would fit into this category, including social media) and the protection from liability for illegal content stored on their platforms, provided that they do not have the *actual knowledge* about this content, and that they act promptly upon obtaining this knowledge.[125] The Directive prohibits the general monitoring of Internet users for the purpose of detecting such content.[126] There is extensive CJEU case law on the matter[127] as well as the related ECtHR jurisprudence on Articles 8 and 10 of the ECHR and the liability of Internet platforms.[128] The Government claims the new regime will be compatible with the Directive, but given the scope and requirements of duty of care, this is uncertain.[129] The prohibition of general monitoring will almost certainly have to be violated; it would be practically impossible to identify and remove all the potentially "harmful" content without monitoring activities of all the users on a certain platform.

6 Meet Ofweb, the UK's New Internet Overlord

The proposed model, placed on statutory footing, will be co-regulation with initial responsibility handed to Ofcom, the United Kingdom's broadcasting regulator, while a new regulator (provisionally called 'Ofweb') is established.[130]

The obvious questions that arises in a legal analysis is whether decisions made by Ofcom or Ofweb are subject to judicial review and what effects judicial review might have. In the UK, 'parliamentary sovereignty' ensures legislation cannot be reviewed by inferior courts; accordingly, this new regulatory system rooted in primary legislation cannot be subjected to judicial review. As the DCMS envisages that Ofcom's responsibilities will eventually be handed to Ofweb, judicial review can only have a role in most limited of circumstances; for example, codes of conduct and enforcement decisions. Furthermore, one can only raise proceedings in the UK on one of four grounds – (a) illegality, (b) procedural unfairness, (c) irrationality, or (d) incompatibility with human rights that are given effect by the Human Rights Act 1998. Therefore, the threshold for raising a judicial review against *any* specialist regulator in the UK is incredibly high. As a result of a failed petition for judicial review in *RT v Ofcom*[131], one commentator was prompted to note: "the judgment was an all-out win for Ofcom. It demonstrated yet again how difficult it is to succeed in a judicial challenge against the decision of a specialist regulator unless it has failed to comply with its own procedures or due process".[132]

Furthermore, judicial review should not be seen as a viable appeals process for decisions. In fact, one cannot apply to the courts if there is an open and valid appeals process that could have been followed, nor can judicial review overturn an earlier decision; it can only determine whether that decisions were illegal, improper, or irrational. Judicial review can only nullify the act (i.e. it never actually happened) rather than overturning it. A regulator like Ofweb would then be free to make the same exact decision having corrected for the earlier, for example, procedural error.

Admittingly, putting platform regulation on a co-regulatory framework has its benefits (e.g. stronger legitimacy than self-regulation, based on powers given by the Parliament, expertise, principle-based regulation, flexibility, cooperation with the industry),[133] yet there is a danger in Ofweb uncritically replicating the existing model of broadcast regulation, which has a very different historical rationale and justification, onto the Internet. Broadcast regulation affects entities who have access to scarce resources, such as spectrum,[134] who produce and distribute content at a large scale, and exercise editorial control with little user-created and/or generated content.[135] The Americans have also rejected this approach. In *ACLU v Reno*,[136] the US Supreme Court famously rejected the argument that regulating the Internet was

[120] Ibid ¶ 41.
[121] Directive 2000/31/EC.
[122] Regulations 17, 18, and 19 of the E-Commerce (EC Directive) Regulations 2002 SI 2002/2013.
[123] In the UK, for example, *Godfrey v Demon Internet Service* [1999] EWHC 244 (QB).
[124] House of Lords Communications Committee Report (n 18) 185.
[125] EU, Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), article 14.
[126] Ibid, article 15.
[127] For example, *Scarlet Extended SA v Societe Belge des Auteurs, Compositeurs et Editeurs SCRL* (SABAM) (C-70/10) [2011] E.C.R. I-11959 (24 November 2011).
[128] *Delfi AS v. Estonia* (Application no. 64569/09); *Tamiz v the United Kingdom* (Application no. 3877/14) ECHR (12 October 2017); *Magyar Jeti Zrt v. Hungary* (Application no. 11257/16), 4 December 2018.
[129] For example, *Tamiz v the United Kingdom* (Application no. 3877/14) [2017] ECHR (12 October 2017).
[130] Online Harms White Paper (n 1) ¶ 5.15.
[131] https://www.judiciary.uk/wp-content/uploads/2020/03/RT-v-Ofcom-approved-judgment-27.3.20.pdf.
[132] https://smab.co.uk/first-court-decision-ofcom-impartiality.
[133] See generally Christopher T. Marsden, 'Internet Co-Regulation and Constitutionalism: Towards a More Nuanced View' (August 29, 2011). Available at SSRN: https://ssrn.com/abstract=1973328 or Marsden (2011), Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace (Cambridge, UK: Cambridge University Press).
[134] The radio spectrum is the part of the electromagnetic spectrum, widely used in modern technology, particularly in telecommunications and broadcasting. Examples of its use include TV, radio, mobile internet etc. see Wikipedia, Radio spectrum, https://en.wikipedia.org/wiki/Radio_spectrum
[135] Rowbottom (n 112) 280-288.
[136] *Reno v. American Civil Liberties Union*, 521 U.S. 844 (1997).

justified on grounds that broadcasting platforms are heavily regulated for content.[137] The Court stated that regulating content was permitted in broadcasting contexts because viewers had little control over what they were exposed to; however, users have to take a series of affirmative action to access the online content they want to see.[138] In addition to this, users also produce different types of content in ways unimaginable for broadcast.

In a case before the European Court of Human Rights, the Court recognized the difficulty applying broadcasting codes to Internet platforms:

> It is true that the Internet is an information and communication tool particularly distinct from the printed media, especially as regards the capacity to store and transmit information. The electronic network, serving billions of users worldwide, is not and potentially will never be subject to the same regulations and control.[139]

The Court also recognizes that the risk of harm online is different to that of broadcast and press media.

> The *risk of harm* posed by content and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press. Therefore, the policies governing reproduction of material from the printed media and the Internet may differ. The latter undeniably have to be adjusted according to the technology's specific features in order to secure the protection and promotion of the rights and freedoms concerned.[140]

Note the Court's two concerns about the risk of harm by content and communications to first, the interference with other rights; and, second, that technology-specific features require adjustments to "secure the protection and promotion of the rights and freedoms concerned."[141]

Regulation over broadcasting has a much smaller impact than on individual speech. There are no scarce resources; however, there is peer-to-peer sharing, user-generated content, and individually created, but non-filtered speech. Furthermore, the press' self-regulatory model is a result of a long and exhausting struggle against historically regulated sectors like the press.[142] On the other hand, the Internet was founded on - and still largely embraces - the libertarian principle of openness.[143] Chapter Six of the House of Lords Communications Committee report suggests a new Parliamentary Joint Committee to ensure the regulator does not act on their own.[144] If implemented, there would be a tripartite regulatory relationship between 'Ofweb' (the regulator), the Government via the Cabinet Office and Parliament via a new Joint Committee. This is not independence; on the contrary, it is government using platforms as proxies to control the Internet.

Beyond the inappropriateness of using broadcasting's model to regulate the Internet, the intricacy of meeting "economically effective and socially just"[145] targets means that there is no uniformly accepted regulatory technique for digital technologies.[146] Rather, numerous possibilities exist within the categories of self-regulation, state regulation and "multi stakeholder co-regulation."[147] Each option carries advantages and disadvantages,[148] satisfying and undermining different notions of legitimacy,[149] such that implementation is fraught and "cynicism is at least partly justified."[150]

Asking a public authority to make specific rules can result in regulatory capture or a climate of resistance between the regulated and regulators and an impediment to higher performance.[151] Moreover, the strong intervention of public authority who represent the overall interests of the state may cause undue influence on the assessment through external factors. Both political and economic considerations can damage the advantages associated with the top-down model of platform regulation. There is also the risk of the most restrictive content laws becoming the norm across multiple platforms, regardless of the audience and user demographics. Schultz calls this the "slowest ship in the convoy problem"– the universal availability of information on the Internet might produce universal effects.[152] All platforms would have to comply with the most restrictive (i.e. the "slowest ship") standard.

7 Recommendations

Placing a duty of care on platforms for user-generated content that may cause harm will chill free expression and conflates the well-established common and statutory duty of care with clear duties and actual injuries. The least we could do is refer to the "duty" as "the duty to comply with existing regulation," or just maintain general terms of legal and regulatory obligations and duties. The government needs to reacquaint itself with the historical rationales for regulating broadcast (initially unregulated with increasing regulation amid scarce resources) or the press (initially heavily regulated with gradual deregulation and strengthening of media freedom).[153]

Whilst it is clear that there are problems with online platforms, their power and different harms that arise as a consequence,[154] most of the identified harms in the White Paper could be remedied with proper co-regulation of actors operating online, and enhanced obligations to cooperate with law enforcement over a variety of existing forms of

[137] Ibid at 845, 870.
[138] Idem at 854.
[139] ECtHR, Judgment of 5 May 2011, *Case of Editorial Board of Pravoye Delo and Shtekel v Ukraine,* (Application No. 33014/05) at ¶ ;See also Judgment of 16 June 2015, Case of *Delfi AS v Estonia* (Application no. 64569/09).
[140] Editorial Board at ¶ 36 (emphasis added).
[141] Editorial Board at ¶ 63.
[142] Rowbottom (n 111).
[143] John Perry Barlow, A Declaration of the Independence of Cyberspace, available at: https://www.eff.org/cyberspace-independence.
[144] House of Lords Communications Committee (n 18) chapter 6.
[145] Brown, I., & Marsden, C. T. (2013). *Regulating code: Good governance and better regulation in the information age.* (Boston: MIT Press) ix; See also Orla Lynskey (2015), *The Foundations of EU Data Protection Law* (Oxford: OUP) at Page 47.
[146] Brown & Marsden (n 145) 1; See also Terry Flew (2018), 'Technology and Trust: The Challenge of Regulating Digital Platforms' (Korean Association for Broadcasting and Telecommunications Studies) 9-11.
[147] Ibid at Page 2.
[148] Ibid at Page 2-3.
[149] Black, J. (2008). Constructing and contesting legitimacy and accountability in polycentric regulatory regimes. *Regulation & governance*, 2(2), 137-164 at Page 145.
[150] Brown & Marsden (n 145) 3.
[151] Baldwin R, Cave M, Lodge M. (2012) *Understanding regulation: theory, strategy, and practice.* (Oxford University Press on Demand)108-110.
[152] Schultz, T. (2008). Carving up the Internet: jurisdiction, legal orders, and the private/public international law interface. *European Journal of International Law*, 19(4), 799-839 at 813 citing Zittrain, ' Be Careful What You Ask For: Reconciling a Global Internet and Local Law', in A. *Thierer and C.W. Crews (eds) (2013), Who Rules the Net? Internet Governance and Jurisdiction* (Cato Institute) 17.
[153] Rowbottom (n 153) 2 -5, 256 – 288.
[154] As the EU recognises and addresses in the ongoing attempt to reform platform liability, inter alia. See e.g. European Commission, 'Shaping Europe's digital future' (COM (2020)0067), 19 February 2020.

criminal speech and behavior. Because of the extent of the impact of regulation on the digital rights of users, judicial oversight is crucial, and any regulator should be independent with pathways for judicial remedies and reviews. Furthermore, it is insufficient to base platform regulation on a handful of user submissions and surveys. Although Ofcom's annual survey is widely cited throughout, albeit quite selectively, any additional harms subject to further regulation need to be based on clear and unambiguous evidence.[155] As our analysis shows, the concept of "online harms" is vague and it should be dropped entirely. Any additional harm criminalized in the future needs to be clearly defined, well evidenced and regulated in the public interest.

Additionally, it is suggested that companies do not rely on technology solely, but human oversight should also be a requirement wherever there are takedown procedures in place. Automated systems and AI are not reliable enough to be used alone, as we have seen in the case of the YouTube Content ID system[156] and the likelihood of errors.[157] Speech assessment includes qualitative questions on whether content should be treated differently to information offline for every individual user (*the parity principle*, for example).[158] For this to happen, the platform will need to understand the context of exchanges between every user on a platform and how people communicate offline with one another.[159] Different platforms have different social norms and communication practices and this should be respected (e.g. it is not realistic to expect the same language on 4Chan, Reddit, and Mumsnet).

Using technology to search for fake news is potentially problematic with false positives potential affecting media pluralism.[160] In *Jersild v Denmark*,[161] the court stated that "the methods of objective and balanced reporting may vary considerably, depending among other things on the media in question."[162] Second, Article 10 ECHR "protects not only the substance of the ideas and information expressed, but also the *form* in which they are conveyed."[163] The observation that "the methods of objective and balanced reporting may vary considerably"[164] takes on increased importance in contemporary times. In the current "post-truth" era, fake news, misinformation and disinformation are widely generated and disseminated by a range of actors (and algorithmic techniques) and they compete fiercely with one another for the public's attention and acceptance.

More generally, the White Paper also lacks the clarity necessary in law. The regulatory framework must be accessible: users "must be given an indication that is adequate in the circumstances of the legal rules applicable to a given case."[165] Secondly, users must be able to moderate their behavior in line with what is reasonably foreseeable.[166] Users "must be able - if need be with appropriate advice - to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail."[167] The White Paper's proposed framework is vague and insufficient and lacks the clarity to "give individuals an adequate indication as to the circumstances in which and the conditions on which the authorities are empowered to resort" to any such measures.[168]

The White Paper falls short by not properly considering alternatives to its proposed measures. We point out some alternatives below with the purpose of demonstrating that alternative recommendations would be conceivable, rather than attempting to develop these fully.

1. **Reform Intermediary liability**. An alternative way to respond to the "online harms" identified in the White Paper is reformation of the liability provisions of the e-Commerce Directive, in line with the Regulations adopted in the last mandate of the European Commission. The general principle of a harmonized, graduated, and conditional exemption continues to be needed as a foundational principle of the Internet. The principle, however, needs to be updated and reinforced to reflect the nature of the services in use today. This could mean that the notions of mere conduit, caching and hosting service could be expanded to explicitly include other services. In some instances, this can amount to codifying existing case law (e.g. for search engines or Wi-Fi hotspots), while in other cases a clarification of its application to collaborative economy services, cloud services, content delivery networks, domain name services, etc. is necessary. Building on concepts like editorial responsibility[169], actual knowledge[170] and degree of control[171], the concept of active/passive hosts should be replaced by more appropriate concepts that reflect the technical reality of today's services.[172]

2. **General monitoring and automated filtering**. While the prohibition of general monitoring obligations should be maintained as another foundational cornerstone of Internet regulation, specific

[155] For a good example of how evidence submitted to the Committee has been bastardized to make a political point, see Goldman, Eric, The U.K. Online Harms White Paper and the Internet's Cable-ized Future (2019). *Ohio State Tech. L.J.*, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3438530 at Page 2.

[156] YouTube, 'How Content ID Works' https://support.google.com/youtube/answer/2797370?hl=en; J Bailey, 'YouTube Beta Testing Content ID For Everyone' (*Plagiarism Today*, 2 May 2018) https://www.plagiarismtoday.com/2018/05/02/youtube-beta-testing-content-id-for-everyone/.

[157] J M Urban, J Karaganis, and B Schofield,'Notice and Takedown in Everyday Practice', *UC Berkeley Public Law Research Paper* No. 2755628. http://dx.doi.org/10.2139/ssrn.2755628.

[158] Online Harms White Paper (n 1)

[159] This is virtually impossible, see Banerjee, S., Chua, A. Y., & Kim, J. J. (2017). Don't be deceived: Using linguistic analysis to learn how to discern online review authenticity. *Journal of the Association for Information Science and Technology*, 68(6), 1525-1538.

[160] Heins, M., & Beckles, T. (2005). *Will fair use survive? Free expression in the age of copyright control*. Marjorie Heins.

[161] *Jersild v Denmark*, ECtHR 23 September 1994 (GC), ECLI:CE:ECHR:1994:-0923JUD001589089, Series A no. 298.

[162] Ibid ¶ 31.

[163] ECtHR, *Autronic AG v Switzerland* (1990) 12 EHRR 485, [47]

[164] *Bladet Tromso and Stensaas v Norway*, 20 May 1999, [59].

[165] ECtHR The Sunday Times v. the United Kingdom (No. 1), 6538/74, 26 April 1979 at ¶ 49.

[166] *Rekvényi v Hungary*, 25390/94, 20 May 1999, At ¶ 34f.

[167] ECtHR *The Sunday Times v. the United Kingdom* (No. 1), 6538/74, 26 April 1979 at ¶ 49.

[168] Ibid, at at ¶ 49; See also *Malone v United Kingdom* Application no. 8691/79, 2 August 1984 at [67].

[169] For example, See The New Definitions Of "Audiovisual Media Service" (Article (1)(a)(i)) and "Video-Sharing Platform Service" (Article (1)(b)(aa) and Article 1(D)(Bb) of Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities.

[170] Note 123, supra.

[171] *L'Oreal v eBay* Case C324/09, 12 July 2011 at ¶ 116, 123, 145; See also Article 14(2), e-Commerce Directive

[172] For some other approaches see, for instance: C. Angelopoulos C. and S. Smet, 'Notice-and-fair-balance: how to reach a compromise between fundamental rights in European intermediary liability', (2016) *Journal of Media Law*, 8(2); S. Stalla-Bourdillon (2017), 'Internet Intermediaries as Responsible Actors? Why It Is Time to Rethink the Ecommerce Directive as Well', in Taddeo M., Floridi L. (eds), *The Responsibilities of Online Service Providers*. Law, Governance and Technology Series, vol 31. (Heidelberg etc: Springer).

provisions governing algorithms for automated filtering technologies - where these are used - should be considered, to provide the necessary transparency and accountability of automated content moderation systems.

3. **Regulating content moderation**. Uniform rules for the removal of illegal content like illegal hate speech should be made binding across the EU. Replacing notice-and-takedown with notice-and-action rules could be tailored to the types of services, e.g. whether the service is a social network, a mere conduit, or a collaborative economy service, and where necessary to the types of content in question, while maintaining the maximum simplicity of rules. The feasibility of introducing thresholds could be examined in this context, taking due account of the size and nature of the service provider and of the nature of the potential obligations to be imposed on them. Building on the Recommendation on Illegal Content,[173] binding transparency obligations would also be at the heart of a more effective accountability framework for content moderation at scale and would complement recently adopted rules under the Audiovisual Media Services Directive[174] or the modernization of the EU copyright rules.[175] Increasing transparency for algorithmic recommendation systems of public relevance like social media news feeds should be examined. At the same time, these rules should prohibit allowing Member States to impose parallel transparency obligations at national level, providing for a simple set of rules that comply with the Manila principles[176] on content moderation and intermediary liability in the European Union.

8 Conclusions: Broad and Flawed

The Internet is not a "safe space," nor was it intended to be. Without a doubt, the Internet is a complicated space; however, it also makes us look at humankind's most unsavory characteristics in a way never imagined before. Accordingly, we should look at platforms as a blessing, not a burden. How else could we know that so many people think like us at the same time as hold such divergent, even abhorrent views? Yet the White Paper goes beyond turning the Internet into a virtual soft play area where everyone has to watch what they say, what they do, and how they act. It burdens the platform with a duty of care to police the speech of its patrons, under the threat of sanctions for what might be offensive or intimidating and might cause harm. This is a prime example of the "chilling effects" of content moderation. Furthermore, platforms have undertaken significant self-regulatory responses to mitigate the threat of co-regulation. For example, Facebook launched an Independent Oversight Board and charter for content moderation on its site.[177]

In what feels like ancient history, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, warned in 2011 against the adverse effects that disproportionate regulation of content might have on free speech.[178] To address this, he recommends:

> States should only seek to restrict content pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy. States should refrain from imposing disproportionate sanctions, whether heavy fines or imprisonment, on Internet intermediaries, given their significant chilling effect on freedom of expression.[179]

Furthermore, there is little evidence that the data published in the Online Harms White Paper or the House of Lords Communications Committee will change anyone's opinion or behavior. Opinion on platform regulation will always remain divided along deeply held beliefs about the constitutional merits of criminal prohibitions in areas like hate speech and the role of private actors in content regulation. Rather than assuming every view different from our own results in harm and place a burden on online services to remove them, we need to develop techniques and strategies for defeating ideology through competition in the marketplace of ideas. With so many special interests competing with each other for the attention of lawmakers, each with their own agenda in protecting identifiable stakeholders, regulation should be forward-thinking and dynamic, and protect the principles of free expression and media pluralism, rather than take action to inhibit and control. In hindsight, it is quite surprising that there was not more emphasis on enhanced cooperation between platforms and law enforcement.

The Internet is for expression – it is for argument, emotion, anger, purchasing, love, sex, and sharing. Expression is its bread and butter. All of the above, of course, come with negative consequences. Arguments can turn into violence, emotions can run high and lead to regret, anger can cause permanent damage, love can turn to heartbreak, sex can lead to objectification and pain, and sharing can be a violation of someone else's rights. We are already well-equipped to deal with this through different forms of online offences such as harassment, revenge porn and other communication offences. Some of these, as suggested by the Law Commission, should be reviewed and consolidated,[180] but this will be dealt with through criminal law reforms, and not the vaguely imposed duty of care that threatens fundamental rights online. Behind all of this is the harm associated with regulatory capture, the protectionist mindset of 'something must be done,' and the problem of regulating the wrong actors.

Imposing a duty of care on platforms inadvertently creates a framework for crushing dissent, plurality, diversity, "British values,"[181] and

[173] Commission Recommendation on measures to effectively tackle illegal content online, Commission Recommendation of 1.3.2018 on measures to effectively tackle illegal content online (C(2018) 1177 final), Available at https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online, (last accessed 19 September 2019).

[174] Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services.

[175] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

[176] Manila Principles on Intermediary Liability Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation, Available at https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online (last accessed 19 September 2019).

[177] Establishing Structure and Governance for an Independent Oversight Board, Available at https://newsroom.fb.com/news/2019/09/oversight-board-structure (last accessed 19 September 2019).

[178] HRC, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 06 April 2019, available at: https://documents-dds- ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf

[179] Ibid 19.

[180] Note 76, Supra 328- 334.

[181] UK, Department for Education, Promoting fundamental British values as part of SMSC in schools, Departmental advice for maintained schools, November 2014, at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/380595/SMSC_Guidance_

ultimately, free speech. Populism, is, by its very definition,[182] wedded to the preservation of the status quo. That said, it is surely a wonderful thing that, for all its faults, there is at least one remaining space in our culture where words still matter and where promises made in the form of written undertakings ("laws") have consequences. However, for the Internet, the trick is getting the law right. A society that stops being governed by the authority and rule of law and reverts to that of the "populist," the priest, or "the people" is not a place where freedom, openness and democracy will long survive. It seems a long way from Frank La Rue, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2011 Report to the UN[183] when he stated:

> The Special Rapporteur believes that censorship measures should never be delegated to a private entity, and that no one should be held liable for content on the Internet of which they are not the author. Indeed, no State should use or force intermediaries to undertake censorship on its behalf.[184] Of concern, Subject to abuse by state and private entities; Risk of liability causes intermediary to err on the side of taking content down; Lack of transparency on decision making practices obscures discriminatory practices or political pressure affecting their decisions; and companies shouldn't be making the assessment of legality of content.[185]

More recently, Catalina Botero Marino strongly endorsed transparency in her 2013 report, stating:

> [w]ith respect to the duty of transparency, intermediaries should have sufficient protection to disclose the requests received from government agencies or other legally authorized actors who infringe upon users' rights to freedom of expression or privacy. It is good practice, in this respect, for companies to regularly publish transparency reports in which they disclose at least the number and type of the request that could lead to the restrictions to users' rights to freedom of expression or privacy.[186]

A flat-earther that has been called an idiot or an imbecile could have a claim of "abuse" and/or intimidation. Empowering users might be a noble objective, but that requires empowering the *right* users and educating everyone.

Maintained_Schools.pdf.

[182] David Molloy, What is populism, and what does the term actually mean?, (*BBC News*, 6 March 2018), https://www.bbc.co.uk/news/world-43301423.

[183] Frank La Rue, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression Human Rights Council, Seventeenth session, Agenda item 3, Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development, Available at https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf (last accessed 11 September 2019).

[184] Ibid ¶ 43.

[185] Ibid ¶ 42.

[186] IACHR Office of the Special Rapporteur for Freedom of Expression (OSRFE), Freedom of Expression and the Internet, (Dec. 31, 2013), Available at https://www.oas.org/en/iachr/expression/docs/reports/2014_04_08_internet_eng%20_web.pdf ¶ 113.

09

Technology and Regulation

Are cookie banners indeed compliant with the law?

Deciphering EU legal requirements on consent and technical means to verify compliance of cookie banners

Cristiana Santos*, Nataliia Bielova** and Célestin Matte***

consent, cookie banners, GDPR, ePrivacy Directive, web tracking technologies

c.teixeirasantos@uu.nl

nataliia.bielova@inria.fr

celestin.matte@cmatte.me

In this paper, we describe how cookie banners, as a consent mechanism in web applications, should be designed and implemented to be compliant with the ePrivacy Directive and the GDPR, defining 22 legal requirements. While some are provided by legal sources, others result from the domain expertise of computer scientists. We perform a technical assessment of whether technical (with computer science tools), manual (with a human operator) or user studies verification is needed. We show that it is not possible to assess legal compliance for the majority of requirements because of the current architecture of the web. With this approach, we aim to support policy makers assessing compliance in cookie banners, especially under the current revision of the EU ePrivacy framework.

1. Introduction

The ePrivacy Directive[1] 2002/58/EC, as amended by Directive 2009/136/EC, stipulates the need for consent for the storage of or access to cookies (and any tracking technology, e.g. device fingerprinting) on the user's terminal equipment, as the lawfulness ground, pursuant to Article 5(3) thereof. The rationale behind this obligation aims to give users control of their data. Hence, website publishers processing personal data are duty-bound to collect consent. Consequently, an increasing number of websites now display (cookie) consent banners.[2]

However, there is no established canonical form for the consent request. It is clear from Recital 17 of the ePrivacy Directive (hereinafter ePD) that a user's consent may be given by any appropriate method. Website operators are free to use or develop consent flows that suit their organization, as long as this consent can be deemed valid under EU legislation.[3] [4] As such, excessive focus is being placed on the manufacturing of consent, taken up by consent management platforms and tools. The most well-known way to collect consent is through "cookie banners", also often referred to as *prompts*, *overlays*, *cookie bars*, or *cookie pop-up-boxes* that pop up or slide atop websites prominently.[5] Their design and functionality differ – the simplest banners merely state that the website uses cookies without any option, whereas the most complex ones allow users to individually (de)select each third-party service used by the website.

Amid information overload and the development of manipulative dark patterns[6] [7] [8] that lead to nudging users to consent, data subjects are

[1] In this paper we will only regard to the recent amended version of the ePrivacy Directive, the Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws (Text with EEA relevance) OJ L 337, 11–36 (hereinafter named "ePD") .

[2] Jannick Sørensen, Sokol Kosta (2019), "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites", *Proceedings of the World Wide Web Conference*, ACM, NY, USA, 1590–1600.

[3] In this paper, we provide many excerpts of the opinions and guidelines of the Article 29 Working Party. For readability and presentation purposes, we convey in the text of the article the abbreviation "29WP", followed by the reference number of each opinion. Even if the European Data Protection Board has endorsed the endorsed the GDPR related WP29 Guidelines, for simplicity purposes, we only mention Article 29 Working Party.

[4] Article 29 Working Party, "Guidelines on consent under Regulation 2016/679" (WP259 rev.01, 10 April 2018).

[5] For example, the French DPA (henceforth named CNIL) decided to remove its cookie banner and to leave no tracer until the user has consented by going actively to the cookie management menu or directly through the content pages. This choice not to use a banner is neither an obligation nor a recommendation for other websites that are free to adopt solutions tailored to their situation, in compliance with Regulations, CNIL (2019), "The legal framework relating to consent has evolved, and so does the website of the CNIL" www.cnil.fr/en/legal-framework-relating-consent-has-evolved-and-so-does-website-cnil accessed 7 May 2020.

[6] Harry Brignull, "What are Dark Patterns?" (2018) https://darkpatterns.org accessed 7 May 2020.

[7] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs (2018), "The Dark (Patterns) Side of UX Design" *Proceedings of the CHI Conference on Human Factors in Computing Systems*, ACM, New York, USA.

[8] CNIL's 6th Innovation and Foresight Report "Shaping Choices in the Digital World, "From dark patterns to data protection: the influence of UX/UI design on user empowerment" (2019) https://linc.cnil.fr/fr/ip-report-shap-

* Cristiana Santos is lecturer and researcher at Utrecht University, the Netherlands.

** Nataliia Bielova is a Research Scientist at PRIVATICS team in Inria, France.

*** Célestin Matte is independent researcher.

Received 15 Dec 2019, Accepted 12 Oct 2020, Published 27 Nov 2020.

not always able to easily understand the outcomes of data collection, and the use of their data.

The assessment as to whether or not cookie banner designs implemented by website operators fulfil all the requirements for valid consent, as stipulated by the General Data Protection Regulation[9] (hereinafter named GDPR), is considered in the guidelines of both the Article 29 Working Party and Data Protection Authorities (hereinafter named 29WP and DPAs). These guidelines provide a useful framework of *what* is a valid consent for cookie banners, but they do not define *how to assess, in practice, their legal compliance*. Though these guidelines have an important interpretative value, in concrete settings, they offer still vague guidance on the consent implementation. Even though Recital 66 of the ePD disposes that "the enforcement of these requirements should be made more effective by way of enhanced powers granted to the relevant National Authorities", this point is still under work, despite the recent guidelines issued by various DPAs. The legislative provisions in the GDPR are purposefully general to cover a range of different scenarios, including unanticipated future developments. The ePD does not sketch procedures to guide the enforcement of its principles, nor provides guidelines to perform systematic audits. Moreover, the lack of automatic tools which can verify whether a website violates the legislative instruments possibly makes it complicated for the deputed agencies to plan systematic audits.

The consequence of not complying with the requirements for a valid consent renders the consent invalid and the controller may be in breach of Article 6 of the GDPR. Hence, the controller may be subject to fines (Article 83).[10]

We consider in this work that there is a need for a technical perspective in the analysis of a valid consent for browser-based tracking technologies (including cookies), as processing operations of web services are *technology intensive*. This means that the use of the technology underlying processing operations is such, that specific guidance on the use of that technology is needed to adequately protect personal data while managing cookies on the server side, the third-party side, and also on the side of designers and/or developers of websites. We state that a *privacy by design* approach, as posited in Article 25 of the GDPR, advocates good technical design which embeds privacy into IT systems and business practices from the outset (and does not just add privacy measures ex-post).

Our **aim** is to identify the requirements for a valid consent to assess compliance of cookie banners, preparing the ground for compliance of cookie consent banners to be automated. Therefore, our final goal is to evaluate to which extent Web Privacy Measurements (WPM) are capable of assessing compliance automatically. To ensure automatic WPM, we need to rely on a combination of law, policy and technology areas to operationalize requirements for consent. Hence, our intention is to contribute to closing the gap between existing legal guidelines and interpretations and technical solutions for consent banners to discern compliant banner designs and to spot invalid ones. This analysis can be useful to compliance officers, regulators, privacy NGOs, law and computer science researchers, web services business owners and other services concerned with the design or operation of web services.

This paper makes the following contributions:

- We identify 22 legal-technical requirements for a valid consent of cookie;
- We show how the 22 requirements for valid consent can be used in practice when performing a compliance audit of consent request (i.e. the banner design).
- We explore to what extent automated consent verification is possible.

We conclude that a fully automatic consent verification by technical means is not possible because the majority of the low-level requirements either require manual inspection, can be evaluated with technical tools only partially, or must be evaluated with user studies to assess users' perceptions and experience with the website's consent implementation.

The remainder of the paper is as follows. Section 2 describes the methodology adopted to construe the requirements for a valid consent for consent banners. Section 3 provides the background knowledge of the paper. Section 4 discusses the scope of browser tracking technologies and analyzes which purposes are subject to the legal basis of consent. Section 5 expounds on each of the requirements and low-level requirements for a valid consent for consent banners, providing compliant and non-compliant examples and the means to verify compliance. Section 6 summarizes different technical solutions that could be applied to detect violations of requirements that depend on natural-language processing and user perception. Section 7 discusses scenarios and consequences of a shared consent. Section 7 discusses a recent draft of the ePrivacy Regulation. Section 8 opens a discussion on the upcoming ePrivacy Regulation. Section 9 compares our work with related work in the area of consent to browser tracking technologies. Section 10 concludes the paper.

2. Methodology

This section presents the methodology used in our work. We propose a methodology based on two steps: a legal analysis (Section 2.1), and a technical analysis (Section 2.2). The definition of requirements emerged from joint interdisciplinary work composed of law and computer science experts in the domains of Data Protection Law and Web Tracking Technologies. The combined expertise was conducive to inspect legal and technical effects and the practical implementation of each requirement.

2.1 On the legal analysis

Bottom-up approach. In our work, we follow a bottom-up approach, using granular content from the elicited legal sources to build the devised requirements. First, we analyzed consent elements separately for general consent, and afterwards, we delved into the specificities of consent dedicated to browser-based tracking technologies (henceforth named BTT), including cookies.

ing-choices-digital-world accessed 7 May 2020.

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1 (hereafter, "GDPR").

[10] The German DPA acknowledges that if consent is required but not effectively granted, the setting or reading of a cookie is unlawful and data controllers face both the prohibition of data processing and fines, LfDI Baden-Württemberg (2019), "On the use of cookies and cookie banners - what must be done with consent (ECJ ruling "Planet49")?" www.baden-wuerttemberg.datenschutz.de/zum-einsatz-von-cookies-und-cookie-bannern-was-gilt-es-bei-einwilligungen-zu-tun-eugh-urteil-planet49/, and "Guidelines for Telemedia Providers", (2019) www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2019/04/Orientierungshilfe-der-Aufsichtsbeh%C3%B6rden-f%C3%BCr-Anbieter-von-Telemedien.pdf accessed 21 November 2019.

Legal sources. We have included in our analysis the following sources:

- legislation: GDPR and ePrivacy Directive (ePD)
- regulatory overview of decisions issued by the European Court of Justice of the EU (CJEU)
- DPA decisions on the use of cookies
- DPA guidelines on the use of cookies
- related works by legal scholars regarding some requirements

Criterion to elicit requirements based on legal effect. By analyzing the legal sources listed above, we have defined the requirements for a valid consent considering their *legal effect*. First, we extract requirements from legal sources with a *binding effect*, which can render legal certainty and predictability which happens with the GDPR and case-law from the CJEU. Table 1 depicts the legal source according to its legal effect.

Table 1 Legal sources according to its legal effect

Legal source	Type	Legal effect
Legislation	**GDPR**	Binding
Case law	**CJEU case law**	Binding
Guidelines	**EDPB**	Non-binding, interpretative effect. These contain persuasive authority, which means that the court is not required to follow the analysis
	DPA guidelines	

We now present requirements with binding and non-binding legal effect that we rely on in this work.

2.1.1 Requirements coming from binding sources.

Standard requirements from GDPR, ePrivacy Directive and CJEU. We include the four cumulative elements for a valid consent prescribed by Articles 4(11), 7(4) of the GDPR which amount to: freely given, specific, informed and unambiguous consent. We also include the requirement of an informed consent mandated in Article 5(3) of the ePD. We consider the Planet 49 ruling of the CJEU.[11] Table 2 shows the binding requirements coming from the GDPR, ePD and CJEU.

Table 2 Standard requirements for a valid consent from binding sources: GDPR Article 4(11) and 7(4), ePD Article 5(3) and CJEU

High-level requirements	Provenance
• Freely given	• Article 4(11) of GDPR • Article 7(4) of GDPR
• Specific	• Article 4(11) of GDPR • CJEU Planet 49
• Informed	• Article 4(11) of GDPR • CJEU Planet 49 • Article 5(3) of ePD
• Unambiguous	• Article 4(11) of GDPR • CJEU Planet 49

GDPR Articles 6 and 7. Besides these mentioned elements, we make salient and autonomous three other requirements:

- Prior
- Readable and accessible
- Revocable

These elements are mentioned in the GDPR, though they are not part of the definitional elements of Article 4(11). However, these three additional requirements, depicted in Article 7, are called as *conditions for validity* of consent, and are also meaningful to be considered for their practical effects in the online environment of consent banners (as explained throughout the paper). In this line, the 29WP (WP 259 rev.01) refers that the GDPR introduces requirements (beyond Article 4(11)) for controllers to make additional arrangements to ensure they obtain and are able to demonstrate valid consent. It refers to Article 7 which sets out these additional conditions of validity for valid consent, with specific provisions on keeping records of consent and the right to easily withdraw consent. In this regard, the Advocate General Spuznar[12] contends that the purpose of Article 7(1) of Regulation 2016/679 requires a *broad interpretation* in that the controller must not only prove that the data subject has given his or her consent but must also prove that *all the conditions for effectiveness* have been met, hence, expressing the practical side of the conditions of Article 7. Table 3 extends Table 2 and depicts these three added high-level requirements and their respective provenance.

Table 3 Additional requirements for valid consent from binding sources: GDPR Articles 6 and 7

High-level requirement	Provenance in the GDPR
• Prior	• Article 6, by the wording "has given" consent
• Readable and accessible	• Article 7 (2) "conditions for consent" • Recitals 32, 42
• Revocable	• Article 7 (3) "conditions for consent"

2.1.2 Requirements coming from non-binding sources

EDPB guidelines. Whenever there is no binding rule, we resort to the agreed-upon and harmonized guidelines coming from the EDPB – as a new EU decision-making body building on the work of the 29 Working Party (29WP). The EDPB has adopted various opinions and guidelines to clarify fundamental provisions of the GDPR and to ensure consistency in the application of the GDPR by DPAs.

DPA guidelines. We also resort to the guidelines of DPAs on consent for browser-based tracking technologies ("BTT"). We give a comparative analysis of the DPA guidelines. The usefulness of these guidelines is twofold:

- they connect to the legal requirements implemented at national level in the light of the GDPR standard requirements for consent,
- they incorporate the recent binding requirements coming from the CJEU.

In the comparative analysis of the existing DPA guidelines on the use of BTT, we discuss to what extent our 22 low-level requirements (see Table 6) are reflected in these guidelines, and where there are diver-

[11] cf. Planet49 Judgment (n 87) Verbraucherzentrale Bundesverband v. Planet49, Case C-673/17, [2019] OJ C 112 (ECLI:EU:C:2019:801) para 75.

[12] Case C61/19 *Orange România SA v ANSPDCP*, 4 March 2020, (ECLI:EU:C:2020:158) http://curia.europa.eu/juris/document/document.jsf?text=&docid=224083&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&=1&cid=272271#Footnote31

gences. We only assess the guidelines under the specific criterion of being comprehensive. For example, the Italian DPA simply presents in its website Frequently Asked Questions (FAQs) on cookies along with basic information and does not specify the requirements for trackers. Even though the Finnish DPA[13] issued guidelines on cookies, it was not possible to analyze even the high-level requirements, while the CNIL, the ICO, the Irish and Greek guidelines provide a more detailed description on each of the GDPR requirements. We have selected these eight guidelines that offer a wide coverage and reasoning upon the requirements for a valid consent: UK, French, Irish, German, Belgium, Danish, Greek, Spanish.

Table 4 extends Table 3, now depicting non-binding sources, thereby consolidating the assumed requirements.

Table 4 Additional requirements for a valid consent from non-binding sources: EDPB (WP29) and DPA guidelines

High-level requirements	Provenance from other sources
Prior	• 29WP on Consent • Article 2 of CNIL Guidance for cookies, 2019 • DPAs: Finnish, German Guidelines
Readable and accessible	• 29WP Guidelines on Transparency; DPAs: Belgium, Spanish, French, ICO
Revocable	• 29WP on Consent • Article 2 CNIL Guidance for cookies, 2019 • CNIL Recommendation for cookies, 2020 • DPAs: French, Greek, Irish, Danish, Spanish, German, Belgium

2.1.3 Requirements coming from our own legal interpretation

We have pursued with our own interpretation regarding some requirements, demarking our explicit positioning. Concretely, we have proposed three new requirements:

- "configurable consent banner" (R12),
- "balanced choice" (explained in R13),
- "no consent wall" (explained in R20).

Additionally, we propose five other low-level technical requirements, as a result of our technical analysis of consent banners. See Section 2.2 for details.

List of requirements. We assert that the complete list of 22 low-level requirements (see Table 6 of Section 5) derived from the high-level requirements presented in this section is exhaustive from a legal perspective. However, given that technologies are constantly evolving, we do not guarantee the exhaustiveness of the low-level technical requirements. To ensure legal compliance today, a consent banner implementation must meet all low-level requirements derived from binding legal sources. However, we strongly encourage that such implementations also comply with other, non-binding requirements presented in this paper. We believe all requirements should become mandatory and binding in the near future.

Presentation of requirements. Whilst deciphering each high-level requirement and the respective low-level requirements, we propose a description thereof, consisting of a concise designation of a requirement (e.g. "Prior to setting cookies"), and followed by its concrete and objective explanation (e.g. consent must be obtained before cookies requiring consent are set). For readability purposes, we additionally extend this description, whenever possible, with further observations.

2.1.4 On the national implementation of the ePrivacy Directive

Implementation of the ePD at national level. Although the ePD stipulates the need for consent for the storage of and/or access to cookies, the practical implementations of the legal requirements vary among website operators across EU Member States (MSs)[14]. Fragmented transpositions of the ePD at national levels create problems and legal uncertainty for European citizens as well as for the digital single market. Interestingly, 25 MSs have not fully updated ePD since the GDPR came into force. On the 30 January 2020, the Commission[15] reported that only three Member States "seem to have properly adapted the provisions" of the ePD following the entry into application of the GDPR. On the 5th of May, the Commission mentioned[16] that it asked MSs for information regarding the implementation of certain provisions. It stated it is still assessing the situation and will take a decision on appropriate measures in accordance with the Treaties. In this line, we do not study the differences of the ePD implementation in each MS due to lack of linguistic skills, volume constraints and also due to the fact that some legislations are quite old and do not contemplate neither the GDPR standard of consent, nor the recent CJEU jurisprudence.

2.2 On the technical analysis

Technical assessment of legal requirements. A technical analysis is performed of the legal requirements by computer scientists – co-authors of the paper, experts in Web Tracking Technologies. They evaluate how each requirement translates into practice for the technology of consent banners. They notably reflect on whether the GDPR requirements are compatible with existing web technologies, as it is not always the case, and technologies need to be adapted. For instance, the "prior to sending cookies" requirement cannot be enforced with the current state of technologies, because unless configured otherwise, cookies are sent automatically by browsers in every request.

Checking consent implementation on websites. We proceed with an empirical step of visiting example websites where a given low-level requirement is respected or violated and investigating the consent banner implementation.

Requirements coming from technical computer science analysis. Other requirements resulted from the domain-expertise of computer scientists. The four technical requirements are the following:

- R2 prior to sending an identifier,
- R14 post-consent registration,
- R15 correct consent registration,
- R22 delete "consent cookie" and communicate to third parties.

For example, we explain how the GDPR's requirement for "revocable

[13] Finish DPA (201), "Guidance on Confidential Communications" www.kyberturvallisuuskeskus.fi/fi/toimintamme/saantely-ja-valvonta/luottamuksellinen-viestinta, accessed 7 May 2020 (our translation).

[14] For example, Germany has not implemented the ePrivacy Directive, though it has formulated guidelines on the use of trackers.

[15] European Parliament, "Implementation of the ePrivacy Directive following the entry into application of the GDPR" (2020) https://www.europarl.europa.eu/doceo/document/E-9-2020-000790_EN.html accessed on 18 June 2020.

[16] Idem.

consent" could be implemented in practice: when consent is revoked, the publisher should delete the consent cookie and communicate the withdrawal to all third parties who have previously received consent. This operation implies the emergence of a new technical requirement: R22 "Delete "consent cookie" and communicate to third parties".

Procedure to verify compliance. For each requirement, we analyzed whether compliance thereto can be verified with technical means, using existing computer science tools or where feasible with state-of-the art technologies. If no such tools exist or seem feasible, we analyzed how each requirement can be manually verified. As a result, for each requirement, we have identified whether its violation can be detected:

- Technically, by an expert using computer tools
- Manually, relying only on a human operator, or
- Performing user studies to evaluate perception of end users.

Additionally, for each requirement where technical means are not possible today, we analyzed which upcoming technologies and possible technical solutions could be implemented.

2.3 Exclusions from this work

Explicit consent. We have excluded the requirement of explicit consent which is required whenever websites deal with: i) special categories of data (listed in Article 9 of the GDPR); ii) data transfers to third countries; and iii) automated decision-making (including profiling). As this requirement should contain a double-layer verification approach following the recommendation by the 29WP (since ticking one box or pressing one button is not enough to ensure an affirmative and explicit act) we decided not to contemplate this added layer verification effort.

Freely given. In the analysis of the element of a freely given consent, we did not consider the cases of *imbalance of power* (Recital 43 of the GDPR) for the same motive as above. This is mostly observed in the context of a public authority, employer, medical service relationship, or wherever there is a dominant position in relation to the data subject. In such contexts, the data subject fearing adverse consequences has no realistic alternative to accept the processing terms.

Informed consent. While considering the information necessary for an informed consent, we excluded the analysis of the purposes of an informed consent, meaning that we do not analyze the meaning of the purposes presented in the cookie banners. We state that in the information page, each purpose should be sufficiently unambiguous and clearly expressed, specific and clear. We nevertheless in general address intelligible and clear expression of information in the "Readable and Accessible" high-level requirement.

Browser settings. This paper does not analyze consent expressed through browser settings. We think that browser settings, as they exist today, do not correspond to the requirements of a valid consent for the following reasons: (a) no purposes are specified; (b) they do not reflect an informed decision; and (c) browser settings do not express an unambiguous consent. The 29WP[17] mentions that browser settings may be considered as a mechanism for expressing consent if they are clearly presented to the user. We do not agree with this statement for the reason that many browser vendors expose cookie settings in browser preferences that are hard to find. Moreover, the location and user interface of such cookie settings changes significantly from one version of the browser to another. Even though cookie settings work in some browsers, this does not generally apply to all tracking technologies. For example, since there is no precise way to detect browser fingerprinting and moreover, the purpose of such fingerprinting is not known, browser preferences are not a meaningful control mechanism for this tracking technology. Due to the complexity of this topic, we have excluded it from this paper.

Children consent. We do not address the specific concerns related to children's consent.

Exceptions. We left exceptions specified in the GDPR out of our study, e.g. cases of medical research conducted in the public interest or for compliance with legal obligations (Recital 51).

3. Background

In this section, we outline a summary of the legal fabric mostly related to cookies and other browser-based technologies, personal data collection and consent as reflected in a consent banner. In Section 3.1, we discuss how web services process personal data through tracking technologies. Section 3.2 presents the applicable rules for consent – the ePrivacy Directive and the GDPR.

3.1 Web services process personal data

The digital economy is increasingly dominated by service providers that collect and process vast amounts of personal data. Web services are a central part of the interface of any organization for the dissemination of information, collection of input and more complex transactions. We assume that web services process personal data.[18] Therefore, these web services must be operated in compliance with the privacy and data protection principles, so that the fundamental rights to privacy and to the protection of personal data are guaranteed. Examples of personal data abound: data that enables users to log in into the web service for authentication and customization purposes, IP addresses, user identifiers, timestamps, URLs of the visited pages and other parameters that enable the user to be singled-out. Usage of cookies for storing identifiers are explicitly mentioned in Recital 30 of the GDPR:

> "Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers. (...) This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them".

It is noteworthy that personal data do not consist only in the data originally collected via the web service, but also in any other information that the controller collected through other means and that can be linked to personal data collected through the web service. It also means any other information inferred that relates to an individual. The European Data Protection Supervisor (hereinafter named EDPS) declares that the use of device fingerprinting can lead to a certain percentage of assurance that two different sets of data collected belong

[17] Article 29 Working Party, "Working Document 02/2013 providing guidance on obtaining consent for cookies" (WP 208, 2 October 2013) 4 (henceforth named 29WP 208).

[18] Personal data means any information relating to an identified or (directly or indirectly) identifiable natural person. In determining whether the information relates to an identifiable individual, website publishers need to consider any means that could reasonably be used by them or any third party to enable the identification of an individual, according to Art. 4(1) and Recital 26 of the GDPR. For a deeper analysis of this concept, see Article 29 Working Party, "Opinion 4/2007 on the concept of personal data" (WP 136, 20 June 2007).

to the same individual.[19] Thus, the GDPR applies to data that can identify users (i.e. when identification of users is likely), whether they are meant or used to track the online activity of such users.

In general, any use of tracking technologies[20] which involves the processing of personal data, whether to identify directly (e.g. an email address) or more often to identify indirectly (e.g. unique cookie identifier, IP address, device identifier or component of the device, device fingerprinting, identifier generated by a software program or operating system) must comply with the GDPR. While many cookies indeed contain unique identifiers, it does not hold to all types of data; for example, some of them carry information which is too coarse to identify users, while several of them can be combined to uniquely identify users. As such, website operators need to consider cookies as storage mechanisms that may potentially contain personal data and therefore protect it accordingly. Cookies used for tracking users' online activities are unique identifiers used to single them out and recognize returning website visitors. As a result, such tracking cookies are personal data as defined in the GDPR, even if the traditional identity parameters (name, address, etc.) of the tracked user are unknown or have been deleted by the tracker after collection.

3.2 Applicable rules for consent: the ePrivacy Directive and the GDPR

The ePD prescribes that websites obtain users' informed consent before using any kind of tracking technology. Article 2(f)[21] and Recital 17[22] of the 2002 ePD define consent in reference to the one set forth in Directive 95/46/EC[23], the GDPR predecessor. The GDPR points out the conditions for obtaining valid consent in Articles 4(11) and 7. Article 4(11) of the GDPR provides for the elements composing a valid consent: "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". The GDPR provides additional guidance in Article 7 and in Recitals 32, 33, 42, and 43 as to how the controller must act to comply with the main elements of the consent requirement.

On websites, consent for cookies is usually presented in a form of cookie banners. A cookie banner is a means for getting user's consent on the usage of cookies and potentially other web application technologies that can store data or use browser attributes to recognize the user's browser, such as browser fingerprinting.[24]

4. Scoping browser-based tracking technologies

This section presents the important elements of tracking technologies: user and/or subscriber, terminal equipment, browser-based tracking technology and provider of an information society service.

This paper focuses on legal requirements relating to the processing of personal data from/onto users' devices through cookies and similar technologies. In particular, within the scope of this work, we refer to the use of cookies, and any similar technologies (browser-based tracking technology) to be stored, executed and read on the user´s terminal device, and thus falling within the scope of Article 5(3) of the ePrivacy Directive, which is worded as follows:

> Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.

Article 5(3) of the ePD applies to providers that store or gain access to information in the *terminal equipment of the subscriber or user*. Account must be taken to these four framing elements below:

- *Subscriber and/or user,*
- *Terminal equipment,*
- *Browser-based tracking technology,*
- *Provider of an information society service.*

Subscriber and/or user. The *subscriber*[25] means the person who pays the bill for the use of the online service. The user is the person using either the computer or any other device to access the online service. In many cases, the subscriber and the user can coincide, e.g. when an individual uses the broadband connection to access a website on his computer or mobile device – this person would be both the "user", as well as the "subscriber", if he or she pays for the connection. However, this is not always the case, since end-users might include employees, tenants, hotel guests, family members, visitors, and any other individuals who are using the service, for private or business purposes, without necessarily having subscribed to it. Following the example given by the UK DPA, if a family member or a visitor visits this subscriber's home and uses his internet connection to access that service from their own device, he would be the user.[26]

The ePD does not specify from whom the consent is required. The legislator did not preview which consent takes precedence (the user's or the subscriber's), nor if that choice should be at the discretion of the entity that stores or gains access to the information.[27] Whilst the web publisher, in principle, is not meant to distinguish between a consent provided by the subscriber or the user, what is relevant is that one of the parties must deliver a valid consent against BTT-related information in the landing page. Surmounting this qualification, the EDPS[28] recommends including a stand-alone definition of end-

[19] European Data Protection Supervisor, "Opinion 6/2017 on the Proposal for a Regulation on Privacy and Electronic Communications (ePrivacy Regulation)", 14 April 2017 (henceforth named EDPS Opinion).

[20] Irene Kamara and Eleni Kosta, "Do Not Track initiatives: regaining the lost user control", (2016) *International Data Privacy Law*, Volume 6, 276–290.

[21] Art. 2(f) reads that "consent by a user or subscriber corresponds to the data subject's consent in Directive 95/46/EC".

[22] Recital 17 provides that "for the purposes of this Directive, consent of a user or subscriber, regardless of whether the latter is a natural or a legal person, should have the same meaning as the data subject's consent as defined and further specified in Directive 95/46/EC".

[23] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ 1995 L 281/31.

[24] Pierre Laperdrix, Nataliia Bielova, Benoit Baudry, and Gildas Avoine, "Browser Fingerprinting: A survey", (2019) *ACM Transactions on the Web (ACM TWEB)* https://arxiv.org/abs/1905.01051 accessed 7 May 2020.

[25] This paper uses "subscriber" and "user" interchangeably.

[26] UK DPA (also known as ICO) (2019), "Guidance on the rules on use of cookies and similar technologies", Privacy and Electronic Communications Regulations, 9 https://ico.org.uk/media/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies-1-0.pdf accessed 7 May 2020 (henceforth named ICO Guidance).

[27] Eleni Kosta, "Peeking into the cookie jar: the European approach towards the regulation of cookies" (2013) *International Journal of Law and Information Technology*, Volume 21, Issue 4, 380–406 https://doi.org/10.1093/ijlit/eat011 accessed 7 May 2020.

[28] cf. EDPS Opinion (n 19) 14.

user in the forthcoming ePrivacy Regulation, for purposes of providing consent, to ensure that it is the individuals effectively using the service, rather than those subscribing to it. In this paper, we use *user* and *data subject* interchangeably.

Terminal equipment. *Terminal equipment* refers to a device where information is accessed or stored, e.g. desk computers, laptop, pads, smartphones, but also other equipment such as wearable technologies, smart TVs, game consoles, connected vehicles, voice assistants, as well as any other object that is connected to an electronic communication network open to the public. Our understanding of the term *web service* refers to any type of information service made accessible over the internet with which users interact usually through web browsers, mobile apps or other client software. IoT web services, accessed by IoT devices, are included.

Browser-based Tracking Technology. A *Browser-based Tracking Technology* (henceforth named BTT), the third element of this quadrant, is commonly acknowledged as any technology which enables tracking of the user while she visits a website using a Web browser. From a legal perspective, BTT is defined as the reading or storing of information from/onto the users' devices for tracking purposes, in line with the text of Article 5(3) of the ePD. From a computer science perspective, BTT is a technology that enables tracking of the user by either depositing identifiers on their computer or using fingerprinting methods to identify them. For a technology to successfully track a browser user, trackers need to have two key capabilities:

1. The ability to store a unique identifier (or to re-create it) on a user's machine,

2. The ability to communicate that identifier, as well as visited sites, back to the domain, controlled by the tracker.

The most common type of BTT is "*stateful*" tracking. A typical example of it are *browser cookies*. They are used to store a unique identifier and communicate it to their owner's domain when they are automatically sent by the browser, or via JavaScript programs running on a visited website. Alternative tracking technologies that rely on other browser storages are also actively used by trackers today – they include HTML5 local storage, browser cache and many others. Most of technologies rely on JavaScript programs, since this is the most convenient and portable way to both store and send the unique identifier to the tracker's domain.

Alternatively, instead of storing an identifier, a tracker can re-create it based on the browser's and machine's properties, accessible via the HTTP protocol and also via JavaScript. Such tracking is called "*stateless*", and for Web browsers is represented by browser fingerprinting.[29]

In this paper, we unify all such technologies under the common terminology of BTT. In the scope of current BTT, the risk to data protection comes from the purpose(s) of processing.[30]

Provider of an information society service. The *provider of an information society service* (i.e. a publisher) provides a website content service, at the request of a user, either paid or unpaid, remotely and electronically.

[29] cf. Laperdrix et al. (n 24).
[30] Article 29 Working Party, "Opinion 04/2012 on Cookie Consent Exemption" (WP194), June 2012 (henceforth named "29WP (WP194)").

4.1 Browser-based tracking technologies requiring or exempted from consent

In this paper, we only refer to the use of BTT requiring consent. According to Article 5(3) of the ePD, consent is not required when the purpose of trackers is:

- **Communication**: used for the sole purpose of enabling the communication on the web; and
- **Strict necessity**: cookies strictly necessary to enable the service requested by the user: if BTT is disabled, the service will not work.

The above mentioned 29WP (WP194)[31] analyzed these two exceptions accordingly (considering browser cookies, but it is extended to all BTTs):

- The *communication exemption* applies when the transmission of the communication is impossible without the use of the BTT (e.g. load-balancing cookie). Hence, using BTT to merely "assist" or "facilitate" the communication is insufficient.
- The *strict necessity exemption* involves a narrow interpretation. It means that the use of BTT must be restricted to what is strictly necessary (and hence essential) to provide a service explicitly requested by a user. Thus, using BTT that is *reasonably necessary* or *important* – this implies that the service provided by the website operator would not function without the BTT. In this regard, the choice of a certain functionality that relies on BTT is not enough to justify the *strict necessity* if the web publisher has a different implementation choice that would work without a BTT.

Both the 29WP and DPAs provide explicit examples of BTT's purposes that require the user's consent. They assert that the following purposes are usually not strictly necessary to the user visiting a website, since they are usually related to a functionality that is distinct from the service that has been explicitly requested: "advertising, and use of the data for marketing, research and audience measurement" are not strictly necessary to deliver a service that is requested by a user (29WP (WP240)).[32]

We will further analyze which BTTs are exempted from consent based solely on their *purpose*, and not on their technical abilities. Ultimately, as the 29WP (WP194) exposes, "it is thus the purpose and the specific implementation or processing being achieved that must be used to determine whether or not a cookie can be exempted from consent". The 29WP (WP194) clarifies further that when applying the exemptions for obtaining consent, it is important to examine what is strictly necessary *from the point of view of the user*, not of the service provider. Regarding *multipurpose* BTT, whenever a BTT covers different purposes, some of which require consent (e.g. can be used for the purpose of remembering user preferences and for the purpose of tracking), the website still needs to seek user consent for such multipurpose BTT. The 29WP recalls that in practice, this should encourage website owners to use a different BTT for each purpose.

For this classification of the purposes of BTT, we relied on the guidance from the 29WP.[33-34] For a comparative analysis, we also consulted the recent guidelines from DPAs (ICO, CNIL, German and Dutch DPA). This classification is shown in Table 5.

[31] cf. 29WP (WP194) (n 30).
[32] Article 29 Working Party, "Opinion 03/2016 on the evaluation and review of the e-Privacy Directive (2002/58/EC)", (WP240, 19 July 2016).
[33] Article 29 Working Party, "Opinion 2/2010 on online behavioural advertising", (WP 171, 22 June 2010).
[34] cf. 29WP (WP194) (n 30).

Table 5 Examples of purposes of BTT exempted and non-exempted of consent.

Purposes exempted of consent	Purposes needing consent
Local Analytics – These are statistical audience measuring tools for providing information on the number of unique visits to a website, how long users stay in the site, what parts and pages of the website they browse, detecting main search keywords, track website navigation issues. The 29WP and the EDPS[35] exempt these from consent insofar they are limited to first party (website owner) anonymized and aggregated statistical purposes, as these are not likely to create a privacy risk. The CNIL[36] points out that certain analytic cookies can be exempted if they meet a list of cumulative requirements. The Irish[37] and Dutch DPAs[38] state that these may have little privacy effects on users.	***Non-local Analytics*** – Even if a website owner relies on self-claims of "strictly necessary" first-party analytics, the 29WP[39] says that they are not strictly necessary to provide a functionality explicitly requested by the user, because the user can access all the functionalities provided by the website when such cookies are disabled. As a consequence, these cookies do not fall under the exemption of consent if they are not limited to the website owner. Moreover, both the ICO[40] and the German DPA[41] held that third-party analytics cookies are not *strictly necessary*. The Greek DPA[42] says that third-party web analytics trackers, such as the Google Analytics service as *not strictly necessary*, hence requiring consent.
Session User input – The 29WP states these are used to keep track of the user's input (session-id) when filling online forms over several pages, or as a shopping cart, to keep track of the items the user has selected by clicking on a button. These BTT are clearly needed to provide a service explicitly requested by the user, for the duration of a session. Additionally, they are tied to a user action (such as clicking on a button or filling a form).	***Advertising*** – The 29WP affirms that third-party advertising BTTs require consent, as well as operational purposes related to third-party advertising, such as frequency capping, financial logging, ad affiliation, click fraud detection, research and market analysis, product improvement and debugging.[43] Even though the 29WP only distinguishes third-party advertising, we believe that the category of purposes should only be called "Advertising". We insist on it because it has been observed that first-party cookies are also often synchronized with third-party cookies[44], and moreover, publishers started hiding advertising content under the first-party content (typical case is with DNS redirection). The ICO[45] posits that while advertising cookies may be crucial in the eyes of a website or mobile app operator as they bring in revenue to fund the service, they are not "*strictly necessary*" from the point of view of the website user and hence, the law. The Dutch DPA[46] names these as tracking cookies and advises companies to request consent to place tracking cookies. The same reasoning holds for the German[47] and Irish[48] DPA.
User-security for a service explicitly requested by the user – The 29WP names these due to their function on providing security functionalities for a service the user has requested (e.g. online banking services) and for a limited duration, e.g. to detect repeated failed login attempts on a website, or other similar mechanisms designed to protect the login system from abuses.	***User-security for a service not explicitly requested by the user*** – The 29WP[49] refers to cookies providing security for content not explicitly requested by the user. For example, if a website uses advertising content that contains user-security cookies, such as those of Cloudflare, then the user consent is required.[50]

35 European Data Protection Supervisor, "Guidelines on the protection of personal data processed through web services provided by EU institutions" (2016) 10-13 (henceforth named "EDPS Guidelines").

36 The CNIL prescribes that these cookies may be exempt from consent if the conditions of Article 5 are met, such as: the user is informed thereof and has the possibility to refuse them; such cookies exclude any form of unique targeting of individuals; the collected data must not be combined or merged with other types of data, nor disclosed to third parties; the use of trackers must be strictly limited to producing anonymous statistics; the trackers may only be used by one publisher and must not enable tracking a user over different websites or mobile apps; an IP address cannot be used to geolocate the user more precisely than the city, otherwise must be deleted or anonymized once the user has been located to avoid this data from being used or combined with other data. See CNIL Guidelines on cookies and other trackers (2019) www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038783337 accessed 7 May 2020.

37 Irish DPA (2020), Guidance note on the use of cookies and other tracking technologies (2020) https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Guidance%20note%20on%20cookies%20and%20other%20tracking%20technologies.pdf (henceforth named "Irish DPA Guidance")

38 An explanation of the legal requirements for cookies (besides tracking cookies) is available on the website of the Netherlands Authority for Consumers and Markets (ACM), "Cookies" (2019) www.acm.nl/nl/onderwerpen/telecommunicatie/internet/cookies accessed 7 May 2020.

39 Regarding first-party analytics, the 29WP (WP194) (n 30) considers that these are not likely to create a privacy risk when they are strictly limited to aggregated statistical purposes and when users are informed thereof and can opt out therefrom.

40 The ICO declares that it is "unlikely that priority for any formal action would be given to uses of cookies where there is a low level of intrusiveness and low risk of harm to individuals" and first party analytics cookies are given as an example of cookies that are potentially low risk, cf. ICO Guidance (n 26).

41 cf. German DPA Guidelines (n 10).

42 Greek Data Protection Authority, "Guidelines on Cookies and Trackers" (2020) http://www.dpa.gr/APDPXPortlets/htdocs/documentSDisplay.jsp?docid=84,221,176,170,98,24,72,223.

43 cf. 29WP (WP194) (n 30) 9-10.

44 Imane Fouad, Nataliia Bielova, Arnaud Legout, and Natasa Sarafijanovic-Djukic (2020). Missed by Filter Lists: Detecting UnknownThird-Party Trackers with Invisible Pixels. In *proceedings on Privacy Enhancing Technologies* (2):499–518, https://petsymposium.org/2020/files/papers/issue2/popets-2020-0038.pdf

45 cf. ICO Guidance (n 26) 39.

46 Autoriteit Persoonsgegevens (2019), "Cookies" https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/internet-telefoon-tv-en-post/cookies#mag-ik-als-organisatie-een-cookiewall-gebruiken-7111 accessed 7 May 2020 (henceforth named "Dutch DPA Guidelines").

47 cf. German DPA Guidelines (n 10).

48 cf. Irish DPA Guidance (n 37).

49 According to the 29WP the consent exemption does not cover the use of cookies that relate to the security of websites or third-party services that have not been explicitly requested by the user, (WP194) (n 30) 7.

50 The purpose of the cookie "__cfduid" is used by Cloudflare for detection of malicious visitors. Such cookie requires consent when it is used by Cloudflare in advertising content or other content not explicitly requested by the user, Cloudflare, "Understanding the Cloudflare Cookies" (2019) https://support.cloudflare.com/hc/en-us/articles/200170156-What-does-the-Cloudflare-cfduid-cookie-do- accessed 2 December 2019.

Social media plugin for a functionality explicitly requested by the user – The 29WP refers that many social networks propose "social plug-in modules" that website owners integrate in their platform, to provide some services than can be considered as "explicitly requested" by their members, e.g. to allow them to share content they like with their "friends" (and propose other related functionalities such as publishing comments). These plugins store and access cookies in the user's terminal equipment in order to allow the social network to identify its members when they interact with them.	***Social media plugin for a functionality not requested by the user*** – The 29WP refers that these "social plug-in modules" can also be used to track users: logged-in, "non-logged-in" users, and also non-members. We conclude however that even logged-in members can be tracked and therefore name this category as "functionality not requested by the user". The German DPA has the same position.[51]
Session Authentication – The 29WP describes these as the ones used to identify the user once he has logged in into websites, for the duration of a session. They allow users to authenticate themselves on successive loads of the website and gain access to authorized content or functionality, such as viewing their account balance, transactions in an online banking website, online shopping. This authentication functionality is an essential part of the service a user explicitly requests.	***Persistent Authentication*** – The 29WP says also that persistent login cookies which store an authentication token across browser sessions are not exempted of consent. This is an important distinction because the user may not be immediately aware of the fact that closing the browser will not clear their authentication settings. They may return to the website under the assumption that they are anonymous whilst in fact they are still logged in to the service.
Short-term User Interface Customization (personalization, preferences) – According to the 29WP, these are used to store a user's preference regarding a service across web pages and not linked to other persistent identifiers such as usernames. These are explicitly enabled by the user, e.g. by clicking on a button or ticking a box to keep a language, display format, fonts, etc. Only session (or short term) cookies storing such information are exempted.	***Long-term User Interface Customization*** – The 29WP says that the addition of information to remember the user's preference for a longer duration will not be exempted of consent.
Load Balancing – The 29WP says that load balancing is a technique that allows distributing the processing of web server requests over a pool of machines instead of just one. Among several techniques, a cookie may be used to identify the server in the pool in order for the load balancer to redirect the requests appropriately. These are session cookies.	
Session Multimedia Content Player – The 29WP clarifies that these apply to BTT used to keep track of the state of audio/video. When the user visits a website containing related text/video content, this content is equally part of a service explicitly requested by the user and is exempted of consent. As there is no long-term need for this information, they should expire once the session ends.	

[51] cf. German DPA Guidelines (n 10).

5. Requirements for valid consent for consent banners

This section presents our interdisciplinary legal and technical analysis of the requirements applied to consent banner design. We present the seven high-level requirements, followed by the definition of 22 low-level requirements.

We convey the respective *legal sources* upon which each requirement is based. The sources are either: binding (GDPR, ePD and CJEU case-law), non-binding (EDPB and DPA guidelines) and grounded in our own subjective interpretation of legal sources (L) or from a technical computer science perspective (CS). We present the *procedure for compliance verification*. For each requirement, we describe the procedure that needs to be put in place in order to detect violations. Such procedure can be assessed in three ways:

- Manual, relying only on a human operator (M);
- Technical, an expert using computer tools able to detect a violation (T);
- Performing user studies to evaluate perceptions of end users (U).

Table 6 describes all the high- and low-level requirements, their provenance, position in the paper that describes them in detail and how they can be assessed. Table 7 depicts the positioning of DPAs (French, UK, Irish, German, Spanish, Greek, Danish, Belgium) in relation to the 22 low-level requirements proposed in this paper.

Every subsequent subsection is structured as follows:

Analyzing legal sources for high-level requirement. We present each high-level requirement, followed by derived low-level requirements presented in a table with the respective sources: binding (GDPR, ePD and CJEU case-law), non-binding (EDPB and DPA guidelines) and grounded in our own subjective interpretation of legal sources (L) or from a technical computer science perspective (CS).

One subsection for each low-level requirement. For every low-level requirement, we present its description and

1. Explain the correspondent violation in a "requirement box" (in a consolidated form, for ease of reading).
2. Provide two example websites: one demonstrating compliance, and one presenting a violation. We illustrate (where possible) screenshots or technical content from each website to explain the technical details of each requirement. Each example is extracted from real-world websites, illustrated in figures duly dated.
3. Describe the procedure (manual, technical, or with user studies) to detect violations.

5.1 Prior consent

Before storing information or gaining access to information on a user's terminal, website publishers need to request prior consent to data subjects in order to guarantee that the user has some control over the processing of their information.[52] Even if no explicit provision

[52] The CNIL recalls that many site publishers have reported difficulties in obtaining prior consent from Internet users before depositing and reading cookies for two main reasons: 1. this would prevent the display of certain advertisements, resulting in a significant loss of income; 2. cookies do not come from their own servers, being linked to the activity of third-party partners, over which they have no control. As a result, publishers alone cannot bear full responsibility for enforcing tracer rules as "third-party cookies" because they originate from third-party companies, "Cookies: CNIL extends its controls beyond site publishers" www.cnil.fr/fr/cookies-la-cnil-etend-ses-controles-au-dela-des-editeurs-de-sites accessed 11 December 2019.

Table 6 Requirements for a valid consent on consent banner design, assessment and source

Requirements		Assessment	Sources at low-level requirement			Location in the paper (page)
High-Level Requirements	Low-Level Requirements	Manual (M), Technical (T) or User study (U)	Binding	Non-binding	Interpretation: Legal (L) or Computer Science (CS)	
Prior	R1 Prior to storing an identifier	M (partially) or T (partially)	√	√	-	101
	R2 Prior to sending an identifier	T (partially)	-	-	CS	102
Free	R3 No merging into a contract	M (fully) or T (partially)	√	√	-	104
	R4 No tracking walls	M (fully)	-	√	-	105
Specific	R5 Separate consent per purpose	M (fully)	√	√	-	108
Informed	R6 Accessibility of information page	M (fully) or T (partially) together with U	-	√	-	111
	R7 Necessary information on BTT	M (fully) or T (partially)	√	√	-	111
	R8 Information on consent banner configuration	M (fully) or T (partially)	-	√	-	113
	R9 Information on the data controller	M (fully) or T (partially)	√	√	-	113
	R10 Information on rights	M (fully) or T (partially)	√	√	-	113
Unambiguous	R11 Affirmative action design	Combination of M and T (partially)	√	√	-	114
	R12 Configurable banner	M or T (partially)	-	√	L	115
	R13 Balanced choice	M (fully)	-	√	L	117
	R14 Post-consent registration	T (partially)	-	√	CS	118
	R15 Correct consent registration	Combination of M and T (partially)	-	√	CS	119
Readable and accessible	R16 Distinguishable	M (fully) or T (partially)	√	√	-	121
	R17 Intelligible	U	√	√	-	121
	R18 Accessible	U	√	√		121
	R19 Clear and plain language	U	√	√	-	121
	R20 No consent wall	M (fully) or T (partially)	-	√	L	122
Revocable	R21 Possible to change in the future	M (fully)	√	√	-	124
	R22 Delete "consent cookie" and communicate to third parties	Not possible	-	-	CS	125

was made manifest both in the GDPR and the ePD, the "prior" timing is confirmed through the combined analysis of both legislative instruments.

Under the GDPR aegis, the 29WP (WP259 rev.01)[53] claims that "prior consent" can be derived from Article 6 by the wording "has given",

> Although the GDPR does not literally prescribe in Article 4(11) that consent must be given prior to the processing activity, this is clearly implied. The heading of Article 6(1) and the wording "has given" in Article 6(1)(a) supports this interpretation. It follows logically from Article 6 and Recital 40 that a valid lawful basis must be present before starting a data processing.

From an ePD stance, such understanding of a "prior consent" is derived from Article 5(3) of the ePD, according to the 29WP guidance,[54-55]

> Article 5(3) contains a specific rule regarding the storing of information or gaining of access to information on a user's terminal, including for the purpose of tracking the user's on-line activities. While Article 5(3) does not use the word prior, this is a clear and obvious conclusion from the wording of the provision. (...) It makes good sense for consent to be obtained prior to the starting of the data processing.

In the light of the above, a consent request needs to be presented before BTT are deployed. Seconding this rule, the 29WP (WP208)[56] asserts that "consent should be sought before cookies are set or read.

As a result, a website should deliver a consent solution in which no cookies are set to user's device (other than those that may not require user's consent) before that user has signaled their wishes regarding such cookies".

53 cf. 29WP (WP259 rev.01) (n 4) 17.
54 Article 29 Working Party, "Opinion 15/2011 on the definition of consent" (WP187, 13 July 2011).
55 cf. WP29 (WP 208) (n 17) 4.
56 cf. 29WP (WP208) (n 17) 4.

Table 7 DPAs positioning in relation to the low-level requirements

Requirements		DPAs positioning							
High-Level	**Low-level**	**French (CNIL)**	**UK (ICO)**	**Irish**	**German**	**Spanish**	**Greek**	**Danish**	**Belgian**
Prior	R1 Prior to storing an identifier	√	√	√	√	-	√	√	√
	R2 Prior to sending an identifier	-	-	-	-	-	-	-	-
Free	R3 No merging into a contract	√	√	√	√	-		√	√
	R4 No tracking walls	√	**?**	√	√	**?**	√	√	√
Specific	R5 Separate consent per purpose	√	√	√	√	√	√	√	√
Informed	R6 Accessibility of information page	√	√	√	√	√	√	√	√
	R7 Information on BTT	√	√	√	√	√	√	√	√
	R8 Information on consent banner configuration	√	√	√	-	√ **decision**	√	√	√
	R9 Information on the data controller	√	√	√	√	√	√	√	√
	R10 Information on rights	√	√	√	√	√	√	√	√
Unambiguous	R11 Affirmative action design	√	√	√	√	**X**	√	√	√
	R12 Configurable banner	√	√	√	√	√ **decision**	√	√	-
	R13 Balanced choice	√	√	√	-		√	√	-
	R14 Post- consent registration	√	√	√	-	√	-	√	√
	R15 Correct consent registration	√	-	-	-	√	√	-	-
Readable and accessible	R 16 Clearly distinguishable	√	√	√	√	√	-	-	√
	R17 Intelligible	√	√	-	-	√	-	√	-
	R18 Easily accessible	√	√	√	-	√	-	-	√
	R19 Clear and plain language	√	√	-	-	√	-	√	-
	R20 No consent wall	**?**	√	-	-	-	-	-	-
Revocable	R21 Possible to change in the future	√	√	√	√	√	√	√	√
	R22 Delete "consent cookie", communicate to third parties	-	-	-	-	-	-	-	-

Moreover, processing is unlawful if carried out before the request for consent due to the lack of legal ground, as denoted by the 29WP (WP147)[57]:

> Otherwise, the processing carried out during the period of time from the moment the processing had started until the moment that consent had been obtained would be unlawful because of lack of legal ground. Furthermore, in such cases, if the individual decided against consenting, any data processing that had already taken place would be unlawful for that reason as well.

Notice that instead of specifying a certain type of BTT, such as cookies, we resort to describe low-level requirements in terms of the usage of user identifiers because BTTs are simply mechanisms that store and/or transfer identifiers, thus allowing tracking the users across the Web. We have therefore subdivided the requirement of "prior consent" into two low-level requirements, as shown in Table 8:

- first, consent must be obtained before user identifier is set or stored (those requiring consent);
- second, consent must be obtained before user identifier is sent, i.e. before the content of the webpage that is associated to such identifier is loaded.

Table 8 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-Level	**Low-level**	**Binding**	**Non-binding**	**Interpretation**
Prior	**R1 Prior to storing an identifier**	4(11), 6(1) (a) GDPR	29WP (almost all DPAs)	
	R2 Prior to sending an identifier	–	–	CS

R1 Prior to storing an identifier

It follows from the foregoing subsection that consent must be collected before an identifier is stored in the user's device (other than those that may not require user's consent). This requirement has often been considered in legal sources as "prior to setting cookies", however cookies is just one example of a stateful BTT. Therefore, we rename this requirement as "prior to storing an identifier" because what technically happens is that a user identifier is stored on her device.

Requirement	**Prior to storing an identifier**
Description	Consent must be obtained before a user identifier is stored
Violation	A user identifier is stored before consent is given

[57] cf. 29WP (WP187) (n 54) 31.

Examples. Figures 1 and 2 depict the case of violation of this requirement. While accessing the eBay webpage, a banner appears affirming that by using the website, the user accepts the use of cookies to enhance their services. This overlay includes a link to "learn more". This consent mechanism does not allow a user to make a choice before an advertising cookie that requires consent, named "IDE", stores a user identifier in the user's browser.

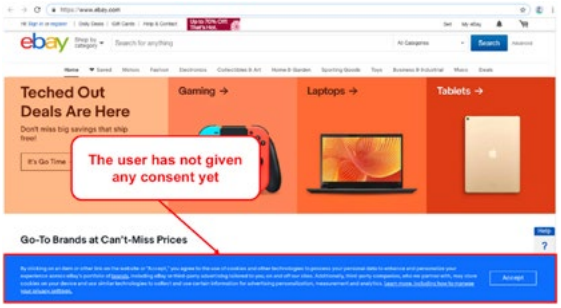


Figure 1 Access to the eBay website (www.ebay.com accessed 27 July 2019)

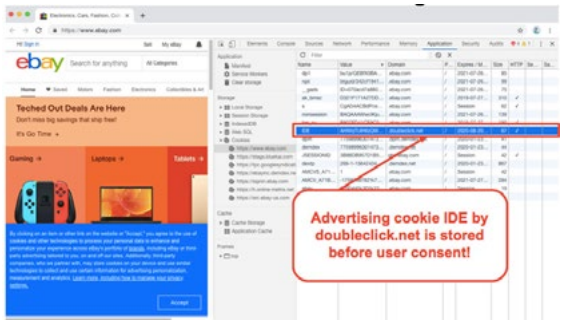


Figure 2 Violation of the requirement "Prior to setting cookies" by eBay website (www.ebay.com accessed 27 July 2019)

How to detect violations? One could detect a violation of the "Prior to storing an identifier" requirement by visiting a website with an empty browser storage (no cookies, empty cache, all other storages are empty) and analyzing all the elements stored in the browser (in all storages) that are set upon visiting the website (as shown in Figure 2). Such verification, however, contains two complex tasks:

1. **Detecting whether a stored element is a user identifier**. It is a very complex question of what constitutes an identifier and whether a specific element stored in a browser storage is indeed an identifier. Computer science researchers resorted to heuristics[58] [59] [60] [61] that included length and expiration date for each stored element in the browser or relied on entropy measures. Nevertheless, it is not possible to know with certainty whether a given string stored in a browser storage is indeed a user identifier.

2. **Analyzing all possible browser storages**. Very specific browser tools would be required to detect such violations when various browser storage mechanisms are used, like Web caching mechanisms. It is possible to detect the setting of cookies by a technical expert with the corresponding browser tools or even fully automatically, but the complexity grows as trackers use other storages or even combinations of them (e.g. storing a piece of an identifier in one storage, and another piece in another storage). Computer science researchers have mostly analyzed cookies, and other basic storages and raised concerns about the usage of more advances techniques, such as HTTP Strict Transport Security (HSTS).[62]

3. **Identifying the purpose of an identifier.** Finally, the *purpose* of each stored identifier needs to be declared and known in order to determine whether consent is required. In general, it is rarely possible to detect a cookie's purpose automatically or with technical tools. Even manually, it is complex to estimate whether a cookie requires consent or not by reading its purpose in the cookie policy.[63] Also, the purposes of cookies described in cookie policies are often not clear, too vague or incomplete.[64] For automatic verification, one would need a self-declaration of the purpose of each cookie in a standard format.

Conclusion: Detection of an identifier storage is a very complex task, for which technical tools do not exist today due to impossibility to detect an identifier, to technical analysis of all browser storages and to the difficulty of identifying the purpose of an identifier. Manual analysis is neither possible for the same reasons. Therefore, this requirement can only be partially assessed with technical or manual analysis.

R2 Prior to sending an identifier

Consent must be obtained before identifiers are sent to the third parties. This requirement originates from cookies that are sent automatically when the third-party content is loaded (hence, cookies are "read"), however we generalize it to sending of identifiers that require consent via any means, including JavaScript code, for example via XMLHTTPRequest,[65] or any request that is performed on a visited website.

Such requirement is not based on any legal source but derives from its technical implementation. Law is based on a general "document

[58] Steven Englehardt and Arvind Narayanan (2016), "Online Tracking: A 1-million-site Measurement and Analysis", *ACM CCS*. https://senglehardt.com/papers/ccs16_online_tracking.pdf accessed June 19, 2020.

[59] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, Edward Felten (2015), "Cookies that Give You Away: Evaluating the surveillance implications of web tracking" https://senglehardt.com/papers/www15_cookie_surveil.pdf accessed June 19, 2020.

[60] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, Claudia Diaz (2014), "The Web Never Forgets: Persistent tracking mechanisms in the wild", *ACM CCS* https://securehomes.esat.kuleuven.be/~gacar/persistent/the_web_never_forgets.pdf accessed June 19, 2020.

[61] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, Norbert Pohlmann, "The Unwanted Sharing Economy: An Analysis of Cookie Syncing and User Transparency under GDPR". *ArXiv e-prints*, November 2018. arXiv:1811.08660.

[62] P. Syverson and M. Traudt. HSTS supports targeted surveillance. In *USENIX Workshop on Free and Open Communications on the Internet* (FOCI), 2018.

[63] Imane Fouad, Cristiana Santos, Feras Al Kassar, Nataliia Bielova, Stefano Calzavara. On Compliance of Cookie Purposes with the Purpose Specification Principle. IWPE, Jul 2020, Genova, Italy. hal-02567022 accessed June 19, 2020.

[64] For instance, the privacy policy on the pubmatic.com website indicates that the "repi" cookie is "a short-lived cookie that is used to determine if repixeling is in progress". This description is obscure and makes it difficult to qualify the purpose of this cookie. Another example is "centerVisitorId" on the learnworlds.com website, whose only description states: "used by site's popups and download forms".

[65] XMLHttpRequest Living Standard, https://xhr.spec.whatwg.org.

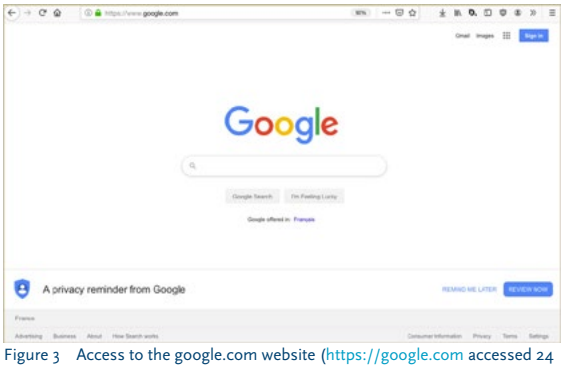


Figure 3 Access to the google.com website (https://google.com accessed 24 September 2019)

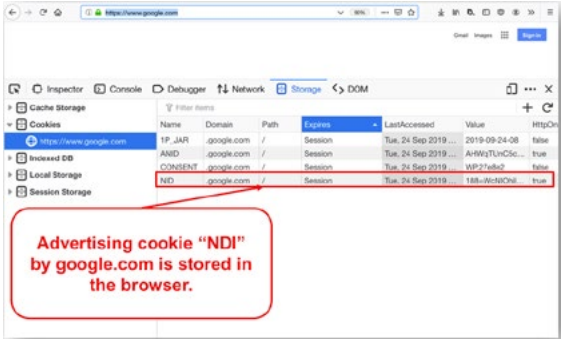


Figure 4 Access to the google.com website: advertising cookie NID is stored in the browser (https://google.com accessed 24 September 2019)

read/access" view which is not in line with how the Web operates technically. Advertisers do not visit the user's browser to read their cookies, but the opposite happens: users visit websites, and their browsers send user identifiers (for example, cookies are sent automatically). Thus, we need to add this supplementary requirement that consent must be obtained before identifiers are sent, and not read. We note that respecting such a requirement demands important adaptation of current technical tools. Browsers automatically attach cookies to requests that fetch Web content, and also run JavaScript code that includes identifiers in requests, which makes it complicated for cookie banners implementation to prevent cookie transmission prior to consent.

Requirement	Prior to sending an identifier
Description	Consent must be obtained before an identifier is sent
Violation	Identifiers that require consent are sent before consent is obtained

Examples. Figures 3 and 4 show how google.com sets cookies in the user's browser. Notice that google.com is a default search engine in most browsers, hence such experience is common to many users. Google.com is setting a "NID" cookie prior to the user's consent – this cookie now belongs to google.com (see Figure 4). After visiting google.com, a user goes to a different website that contains some content from google.com. Figure 5 shows an example website https://www.w3schools.com/, commonly consulted by Web developers. While accessing this website (with Firefox 69.0.1 in our experiment), no banner is shown to the user, however requests are sent to cse.google.com in order to fetch Google Customized Search Engine that helps the user to search inside this website. Figure 6 shows a violation of the 'Prior to sending an identifier' requirement because the NID cookie (see (1)) that contains a user identifier is now sent to cse.google.com (2) without user's consent while fetching some (supposedly functional) content from cse.google.com (3).

How to detect violations? Detecting violations of this requirement is a very complex task, and no technical solutions exist today that are able to assess this requirement. Apart from the difficulties we have raised in Section R1, the assessment of this requirement requires further technical investigations:

1. **Extensive testing of all browser storages with all possible identifiers set by various domains is needed**. Even for HTTP cookies, one would need to test the website with the corresponding cookies already set in the browser and analyze all the loaded content in order to detect what content is sending such cookies. This procedure might sound easy when cookies are simply attached by the browser when the content is loading. However, computer science researchers[66] have shown that when cookies are sent via JavaScript requests, they are often encrypted or obfuscated. To the best of our knowledge, as of June 2020, no one has measured whether companies encrypt and send identifiers from other storages to third parties.

Figure 5 Access to the W3Schools.com website (https://www.w3schools.com accessed 24 September 2019)

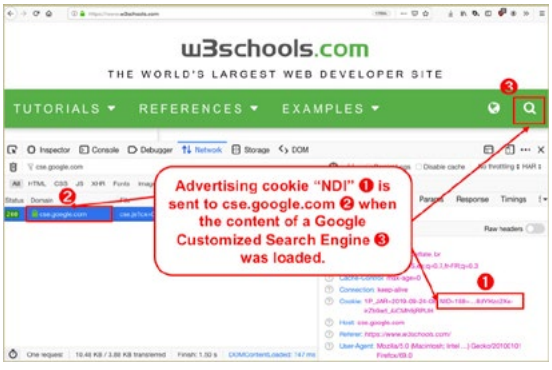


Figure 6 Violation of the requirement "Prior to sending an identifier" by W3Schools.com website (https://www.w3schools.com, accessed 24 September 2019)

[66] Panagiotis Papadopoulos, Nicolas Kourtellis, and Evangelos Markatos, "Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask". In *The World Wide Web Conference 2019*, pp. 1432-1442.

2. **Language-based security analysis[67] for JavaScript[68], such as taint-tracking[69] and information flow monitors[70] can be used** to monitor when identifiers are read and further sent to other third parties. However, these technologies can be used only if it is known how to detect identifiers and what is their purpose (and hence it is clear whether such identifiers require consent).

3. **Browser fingerprinting also falls into this requirement**: no information is explicitly stored in the user's browser; however, a unique identifier built from a browser fingerprint can be constructed and sent. It is well known in the computer science research community that detection of fingerprinting is a complex challenge and as of today, there is no technique to detect browser fingerprinting accurately, as summarized in a recent extensive survey by computer scientists.[71] Similar to browser storages, when browser fingerprinting is used, the purpose must be clearly defined and it must be clear when an identifier is created from a browser fingerprint. As a result, it must be clear whether browser fingerprinting is used for a purpose that requires consent, or is exempted of consent (for example, when fingerprinting is used for a security purpose, such as enhanced authentication).

5.2 Freely given

Consent must be freely given, as prescribed in the GDPR in Article 4(11) and further specified in Article 7(4). The request for consent should imply a voluntary choice to accept or decline the processing of personal data, taken in the absence of any kind of pressure or compulsion[72] on the user in persuading to give his consent.

The same holds for processing personal data through BTT. The 29WP (WP208)[73] refers to this "freedom of choice" of the users in choosing cookie settings; it asserts that "the user should have an opportunity to freely choose between the option to accept some or all cookies or to decline all or some cookies and to retain the possibility to change the cookie settings in the future." The Finnish DPA[74] adds that consent is not freely given when there is "any undue pressure on or influence on the user's free will to consent".

As a consequence of not having a freely given consent, the request becomes invalid, as cautioned by the WP29 (WP187): "any pressure or inappropriate influence exerted on the person (in different ways) preventing them from exercising their will shall invalidate consent", and "cannot be claimed to be a legitimate ground to justify the processing".

Forced consent is decomposed in the 29WP guidelines considering three elements: imbalance of power[75], unconditional and non-detrimental. In this paper, we analyze both the unconditional and the non-detrimental elements, as shown in Table 9. Imbalance of power is a subjective requirement that can be only evaluated in a case-per-case manner and is dependent on a specific context when consent is given. Hence, we excluded this analysis, as explained in Section 2.3.

Table 9 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-level	**Low-level**	**Binding**	**Non-binding**	**Interpretation**
Free	R3 No merging into a contract	7 (2) (4), Recital 43	29WP; DPAs: Danish, French, UK, Irish, Belgium	–
	R4 No tracking wall	–	Recital 42 GDPR, Recital 25 ePD; EDPB, EDPS; BEUC; EU Parliament; DPAs: Dutch, French, German, Danish, Greek, Irish, Belgian	–

R3 Unconditionality related to a contract

Article 7(4) and Recital 43 of the GDPR confer a presumption of a not freely given consent in the presence of a contract or service. Article 7(4) reads as: "When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent". Recital 43 recites as "consent is presumed not to be freely given (...) if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance".

The purpose of these provisions is to ensure that services are not offered upon the condition that users give personal information which are not necessary for the offering of these services. Article 7(4) prohibits any form of bundling of a service with a request for consent, when the consent is not necessary for the delivery of that service. For example, if a website makes online transactions (together with marketing purposes) dependent on the user consent for processing personal data that is not necessary for these purposes, it can be reasonably assumed that consent is forced. As a result of the established presumption, any controller has to prove that consent was freely given.

In practice, this requires consent for processing to be clearly distinguishable (untied, unbundled) from contracts or agreements[76] or pri-

67 Andrei Sabelfeld, Andrew C. Myers. Language-based information-flow security. IEEE J. *Sel. Areas Commun.* 21(1): 5-19 (2003)

68 Daniel Hedin, Luciano Bello, Andrei Sabelfeld. Information-flow security for JavaScript and its APIs. *J. Comput. Secur.* 24(2): 181-234 (2016)

69 B. Livshits. Dynamic taint tracking in managed runtimes. *TechnicalReport MSR-TR-2012-114*, Microsoft, November 2012.

70 Jonas Magazinius, Daniel Hedin, Andrei Sabelfeld: Architectures for Inlining Security Monitors in Web Applications. *ESSoS* 2014: 141-160, 2012.

71 cf. Laperdrix (n 24).

72 The 29WP opinions provide examples of a non-freely given consent can reveal different conducts: compulsion, pressure or inability to exercise free will; being put under pressure, be it social, financial, psychological or other; deception; intimidation; inappropriate influence; coercion; significant negative consequences if he does not consent (e.g. substantial extra costs), 29WP (WP187 (n 54), and WP259 rev.01 (n 4)).

73 cf. 29WP (WP208) (n 17) 5.

74 Finish DPA (n 13).

75 Recital 43 of the GDPR clarifies situations in which consent cannot be seen as freely given "where there is a clear imbalance between the data subject and the controller (...) and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation." The Recital concerns authorities, but also corporations in a dominant market position (e.g. in the area of social networking service of relevance, as in the case of Facebook), and/or in a closed and proprietary network where the data subject is factually forced to join or maintain a profile with the controller, to be able to interact with persons that are not available on other services. A representative related complaint on forced consent was issued by NOYB against Facebook, See NOYB, "Complaint filed against Facebook Ireland Ltd." (2018) https://noyb.eu/wp-content/uploads/2018/05/complaint-facebook.pdf accessed 7 May 2020.

76 An illustrative example is the complaint filed by NOYB against Google that we transcribe for the practical relevance of this requirement "bundling happens when the controller requires the data subject to consent to the privacy

vacy policies and terms of contract (as posited in Article 7(2) GDPR). The Danish DPA[77] reiterates that "all-purpose acceptance of general terms and conditions cannot be taken as a clear affirmation whereby the user consents to the processing of personal data". Consent would be deprived of any meaning if services are only offered in exchange for mandatory consent to the exploitation of personal data. As the 29WP (WP159 rev.01)[78] reasserts, the "GDPR ensures that the processing of personal data for which consent is sought cannot become directly or indirectly the counter-performance of a contract".

From Article 7(4), we denote the words *conditionality* and *inter alia* (i.e. "among others"). It follows therefrom that the European legislator chose to explicitly list "conditionality" as an instructive example of a non-freely given consent. In addition, the word "*inter alia*" refers to other cases rather than the case of conditionality.

Drawing on this guidance, we deduce the requirement that the consent request should not be merged into a contract or terms of service, as depicted in the requirement box.

Requirement	No merging into a contract
Description	A request for consent cannot be merged into a contract
Violation	When both consent and a contract (for which consent is not needed) are merged

Example. Figure 7 represents a case of a bundled consent request where the website offers news service, provided by the Washington Post website, and requests consent of the user.

Figure 7 Violation of the requirement "No merging into a contract" by Washington Post website (homepage www.washingtonpost.com/gdpr-consent/?noredirect=on&utm_term=.3161abcd072b accessed 17 May 2019)

How to detect violations? Please see section 6, where we describe automatic means that can be used to assess all the language-based requirements, including "No merging into a contract".

R4 Non detrimental – the case of tracking walls

A freely given consent also implies the consent request to be non-detrimental. "Detrimental consent" refers to the case where the data subject is unable to refuse or withdraw consent without detriment, which means facing significant negative consequences (Recital 42 of the GDPR). For the purposes of this paper, detrimental practices occur is different situations, suchlike:

- When users, even before expressing any choice, face a *tracking wall* blocking access to an online service's content (e.g. stating: "to access our site you must agree to our use cookies");
- When users, after refusing tracking cookies, have denied access to the webpage they want to consult, or the user is redirected to another website, or service is downgraded[79];
- Paid services or extra costs.[80]

The first listed practice refers to the appearance of a barrier page and is known by the designation of *tracking wall* (also known as *cookie wall, take-it-or-leave-it-choices* approaches). A tracking wall means that users who do not accept tracking across other sites will be denied access to the websites they seek to access.[81] It occurs when both of the two following conditions hold:

- the consent banner blocks access to the website,
- the banner gives only the option to accept, without any option to refuse.

However, users should have the possibility to refuse cookies and still be able to browse the page.[82] As mentioned in Section 5.2, if certain cookies are not necessary for the services requested and only provide for additional benefits of the website operator, the user should be in a position to refuse them (29WP 208).[83]

Recently, the EDPB[84] (05/2020) established that tracking walls are invalid and complements this claim with an example:

> "In order for consent to be freely given, access to services and functionalities must not be made conditional on the consent of a user to the storing of information, or gaining of access to information already stored, in the terminal equipment of a user (so called cookie walls). Example: A website provider puts into place a script that will block content from being visible except for a request to accept cookies and the information about which cookies are being set and for what purposes data will be processed. There is no possibility to access the content without clicking on the "Accept cookies" button. Since the data subject is not presented with a genuine choice, its consent is not freely given. This does not constitute valid consent, as the provision of the service relies on the data subject clicking the "Accept cookies" button. It is not

policy and to the terms as a whole, which in fact cover all the "services", that the controller offers e.g. YouTube, Chrome Browser, Google Services, Google Maps, Google Search, Google News, Gmail, AdWords, as well as several other services", NOYB, "Complaint filed against Google LL" (2018) https://noyb.eu/wp-content/uploads/2018/05/complaint-android.pdf accessed 7 May 2020.

77 Danish DPA, "Guide on consent" (2019) www.datatilsynet.dk/media/6562/samtykke.pdf accessed 7 May 2020 (hereafter named "Danish DPA Guide").

78 cf. 29WP (WP259 rev.01) (n 4) 8.

79 cf. 29WP (WP259 rev.01) (n 4) 11.

80 Regarding extra costs, such an obligation could foster social/economic discrimination (i.e. the rich, who can pay to protect their privacy, and the poor, who cannot) which would run against the universal nature of the fundamental rights to privacy and data protection. Forcing websites to offer a paid subscription service could also interfere with the development of new innovative business models that might be advantageous to consumers.

81 cf. EDPS Opinion (n 19) 17.

82 Ronald Leenes, "The Cookiewars: From regulatory failure to user empowerment?" (2015) in M. van Lieshout, & J-H. Hoepman (Eds.), *The Privacy & Identity Lab: 4 years later*, 3, The Privacy & Identity Lab, Nijmegen (2015) 31-49.

83 cf. 29WP (WP208) (n 17) 6.

84 EDPB, "Guidelines 05/2020 on consent under Regulation 2016/679" (2020) https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf (henceforth called "EDPB 05/2020").

presented with a genuine choice."

The ePrivacy Directive refers to the "*conditional access to website content*" in Recital 25. It states: "access to specific website content may be made conditional on the well-informed acceptance of a cookie or similar device, if it is used for a legitimate purpose". A literal interpretation of this excerpt apparently legitimizes conditional access to a website and this literal reading is sometimes used to justify the use a cookie wall.[85]-[86] Notably, this interpretation derives from an incorrect analysis of this Recital, for it makes access to a website conditional on the acceptance of cookies[87], and such conditionality renders a non-freely given consent. In this regard, the 29WP (WP126)[88] recommends clarification or review of this Recital. In the 29WP (WP 240) understanding, these take it or leave it approaches rarely[89] meet the requirements for freely given consent. It specifically stated that "if the consequences of consenting undermine individuals" freedom of choice, consent would not be free. The Working Party invites the EC to develop a specific prohibition on such "*take it or leave it*" choices with regard to electronic communications, where such choices would undermine the principle of freely given consent."

The resulting analysis, also consolidated by the positioning of the majority of the stakeholders shown in the next Section 5.2.1, sustains that websites need to give access to content when a user does not consent to BTT beyond strictly necessary to provide the service, and hence, consent requests should not present a tracking wall.

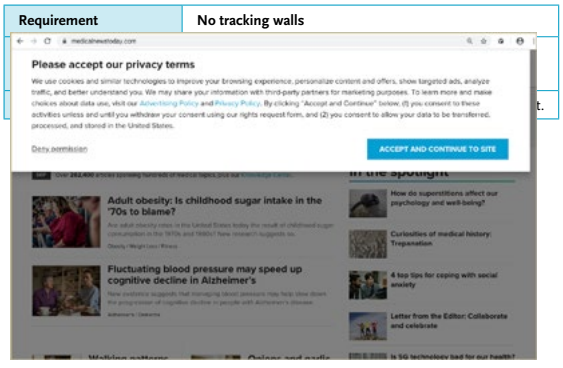


Figure 8 Violation of the requirement "No tracking wall" by the MedicalNewsToday website (www.medicalnewstoday.com accessed on 25 September 2019)

[85] cf. Kosta (n 20) 1.

[86] Frederik Borgesius, Sanne Kruikemeier, Sophie Boerman and Natali Helberge, "Tracking Walls, Take-It-Or-Leave-It Choices, the GDPR, and the EPrivacy Regulation" (2017) *European Data Protection Law Review*, Volume 3, Issue 3, 353-368.

[87] cf. Leenes (n 82).

[88] The 29WP states that "the last paragraph of Recital 25, stipulating that access to specific website content may be made conditional on the acceptance of a cookie, might be contradictory with the position that the users should have the possibility to refuse the storage of a cookie on their personal computers and therefore may need clarification or revision", Article 29 Working Party, "Opinion 8/2006 on the review of the regulatory Framework for Electronic Communications and Services, with focus on the ePrivacy Directive" (WP 126, 26 September 2006) 3.

[89] The 29WP identifies five circumstances in which forced consent should be specifically prohibited, namely: 1. Tracking on websites, apps and or locations that reveal information about special categories of data. 2. Tracking by unidentified third parties for unspecified purposes. 3. All government funded services; 4. All circumstances identified in the GDPR that lead to invalid consent; 5. Bundled consent for processing for multiple purposes. Finally, the 29WP alerts to the position of news media, since they seem to be the heaviest users of tracking cookies and cookie walls, see 29WP (WP240) (n 32) 17.

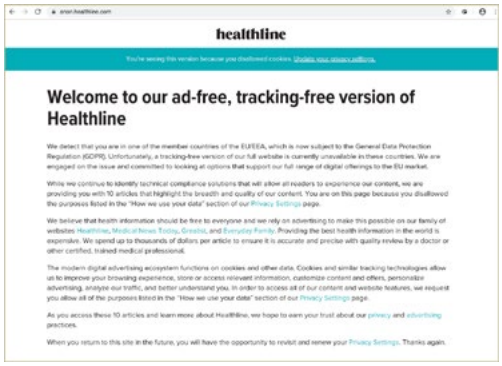


Figure 9 Result of denying consent on MedicalNewsToday website (www.medicalnewstoday.com/ accessed 25 September 2019)

Examples. Figure 8 shows an example of a cookie wall on the MedicalNewsToday website. When the page first loads, the website prevents a visitor from viewing any other page unless the user clicks the displayed Accept and continue to site button. Figure 9 shows the resulting page after the user clicked on the Deny permission link: the website only provides access to 10 articles, preselected by the website (and not the article requested by the user). The banner above reminds the user that he has a limited access to the website because he disallowed cookies and proposes to update the privacy settings.

How to detect violations? Detection of such a violation is possible manually. The user needs to understand whether a website allows the user to access it without expressing consent and whether there is an option to refuse consent. We also consider a violation of this requirement when refusing consent leads to a restrictive access to the service, as in the example of MedicalNewsToday website.

5.2.1 Stakeholders positioning on tracking walls

There is some inconsistency in the positions taken by EU DPAs and other stakeholders on whether a tracking/cookie wall consists in a violation of a valid consent. The European Data Protection Supervisor,[90] the European Parliament[91], and the Bureau Européen des Unions de Consommateurs[92] (BEUC) are of the opinion that tracking walls and any other type of detrimental rendering of consent should be forbidden, as the GDPR mandates. Noyb.eu[93] filed four complaints over forced consent against Google, Instagram, WhatsApp and Facebook.

The CNIL in its current guidelines[94] for cookies (article 3) considers that consent can only be valid if the concerned person is able to

[90] cf. EDPS Opinion (n 19) 17.

[91] In the Proposal for the ePrivacy Regulation of the European Parliament, it is proposed that "the Regulation should prevent the use of so- called "cookie walls" and "cookie banners" that do not help users to maintain control over their personal information and privacy or become informed about their rights", Draft European Parliament Legislative Resolution www.europarl.europa.eu/doceo/document/A-8-2017-0324_EN.html?redirect accessed 7 May 2020.

[92] BEUC Position Paper, "Proposal For A Regulation On Privacy And Electronic Communications (E-Privacy)" (2017) www.beuc.eu/publications/beuc-x-2017-059_proposal_for_a_regulation_on_privacy_and_electronic_communications_e-privacy.pdf accessed 7 May 2020.

[93] NOYB, "GDPR: noyb.eu filed four complaints over "forced consent" against Google, Instagram, WhatsApp and Facebook" (2018) https://noyb.eu/wp-content/uploads/2018/05/pa_forcedconsent_en.pdf accessed 7 May 2020.

[94] CNIL, Guidelines on cookies and other tracers (n 36).

validly exercise his choice and does not suffer major inconvenience in the absence or withdrawal of consent. In its draft recommendation on the use of cookies[95], this authority proposes that users should not be exposed to any prejudice should they decide not to accept cookies. It states that "the ability to express refusal as easily is indeed the counterpart of the ability to express free consent".

The Greek DPA[96] confirms that the website should not provide for "no option for declining/rejecting cookies and trackers".

The Irish DPA[97] confirms that a banner merely giving "the user the option to click "accept" to say yes to cookies and which provides no other option is not compliant. This means banners with buttons that read "ok, got it!" or "I understand", and which do not provide any option to reject cookies or to click for further, more detailed information, are not acceptable and they do not meet the standard of consent required".

The ICO[98] in its recent guidance states that consent which is forced via a cookie wall is "unlikely to be valid". However, it also notes that the GDPR must be balanced against other rights, including freedom of expression and freedom to conduct a business. The ICO seems to adopt a wait and see approach, as it argues that:

> In some circumstances, this approach is inappropriate; for example, where the user or subscriber has no genuine choice but to sign up. (...) If your use of a cookie wall is intended to require, or influence, users to agree to their personal data being used by you or any third parties as a condition of accessing your service, then it is unlikely that user consent is considered valid.

The Dutch DPA[99] published on its website in December 2019 its viewpoint that websites must remain accessible when refusing tracking cookies and that cookie walls are not permitted under the GDPR. It adds that with a cookie wall, websites, apps or other services cannot receive valid permission from their visitors or users. The regulator explains that the inspected websites are involved in an ongoing investigation into cookie walls. Alongside, the Minister for Legal Protection[100] of the Netherlands adverts that when a website is visited, the visitor cannot be denied access to the content of the website if he does not agree with the placement of the cookies (cookie wall). Only functional cookies and non-privacy sensitive cookies do not need permission. It states further that the government is arguing in the European Council for a ban on cookie walls in the new ePrivacy Regulation.

In the same light, the Belgian DPA[101] states that blocking a user's access to a website, on the basis that the user did not consented to cookies, is not a compliant solution.

The German DPA[102] contends that a visit to a website should still be possible if data subjects decide against the setting of cookies. The same reasoning is upheld by the Danish DPA.[103]

Conversely, the Austrian DPA[104] issued a decision on 30 November 2018, pronouncing that consent was freely given via a cookie wall in the case of an Austrian newspaper, "*Der Standard*", that gave users the option to either: i) accept cookies and receive full access to the website; ii) refuse cookies and receive a limited access to the website; or iii) pay a fee for a monthly subscription without accepting cookies. The authority indicated that cookie walls are not prohibited because the newspaper's own settings provide a degree of choice. First, *Der Standard* only places cookies after the user makes an informed decision to allow the placement of cookies. Second, the individual can withhold consent by either entering into a paid subscription or leaving *Der Standard*'s website. Third, the DPA considered *Der Standard*'s prices to be "not unreasonably high." In fact, giving consent to cookies results in a positive outcome for the individual, because they gain unlimited access to the newspaper's articles. The Austrian DPA did not, however, discuss what would happen if an individual withdrew their consent to the usage of cookies.

The Spanish DPA[105] recognizes as a valid practice the blocking access to the website if a user rejects consent, as depicted in the excerpt below (when information duties were duly complied with). We consider this scenario being equivalent to a tracking wall because the user is not able to access the service unless she gives her consent:

> "In certain cases, not accepting cookies shall entail being entirely or partially prevented from using the service; users must be appropriately informed of this situation. However, access to services may not be denied due to cookie refusal in those cases in which such refusal prevents the user to exercise a legally recognised right, since such website is the only means provided to users to exercise such rights".

Borgesius et al., in their commissioned study[106] on the Proposal for the ePrivacy Regulation, mentioned a circumstance catalogue composed of a non-exhaustive *blacklist* of circumstances in which tracking walls are banned (list of illegal practices), supplemented with a *grey list* (practices presumed to be illegal). The study refers that if a situation is on the grey list, there is a legal presumption that a tracking wall makes consent involuntary, and therefore invalid. Hence, the legal presumption of the grey list shifts the burden of proof. For example, for situations on the grey list, it is up to the company deploying the cookie wall to prove that users gave a free consent,

95 CNIL, On the practical procedures for collecting the consent provided for in article 82 of the French data protection act, concerning operations of storing or gaining access to information in the terminal equipment of a user (recommendation "cookies and other trackers") https://www.cnil.fr/sites/default/files/atoms/files/draft_recommendation_cookies_and_other_trackers_en.pdf (January 2020), (hereafter named "CNIL draft recommendation 2020").

96 cf. Greek DPA (n 42).

97 cf. Irish DPA Guidance (n 37).

98 cf. ICO Guidance (n 26) 31.

99 cf. Dutch DPA "Cookies" (n 46); and "Many websites incorrectly request permission to place tracking cookies" (2019) https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-veel-websites-vragen-op-onjuiste-wijze-toestemming-voor-plaatsen-tracking-cookies accessed 7 May 2020.

100 House of Representatives of the Netherlands, "Answer to questions from members Middendorp and Van Gent about a possible cookie wall ban" (2019) www.tweedekamer.nl/kamerstukken/kamervragen/detail?id=2019D49667&did=2019D49667 accessed 7 May 2020.

101 Belgian DPA, (2020) "Guidance Materials and FAQs on Cookies and Other Tracking Technologies" https://www.autoriteprotectiondonnees.be/recueillir-valablement-le-consentement-des-personnes-concernees, accessed 7 May 2020 (henceafter named "Belgian DPA Guidance").

102 cf. German DPA Guidelines (n 10).

103 cf. Danish DPA Guide (n 77).

104 Austrian DPA decision on the validity of consent (2018) www.ris.bka.gv.at/Dokumente/Dsk/DSBT_20181130_DSB_D122_931_0003_DSB_2018_00/DSBT_20181130_DSB_D122_931_0003_DSB_2018_00.pdf accessed 7 May 2020.

105 Spanish DPA "Guide on the use of cookies" (2019) www.aepd.es/media/guias/guia-cookies.pdf accessed 7 May 2020 (author's translation of the Spanish version) (henceforth named "Spanish DPA Guide").

106 Frederik Borgesius, Joris van Hoboken, Ronan P. Fahy, Kristina Irion, Max Rozendaal, "An Assessment of the Commission's Proposal on Privacy and Electronic Communications" (Study for the LIBE Committee. Brussels: European Parliament, Directorate-General for Internal Policies, Policy Department C: Citizens' Rights and Constitutional Affairs, Chapter 3.5.5, 2017) www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2017)583152 accessed 7 May 2020.

even though the company installed a tracking wall. We are instead of the opinion of a complete ban to cookie walls.

Further developments need to be consolidated through case law from the European Court of Justice. In addition, businesses using tracking/cookie walls to obtain consent may want to consider preemptively streamlining their method for obtaining consent (e.g. by switching to a cookie banner that allows to refuse consent). Table 10 summarizes the different positionings made public from some stakeholders.

Table 10 Positioning of stakeholders on cookie walls

Stakeholders	Positioning on tracking wall
EDPB, EDPS, BEUC, EU Parliament, DPAs: Dutch, French, German, Danish, Greek, Irish, Belgian	Violation of a freely given consent
UK, Spanish DPAs	Not clear
Austrian DPA	Valid consent

5.3 Specific

Specific consent involves granularity of the consent request in order to avoid a catch-all purpose acceptance. In Table 11 and the following subsections we further specify this requirement.

Table 11 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-Level	**Low-level**	**Binding**	**Non-binding**	**Interpretation**
Specific	R5 Separate consent per purpose	4(11), 6(1)(a); Planet 49 ruling	29WP, Recitals 32, 43, all DPAS	–

R5 Separate consent per purpose

The request for consent should be granular in the options for consenting to cookies, so that the user is able to give consent for an independent and *specific purpose* (29WP WP208).[107] This reasoning is given by the following recitals of the GDPR. Recital 43 clarifies the need for a separate consent for different processing operations. Recital 32 of the GDPR states that consent should be given per purpose (or set of purposes). The provision is worded as follows: "consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them".

This element of a specific consent relates to the *purpose limitation principle* observed in Article 5(1)(b) of the GDPR. Therein rely two elements: i) data must be collected for specified, explicit and legitimate purposes only; and ii) data must not be further processed in a way that is incompatible with those purposes. This Article reads: "personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes [...] ("purpose limitation")".

In this same line, the 29WP (WP 203)[108] analyzes this principle of "purpose limitation" and explains that any purpose must be *specified*, i.e. be precisely and fully identified. The 29WP (WP259 rev.01)[109] additionally comments on the needed consent for each purpose to comply with the conditions of a valid consent:

> "Data subjects should be free to choose which purpose they accept, rather than having to consent to a bundle of processing purposes. (...) If the controller has conflated several purposes for processing and has not attempted to seek separate consent for each purpose, there is a lack of freedom. This granularity is closely related to the need of consent to be specific. (...) When data processing is done in pursuit of several purposes, the solution to comply with the conditions for valid consent lies in granularity, i.e. the separation of these purposes and obtaining consent for each purpose".

The 29WP (WP259 rev.01) instructs further that "a controller that seeks consent for various different purposes should provide a separate opt-in for each purpose, to allow users to give specific consent for specific purposes".

Planet49 Judgment of the Court of Justice of the EU[110] determined that *specific* consent means that "it must relate specifically to the processing of the data in question and cannot be inferred from an indication of the data subject's wishes for other purposes". This means that consent should be granular for each purpose of processing.

The resulting analysis sustains that the banner should present each purpose separately (but also, it should allow accepting or rejecting each purpose separately), as depicted in the requirement box.

Requirement	Separate consent per purpose
Description	Consent should be separately requested for each purpose.
Violation	General consent request under conflated or bundled purposes; a user shall not agree or disagree to all at once (Danish DPA[111])

Examples. Figure 10 shows the wordreference.com website, where a user cannot give consent per purpose, but instead is presented with a "Learn More & Set Preferences" link that only allows to give consent per third party. Figure 11 shows (a part of) the list of vendors, which is several-screen-long and is obviously overwhelming and not usable for an average user. Figure 12 outlines the Dailymail website banner, which conflates together different data processing purposes (e.g. personalization, ad selection, content selection and measurement) under a single acceptance request, therefore violating the requirement that consent should be given per purpose. On the other hand, Figure 13 depicts a compliant design banner from the senscritique.com website.

How to detect violations? A human operator can observe violations with no technical support. However, it would be possible to detect such violations automatically if the user interface of consent banner had a standardized design, which is not the case nowadays.

5.3.1 Consent not required per cookie, per publisher, per third party

Under the following three subheadings we add the observation that a request for consent per purpose *does not* include a request: *per cookie, per publisher, nor per third-party*, for the reasons explained below.

Not per cookie. We argue that the requirement of granular purposes does not mandate that the consent request should be provided for each cookie. We claim that the consent request for each cookie is not user-friendly and it might be too overwhelming for users.

[107] cf. 29WP (WP208) (n 17) 3.
[108] Article 29 Working Party, "Opinion 03/2013 on purpose limitation" (WP 203, 2 April 2013).
[109] cf. 29WP (WP259 rev.01) (n 4) 11.
[110] Planet49 Case (n 11).
[111] cf. Danish DPA Guide (n 77).

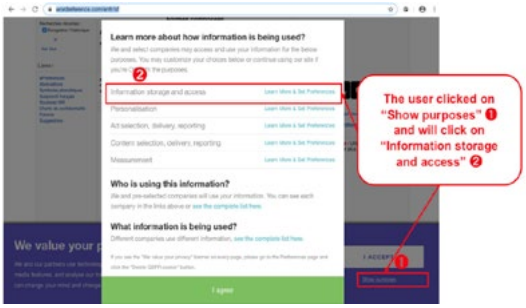


Figure 10 A settings accessible from the cookie banner (www.wordreference.com/enfr/sf accessed on 24 September 2019)

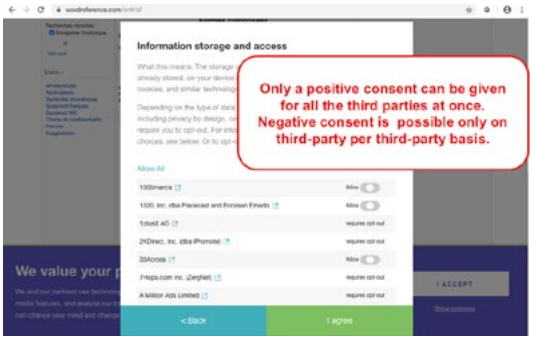


Figure 11 The cookie banner does not allow to refuse consent for all third parties at once, only on a "per third party" basis. (www.wordreference.com/enfr/sf accessed 24 September 2019)

Moreover, few users are familiar with the concept of cookies and tracking technologies. Therefore, this may lead to certain choices as a consequence of users' lack of knowledge. We derive this conclusion from several bases. The text of the Recital 25 of the ePD states that the cookie consent request covers its further uses, insofar as these uses are compatible with the initial purposes for which the consent is provided. The 29WP (WP208)[112] mentions that each website could prominently display a link to a location where all cookies used by the website are presented through types (and hence, not per cookie). In the same line, the ICO[113] gives the same reasoning when referring to cookie categories:

> Some sites might use tens or even hundreds of cookies and therefore it may also be helpful to provide a broader explanation of the way cookies operate and the categories of cookies in use. For example, a description of the types of things you use analytics cookies for on the site will be more likely to satisfy the requirements than simply listing all the cookies you use with basic references to their function.

The Belgian[114], Irish[115] and Danish DPAs[116] accord that consent does not need to be given per cookie, but instead per purpose. The latter refers an example of a specific consent per purpose (and not per cookie):

> [A] website has a cookie pop-up in which the user can accept or decline cookies by purpose, i.e. the user can freely decide whether he or she wants functional, statistical and/or marketing cookies to be set by the website. The user can easily toggle cookies by purpose on and off. Then the website's cookie consent is specific.

Not per publisher. The need of a separate and renewed consent per publisher is also discussable: if one publisher receives consent, it is questionable that it might share the consent with other publishers. In this regard, we refer to the case law of the European Court of Justice and adapt its reasoning to our consent-cookie request context. The Court (in its two decisions of Tele 2 and Deutsche Telekom[117] in the context of electronic public directories), refers to the extension of the initial consent to the subsequent processing of the data by third-party companies, provided that such processing pursues that same purpose, and that the user was informed thereof. The Court holds that where a user consented to the passing of his personal data to a given company, the passing of the same data to another company, with the same purpose and without renewed consent from that user, does not violate the right to protection of personal data. The Court adds that a user will generally not have a selective opinion to object to the sharing of the same data through another, yet similar, provider. From these arguments, we conclude that there is no need for a separate and renewed consent per publisher whenever further processing follows that same purpose, and the user was informed thereof. In these cases, consent could be shared with other publishers.

Not per third party. We believe a fine-grained customization per third party is not required. In fact, showing the full advertisers list configures a deceptive design. The 29WP (WP259 rev.01)[118] suggests that the categories of third parties who receive personal data and wish to rely upon the original consent should be listed by category (or be individually named).

It is possible to conclude that a consent request does not require the user's consent for third-party cookies, but only aims to inform users of third-party cookie usage, or third party access to data collected by the cookies on the website: "necessary information would be the purpose(s) of the cookies and, if relevant, an indication of possible cookies from third parties or third party access to data collected by the cookies on the website" 29WP (WP208).[119]

The Italian DPA[120] adopted the same reasoning and postulated that "publishers may not be required to include, on the home page of their websites, also the notices relating to the cookies installed by third parties via the publishers' websites".

5.4 Informed Consent

Whenever Browser-based Tracking Technology (BTT) are accessed or stored on a user's device, the user must be given clear and comprehensive information, and the content information must comprise what is accessed or stored, the purposes and means for expressing

[112] cf. 29WP (WP208) (n 17) 3 and 5.
[113] cf. ICO Guidance (n 26) 10.
[114] cf. Belgian DPA Guidance (n 101).
[115] cf. Irish DPA Guidance (n 37).
[116] cf. Danish DPA (n 77).
[117] C543/09 Deutsche Telekom AG v Bundesrepublik Deutschland [2011] EU:C:2011:279, para 62 to 65; and C-536/15 Tele2 (Netherlands) BV and Others v Autoriteit Consument en Markt (ACM) [2017] ECLI:EU:C:2017:214.
[118] cf. 29WP (WP259 rev.01) (n 4) 14.
[119] cf. 29WP (WP208) (n 17) 3 and 5.
[120] Italian DPA, "Simplified Arrangements to Provide Information and Obtain Consent Regarding Cookies" (2014) www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3167654 accessed 7 May 2020.

their consent, pursuant to Article 5(3) of the ePD.

The need to present information on the processing operations is triggered by the principles of lawfulness, fairness, and transparency depicted in Article 5(1)(a) and the recitals of the GDPR. In particular, Recital 60 explains that "a data controller should provide a data subject with all information necessary to ensure fair and transparent processing, taking into account the specific circumstances (...)".

The rationale behind the requirement to provide information relies in the premise that providing it puts the user in control of the data on their own device. As argued by the General Advocate Szpunar,[121] the data subject must be informed of all circumstances surrounding the data processing and its consequences: "crucially, he or she must be informed of the consequences of refusing consent", including a reduced service. He proceeds by asserting that "a customer does not choose in an informed manner if he or she is not aware of the consequences". The 29W (WP131)[122] envisioned that the data subject's consent is "based upon an appreciation and understanding of the facts and implications of an action". The judgment of the Court of Justice of the EU on the Planet49 case[123] elucidated that providing "clear and comprehensive" information means "that a user is in a position to be able to **determine easily the consequences of any consent** he might give and ensure that the consent given is well informed". It follows therefrom that information must be also "**clearly comprehensible** and **sufficiently detailed** so as to enable the user to comprehend the **functioning of the cookies** employed".

Regarding the *timing* to deliver information, it should be concomitant to the time and place when consent is requested. As posited by the 29WP (WP208), information should be provided "at the time and place where consent is sought, for example, on the webpage where a user begins a browsing session (the entry page). As such, when accessing the website, users must be able to access all necessary information".

From the analysis of the legal provisions, the 29WP guidance and the mentioned case-law, we derive both the approach and the content of the information:

- the *approach* to disclose information, which happens generally with the presence of a privacy or cookie policy (section R6);
- the *content* of the information to be given on BTT. We focus on the information requirements where personal data is collected from the data subject (Article 13 (1) (2) GDPR). The position of the 29WP (260 rev.0.1) is that there is no difference between the status of the information to be provided under sub-article 1 and 2 of Article 13 and therefore all of the information across these sub-articles is of equal importance and must be provided to the data subject. For readability purpose, we decompose the information requirements:
  - necessary information on BTT (section R7),
  - information on consent banner configuration (section R8),
  - information on the data controller (section R9),
  - information on rights (section R10).

Table 12 depicts the information low-level requirements.

[121] Opinion of Advocate General Szpunar. Opinion of the Case C-61/19 Orange România SA v Autoritatea Na ional de Supraveghere a Prelucr rii Datelor cu Caracter Personal (ANSPDCP), 2020. ECLI:EU:C:2020:158.

[122] 29WP Working Document on the processing of personal data relating to health in electronic health records (EHR) (WP 131, 15 February 2007) 8.

[123] cf. Planet49 Case (n 11) para 74.

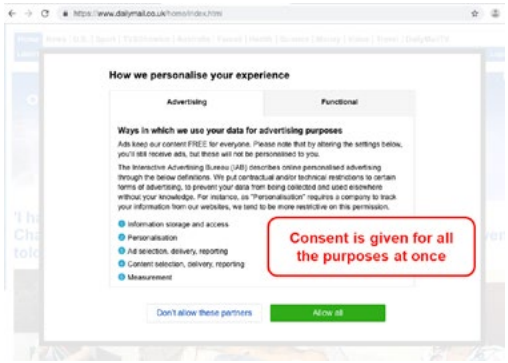


Figure 12 Non-compliance with the "separate consent per purpose" requirement (www.dailymail.co.uk/home/index.html accessed on 17 May 2019)

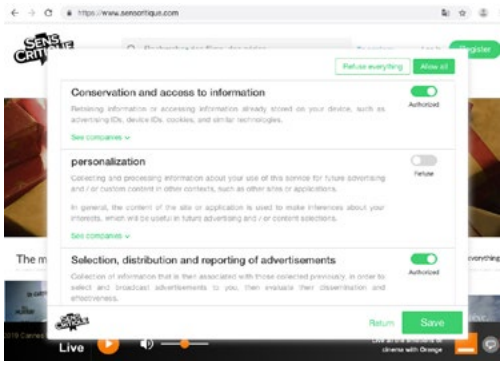


Figure 13 Compliance with the "separate consent per purpose" requirement (www.senscritique.com/ accessed on 18 May 2019)

Table 12 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-level	**Low-level**	**Binding**	**Non-binding**	**Interpretation**
Informed	R6 Accessibility of the information page	-	29WP, DPAs: Irish, German, Belgium, Finnish	-
	R7 Necessary information of BTT:		DPAs: German	
	R7a Identifier name	-	29WP; DPAs: Irish Danish	-
	R7b Purposes	Art.13 (1)(c)	(most DPAs)	-
	R7c Third parties with whom cookies are shared	Art. 13 (1)(e), Planet 49	(most DPAs)	-
	R7d Duration of cookies	Art. 13 (2)(a); Planet 49	DPAs: Greek, CNIL, Finnish	-
	R8 Information on consent banner configuration		29WP, DPAs: UK, Danish, Irish	-
	R9 Information on the data controller	Art. 13 (1)(a)(b)	DPAs: Danish, Irish	-
	R10 Information on rights	Art. 13 (1)(f), (2) (a-f)	almost all DPAs	-

R6 Accessibility of information page

On the recommended approach, the 29WP (WP208) proposes a visible notice displaying a link to an information page (also known as cookie policy, or entry page) where information on BTT is presented (preferably through a layered approach). The 29WP considers the following built-in possibilities for rendering information:

- the mechanism should provide a visible notice on the use of BTT;
- a link to a designated location where all types of BTT used by the website are presented should be prominently displayed;
- Information should be provided in a layered approach[124], typically through a link (or series of links), where the user can find out more about types of BTT being used.

As for the DPAs' understanding, the Finnish DPA[125] states that no separate pop-up window is required for informing the user, and a cookie policy must also be mentioned on the website so that the user can learn more about it. The Irish DPA[126] advises website publishers to include a link or a means of accessing further information about the use of cookies. We have defined the requirement prescribing that the information page (cookie policy) on the use of BTT should be accessible through a banner, with a clickable link.

Requirement	Accessibility of information page
Description	The information page should be accessible through a cookie banner, via a visible link or a button
Violation	Inexistence of an information page

Examples. Figure 14 shows a compliant cookie banner example, where the "information page" is accessible through a link.

Figure 14 Compliant example. The "information page" should be accessible through a cookie banner (via a link or a button) (https://www.lemonde.fr accessed on 18 May 2019)

How to detect violations? A manual analysis of a cookie banner content is enough to identify whether a link to the privacy policy is accessible, however such analysis is not scalable when numerous websites must be audited. Moreover, different users may find it more or less difficult to find the same link on a concrete cookie banner. Hence, user studies are needed to evaluate how accessible information for a given target audience is. We provide more details on user studies in Section 6.

From a technical perspective, it is possible to crawl all the links present on a banner interface and detect whether such links lead to a privacy policy with a rather simple detection based on keywords. While this method could produce inaccurate results, it could be usable in a legal procedure after a manual verification. The same analysis holds for all the following requirements in this section.

Computer science researchers have already used keyword-based approaches: Degeling et al.[127] analyzed the availability of privacy policies on the top 500 websites before and after GDPR came in force, while Libert[128] detected and analyzed over 200,000 websites' privacy policies. More sophisticated approaches based on Natural language processing (NLP) can also be applied to analyze whether a page is a privacy policy. We refer to further discussion on NLP in Section 6.

R7 Necessary information on BTT

Regarding the content of the information to be given on BTT, both the 29WP (WP208) and the recent Planet 49 judgment[129] set the necessary[130] information to be disclosed to ensure fair and transparent processing. Planet 49 (in paragraph 76) explicitly quotes Article 13 of the GDPR on the list of information to be provided. The necessary information specific to BTT is the following:

Purposes of processing (Article 13 (1)(c)), and their legal basis for each specific processing (under Article 6 GDPR);[131]

Recipients, or categories of recipients, with whom personal data is shared (Articles 4(9) and 13(1)(e)), which can consist of other data controllers, joint controllers, processors and third-party recipients. Planet 49 ruling determines (in paragraph 80) "third parties with whom the cookies are shared with"). The 29WP (WP202)[132] adds that information is needed on whether the data may be reused by other parties, and if so, for what purposes. Regarding the categories of recipients, the 29WP (260 rev.0.1) suggests that if controllers opt to provide the categories of recipients, the information should be as specific as possible by indicating the type of recipient (i.e. by reference to the activities it carries out), the industry, sector and sub-sector and the location of the recipients;

[124] Interestingly, the EDPS recommends a layered approach, where the information is given at different stages, providing greater detail. The essential information should be present at a sufficient level of detail to already put the user in control at the first layer. A notice providing (the reference to) the first level of information on cookies must be clearly visible to web service users whatever their landing page is. Further, the EDPS strongly recommends that EU institutions provide information on cookies on the web service under their control and not rely on external sources. If, for some reasons, the institution uses external sources, they should set up measures to manage relevant risks, where possible, cf. EDPS Guidelines (n 35) 15.

[125] cf. Finnish DPA (n 13).

[126] cf. Irish DPA Guidance (n 37).

[127] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz, "We value your privacy ... now take some cookies: Measuring the GDPR's impact on web privacy," in *NDSS*, 2019.

[128] Timothy Libert, "An Automated Approach to Auditing Disclosure of Third-Party Data Collection in Website Privacy Policies" in *Proceedings of the 2018 World Wide Web Conference*, p 207-216.

[129] cf. Planet49 Judgment (n 11).

[130] Article 13 sets a distinction between necessary/essential information (1) and possible "further information" (2) which should be provided only to the extent that is necessary to guarantee fair processing having regard to the specific circumstances in which the data are collected. In case of BTT, further informational elements are required as necessary information (cookie name and their duration, third party sharing and their purposes) to ensure transparent processing. Such distinction between necessary and further information is analyzed further in the 29WP Opinion 10/2004 (WP 100) on More Harmonised Information Provisions.

[131] Article 13 (1) (d) posits that when the controller relies on legitimate interests as a legal basis for processing, it should inform the data subject about the interests. The 29WP (29WP 260 rev.01 n 4) adds that at least upon request, provide data subjects with information on the balancing test.

[132] 29WP Opinion 02/2013 (WP202) on apps on smart devices, adopted on 27 February 2013.

Storage period, interpreted from Article 13 (2)(a). It is explicitly stated in the Planet 49 ruling (paragraph 79) that the "period for which the personal data will be stored, or if that is not possible, to the criteria used to determine that period". The 29WP (260 rev.0.1) declares that this information is linked to the data minimization principle in Article 5(1)(c) and storage limitation requirement in Article 5(1)(e). It adds that the different storage periods should be stipulated for different categories of personal data and/or different processing purposes, including where appropriate, archiving periods;

Identifier name and a responsible party for setting it, as proposed by the 29WP (WP208).[133]

Some DPAs have defined the duration/lifespan for BTT, such as the ones depicted in Table 13:

Table 13 Positioning of stakeholders on the lifespan of BTT

DPAs	Lifespan of BTT
29WP (WP194)	"Cookies exempted of consent should have a lifespan that is in direct relation to the purpose it is used for and must be set to expire once it is not needed, taking into account the reasonable expectations of the average user or subscriber".
CNIL	"analytic tracers must not have a lifespan exceeding thirteen months and this duration should not be automatically extended during new visits. The information collected through the tracers must be kept for a maximum of twenty-five months".
ICO	Lifespan of cookies must be proportionate in relation to the purpose and limited to what is necessary to achieve it.
Belgium	Cookies should be set to expire as soon as they are no longer needed, taking into account the reasonable expectations of the user. Cookies exempt from consent will therefore probably be set to expire when the browsing session ends, or even before.
Spain	Lifespan of cookies must be proportionate in relation to the purposes for which they are intended.
Irish	Lifespan of a cookie must be proportionate to its function.

The Greek DPA[134] reiterated that information should include the used tracking categories. The banner (either in the form of a pop-up window or otherwise) should provide specific information for each tracker or tracker category of the same purpose.

The CNIL adds that the categories of data collected through trackers could be specified for each purpose in a way that is easily accessible to the user. This becomes particularly important when the special categories of personal data are used.

Some DPAs advocate that all cookies should – as a best practice – declare their purpose. The UK, Greek, Finnish and Belgian DPAs endorse as a good practice disclosure of clear information about the purposes of cookies, including strictly necessary ones. The guidance of the 29WP (WP188)[135] notes that although some cookies may be exempted from consent, they are part of a data processing operation, therefore publishers still have to comply with the obligation to inform users about the usage of cookies prior to their setting.

After listing the necessary information for an informed consent based on the legal sources, we define the low-requirement below:

Requirement	Information on BTT should contain:
Description	• Purposes • Recipients or categories of recipients with whom personal data is shared with and for what purposes • Storage period • Identifier name
Violation	Absence of any of these elements in the information page

Examples. Figure 15 renders a partially compliant example. The "information page" contains some of the required information for each cookie: cookie names, party who dropped the cookie, purposes and retention period. However, it does not show with which parties the cookies are further shared., Besides, for some cookies (e.g. _ga or _gid), the provided purposes are not explicit enough – i.e. description "to identify the user" does not explain for what concrete purpose this identification will be used. Figure 16 depicts a non-compliant example wherein the "information page" contains only groups of purposes (analytics, social, etc.) of cookies but does not provide detailed information on each cookie.

How to detect violations? Please see Section 6, where we describe automatic means that can be used to assess language-based requirements.

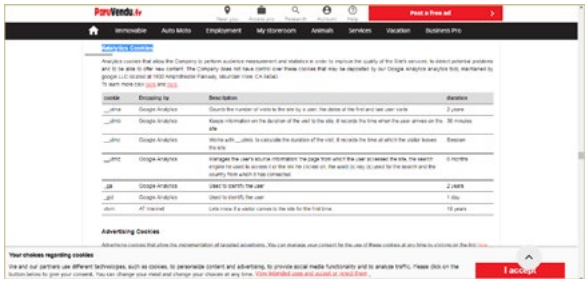


Figure 15 Partially-compliant example www.paruvendu.fr/communfo/default-communfo/defaultcommunfo/infosLegales#cookies (accessed on 18 May 2019)

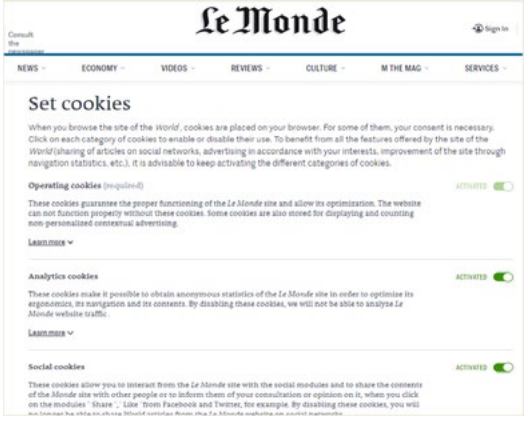


Figure 16 Non-compliant example: the information page contains only groups of purposes (analytics, social, etc.) of cookies but does not provide detailed information on the purpose for each cookie www.lemonde.fr/gestion-des-cookies (accessed on 18 May 2019)

[133] cf. 29WP (WP208) (n 17).
[134] cf. Greek DPA (n 42).
[135] Opinion 16/2011 on EASA/IAB Best Practice Recommendation on Online Behavioural Advertising, adopted December, 2011.

R8 Information on consent banner configuration

The 29WP (WP208)[136] instructs that information should refer to how the user can express his choice by accepting all-some-or-none BTT and how to change this choice afterwards through the settings. It states that "the ways they can signify their wishes regarding cookies i.e. how they can accept all, some or no cookies and to how change this preference in the future (...) and how to later withdraw their wishes regarding cookies". Accordingly, information on the possibility of a configuration of the user's preferences is designed as a low-level requirement presented in the requirement box.

Requirement	Information on consent banner configurations
Description	The banner or the "information page" should explain how the user can accept all, some or no BTT and how to change this preference in the future. For example, via banner's buttons or links.
Violation	Non-existence of information on configuration possibilities

Examples. Figure 17 depicts a non-compliant banner example. This banner, besides showing general purposes, does not give any information on how the user can accept all, some or no cookies and how to change this preference in the future. Figure 18 shows a compliant banner that explains how the user can configure his choices.

How to detect violations? Please see Section 6, where we describe automatic means that can be used to assess language-based requirements.

R9 Information on the data controller

Article 13 (1) (a) (b) of the GDPR establishes that the identity and contact details of the controller[137] and the Data Protection Officer (DPO)[138] (when legally obliged to appoint one) are part of the information list to be provided when personal data are collected from the data subject to enable the exercise of the data subject's rights toward the controller (or its representative).

We defined as a low-level requirement the need for the information page to incorporate the identity of the controller, contact details and whenever applicable, the representative.

Requirement	Information on the data controller
Description	The "information page" should contain, for each data controller: its identity, contact details, contact of the Data Protection Officer (DPO).
Violation	Absence of any reference about the data controller

How to detect violations? Please see section 6, where we describe automatic means that can be used to assess language-based requirements.

R10 Information on rights

Article 13(2) (a-f) of the GDPR stipulates the need to provide information on the rights of the users: access, rectification, erasure,

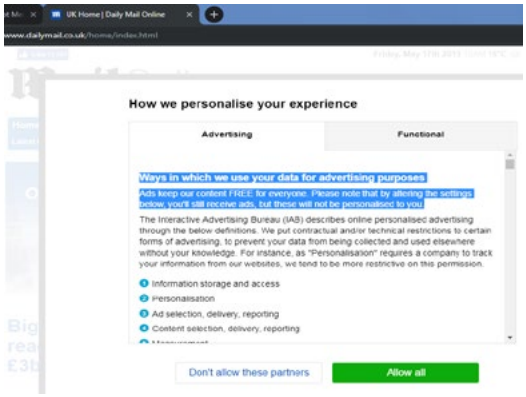


Figure 17 Violation example: this "information page" only renders general purposes (not requesting consent per purposes) nor any information on how the user can accept all, some or no cookies and how to change this preference in the future www.dailymail.co.uk/home/index.html (accessed on 18 May 2019)

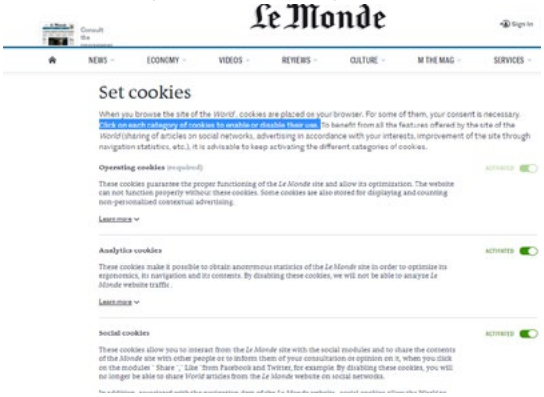


Figure 18 Compliant example banner: this "information page" explains how the user can configure his choices www.lemonde.fr/gestion-des-cookies (accessed on 18 May 2019)

restriction, object, portability, withdraw consent, lodge a complaint with a DPA, right to be informed about the use of data for automated decision-making and data transfers to a third country or an international organization (and the corresponding safeguards).

We have reproduced these rules into another low-level requirement as shown in the requirement box.

Requirement	Information about the data subject rights
Description	The "information page" should contain the user's rights: • access • rectification • erasure • restriction on processing • objection to processing • portability • withdraw consent • lodge a complaint with a DPA • informed on the use of data for automated decision-making • informed of data transfers to a third country or an intenational organization
Violation	Absence to any reference on the rights

[136] cf. 29WP (WP208) (n 17).

[137] The 29WP (WP 260) lists different forms of contact details of the data controller (e.g. phone number, email, postal address, etc.), and an electronic contact as well, as posited by NOYB, as the service provided is digital, see NOYB, é Report on privacy policies of video conferencing services" (2020) https://noyb.eu/sites/default/files/2020-04/noyb_-_report_on_privacy_policies_of_video_conferencing_tools_2020-04-02_v2.pdf

[138] Article 29 Working Party "Guidelines on Data Protection Officers ("DPOs")", WP243 rev.01, adopted in 2017.

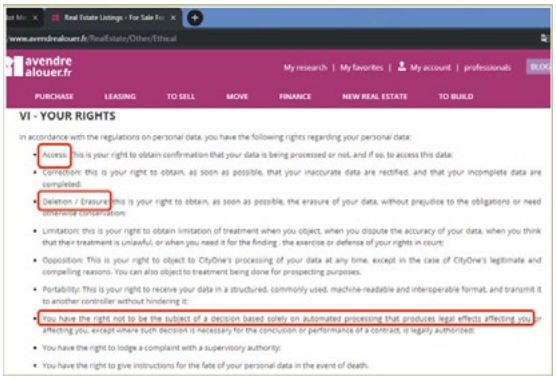


Figure 19 Compliant example on informed consent: the page refers to the rights of the data subjects, such as the right of access or deletion (www.avendrealouer.fr/RealEstate/Other/InfosCookies accessed on 18 May 2019).

Examples. Figure 19 shows an information page in which the rights of the subjects are illustrated. However, as shown on Figure 20, it does not provide for all the informative elements, such as the risks of transfers of data.

How to detect violations? Please see Section 6, where we describe automatic means that can be used to assess language-based requirements.

5.5 Unambiguous consent

For the consent to be valid, the user must give an "unambiguous indication" through a "clear and affirmative action" (Article 4(11) of the GDPR). In the following subsections we further decompose this requirement of an unambiguous consent into five low-level requirements: affirmative action design, configurable banner, balanced choice, post-consent registration and correct consent registration, as shown in Table 14.

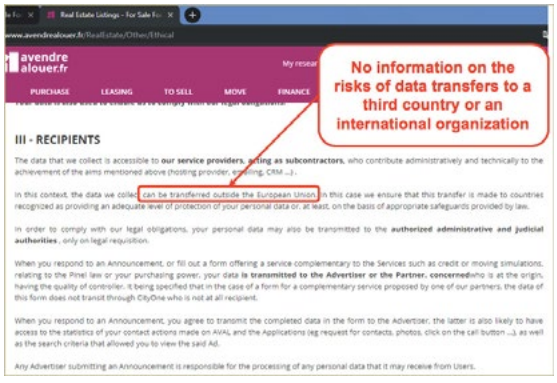


Figure 20 Non-compliant example of informed consent: the information page does not show the risks of data transfers to a third country (www.avendrealouer.fr/RealEstate/Other/InfosCookies accessed on 18 May 2019).

Table 14 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-Level	**Low-level**	**Binding**	**Non-binding**	**Interpretation**
Unambiguous	R11 Affirmative action design	Planet 49	29WP; Recital 32 DPA: Danish, German	-
	R12 Configurable banner	-	29WP; DPAs: French, Spanish, Irish, ICO, Greek, Danish, German	L based on 7(4)
	R13 Balanced choice	-	DPAs: French, Danish, UK, Greek, Irish	L based on 7(4)
	R14 Post-consent registration	-	DPAs: French, Greek	CS 7(1)
	R15 Correct consent registration	-	DPAS: French, Greek, Spanish	CS 7(1)

R11 Affirmative Action Design

Unambiguous consent refers to an active behavior of the user through which he indicates acceptance or refusal of BTT (Article 5(3) and Recital 66 of the ePD).

The 29WP (WP208) explains this active behavior:

> "Active behaviour means an action the user may take, typically one that is based on a traceable user-client request towards the website. (...) The process by which users could signify their consent for cookies would be through a positive action or other active behaviour [...] The consent mechanism should present the user with a real and meaningful choice regarding cookies on the entry page".[139]

The Advocate General[140] points (in paragraph 44), that the requirement of an "indication" of the data subject's wishes necessitates that the data subject enjoys *a high degree of autonomy* when choosing whether or not to give consent.

Planet49 Judgment[141] made even more precise this requirement. The ruling asserts that "**only active behavior on the part of the data subject with a view to giving his consent may fulfil that requirement", and this** wording ("with a view to") denotes the element of volition and willfulness towards giving an affirmative consent.

An active behavior leaves *no scope for interpretation of the user's choice*, which must be distinguishable from other actions. As such, behaviors presenting a margin of doubt do not deliver a choice and therefore are void[142]. To netter ascertain in practice how to distinguish an unambiguous from ambiguous practices, we document examples of both herein. Both Recital 32 GDPR and the 29WP (WP208)[143] provide for concrete and clear examples of an active behavior:

[139] cf. 29WP (WP208) (n 17) 4 and 5.
[140] cf. Opinion of Advocate General Szpunar (n 120).
[141] cf. Planet49 Judgment (n 11) para 54.
[142] cf. 29WP (WP187) (n 54) 35.
[143] The 29WP refers that opt-in consent is the mechanism most aligned to Art. 5(3) of the ePD: "in general users lack the basic understanding of the collection of any data, its uses, how the technology works and more importantly how and where to opt-out. As a result, in practice very few people exercise the opt-out option", cf. 29WP (WP208) (n 17) 4 and 5.

"clicking on a link, or a button, image or other content on the entry webpage, ticking a box in or close to the space where information is presented (...) or by any other active behavior from which a website operator can unambiguously conclude it means specific and informed consent".

Stakeholders also pinpoint instances of ambiguous behaviors, such as:

- presumed or implied consent from inactivity (or silence) of the data subject[144] (as signed in Recital 32 of the GDPR), e.g.:
    1. Actions such as browsing, scrolling on a website, swiping a bar on a screen, waiving in front of a smart camera, turning a smartphone around clockwise, or in a figure eight motion (29WP WP 259 rev1.; EDPB 05/2020);
    2. Actions like clicking on a "more information" link:[145] "By continuing to use the site we assume you consent to this", or "By accessing the website, you give your consent to our use of cookies". This practice is still acceptable by the Spanish DPA.
    3. Disappearance of the cookie banner without an affirmative action of the user, and a positive consent is registered by the fact that the user scrolled the website, visited other pages, clicked on links or other actions on a website;[146]
- when the user's browser is configured to receive cookies;[147]
- pre-ticked boxes[148]. The Advocate General[149] refers that unticking a pre-ticked checkbox on a website is considered too much of a burden for a customer. The ICO[150] stated that pre-selecting any cookie needed of consent, without the user taking a positive action before it is set on their device, does not represent valid consent.

The EDPD[151] (05/2020) establishes that such actions or similar user activities may be difficult to distinguish from other activities or interactions by a user and therefore determining that an unambiguous consent has been obtained will also not be possible.

In the light of the above, we define the "Affirmative Action Design" low-level requirement to make prominent this positive action. The consent must be registered by the controller only after an affirmative action of a user, like clicking on a button, checking a box, or actively selecting settings.

Requirement	Affirmative action design
Description	Consent must be registered only after an affirmative action of a user, like clicking on a button or checking a box.
Violation	The action of closing a cookie banner considered as consent. Allowing only closing the banner and forcing agreement to consent. Pre-ticked boxes. Disappearance of the cookie banner without an affirmative action of the user with a positive consent registered.

Examples. Figure 21 gives a non-compliant example of the requirement of "Active Action Design". It shows the Twitter account of the European Data Protection Board and a cookie banner provided by twitter.com, wherein it is not possible to exercise an active consent since the only possible action is to close the cookie banner, while agreeing to the use of cookies.

Figure 21 Violation of the "Affirmative Action Design" requirement (https://twitter.com/eu_edpb?lang=en accessed on 24 September 2019)

How to detect violations? To detect a violation of this requirement, one needs to perform an action on the website, like closing the banner or scrolling the website and verify whether a positive consent has been registered. While an action on a website must be done by a human operator (because there is no standard design of closing banners that can be automated), verification of a registered consent can be done only with technical means.

However, verification of a registered consent is only possible if it is known a priori which standard or specification is used to store the consent in the user's browser. For instance, this is the case if the publisher is using IAB Europe's Transparency and Consent Framework, as demonstrated by Matte et al.[152] By using the "Cookie Glasses" browser extension[153] (developed by the same authors), a human operator can observe whether the consent is stored in the browser before making an affirmative action in the banner, or upon scrolling the website.

R12 Configurable banner

Several customization implementations are possible, such as the ones proposed by different stakeholders in Table 15.

[144] On implied consent, the ICO observes that statements such as "by continuing to use this website you are agreeing to cookies" should not be used as they do not meet the requirements for valid consent required by the GDPR. Pre-ticked boxes or any equivalents, such as sliders defaulted to "on", cannot be used for non-essential cookies. Users must have control over any non-essential cookies, and they must not be set on landing pages before consent is obtained". cf. ICO Guidance (n 26).

[145] Controllers must avoid ambiguity (...). Merely continuing the ordinary use of a website is not conduct from which one can infer an indication of wishes by the data subject to signify his or her agreement to a proposed processing operation, cf. 29WP (WP187) (n 54) 17.

[146] cf. 29WP (29WP208) (n 17) 5.

[147] cf. Greek DPA (n 42).

[148] cf. Planet49 Case (n 11).

[149] Opinion of Advocate General Szpunar (n 120).

[150] cf. ICO Guidance (n 26).

[151] cf. EDPB 05/2020 (n 84).

[152] Celestin Matte, Nataliia Bielova, Cristiana Santos, "Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework" (*IEEE Symposium on Security and Privacy* (IEEE S&P 2020). To appear. http://www-sop.inria.fr/members/Nataliia.Bielova/papers/Matt-etal-20-SP.pdf

[153] "Cookie Glasses" extension available for Firefox: https://addons.mozilla.org/fr/firefox/addon/cookie-glasses/ and Chrome / Chromium : https://chrome.google.com/webstore/detail/cookie-glasses/gncnjghkclkhpkfhghcbobednpchjifk

Table 15 Positioning of the 29WP and DPAs on the configuration of consent dialogs

Stakeholders	Configurations of consent dialogs (configuration and web design)
29WP (WP208)	"The user should have an opportunity to freely choose between the option to accept some or all cookies or to decline all or some cookies and to retain the possibility to change the cookie settings in the future".
French DPA[154]	"By its presentation, the mechanism for obtaining consent must enable the data subject to be aware of the goal and scope of the act enabling him or her to signify his or her agreement or disagreement." (Parag. 51)
Spanish DPA[155]	"a management system (or cookie configuration panel) that allows the user to choose in a granular way to manage his preferences, by: enabling a mechanism or buttons to reject all cookies, another to enable all cookies, or to be able to do it in a granular way."
Danish DPA[156]	"users are given the option to accept or decline cookies either by an "accept" or "reject" button, or by toggles to accept or reject specific cookie purposes. The option to decline cookies is as easy as it is to accept cookies".
Irish DPA[157]	- "if you use a button on the banner with an "accept" option, you must give equal prominence to an option which allows the user to "reject" cookies, or to one which allows them to manage cookies and brings them to another layer of information in order to allow them to do that, by cookie type and purpose". - "manage cookies and brings them to another layer of information in order to allow them to do that, by cookie type and purpose".
UK DPA[158]	"A consent mechanism that does not allow a user to make a choice would also be non-compliant, even where the controls are located in a "more information" section".
Greek DPA[159]	"the option to decline the use of trackers is only given at a second level, i.e. following the selection of a hyperlink to "more information" or "settings".

We believe that a sufficient level of granularity of choice is demanded in the consent banner design. We interpret that a consent banner must give the user an option such as: i) one "configure" button; or ii) "accept", "reject" and "customize/configure".

Requirement	Configurable banner
Descripion	A banner must give the user an option to customize his consent. Several implementations are possible: • One "Configure" button • "Accept" and "Configure" buttons • "Accept", "Reject" and "Configure" buttons
Violation	• A banner does not provide a choice in the settings/configuration, • the customization options are not emphasized as a link in the first layer, • the only existing box is "Accept and close", • invisible "Parametrize" button.

Examples. Figure 22 illustrates a non-compliant banner design, curiously from one of the flagship security and privacy conferences (IEEE Symposium on Security and Privacy). In this banner, the only available option is to accept and close the banner, not offering a sufficient level of granularity of choice demanded by the GDPR. This example, however, presents a violation only if cookies that require consent are used on this website, while the banner does not provide an explicit enough purpose (i.e. "to give you the best user experience") to determine whether consent is needed in this case. Figure 23 shows a banner design which is closer to be compliant with the "Configurable" consent requirement. Through the indication of the "yes" and "configure" buttons, it is possible to accept and customize consent. Configuring the choice at any time in the privacy center is

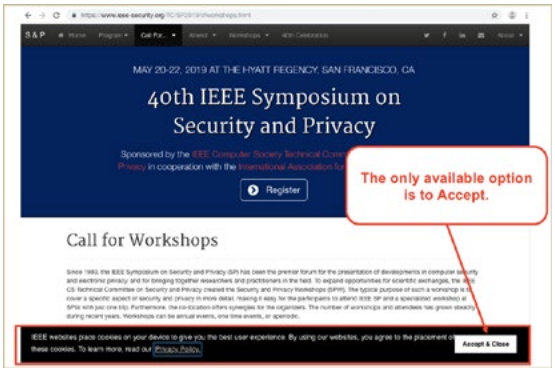


Figure 22 Violation of the "Configurable consent" requirement www.ieee-security.org/TC/SP2019/venue.html (accessed on 17 May 2019)

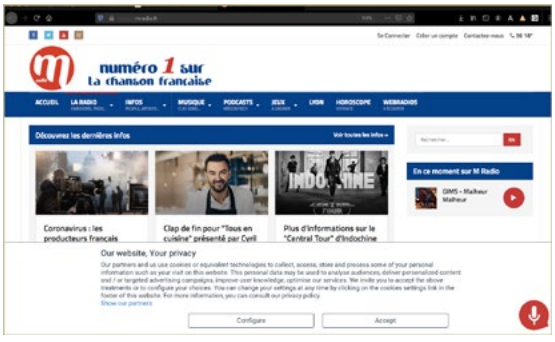


Figure 23 Example of compliance with the "Configurable consent" requirement (https://mradio.fr accessed on 27 May 2020). It is possible to accept or configure, however this banner does not provide a balanced choice.

[154] cf. CNIL draft recommendation 2020 (n 95).
[155] The Spanish DPA decision reads accordingly, "III. It does not provide a management system or cookie configuration panel that allows the user to eliminate them in a granular way. To facilitate this selection the panel may enable a mechanism or button to reject all cookies, another to enable all cookies or do so in a granular way to manage preferences. In this regard, it is considered that the information offered on the tools provided by several browsers to configure cookies would be complementary to the previous one, but insufficient for the intended purpose of allowing to set preferences in a granular or selective way", See Spanish DPA decision, "Procedimiento PS/00300/2019" (2019) www.aepd.es/resoluciones/PS-00300-2019_ORI.pdf?utm_source=POLITICO.EU&utm_campaign=fc1f5e664f-EMAIL_CAMPAIGN_2019_10_17_04_52&utm_medium=email&utm_term=0_10959edeb5-fc1f5e664f-190359285 accessed 7 May 2020 (our translation).
[156] cf. Danish DPA Guide (n 77).
[157] cf. Irish DPA Guidance (n 37).
[158] cf. ICO Guidance (n 26).
[159] cf. Greek DPA (n 42).

also possible. As the customization of the preferences is easy and user-friendly, the banner seems to comply with the above-mentioned requirement.

How to detect violations? To detect violations of this requirement, a human operator needs to evaluate whether a banner gives a set of options to the user. As of today, verifying this requirement fully automatically is not possible because of a lack of standards in cookie banner design. Nevertheless, some technical developments have been made to automatize the process of either fully accepting or fully refusing consent when such option exists in the configuration settings. The "Consent-o-Matic" Web browser extension[160] (and previously the Cliqz browser[161]) implements such functionality to automatically interact with the HTML content of a cookie banner to either refuse or accept consent.

R13 Balanced choice

From Article 7(4) of the GDPR which states that withdrawing consent should be as easy as giving it, we additionally interpret that the choice between "accept" and "reject" BTT must be consequently balanced (or equitable). Our interpretation is also sustained by the recent opinion of the CJEU's Advocate General and by some DPAs, as listed in Table 16.

Table 16 Positioning of stakeholders regarding the requirement of a balanced banner

Stakeholders	Positioning regarding the "balanced banner" requirement
CJEU's Advocate General[162]	emphasized the need that actions, "*optically in particular, be presented on an equal footing*"
French DPA[163]	Parag. 39. 'interfaces should not use potentially misleading design practices, such as the use of visual grammar that might lead the user to think that consent is required to continue browsing or that visually emphasizes the possibility of accepting rather than refusing. Parag. 40. The user may also have the choice between two buttons presented at the *same level and in the same format*, with "accept" and "refuse", "allow" and "forbid", or "consent" and "do not consent", or any other equivalent wording that is sufficiently clear to the user. Parag. 51. Thus, this mechanism should not involve potentially misleading design practices, such as the use of visual grammar that impedes the user's understanding of the nature of his or her choice.
Danish DPA[164]	"users are given the option to accept or decline cookies either by an "accept" or "reject" button, or by toggles to accept or reject specific cookie purposes. The option to decline cookies is as *easy as* it is to accept cookies".
UK DPA[165]	refusal of trackers should be at the *same level* as the "accept" button. "[A] consent mechanism that emphasizes "agree" or "allow" over "reject" or "block" represents a non-compliant approach, as the online service is influencing users towards the "accept" option".
Greek DPA[166]	Parag. 4: The user must be able, with the same number of actions ("click") and from the *same level*, to either accept the use of trackers (those for which consent is required) or to reject it, either all or each category separately. Parag. 7: "To ensure that the user is not affected by website designs favoring the option to consent vis-à-vis the option to decline, buttons of the *same size, tone and color* ought to be used, so as to provide the same ease of reading to the attention of the user". Parag. 6: "The size and colour of the "accept" or "consent" button strongly urges the user to choose, e.g. is very large and / or in bold and / or is pre-ticked."
Irish DPA[167]	"no use of an interface that "nudges" a user into accepting cookies over rejecting them. Therefore, if you use a button on the banner with an "accept" option, you must give *equal prominence* to an option which allows the user to "reject" cookies, or to one which allows them to manage cookies and brings them to another layer of information in order to allow them do that, by cookie type and purpose.

Potential violations of a balanced consent banner requirement normally happen when there is unbalance in the design choices given. Design choices related to an unbalanced choice in a consent banner can consist, for example, of "False Hierarchy" and "Aesthetic manipulation".

According to Gray et al.,[168] *False Hierarchy*

> "gives one or more options visual or interactive precedence over others, particularly where items should be in parallel rather than hierarchical. This convinces the user to make a selection that they feel is either the only option, or the best option".

Some examples of false hierarchy in consent banners are illustrated below, both at the first and second layers of the banner:

- far-away approach: when the sliders or the menu settings are far, e.g. website forwards the user to click on a link for opt-out tools offered by DAA, NAI, and Google;
- click-away approach: requires more clicks and diligence to reach to the parametrization and refuse consent (more than two sub-menus)
- box with a bigger "OK" button and a small, less visible "Configure" button, which gives a higher hierarchy to "OK";
- banner presenting 2 options: "accept all" and "reject all" whereas the "accept all" option comes first, has green color, while "reject all" is in white, indicating some desirability in choosing this one;
- box with "I consent" emphasized in a black box, and "More Options" link on the corner of the banner;
- banner in which the option to refuse is only given at a second level, i.e. following the link to "more information";
- button for "Refuse" or "Preferences" is a text, while "Accept" is a button;
- use of the same color with dark/light differences, e.g. light-blue and dark-blue sliders to signify accept and refuse;
- the legend in the banner always labels "Activate", whether the slider is activated or not;
- the same button corresponds to "Activate all" and "Deactivate all" (and the meaning of the slider is unclear), it is not possible to see what is was chosen;
- the only button to disable purposes in barely visible;

[160] https://github.com/cavi-au/Consent-O-Matic
[161] https://github.com/cliqz-oss/autoconsent
[162] Opinion of Advocate General Szpunar on the case of Planet 49, delivered on 21 March 2019, http://curia.europa.eu/juris/document/document.jsf?docid=212023&doclang=en, accessed 18th June 2020.
[163] cf. CNIL draft recommendation 2020 (n 95).
[164] cf. Danish DPA Guide (n 77).
[165] cf. ICO Guidance (n 26).
[166] cf. Greek DPA (n 42).
[167] cf. Irish DPA Guidance (n 37).
[168] cf. Gray et al. (n 7).

Aesthetic manipulation consists of an interface design that nudges users into clicking the "accept" button rather than the "refuse" button by using design, colors, hovering properties. The CNIL[169] names it "*Attention Diversion*" referring to these design choices that draw attention to a point of the site or screen to distract or divert the user from other points that could be useful. The CNIL states that visual saliency is effective and abundant. It gives a concrete example, working on the color of a "continue" green button while leaving the "find out more" or "configure" button smaller or grey, users may perceive green as the preferable choice. This holds particularly if they are conditioned by the traffic light metaphor bias used by designers that assign colors according to the flow of information ("green" = free flowing"; "red"=stop"), which bring ambiguity to the choice. Some examples of false manipulation in consent banners are illustrated below, both at the first and second layers of the banner:

- hovering over a button: "accept" has a hover background while "reject" does not
- bright and attractive "accept" button and grey/white "reject" button
- all information is written in a very small window difficult to navigate size and colour of the "accept" or "consent" button strongly urges the user to choose it, e.g. is very large and / or in bold and / or is pre-ticked.

Requirement	Balanced choice
Description	A banner must present a fair or balanced design choice
Violation	A banner does not provide a fair choice.

Examples. Figure 24 presents an example of non-compliant banner. Even though this banner provides an option to configure privacy settings, the provided choice is not balanced – it contains "Accept" and "Configure" buttons – and hence guides the user towards acceptance. Figure 25 provides an example of a balanced choice since both options, "Yes" and "No" are present in the banner interface.

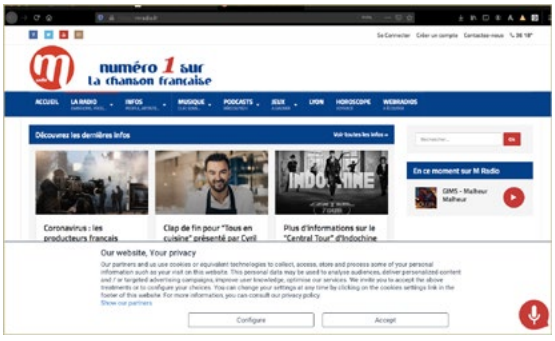


Figure 24 Violation of the requirement "Balanced choice" (https://mradio.fr/ accessed 27 May 2020).

How to detect violations? To detect violations, a human operator needs to evaluate whether a banner gives fair and balanced choices. As of today, it is not possible to verify this requirement automatically because of lack of standards in cookie banner design.

[169] cf. CNIL report (n 8).

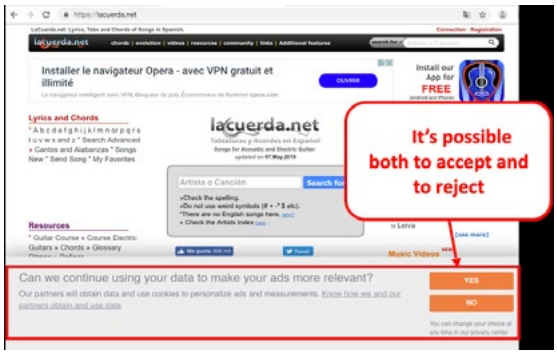


Figure 25 Compliant example for "Balanced choice" requirement (https://lacuerda.net accessed 17 May 2019)

R14 Post-consent registration

The GDPR mandates in Article 7(1) and Recital 42 that controllers have the obligation to demonstrate that the data subject has consented to processing of his personal data. Further, Article 30 requires that each shall maintain a record of processing activities under its responsibility (which includes consent). These provisions constitute a specification of the principle of accountability, enshrined in Article 5(2) of the GDPR.

This auditable legal obligation entails a *technical* side that is relevant to consider: after a certain user action done via the user interface (like clicking on a button, checking a box, etc.), the user's choice (acceptance or refusal) needs to be "registered" or "stored" in the user's device (a browser in our case). We therefore use the noun "registration" to mean that the consent choice is stored.

Both the need for an auditable consent and an adequate procedure thereto are emphasized by the 29WP and DPAs:

- *Consent registration*: both the Spanish, the Greek[170] and the Danish DPAs asserts that consent must be stored for documentation (in case of inspections by DPAs).
- *Mechanisms for consent registration*:
  - 29WP: an auditable consent can be achieved by keeping a record of the received consent statements, so the controller can show how/when consent was obtained. Consent receipt mechanisms can be helpful in automatically generating such records;
  - CNIL[171] (paragraphs 36 and 37): points out that consent (either its acceptance or refusal) should be registered. A tracker may be used with the sole purpose of storing consent or refusal;
  - Italian[172] and Irish[173] DPAs: affirm that a special technical cookie is normally used to store and keep track of the acquired consent. The Irish DPA further states that "any record of consent must also be backed up by demonstrable organizational and technical measures that ensure a data subject's expression of consent (or withdrawal) can be effectively acted on";
  - Belgium DPA: illustrates that companies must be able to demonstrate that consent was collected by using logs.

[170] cf. Greek DPA (n 42).
[171] cf. CNIL draft recommendation 2020 (n 95).
[172] Italian DPA, FAQs on cookies (2020) https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3585077
[173] cf. Irish DPA Guidance (n 37).

Additionally, some DPAs opted to define or comment on the storage period of the user's choice, as shown in Table 17.

Table 17 DPAs' positioning about the storage period of the user's choice

DPAs	Positioning about the storage period of the user's choice
French	6 months
Spain	consent should be renewed after 24 months
Danish	storage period of 5 years
Irish	no longer than 6 months, after which the consent request must be renewed

We include a technical low-level requirement "post-consent registration", as depicted in the requirement box below. Notice that closing a consent banner without a consent being registered as "positive" does not configure a violation of this requirement.

Requirement	Post-consent registration
Description	Consent should be registered (e.g. stored on a terminal equipment) in a "consent cookie" (or any other browser storage) only after an affirmative action of the user.
Violation	A consent registered without any user action.

Example. Figure 26 refers to an example of non-compliant design of the "Post-consent registration" requirement. When accessing the tpi.it website, it is possible to check that the user's consent was registered before the user has made his choice.

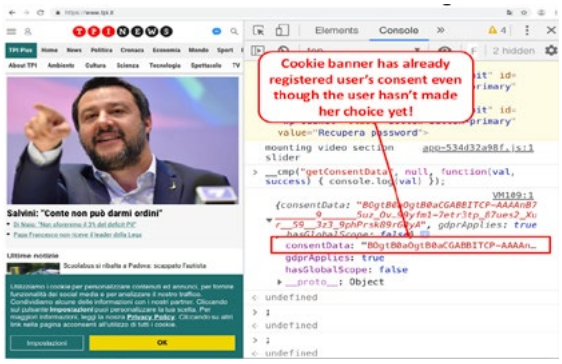


Figure 26 Violation of the "Post-consent registration" requirement www.tpi.it/ (accessed 17 May 2019).

How to detect violations? Detecting violations is only possible with technical means, but only on websites where it is known how the consent is registered by the publisher (e.g. for websites using the IAB's Transparency and Consent Framework, as demonstrated by Matte et al.).[174]

For the majority of websites, it is not the case and therefore, detecting violations without standardizing the storage of consent is not possible. Two elements would be needed to standardize consent: the data structure for consent storage, and a concrete browser storage (cookie, localStorage, or any other) used to store consent. On websites where storage and structure of consent is known (such as on the IAB TCF), the verification procedure is the following: in a browser with an empty session, open the target webpage. Before doing any interaction with the cookie banner, verify whether the browser storage contains a positive consent.

In order to assure that the violation does not occur, website publishers or other entities that register consent on a website (such as CMP in case of IAB TCF) have to be able to prove that they have rightfully obtained consent after user's interaction with the cookie banner. Technically, the proof of consent could be done through the use of cryptographic primitives, or secure hardware. Additionally, publishers need to record the user's interaction with the banner via screenshots or video recording. As of May 2020, no technical solution exists to handle the problem of proving that consent has been obtained only after the users has interacted with the banner.

R15 Correct consent registration

Another technical side deriving from the obligation to register consent (Article 7(1)) refers to the fact that the decision made by the user in the banner interface should be identical to the consent that gets registered/stored by the website. If consent is correctly registered, the user will not be pressured to choose again by the same website. To this scope, nagging[175] practices (identified as dark patterns) seems to be related to the functionality of consent dialogs being correctly registered.

Some DPAs express concerns when consent is not registered correctly:

1. The CNIL draft recommendation[176] (paragraph 36): states that as a result of a non-registration of the user's choice, the user will be pressured to accept out of weariness:
   "Thus, the choice expressed by the user, be it consent or refusal, should be recorded in such a way that the user's consent is not sought again for a certain period of time. Indeed, failure to register the refusal would prevent it from being considered in the long term, in particular during new visits. If the choice that the user has expressed is not registered, he or she would be asked again to consent. This continued pressure would be likely to cause the user to accept out of weariness. Failure to record the refusal to consent could therefore have the consequence of exerting pressure that could influence his or her choice, thus calling into question the freedom of the consent he or she expresses".
2. Greek DPA:[177] emphasizes the consequences of an incorrect registration of the user's choice: "in case trackers are rejected, the user is constantly requested to register a new choice through the perseverance of pop-up windows, whereas, in case trackers are accepted, this choice is maintained for a longer period of time than the choice of rejection".
3. Spanish DPA: the users' preferences may be stored so they are not asked to set them up again every time they visit the relevant page.

In this line, we have devised the technical low-level requirement of "correct consent registration". Notice that the consent should be *registered both in the user's browser, and also on the server* of the entity that proves collection of consent.

Requirement	Correct consent registration
Description	The registered consent must be identical to the user's choice in the user interface
Violation	A registered consent is different from the user's choice.

[174] cf. Matte et al. (n 151).

[175] cf. Gray et al. (n 7).

[176] CNIL draft recommendation 2020 (n 95).

[177] cf. Greek DPA (n 42).

Example. In Figure 27, by using Matte et al. Cookie Glasses tool,[178] it is possible to verify whether consent is correctly registered by cookie banners of IAB Europe's Transparency & Consent Framework.

Figure 27 Browser extension "Cookie Glasses" showing consent registered by cookie banners of IAB Europe's Transparency & Consent Framework

How to detect violations? Similar to detecting violations of R14 "Post-consent registration" requirement, assessing this requirement is possible only by a combination of manual and technical means and only on websites where it is known how the consent is registered by the publisher.

Manual verification is needed for the evaluation of the user interface in the banner. For example, a human operator can decide to refuse consent and then compare this choice to the consent registered in the browser. However, since there is no standardized way to structure and store consent, it is not possible to detect violations automatically on all websites.

Matte et al.[179] demonstrated verification of correct consent registration requirement on websites that contain banners of the IAB Europe's Transparency and Consent Framework. Matte et al. used the following procedure: in a browser with an empty session, open the target webpage. After giving a consent on the cookie banner interface, analyze whether the consent given in the user interface of the banner is consistent with the consent (a) stored in the dedicated browser cookie, or (b) collected from querying the Consent Manager Provider (cookie banner provider in the terminology of IAB Europe TCF).

Notice that technical analysis of stored consent is only possible for consent stored on the "client-side", that is in the user's browser. Website publishers also need to store consent on their own servers, in order to prove consent collection upon request. Server-side storage of consent cannot be verified due to absence of access to such servers.

5.4 Readable and accessible

The requirements explained in this section refer to the consent request (i.e. how consent should be collected by the data controller). We derived the requirement "readable and accessible" consent request from our analysis of the following provisions: Article 7(2) GDPR, and its further articulation in Recitals 32 and 42 of the GDPR. Herewith we transpose their excerpts for legibility purposes. In the wording of Recital 32, "if the data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided". Pursuant to Recital 42, "a declaration of consent pre-formulated by the controller should be provided in an intelligible and easily accessible form, using clear and plain language and it should not contain unfair terms". From these three precepts, we derive that the request for consent should be clearly distinguishable from any other matters, in an intelligible and easily accessible form, using clear and plain language. These requirements are shown in Table 18.

The GDPR (Article 7(2)) mandates that a failure to comply with these requisites constitutes an infringement and renders a non-binding consent, which signals the *practical effect* of these validity conditions. These four requisites were mostly elaborated in the 29WP Guidelines on Transparency[180] and relate to how information should be disclosed.[181]-[182] In this work, we adapt the content of these four elements described in these guidelines and apply them as low-level requirements for a valid consent request, as explained in the following subsections.

We included the "readable and accessible" (and the low-level) requirement for a consent request considering two main factors:

1. The reasonable expectations of data subjects (which are, in general, laymen), as evoked by the recent jurisprudence of the European Court of Justice, in Planet49 judgment[183] that reads: "due to the technical complexity of cookies, the asymmetrical information between provider and user and, more generally, the relative lack of knowledge of any average internet user, the average internet user cannot be expected to have a high level of knowledge of the operation of cookies"; [184]
2. The average user needs specific information to easily determine the consequences of any consent he might give, in an intelligible, clear way, where layered[185] information is amenable.

[178] cf. Matte et al. (n 152).
[179] cf. Matte et al. (n 151).

[180] Article 29 Working Party, "Guidelines on transparency under Regulation 2016/679" (WP260 rev.01, 29 November 2017).
[181] "The GDPR puts several requirements for informed consent in place, predominantly in Article 7(2) and Recital 32. This leads to a higher standard for the clarity and accessibility of the information", cf. 29WP (WP259 rev.01) (n 4) 14.
[182] The Article 29WP mentions that "transparency requirements in the GDPR apply irrespective of the legal basis for processing and throughout the life cycle of processing", 29WP (WP260 rev.01) (n 180) 6.
[183] cf. Planet49 Judgment (n 11) para 114.
[184] On the terminology in the area of consumer protection, see Directive 2011/83/EU on consumer rights, ELI: http://data.europa.eu/eli/dir/2011/83/2018-07-01. See, by way of example, the judgments of the following cases: Case C485/17 Verbraucherzentrale Berlin eV v Unimatic Vertriebs GmbH [2018] ECLI:EU:C:2018:642, para 44; Case C44/17 Scotch Whisky Association v Michael Klotz [2018] ECLI:EU:C:2018:415, para 47; Case C210/96 Gut Springenheide and Tusky v Oberkreisdirektor des Kreises Steinfurt [1998] ECLI:EU:C:1998:369, para 31.
[185] The Handbook on European Data Protection Law refers to the "accessibility in an online environment", as follows:
"The quality of the information is important. Quality of information means that the information's language should be adapted to its foreseeable recipients. Information must be given without jargon, in a clear and plain language that a regular user should be able to understand. Information must also be easily available to the data subject (...). Accessibility and visibility of the information are important elements: the information must be clearly visible and prominent. In an online environment, layered information notices may be a good solution, as these allow data subjects to choose whether to access concise or more extensive versions of information", European Agency for Fundamental Rights, *Handbook on European Data Protection Law* (2018 edition) (Publications Office of the European Union, 2018) 147.

Table 18 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-Level	**Low-level**	**Binding**	**Non-binding**	**Interpretation**
Readable and accessible	R16 Distinguishable	7(2)	29WP, Recital 42; DPAs: Belgium, Spanish, UK, Irish	-
	R17 Intelligible	7(2)	29WP, Recital 42. DPAs: UK, Spanish, Danish	-
	R18 Accessible	7(2)	29WP, Recital 42; DPAs: UK, Irish, Spanish, Belgium	-
	R19 Clear and plain language	7(2)	29WP, Recital 42; DPAs: UK, Spanish, Danish	-
	R20 No consent wall	-	DPAs: UK, French	L Based on 7(2) and recitals 32, 42

R16 Distinguishable

The requirement "distinguishable", according to Article 7(2) and the 29WP Guidelines on Transparency (WP260 rev.01),[186] means that the consent request should be clearly differentiated from other non-related information, such as contractual provisions or general terms of use, warning boxes, etc.

R16 is related to R3 ("No merging into a contract"), however, R16 is more general (distinguishable from the other matters), while R3 is an instantiation of one such matter – the existence of a contract.

Requirement	**Distinguishable**
Description	The consent request should be clearly differentiated from other non-related information.
Violation	When the consent request is mixed with other matters, like terms of use, warning boxes, among others.

How to detect violations? An example of violation shown in R3 shows also a violation of this requirement. Both R16 and R3 can be verified either manually or by using NLP tools that analyze structural properties of the text – we detail the discussion on their verification in section 6.

R17 Intelligible

Intelligible means, in the context of a consent request, that the collection of consent "should be understood by an average member of the intended audience" (WP260 rev.01).

Requirement	**Intelligible**
Description	The consent request should be understood by any user.
Violation	When the consent request is not understood by average users.

How to detect violations? Intelligible consent is dependent on the understandability of the target audience – composed by users who try to make their choice in the consent banner interface. Therefore, such requirement can only be analyzed by means of user studies, questioning users about their understanding of various types of explanations in cookie banners. We provide a deeper analysis of verification for this requirement in section 6.

R18 Accessible

The consent request should be easily accessible, which means that "the data subject should not have to seek out" for the settings to customize her preferences; and "it should be immediately apparent to them where and how this information can be accessed" (WP260 rev.01).

Violations of R18 can be noted, for example, whenever there is a decoupled choice, meaning that a consent request and the settings are located far from the primary interaction with the banner, or when it is required unnecessary user effort (such providing different opt-out links) for users to make a choice.

Requirement	**Accessible**
Description	The consent request should be easily accessible to the data subject.
Violation	When the sliders or settings are far from the settings of a banner.

How to detect violations? This requirement is also subjective to the capacities to find information of the target audience, and therefore can only be verified with user studies. We discuss it in detail in section 6.

R19 Clear and plain language

The information of a consent request should use clear and plain language. The 29WP (WP260 rev.01) reads that information should be presented:

> "in as simple a manner as possible, avoiding complex sentence and language structures. The information should be concrete and definitive; it should not be phrased in abstract or ambivalent terms or leave room for different interpretations. In particular the purposes of, and legal basis for, processing the personal data should be clear. (...) The information provided to a data subject should not contain overly legalistic, technical or specialist language or terminology".

While R17 is related to R19 (consent request should be intelligible, in the sense of being understood by an average user), however, R19 refers concretely to the language used in the text of a consent request. We have observed that the text within cookie banners is generally prone to the existence of manipulative dark patterns. We name a few violations of this requirement for illustration purposes:

1. Questioning the choice of refusing tracking, e.g. "Would you re-consider?", or "Are you sure?",
2. Use positive framing regarding one choice (to accept tracking), while glossing over any potentially negative aspects of that same choice, e.g. "We value your privacy","We care about your privacy","Go to the website",
3. Use of technical language and legal jargon, e.g. "This site uses cookies", instead of "This site collects your data",
4. Use of compliance wording might influence the user towards accepting consent, e.g. mentioning a Data Protection Authority recommendation, or making salient regulations.

[186] cf. 29WP (WP260 rev.01) (n 180).

Requirement	Clear and plain language
Description	Concrete, explicit, clear
Violation	Use of positive framing, technical and overly legalistic language, use of compliance wording, questioning the choice of refusal, use of abstract and ambivalent terms.

How to detect violations? This requirement is also subjective to the capacities to find information of the target audience, and therefore can only be verified with user studies. We discuss it in detail in section 6.

In this section 5.4, we do not present examples of compliant examples and violations of requirements R17, R18, and R19 because they are subjective to the perception and even biases of the target audience, which are users.

R20 No "consent wall"

Recital 32 of the GDPR states that the consent request should not be unnecessarily disruptive to the use of the service for which it is provided. In our own opinion of this Recital, *unnecessary disruption* to the use of a website/app reflects a common practice that we name "*consent walls*". Consent walls consist of a mechanism that forces the user to make a choice, by blocking access to the website/app until the user expresses her choice regarding consent.

For publishers that provide a free access to the website independently of the user's choice, a consent wall is discouraged because it forces the user to make a choice that does not influence her.

Differently from tracking wall, this practice allows the user to make a choice between acceptance and refusal. A consent wall appears to be unnecessarily disruptive to the use of a website.

Notice that "on interpreting "unnecessarily disruptive" consent request: "it may be necessary that a consent request interrupts the user experience to some extent to make that request effective". In line with Leenes,[187] this disruption could merely occur depending on the user's choice, e.g. a certain functionality may be lacking, such as a forum if the user does not accept social media cookies, or be replaced by other content, such as behavioral advertisements being replaced by other types of advertisements.

We take the view that if there are other ways to display the overlay without blocking the access to the service, then such banner is preferred to a consent wall. In practical settings, the website should still be accessible even if the user did not respond to the consent request. If there are other ways to show the banner without blocking (disturbing) the access to the service, or disrupting the user experience, then it is preferred to a consent wall. Thus, we argue that consent walls do not configure a valid design for consent mechanisms they are confusing and unnecessarily disruptive of the user experience. Other consent design implementations could be sought while engaging the users.

This requirement has even stronger practical significance with mobile devices. Its small configuration implies that consent walls can be more obvious while users do not consent. Relevantly, the ICO[188] emphasizes the user experience along with the electronic consent request implementation:

> Message boxes such as banners, pop-ups, message bars, header bars or similar techniques might initially seem an easy option for you to achieve compliance. However, you need to consider their implementation carefully, particularly in respect of the implications for the user experience. For example, a message box designed for display on a desktop or laptop web browser can be hard for the user to read or interact with when using a mobile device, meaning that the consents you obtain would be invalid (...) so you need to consider how you go about providing clear and comprehensive information without confusing users or disrupting their experience.

The CNIL (in paragraph 38 of its draft recommendation for the use of trackers) states that in the absence of any manifestation of choice to either accept or reject, no trackers should be written. Even if the statement below does not explicitly discourage the use of consent walls, it seems to be inclined thereto.

> "nothing prohibits the person responsible for the processing operation(s) to provide the user with the possibility of not making any choice and delaying his or her decision, as long as the user is given the choice between acceptance and refusal. The situation in which the user does not express any positive choice must be distinguished from the situation of refusal. In the absence of any manifestation of choice (neither acceptance nor refusal), no trackers requiring consent should be written. The user could then be asked again as long as he or she does not express a choice".

The requirement box summarizes the "No consent wall" requirement.

Requirement	No consent wall
Description	The website needs to be accessible even if the user did not respond to request for consent. If there are other ways to show the banner without blocking (or disturbing) the access to the service, then it is preferred than a consent wall
Violation	"*Consent wall*" that blocks the service before the user accepts or rejects consent

Example. Figure 28 depicts a consent wall displayed by the website fandom.com. This consent wall allows to accept or reject consent. Figure 29 shows a cookie banner that is compliant with the "No consent wall" requirement on a desktop version of the website but becomes non-compliant on a mobile device because the cookie banner covers the majority of the screen. Moreover, as the cookie banner of LBC website only proposes to accept consent, it is non-compliant

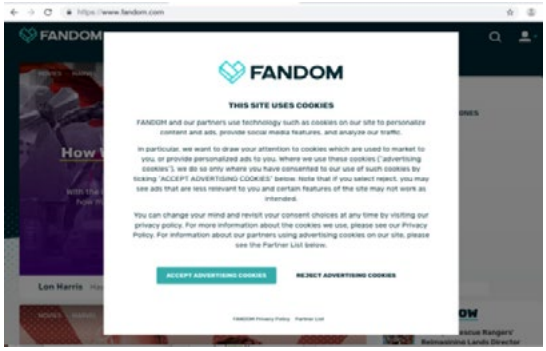


Figure 28 Violation of the "No consent wall" requirement (www.fandom.com/ accessed 17 May 2019).

[187] cf. Leenes (n 82).
[188] cf. UK DPA (n 26) 28.

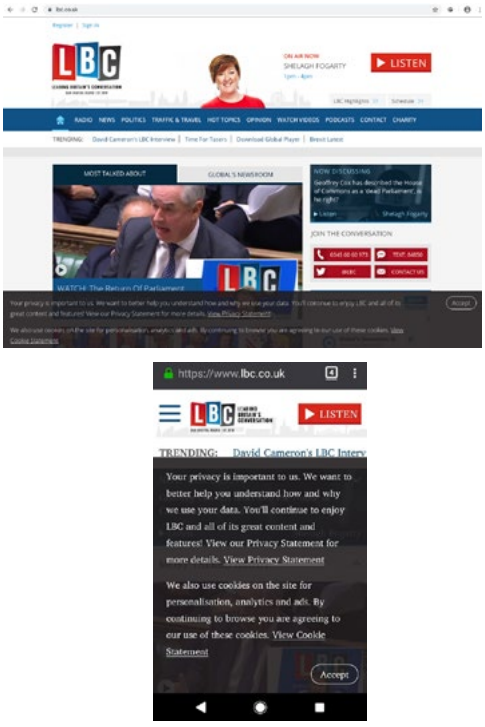


Figure 29 The Desktop version (top) of the LBC website does not violate the "No consent wall" requirement, however the mobile version (bottom) of the same website does (www.lbc.co.uk/accessed 25 September 2019).

with the "Configurable banner" requirement (Unambiguous consent). As a result, the mobile version of the LBC website has a cookie banner that forces the user to accept the data collection and at the same time blocks access to the website, which violates the "No tracking wall" requirement.

How to detect violations? Detection of such violation is possible manually, by evaluating whether the cookie banner blocks access to the website or not. Currently, technically detecting this violation is challenging because there is no specification that defines which part of the website is a cookie banner. However, techniques based on separation of HTML elements[189] and keyword searches in the cookie banner text could be envisioned. Also, crowd-sourced lists for banners blocking such as the Easylist Cookie List[190] or the Consent-o-Matic tool[191] could be used for that purpose.

The procedure for detection would be the following:

1. Detect whether the element responsible for the cookie banner is displayed;
2. Detect its size relative to the screen size, on different screen sizes (to account for both desktop and mobile)
3. Detect whether the other elements of the page are reachable, or e.g. blocked by an overlay.

We therefore conclude that the technical development of automated tools is not complex, however its accuracy has to be measured at scale to evaluate its effectiveness.

5.5 Revocable

The GDPR establishes the right of the data subject to withdraw consent in Art. 7(3). We have made the "withdrawal of consent" an additional explicit requirement due to the practical implications of this right.

Primarily, Article 7(3) explicitly states this right as one of the "conditions for consent", or condition for consent validity. Among other provisions, Recital 42 mentions the revocability of consent. The 29WP (WP259 rev.01)[192] confirms that the GDPR gives a prominent place to the withdrawal of consent. The German DPA[193] [194] also makes salient the requirement for revocability. It states that "anyone using cookies to analyze and track user behavior for advertising purposes or have them analyzed by third parties generally requires the informed, voluntary, prior, active, separate and revocable consent of the user".

Regarding the easiness to withdraw consent, Article 7(3) declares that it shall be as easy to withdraw as to give consent. This easiness attribute also relates to withdrawing consent without detriment. Easiness and without detriment entail an obligation of the controller at different levels: cost, simplicity of the procedure and finally, service level.

1. *Cost*: free of charge;[195]
2. *Service level*: without lowering service levels (29WP WP259 rev.01);[196]
3. *Simplicity of the procedure*: using simple and easily accessible mechanisms, not burdensome.[197] We claim that withdrawal should be done by the same means it was obtained in the first place, without the need to ask the user to state the reason for withdrawing consent.[198] We additionally suggest that the "withdrawal tool" should be named appropriately and should be standardized for all environments (including web and mobile). This positioning on the same means was already endorsed by the 29WP and other DPAs, as reflected in Table 19.

An example of easiness without detriment is the case of a recent decision[199] of the Polish DPA against the company ClickQuickNow referred to a GDPR violation due to the fact that the mechanism for consent withdrawal, involving the use of a link included in the commercial information, did not result in a quick withdrawal. After the link was set up, messages addressed to the user were misleading. Moreover, the company forced stating the reason for withdrawing consent. Furthermore, failure to indicate the reason resulted in discontinuation of the process of withdrawing consent.

The right to withdrawal does not have retroactive effects, meaning that it does not apply for processing that had taken place before

[189] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, Arvind Narayanan. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *ACM CSCW 2019*.
[190] https://fanboy.co.nz/fanboy-cookiemonster.txt
[191] Nouwens, Midas, Ilaria Liccardi, Michael Veale, David R. Karger and Lalana Kagal. "Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence." *ACM CHI 2020*.
[192] cf. 29WP (WP259 rev.01) (n 4) 21.
[193] cf. German DPA (n 10).
[194] "Since a consent is revocable, a corresponding option for revocation must be implemented. The revocation must be as easy as the granting of consent, Art. 7 (3) sentence 4 GDPR" (our translation), cf. German DPA (n 10).
[195] Article 29 WP Opinion 4/2010 (29WP 174) on the European code of conduct of FEDMA for the use of personal data in direct marketing, adopted on 13 July 2010.
[196] ibid.
[197] cf. 29WP Opinion 02/2013 (WP202) (n 132).
[198] cf. 29WP Opinion 4/2010 (WP 174) (n 195).
[199] Polish DPA, "Polish DPA: Withdrawal of consent shall not be impeded" (2019) https://edpb.europa.eu/news/national-news/2019_en accessed 7 May 2020.

withdrawal. Revocation cannot affect nor devalue already conducted research, decisions or processes previously taken on the basis of this data. This reasoning is supported by Article 7(3) of the GDPR that lays down that "the withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal". Moreover, the 29WP (WP187)[200] supports also this view that "withdrawal is exercised for the future, not for the data processing that took place in the past, in the period during which the data was collected legitimately".

Table 19 Positioning of the 29WP and DPAs on the easiness to revoke consent

Stakeholders	Positioning regarding the easiness of the procedure to revoke consent
29WP (WP259 rev.01)201	"the GDPR does not say that giving and withdrawing consent must always be done through the same action. However, when consent is obtained via electronic means through only one mouse-click, swipe, or keystroke, data subjects must, in practice, be able to withdraw that consent *equally as easily*. Where consent is obtained through use of a service-specific user interface (for example, via a website, an app, a log-on account, the interface of an IoT device or by e-mail), there is no doubt a data subject must be able to withdraw consent via *the same* electronic interface, as switching to another interface for the sole reason of withdrawing consent would require undue effort."
French DPA[202]	proposes as criteria for easiness: i) the time spent, and ii) number of actions required.
Greek DPA[203]	a user "must be able to withdraw his consent in the same manner and with the same feasibility with which he has given it". It refers that revoking should not be cumbersome and illustrates an unlawful practice "following the user's consent or decline, the user is not given any opportunity to change his / her preferences or user preferences may only be changed through his / her web browser settings".
Irish DPA[204]	use of an "easy tool, such as a "radio button" on your website which allows users to control which cookies are set and to allow them vary their consent at any time"
Danish DPA[205]	"when a website has a valid cookie consent solution with a GDPR valid cookie pop-up, the user can simply reject/decline cookies by reopening the cookie pop-up and thus the consent to cookies is withdrawn (hence no cookies are further set in the browser)".
Spanish DPA	a system shall be considered easy to use, for example, when users may access easily and permanently to the cookie setup or management system.

Moreover, revocability offers also a possibility for the user to make subsequent changes/configurations to his preferences, at any time. In this line, the 29WP (WP 208)[206] mentions that revocability is "an option for the user to subsequently change a prior preference regarding cookies". In another opinion, the 29WP (WP 259 rev.01)[207] ascertains that "consent is a reversible decision".

In this line, we decompose the revocability requirement in Table 20, explicitly emphasizing the need to communicate withdrawal of consent to all the parties that have previously received it.

Table 20 Derived low-level requirements and their sources

Requirements		Sources at low-level requirement		
High-Level	Low-level	Binding	Non-binding	Interpretation
Revocable	R21 Possible to change in the future	7(3)	29WP; DPAs: French, Greek, Irish, Danish, Spanish, German	-
	R22 Delete "consent cookie", communicate to third parties	-	-	CS

R21 Possible to change in the future

Under the revocability requirement, we define the low-level requirement on the possibility to withdraw consent in the future.

Requirement	Possible to change in the future
Description	The website should give an opportunity to withdraw consent after it has been given. The banner should allow the user to change the consent with easiness, by the same means, without detriment
Violation	It is not possible to withdraw consent by the same means it was asked; It is cumbersome to revoke – the means of withdrawing are more complex that initial consent; It is rather complex to understand for an average user how to remove cookies, and it is only accessible to the technical experts if other browser storages, such as HTML5 localStorage or cache should be cleaned. Moreover, there are no means to withdraw from browser fingerprinting. Revoking poses a delay, while positive consent was instantaneous.

Examples. Figures 30 and 31 show compliant banners to this requirement based on the possibility to change preferences in the future. The banner from the faktor.io website offers users the possibility to

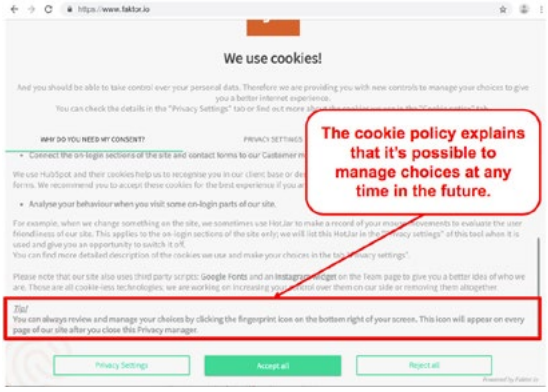


Figure 30 Compliance with "Possible to change in the future" requirement (https://www.faktor.io accessed 17 May 2019).

[200] cf. 29WP (WP187) (n 54) 33.
[201] cf. 29WP (WP259 rev.01) (n 4) 21.
[202] CNIL draft recommendation 2020 (n 95).
[203] Greek DPA, Guidelines on Cookies and Trackers (n 42).
[204] cf. Irish DPA Guidance (n 37).
[205] Danish DPA, "Guide on consent" (n 77).
[206] cf. 29WP (WP208) (n 17) 2.
[207] cf. 29WP (WP 259 rev.01) (n 4) 5.

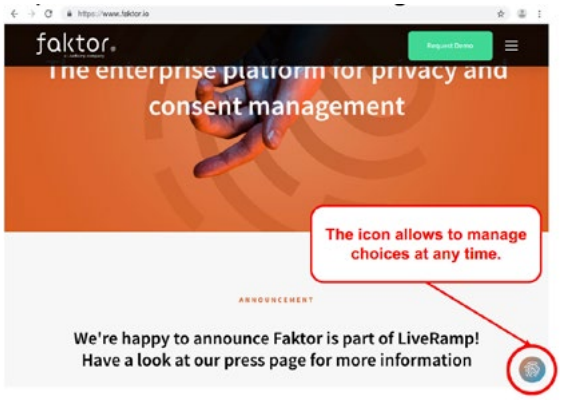


Figure 31 Compliance with the "Possible to change in the future" requirement (https://www.faktor.io accessed 17 May 2019).

review and manage their choices by clicking the fingerprint icon on the bottom right of the screen. This icon is available on every page of the site.

How to detect violations? Detection of this violation requires a manual analysis of the banner's interface, by evaluating whether there is a mean to change the consent after it has been given and how easy it is to revoke consent. Only standardized consent design can enable technical means to detect violations.

R22 Delete "consent cookie" and communicate to third parties

When the user revokes his consent, no BTT cab be further stored/read in the browser. Hence, revoking consent has two technical consequences: blocking and posterior deletion of cookies[208]-[209] in the user's browser, and as such, data processing will no longer occur. The CNIL[210] states that once the consent is revoked, both the reading and the deposit of new cookies should be blocked. The 29WP (WP 259 rev.01)[211] reasons in the same line,

> As a general rule, if consent is withdrawn, all data processing operations that were based on consent and took place before the withdrawal of consent - and in accordance with the GDPR - remain lawful, however, the controller must stop the processing actions concerned. If there is no other lawful basis justifying the processing (e.g. further storage) of the data, they should be deleted by the controller (art. 17(1)(b) and (3) GDPR. (...) Controllers have an obligation to delete data that was processed on the basis of consent once that consent is withdrawn, assuming that there is no other purpose justifying the continued retention.

Pursuant to the above analysis, we defined the technical low-level requirement that the publisher should delete the registered consent and communicate this withdrawal to all the third parties who have previously received consent.

Requirement	Delete "consent cookie" and communicate to third parties
Description	When consent is revoked, the publisher should delete the "consent cookie" and communicate the withdrawal to all the third parties who have previously received consent.
Violation	When the "consent cookie" is not deleted, and the publisher does not communicate to third parties that have received the consent

Example. We cannot provide an example for this requirement. As of June 2020, cookie banners rarely give users a way to modify their choice, and when they do, it is still unclear whether this change is actually communicated to a third party.

How to detect violations? Detection of such violation is a complex task because it requires checking whether the publisher has communicated the withdrawal of consent to all the third parties who have received it in the first place. As of today, there is no system that would be able to certify this because communication of consent (and of its withdrawal) does not have a standard technical implementation. If consent storage and communication is standardized and is observable in the web browser, technical tools could be devised for complete transparency and verification of this requirement.

6 Detection of violations for requirements based on Natural Language Processing and on user perception

This section refers on the mechanisms for detection of violations of requirements that depend on natural language processing (NLP) and user perception of the statements in natural language. We merge these requirements into three groups based on types of techniques that can be used to assess them:

1. Requirements based on the presence of information
   - R6 Accessibility of information page
   - R7 Necessary information on BTT
   - R8 Information on consent banner configuration
   - R9 Information on the data controller
   - R10 Information on rights
2. Requirements that rely on the distinguishability and structure of information
   - R3 No merging into a contract
   - R16 Distinguishable
3. Requirements that can be evaluated based on user perception and understanding.
   - R6 Accessibility of information page
   - R17 Intelligible
   - R18 Accessible
   - R19 Clear and plain language

Detection of violations for requirements based on the *presence* of information and distinguishability and structure of information.

The detection of violations related to the presence of information can be done manually but can be extremely time-consuming in case the required information is scattered across several pages of a privacy policy text, and privacy policies of included third parties. Libert[212] has measured that the average time to read both a given site's policy and

[208] The European Commission, it its portal, states that "data is deleted unless it can be processed on another legal ground (for example storage requirements or as far as it is a necessity to fulfill the contract", "What if somebody withdraws their consent?" https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/grounds-processing/what-if-somebody-withdraws-their-consent_en accessed 7 May 2020.

[209] It is noticeable that the request for revoking consent does not imply data erasure. For the data to be erased, the data subject needs to exercise this right to erasure. However, revoking consent should imply deletion of data as an immediate consequence.

[210] cf. CNIL (n 36).

[211] cf. 29WP (WP 259 rev.01) (n 4) 22.

[212] cf. Libert (n 128).

the associated third-party policies exceeds 84 minutes. Additionally, the text of privacy policies is often written with complex linguistic structured. For instance, Sanchez et al.[213] used the Flesch Reading Ease Score (FRES) and the Flesch-Kincaid Reading Level (FKRL), both used by legislators and government agencies, to measure the readability of privacy policies. Libert[214] proposed a formula to compute the average time required by users to read a policy. Therefore, manual analysis of privacy policies should be discouraged and instead automatic means to analyze privacy policies must be considered.

Natural Language Processing (NLP) tools have been conceived to analyze privacy policies. For instance, Libert[215] used an approach based on keywords to extract information from privacy policies. Following Brodie et al[216], The Usable Privacy Policy project[217] combines technologies, such as crowd sourcing, to develop browser plug-in technologies to automatically interpret policies for users. Ammar et al.[218] performed a pilot study, thus deriving and collecting a corpus of website privacy policies. This corpus has later been used by Harkous et al.[219] in a more complex deep-learning-based method within the Polisis tool that automatically extracts information flows described in privacy policies. A similar approach has been taken by Zaeem et al.[220] proposing PrivacyCheck browser extension that analyzes privacy policies with data mining. A further recent report compares the results of Polisis and PrivacyCheck on hundreds of privacy policies.[221] For a complete overview of modelling of privacy policies and automatic analysis of privacy policies at scale, a recent survey by Morel and Pardo[222] describes recent tools to analyze privacy policies at scale. Nevertheless, NLP tools have not been developed so far in the context of consent requirements and further research by NLP experts is needed to approach legal requirements presented in this paper.

To conclude, techniques using NLP and data mining are extensively tested today in order to process and structure privacy policies. Various NLP approaches are usually compared with respect to precision and recall metrics, however this evaluation heavily depends on (a) the "ground truth", that is how well the underlying corpus is labeled, and (b) the structure of privacy policies that are often not well organized. Nevertheless, these techniques have not been applied to assess legal requirements on consent presented in this work. Further investigation by computer scientists and legal experts are needed to assess whether these requirements are verifiable by automatic means or whether a new standardized format is required to display this information. In the NLP domain, such standardized format is often called by "controlled natural language, template, or pattern". With such a standardized format, a rather simple algorithm could verify the presence of the different information that is required by valid consent. Additionally, requirements that are based on distinguishability and structure of information need additional set of NLP tools that can answer the question whether consent is bundled with other types of information, such as a contract. We therefore conclude that such requirements can be analyzed manually (however with significant effort), or can be partially analyzed with technical means, whose efficiency is still to be evaluated by the experts.

Detection of violations for requirements based on *user perception* and *understanding*.

Several requirements for valid consent rely on user understanding and is heavily dependent on the target audience of a dedicated website. For example, requirement R6 "Accessibility of information page" depends on the usability of the website in question, but also on the technical ability of the user to find the information. Requirements R17, R18 and R19 are directly related to understandability of the users, but also on the technical background of users, e.g. when they are presented with statements such as "Can I have some cookies?", "If you do not allow cookies, website functionality will be diminished" or "Opting in to data collection will enable new and easier functionality". Such statements are often confusing to users, however in order to quantify whether these statements are intelligible (R17) and whether the language is clear and plain (R19), more structured evaluation of users' perception and understanding is needed.

Usable security and privacy research area has established the standards and techniques in building users' surveys and interviews in order to evaluate privacy perceptions, understanding and motivations of end users when it comes to privacy settings in online environment. For example, Utz et al.[223] have designed several types of cookie banners and ran a user survey to evaluate how users would interact with them depending on banners' text, position and provided options. Nouwens et al.[224] have also investigated users' engagement when placing controls in first- vs second-layer of a cookie banner. However, this research direction is only at its beginning, and more user studies are needed to evaluate whether cookie banners are indeed clear and well-understood by the target end users.

7 Discussion on shared consent

In this section, we discuss compliant scenarios related to the possibility of a shared consent. Section 7.1 refers to the shared responsibility between publishers and third parties. Section 7.2 discusses the scenario when consent itself in shared between them.

[213] Iskander Sanchez-Rola, Matteo Dell"Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos, "Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control" (*ACM Asia Conference on Computer and Communications Security* (AsiaCCS "19), Auckland, New Zealand, 2019).

[214] cf. Libert (n 128).

[215] cf. Libert (n 128).

[216] Carolyn A. Brodie, Clare-Marie Karat, and John Karat. 2006. An empirical study of natural language parsing of privacy policy rules using the SPARCLE policy workbench. In *Proceedings of the second symposium on Usable privacy and security* (SOUPS '06). Association for Computing Machinery, New York, NY, USA, 8–19.

[217] Wilson Shomir, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard H. Hovy, Joel R. Reidenberg and Norman M. Sadeh. "The Creation and Analysis of a Website Privacy Policy Corpus." *ACL* (2016).

[218] Waleed Ammar, Shomir Wilson, Norman Sadeh, Noah A. Smith, "Automatic categorization of privacy policies: A pilot study," Tech. Rep.

[219] Harkous, Hamza, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin and Karl Aberer. "Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning." *USENIX Security Symposium* (2018).

[220] Razieh Nokhbeh Zaeem, Rachel L German, and K Suzanne Barber, "Privacycheck: Automatic summarization of privacy policies using data mining". *ACM Transactions on Internet Technology* (TOIT), 18(4):53, 2018.

[221] Razieh Nokhbeh Zaeem, Suzanne Barber, "Government Agencies and Companies: a Study Using Privacy Policy Analysis Tools", UTCID Report (2020) https://identity.utexas.edu/assets/uploads/publications/Comparing-Privacy-Policies-of-Government-Agencies-and-Companies-a-Study-Using-Privacy-Policy-Analysis-Tools.pdf accessed 18th June 2020.

[222] Victor Morel, Raúl Pardo, "Three dimensions of privacy policies," Working Paper (2019)https://hal.inria.fr/hal-02267641 accessed 7 May 2020.

[223] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz, "(Un)informed Consent: Studying GDPR Consent Notices in the Field", *ACM SIGSAC Conference on Computer and Communications Security* (CCS'19), 2019, 973-990.

[224] cf. Nouwens et al. (n 178)

7.1 Shared responsibility between publishers and third parties

According to Article 5(3) of the ePD, when a main website content, fully controlled by the publisher, is setting cookies in his web domain (first-party cookies), then it will be primarily responsible for complying with the requirement to obtain an informed consent. We question if only the website publisher is obliged to display information and collect consent when third-party services are used as processors, or when third-party services act as controllers. The recurrent scenario in which multiple entities are involved in the installation of and access to an information by means of BTT is advanced by the 29WP, the ICO and the CNIL, while referring to BTT as "cookies" or "trackers":

- 29WP (WP171)[225] contends that a website publisher that allows third parties to place cookies shares the responsibility for information and consent (joint controllership);
- ICO[226] takes the view that where the website publisher sets third-party cookies, this same publisher and the third party are jointly responsible for ensuring that users are clearly informed about cookies and for obtaining consent. This means they are both determining the purpose and means of the processing of personal data of any user that visits the landing website. In substance, it is considerably more difficult for a third party, which has less direct control on the interface with the user, to achieve this. The ICO further instructs the need to include a contractual obligation into the agreements between publishers and third-parties on the allocation of responsibility to provide information about the third-party cookies and to obtain consent;
- CNIL[227] observes that when several parties may be involved in the user of trackers (e.g. a publisher and an advertising agency), publishers (of mobile sites or applications) are in the best position to inform users of the information on deposited trackers and to collect their consent, because of the control they exercise over the consent management interface, and the direct contact they have with the user.

Where controllers determine jointly the purposes and means of the processing, they must enter into a joint controllership agreement. Article 26 of the GDPR stipulates that both shall, in a transparent manner, determine their respective responsibilities, including which party provides information and obtains consent from the users.

Lastly, a data processor, in this context, is defined as an entity which installs information and/or has access to information stored on a user device exclusively on behalf of a data controller, without re-using the data collected via the tracker for the processor's own purposes. In such case, the parties must enter into a data processing agreement under Article 28 of the GDPR.

Real-Time Bidding (RTB) scenarios constitute a grey area: in RTB, publishers often have no knowledge about the specific third parties that a data subject might see, as they change between each website request. The CNIL (paragraph 26) states that a publisher does not have to show a banner again every time the list of third parties changes, when changes are not significant. On the other hand, the ICO[228] strongly criticized RTB because of the lack of transparency, due to the fact that publishers cannot have information about all the third parties they're sending user information to, and because of the opaque nature of the industry itself.

7.2 Shared consent between publishers and third parties

It is apparent from the case law from the CJEU (in its two decisions of Tele 2 and Deutsche Telekom, that we adapt to this context) that consent can be shared among publishers, insofar the processing operations pursues the same purposes, and that the user was informed thereof, as analyzed in section 5.3.1 (in point ii. consent "not required per publisher"). From these legal sources, we reason that if consent is collected in a lawful way, consent can be shared. Note that this observation is not explicitly prohibited by the law-maker.

From practical side, however, such reasoning raises questions about shared responsibility of the data controllers and, most importantly, implies reliance and trust on the way consent was collected by either other publishers or providers of third party content. Below we foresee a practical scenario of a website, in which third-party content a priori does not require consent, but merged with BTT that requires consent.

Example of shared consent. Imagine a user visiting two hypothetical websites: search.com and info.com.

Figure 32 depicts user's browser and its interaction with the server of search.com:

1. the user visits a website search.com, where a cookie named SID of search.com is placed in the user's browser. This cookie is used for advertising purposes, and hence requires consent. Let's assume a valid consent was collected by search.com before placing of cookies in the user's browser.
2. the user visits the website info.com, and it contains a customized search engine from search.com. Therefore, while visiting the website info.com, the user's browser automatically sends a request to search.com to fetch the needed functional content, i.e. the customized search engine. Upon this request, the browser also automatically attaches the cookie SID of search.com. Therefore, search.com receives its advertising cookie SID when the user visits the website info.com.

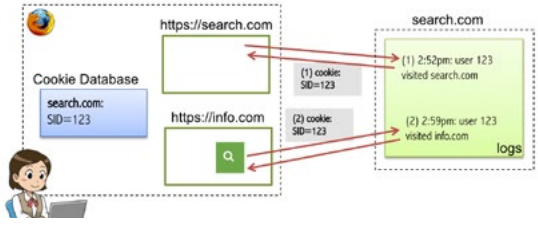


Figure 32 Example of shared consent.

In this hypothetical scenario, let us analyze how info.com can be compliant with the requirements of a valid consent:

- The publisher of the website info.com decides to collect its own consent for the search.com's advertising cookie SID. To be compliant, the consent should be collected before cookies are sent (see R2 "Prior to sending an identifier" requirement).
- This practice, however, prevents the loading of website's functional content before consent is given (the customized search engine from search.com is not loaded before consent is given), and hence violates requirement R4 "No tracking walls". The

[225] Article 29 Working Party, (WP 171) (n 33).
[226] cf. UK DPA (n 26).
[227] cf. CNIL (n 36).
[228] UK DPA, "Update report into adtech and real time bidding" (2020) https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf

publisher of infor.com, therefore cannot collect a valid consent by itself without violating legal requirements and thus, has to rely on the consent collected by search.com.
- The publisher of info.com relies on the consent already collected by search.com, hence info.com has to place full trust in how search.com has collected consent. This scenario will have practical and legal consequences. If the consent is not obtained in a valid way through search.com, then the website info.com will become jointly responsible for the non-compliant consent collection.

Therefore, when a third party merges content that does not require consent (e.g. functional content, such as customized search engine of search.com) with the content that requires consent (e.g. advertising cookies of search.com), this forces the website publisher to rely on the consent collected by the third party.

Conclusion: **only a negative consent can be shared**. We expand our discussion towards scenarios where a publisher relies on other publishers, previously visited by the user, for the collection of valid consent, or on third parties (as in section 7.1). If the user gives a positive consent (i.e. allows at least one type of data processing for at least one purpose), then if the consent collection has violated at least one of the requirements on valid consent (see Table 6), then the publisher can also be claimed responsible for such unlawful consent collection. This triggers a heavy responsibility burden on the publisher side, because he has no control over all the publishers or third parties on the way in which they collect consent (also, websites are very dynamic and quickly change over time, hence even if a publisher has verified consent collection in the past, such evidence might not hold upon a consequent visit to the same website). We underline that such model is not sustainable and very hard to manage for the publishers.

Nevertheless, consent can be shared if the user gives a negative consent (if the user refused all types of data processing for all purposes). In this case, the publisher can safely rely on this consent collected by other parties. Even in the case of an invalid consent in which the user gives a positive consent, but the collected consent is registered as a negative consent, the publisher would be complaint: he would respect a negative consent, and hence would not process any data (processing less data than allowed by the user's consent is always valid).

7.3 Summary on shared consent

Given the concerns raised in section 7.1, we believe that the legislator, when updating the EU ePrivacy framework, should clarify that **content not requiring consent must not be merged or served with BTTs that require consent**. Otherwise that publisher either violates one of the requirements on valid consent or is forced to rely on the way consent was collected by other parties. As we discussed in 7.2, relying on consent collected by other parties is not sustainable in practice and therefore puts a publisher in a weak and at the same time liable position for consent collection. Shared consent should be acceptable in practice only when the consent is negative, however a positive shared consent places again a publisher in a complex and liable position at the same time.

8 Discussion on the ePrivacy Regulation

This section presents a short evolution of the proposal of the ePrivacy Regulation (ePR) (section 8.1) and discusses whether it includes the proposed requirements for a valid consent (section 8.2).

8.1 Brief evolution

The ePR was first introduced by the European Commission in 2017 intended to replace the existing ePrivacy Directive 2002/58, as well as updating the current ePrivacy framework in the EU. The European Parliament adopted its position in October of the same year, but a stalemate still resides at the Council of the EU (in the Working Party on Telecommunications and Information Society - WP TELE). The text has been submitted through iterations of the seven EU presidencies (the holder rotates every six months) which failed to find a compromise between Member States. A Progress Report issued by the Council (Finnish Presidency doc. 14054/19[229] of 18 November 2019) noted that the Regulation continues to divide Member States, and many amendments have been suggested and debated so far and a compromise was not found.

The Croatian Presidency of the Council presented a revised proposal[230] in March 2020 and is still under discussion with the aim of getting a common position agreed. Before the revised Regulation can take effect, it will need to pass through trialogue negotiations among the Parliament, Council and the Commission, after which a compulsory grace period of a maximum of two years will apply to allow EU Member States to implement the Regulation. The Council Presidency also indicated that it is currently reflecting on additional revisions and intends to issue an additional document to be discussed during these meetings – a document which we are not knowledgeable as of today. In order to avoid the legal uncertainty created by some foundational elements of the current proposal, our analysis of the discussions in the Council of the EU specifically and is based on the consolidated text circulated by the Croatian Presidency, 5979/20, of 21 February 2020.

8.2 Requirements for valid consent in the recent draft proposal of the ePrivacy Regulation

Tracking walls. The European Parliament's draft of 26 October 2017(Article 8(1)(1)(b) and recital 22) called for an explicit ban on tracking walls for the first time (we state it as a requirement R4). However, Recital 21 of the Finnish draft of the ePR proposal of 2019 addressed indirectly the case of legitimizing tracking walls for advertising purposes. This indirect indication reveals that it is a topic of political controversy between the stakeholders. These draft signals that consent is valid (freely given) when the processing related to a service the user requested has advertising purposes. The Recital reads,

> "[I]n some cases the use of cookies may also be necessary for providing a service, requested by the end-user, such as services provided to safeguard freedom of expression and information including for journalistic purposes, such as online newspaper or other press publications (...), that is wholly or mainly financed by advertising provided that, in addition, the end-user has been provided with clear, precise and user-friendly information about the purposes of cookies or similar techniques and has accepted such use".

In the event the upcoming draft of the ePR becomes enforced thereby legitimizing the use of tracking walls, it would impact our results regarding requirement "no tracking wall" (R4, explained in section 5.2).

[229] Proposal for a Regulation on Privacy and Electronic Communications (2019) https://data.consilium.europa.eu/doc/document/ST-14068-2019-INIT/en/pdf accessed on 19 June 2020.
[230] Proposal for a Regulation on Privacy and Electronic Communications (2020) https://www.parlament.gv.at/PAKT/EU/XXVII/EU/01/51/EU_15125/imfname_10966469.pdf accessed on 19 June 2020.

Grounds to the processing and storage of BTT. The ePrivacy Regulation might impact the scope of the analysis carried out in this work, that is when a Browser-based Tracking Technology requires or is exempted of consent, summarized in Table 5 of Section 4.1. Specifically, Article 8 of the draft proposal of the ePR points to the following grounds for storage of and access to information from the from end-users' terminal equipment. We compare these with the purposes needing and exempted of consent that we analyzed in Table 5 of Section 4.1:

1. communication purposes (such as load balancing BTT). As we have analyzed, communication purposes would be exempted of consent.
2. consent;
3. necessary for providing a service requested by the end-user. Herewith we transpose the reasoning upheld in Table 5 about which purposes are necessary for providing a service requested by a user. The following purposes are considered to be necessary for the provision of a service (and therefore exempted of consent, requiring another legal basis): users input, local analytics/measurement, user security for service requested by the user, social media functionality requested by the user, authentication that is session-based, customization that is short-termed. However, other purposes are not considered necessary for providing a service to the end-user and would be subject to consent, according to the observations held in Table 5. These are:
    - advertising
    - measurement/non-local analytics
    - user security service not requested by the user
    - social media service not requested by the user
    - authentication that is persistent
    - customization that is persistent
4. necessary for audience measuring, provided that such measurement is carried out by the provider of the information society service requested by the end-user or by a third party, or by third parties jointly. This means that either local and local analytics might be used without consent (and hence subject to another legal basis). In this regard, if such rational will be enforced in the ePR, this would impact our analysis concerning our distinction between local and non-local analytics referred to in Table 5.
5. necessary for emergency communication;
6. necessary for the legitimate interests pursued by a service provider, except when such interest is overridden by the interests or fundamental rights and freedoms of the end-user, such as i. when the end-user is a child; ii. profiling; iii. special categories of personal data.

9 Related work

In this section, we give the reader a summary of the current context on legal compliance to BTT. Notably, consent for BTT deployment has been analyzed through different prisms that we regard in this paper: audits to websites in order to promote responsible behavior of web publishers; through guidance policy from stakeholders; through enforcement actions and decisions of the Court of Justice and DPAs; and finally, through legal scholarship literature. For readability issues, this section is divided in four parts. Section 9.1 refers to relevant audits on websites. Section 9.2 considers DPAs guidance on the elements for a valid consent. Section 9.3 explains some of the issued decisions related to valid consent. Section 9.4 shows the related work on consent analysis portrayed by legal scholarship and automatic auditing of websites by computer scientists. Table 21 presents a summary of the audits, guidelines and enforcement actions related to consent to BTT that will be presented with further detail in the following subsections.

Table 21 Summary of the audits, guidelines and enforcement actions related to consent to consent to BTT

Audits on websites	Guidelines	Enforcement actions
29WP, 2015	29WP/EDPB, 2020	Planet49 Judgment of the CJEU, 2019
EDPS inspection, 2019	EDPS, 2016	French DPA decision, 2018
Bavarian Audit, 2019	UK DPA, 2019	Spanish DPA decisions, 2019
Dutch Check, 2019	German DPA, 2019	Belgium DPA decision, 2019
Irish Sweep, 2020	Finnish DPA, 2019	
Greek Sweep, 2020	Spanish DPA, 2019	
	French DPA, 2020	
	Greek DPA, 2020	
	Belgian DPA, 2020	
	Irish DPA, 2020	

9.1 Recent audits on websites

Recurrent audits on websites are aimed at information-gathering assessment of the state of cookie (and related technologies) consent compliance at scale. Such auditing initiatives reveal the current playing level field of websites that still struggle to comply with the GDPR and ePrivacy rules on consent. In contrast, our paper analyzes legal documents to define more precise requirements for these banners.

Table 22 shows the audits of consent banners performed by stakeholders and the related requirements for consent banners.

Table 22 Stakeholder's audits on websites per sector and the related requirements

Stakeholders	Tested websites on the use of cookies and related technologies	Related requirements
29WP, 2015	478 websites in the e-commerce, media and public sectors across 8 Member States	Unambiguous, revocable, informed
EDPS inspection, 2019	websites of major EU institutions and bodies, e.g. European Council, Council of the EU, Commission, CJEU, Europol, European Banking Authority, EDPS, EDPB, 2018 International Conference of Data Protection and Privacy Commissioners (ICDPPC 2018)	Prior, informed
Bavarian Audit, 2019	40 Bavarian providers (online stores, media companies, insurance companies, banks, sports teams	Informed, freely given, unambiguous
Dutch Check, 2019	175 websites of web shops, municipalities and media	Unambiguous, freely given
Irish Sweep, 2020	38 Irish-based websites of sectors, including media and publishing, the retail sector, restaurants and food ordering services, insurance, sport and leisure and the public sector	Unambiguous, freely given, revocable, prior
Greek Sweep, 2020	audit of the use of cookies by popular Greek websites	Non-specified

29WP Cookie Sweep Combined Analysis Report, 2015.[231] This sweep included 478 websites in the e-commerce, media and public sectors across 8 Member States. Both the automated scan and manual review provide the results, thusly: 74% of studied websites displayed banners, 54% thereof did not request user's consent but were merely informative. 70% of the 16,555 cookies stored were third party cookies. More than half of the third party cookies were set by just 25 third-party domains. The sweep showed that a banner was a popular method of informing visitors on the use of BTT in addition to a link in the header or footer to more information. Only 16% of sites offered a configurable banner. The majority relied on browser settings or an opt-out tool provided on a third-party site (e.g. a third-party advertising site). Amongst those sites which set the highest number of cookies, most had taken some steps to inform users about the use of cookies through a banner which was either permanent (requiring an active click from the user within the banner), a banner which disappears on the next user click anywhere on the page or timed to disappear after a certain length of time. In comparison to our work, three requirements were analyzed: unambiguous, revocable, and informed.

EDPS inspection, 2019.[232] This inspection was carried out on the websites of major EU institutions and bodies, e.g. the shared website of the European Council and the Council of the EU, the Commission, the Court of Justice of the EU, Europol and the European Banking Authority. The EDPS also inspected the website of the European Data Protection Board (EDPB), the 2018 International Conference of Data Protection and Privacy Commissioners (ICDPPC 2018) and the EDPS website itself. The EDPS developed a tool that automatically collects information on personal data processed by websites via the use of cookies, web beacons, page elements loaded from third parties and security of encrypted connections. The inspection revealed that several of the websites were not compliant with the Regulation nor with the ePD. One of the issues encountered was third-party tracking without prior consent. Other issues encountered included the use of trackers for web analytics without visitors' prior consent. In comparison to our work, two requirements were analyzed: prior and informed.

Bavarian State Office for Data Protection Supervision Audit, 2019.[233] This audit found that forty Bavarian providers (online stores, media companies, insurance companies, banks, sports teams, etc.) use trackers, but only a quarter of the websites inform users about the use of these tools. The remaining providers either did not inform users at all or only informed them insufficiently about the use of tracking tools as part of their privacy policies. Regarding the use of cookie banners, 20% of websites failed to ask users to consent to the use of cookies. Consent obtained were either not given in advance, they were given uninformed, or there was a lack of voluntariness. In comparison to our work, three requirements were analyzed: informed, freely given and unambiguous.

Dutch DPA Check, 2019.[234] This DPA carried out a check on approximately 175 websites of web shops, municipalities and media, etc. to determine whether they meet the requirements for placing tracking cookies. Some violations were detected, such as preselected boxes and tracking walls. All checked websites are not compliant. The organizations behind these websites have received a letter from the AP calling on them to adjust their working methods accordingly. In comparison to our work, two requirements were analyzed: unambiguous and freely given.

Irish DPA, 2020.[235] This DPA conducted a sweep between August 2019 and December 2019 on a selection of 40 websites across a range of sectors, including media and publishing, the retail sector, restaurants and food ordering services, insurance, sport and leisure and the public sector to check compliance with the ePD and the GDPR on the use of cookies and other tracking technologies. The 38 respondents signaled either that they were aware of non-compliant practices with the existing rules, or that they had identified improvements that they could make to their websites in order to demonstrate compliance; some detected practices were: implied consent, non-necessary cookies set on landing. A lack of tools for users to vary or withdraw their consent choices, badly designed—or potentially even deliberately misleading—cookie banners and consent-management tools, and prechecked boxes. In comparison to our work, two requirements were analyzed: unambiguous, freely given, revocable, and prior.

Greek DPA, 2020. In February 2020, this authority performed a sweeping audit of the use of cookies by popular Greek websites, in which the Authority found that non-compliance with the GDPR was widespread. So far, we could not have access to the content of the sweep.

9.2 Guidance on a valid consent for BTT

The analysis of the requirements for a valid consent is contained in the guidelines of the 29WP and the EDPS. Other DPAs provide guidance on obtaining consent specifically for trackers. In this section, we give a brief account of the significant aspects of the most comprehensive guidelines. In general, there is broad agreement that a consent banner must contain information on the BTT used, either directly in the banner or via corresponding links. However, the requirements of the authorities vary with regard to the concrete form in which the user may provide consent, ranging from allowing any action from the user, to opinions requiring a concrete banner design configuration. We perform an in-depth comparative analysis of the current state of the art DPA guidelines to the low-level requirements proposed in this paper in Table 7 (see page 15).

EDPS Guidelines on the protection of personal data processed through web services provided by EU institutions, 2016.[236] While these Guidelines are in principle aimed at the EU institutions, anyone or any organization interested in data protection and web services might find them useful. The main topics covered in these Guidelines that are useful for this paper are: the use of cookies, scripts and any other tools to be stored or executed on the user terminal device; server-side processing of personal data and the wider issue of tracking.

29WP Guidance on cookies. Per the 29WP guidelines, the following guidance documents were observed in our study, for they interpret closely the consent requirements in respect of cookies and BTT and were quoted alongside this paper.

[231] Article 29 Working Party, "Cookie sweep combined analysis – Report" (WP229, 3 February 2015).

[232] European Data Protection Supervisor, "EDPS flags data protection issues on EU institutions' websites" (2019) https://edps.europa.eu/sites/edp/files/edpsweb_press_releases/edps-2019-04-website_inspections_en.pdf accessed 7 May 2020.

[233] Bavarian DPA, "Safe on the Internet – Data Protection Check on Digital services" (our translation) (2019) www.lda.bayern.de/media/sid_ergebnis_2019.pdf accessed 7 May 2020.

[234] Dutch DPA, "Many websites incorrectly request permission to place tracking cookies" (2019) (n 99).

[235] Irish DPA, "Sweep conducted between August 2019 and December 2019" (2020) https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Data%20Protection%20Commission%20cookies%20sweep%20REVISED%2015%20April%202020%20v.01.pdf

[236] cf. EDPS Guidelines (n 35).

- Opinion 2/2010 on online behavioral advertising, (WP171, June 2010);
- Opinion 15/2011 on the definition of consent (WP187, July 2011);
- Opinion 04/2012 on Cookie Consent Exemption (WP194, June 2012);
- Working Document 02/2013 providing guidance on obtaining consent for cookies (WP208, October 2013);
- Opinion 9/2014 on the application of Directive 2002/58/EC to device fingerprinting (WP224, November 2014).
- Guidelines on consent under Regulation 2016/679 (WP259 rev.01, April 2018), which was extended by Guidelines 05/2020 on consent under Regulation 2016/679 (May 2020);

French DPA Guidelines on cookies and trackers, 2019.[237] The CNIL published guidelines on cookies and trackers. Hereby we consider the most relevant points related to our work on consent for cookies.

Consent. Continuing to browse a website after its cookie banner is displayed will no longer be considered to be valid consent for cookie use.

Auditable: Operators using trackers have to be able to prove that they have obtained affirmative consent from the user, at all times.

Scope: The guidelines apply to all types of operations involving cookies and trackers on any type of device, including smart phones, computers, connected vehicles and any other object connected to a telecommunications network open to the public.

Cookie Wall: The user should not suffer any major inconvenience if they refuse to give or withdraw their consent. The practice of blocking access to a website or a mobile application unless consent is provided does not comply with the GDPR.

Revocable: Users should be able to withdraw their consent at any time. User-friendly solutions must therefore be implemented to allow users to withdraw their consent as easily as they have given it.

Operator´s Roles and Responsibilities: An operator using cookies and trackers is considered to be a controller and is therefore fully responsible for obtaining valid consent.

In 2020 the CNIL launched a public consultation on a draft recommendation on cookies and other trackers[238] in order to adapt the GDPR rules to trackers and consent. As main takeaways, this document proposes designs of cookie banners; enunciates best practices for legal compliance, practical arrangements for implementation, and examples of how to comply with the applicable rules. It suggests means to define specific purposes for processing; proposes neutral design interfaces and design patterns to avoid misleading design practices; provides examples of proof of consent; and finally advocates for the development of standardized interfaces operating in the same way and using a standardized vocabulary to make it easier for users to understand when navigating from one site to another.

UK DPA Guidance on the rules on use of cookies and similar Technologies, 2019.[239] The ICO updated its guidance on the use of cookies and other similar technologies. Some of the key points to note from the guidance are herewith described. Cookie walls may not comply with the cookie consent requirements and it states these as inappropriate if the use of a cookie wall is intended to require, or influence, users to agree to their personal data being used as a condition of accessing its service, as a user has no genuine choice but to accept cookies. The authority clarifies that implied consent conveyed through statements such as "by continuing to use this website you are agreeing to cookies", pre-ticked boxes or any equivalents, such as sliders defaulted to "on", cannot be used for non-essential cookies. Consent mechanisms incorporating a "more information" section, rather than as part of the initial banner are also deemed non-compliant on the basis that they do not allow users to make a choice before non-essential cookies are set. On the types of cookies, the ICO enunciates that advertising and analytics cookies are not "strictly necessary" and are subjected to consent rules.

German DPA Guidance, 2019. The German DPA published the "Guidelines for Telemedia Providers"[240] and Frequented Asked Questions (FAQ)[241] about web tracking and cookie banners. According to the guidance, a cookie banner is only necessary if cookies are set through the website that require data protection consent; if a website only sets cookies for which the site operator does not require consent, the guidance considers the banner avoidable. In this guidance, consent is needed when a web service uses web services on its website that analyze the user across several domains, e.g. social media plugins, advertising networks or analysis tools such as Google Analytics. The regulator alerts that consent to the use of cookies must not be preselected and does not consider the opt-out procedure to be sufficient. The authority published also a note[242] on the use of cookies and cookie banners – "what must be done with consent (ECJ ruling "Planet49")?".

Finnish DPA Guidance on Confidential Communications, 2019.[243] The NCSC-FI at Traficom mentioned in the guidelines that consent can be requested by using any preferred method (e.g. browser/application setting or pop-up window) as long as it is not requested by using a pre-ticked checkbox. The use of cookies and the related practices must also be indicated on a website in such a manner that a user can obtain additional information about them.

Spanish DPA Guide on the Use of Cookies, 2019.[244] The AEPD published new Guidelines on the Use of Cookies and similar technologies, which were prepared in collaboration with different organizations in the marketing and online advertising industries (e.g. Adigital, IAB Spain, etc.). The Guidelines provide factors for categories of cookies:

- *Who manages cookies* (proprietary or third-party);
- *Purpose* (technical, customization, analytical, and behavioral advertising); and
- *Duration* (session or persistent).

The AEPD provides the following examples of actions that could be considered an affirmative action: the use of the scroll bar, insofar as the information on cookies is visible without using it; clicking on any link contained in the site other than those in the second layer of information on cookies or the privacy policy link; on devices such as mobile phones or tablets, by swiping the initial screen and accessing the content. Even if the Planet49 Judgment[245] ruled otherwise, the AEPD Guidelines state:

[237] cf. CNIL Guidelines (n 36).
[238] CNIL draft recommendation 2020 (n 95).
[239] cf. ICO Guidance (n 26).
[240] cf. German DPA Guidelines (n 10).
[241] German DPA, "FAQ about Cookies and Tracking" (2019) www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2019/04/FAQ-zu-Cookies-und-Tracking.pdf accessed 7 May 2020.
[242] cf. German DPA (10).
[243] Finnish DPA Guidance (n 13).
[244] cf. Spanish DPA Guide (n 82).
[245] cf. Planet49 Judgment (n 11).

> For the action of continuing browsing to be deemed a valid consent, the information notice must be displayed in a clearly visible place, so that due to its shape, color, size or location, it can be secured that the notice has not gone unnoticed to the user. Additionally, it will be necessary, for the consent to be deemed granted, that the user performs an action that can be qualified as a clear affirmative action. For instance, a clear affirmative action may be considered to browse to a different section of the website (other than the second layer of information on cookies or the privacy policy), to slide the scroll bar, closing the first layer notice or clicking on any content of the service. The mere fact of viewing the screen, moving the mouse or pressing the keyboard cannot be considered an acceptance.[246]

Greek Data Protection Authority, 2020.[247] This authority issues in February 2020 its Guidelines on the use of internet cookies and trackers, following the completion of a sweeping audit of the use of cookies by popular Greek websites, in which the Authority found that non-compliance with the GDPR was widespread. The practical guidance provides specific recommendations on as well as practices that should be avoided.

The Irish DPA, 2020. [248] This authority released in April 2020 a comprehensive guidance note on cookies and other tracking technologies. The main takeaways are: implied consent is not valid; the use of prechecked boxes and sliders set to "on" as default are non-compliant. A consent banner must not obscure the text of the privacy or cookie notice. It explains that a website operator must take accessibility into account in designing interfaces. Uniquely, the Guidance advises users should not be "nudged" into accepting trackers and should be given the opportunity to consent on a granular basis. The Guidance also says that there should be given equal prominence between "accept" and "reject" buttons.

9.3 Enforcement actions of consent by the Court of Justice of the European Union (CJEU) and by DPAs

In this section, we show the enforceable decisions in connection to consent requirements for BTT referred in judgements of the Court of Justice of the EU (CJEU) and administrative decisions issued by DPAs. Table 23 shows the enforcement actions by the CJEU and DPAs in relation to the analyzed requirements.

Table 23 Enforcement actions by the CJEU and DPAs in relation to the analyzed requirements

Enforcement actions	Requirements
Planet49 Judgment of the CJEU, 2019	Specific, informed, unambiguous
French DPA decision, 2018	
Spanish DPA decisions, 2019	Unambiguous (configurable banner), informed, prior, revocable
Belgian DPA decision, 2019	Informed, revocable, unambiguous

Planet49 Judgment of the CJEU, 2019.[249] On October of 2019, the CJEU decided that the consent which a website user must give to the storage of and access to cookies on his or her equipment is not validly constituted by way of a prechecked checkbox which that user must deselect to refuse his or her consent. The Court notes that consent must be specific so that the fact that a user selects the button to participate in a promotional lottery is not sufficient for it to be concluded that the user validly gave his or her consent to the storage of cookies. Furthermore, according to the Court, the information that the service provider must give to a user includes the duration of the operation of cookies and whether or not third parties may have access to those cookies. The German Federal Court of Justice (BGH) followed the CJEU's preliminary ruling in the "Planet 49 case" which determined that the request for consent by a preselected tick box constitutes an "unreasonable disadvantage to the user".[250]

Spanish DPA decisions, 2019. This DPA[251] in October of 2019 fined Vueling for failing to provide a compliant consent banner. The poorly constructed banner did not provide a configuration panel that allows the user to delete cookies in a granular way. It was considered that the information was insufficient for the intended purpose of allowing users to configure preferences in a granular or selective form.[252] This DPA also fined IKEA[253] for placing cookies before users clicked the only option in the banner: the "OK" button. Users were prompted with a cookie banner stating that "IKEA website uses cookies that make browsing much easier. More information about cookies". Initially, users were instructed to block cookies through browser settings, also including "strictly necessary" cookies like e.g. shopping cart cookies rendering the website basically impossible to use. It did not identify the purposes of the different cookies used, nor informed about the possibility of setting the usage preferences of the cookies. It did not provide a link to the panel or cookie configuration system enabled to select them in granular form. It did not include a specific button or mechanism for rejecting all cookies. The warning that "If you do not change your browser settings, we will understand that you agree to receive all cookies from the IKEA website" breaches consent requirements. It did not report on how to revoke the consent given.

Belgian DPA decision, 2019. On December 2019, this DPA imposed a fine of € 15,000 on an SME operating a legal information website for their noncompliant cookie management and privacy policy. It found that their privacy policy lacked transparency and infringed the rules on information to be provided. In particular, it provided insufficient information about the cookies deployed on the website (e.g. the list of cookies used, their purpose, the identity of third parties concerned, and the lifespan of the cookies) and did not properly identify the controller. Moreover, the cookie policy was only available in English, whereas the website targeted Dutch and French-speaking readers. The website did not obtain opt-in consent for certain types of cookies used, including first-party analytics cookies. There was no easy way for users to withdraw consent.

Decisions and complaints against the IAB Transparency and Consent Framework (TCF). The TCF of the IAB Europe implements consent solutions for parties in the digital advertising chain. Herewith we report the French DPA decision. The CNIL, in 2018, sued an advertisement company Vectaury using the IAB framework, invoking a

[246] Author's translation from the Spanish version.
[247] Greek DPA, Guidelines on Cookies and Trackers (n 42).
[248] cf. Irish DPA Guidance (n 37).
[249] cf. Planet49 Judgment (n 11).
[250] German Federal Court of Justice for consent to telephone advertising and cookie storage (2020) https://www.bundesgerichtshof.de/SharedDocs/Pressemitteilungen/DE/2020/2020067.html?nn=10690868 accessed 18th June 2020.
[251] cf. Spanish DPA decision (n 155).
[252] EDPB press release, "The Spanish Data Protection Authority fined the company Vueling for the cookie policy used on its website with 30,000 euros" (2019) https://edpb.europa.eu/news/national-news/2019/spanish-data-protection-authority-fined-company-vueling-cookie-policy-used_en accessed 7 May 2020.
[253] Spanish DPA decision, "Procedimiento PS/00127/2019" (2019) www.aepd.es/resoluciones/PS-00127-2019_ORI.pdf accessed 7 May 2020.

lack of informed, free, specific and unambiguous consent.[254] For the CNIL, the consent text was not clear enough regarding the final use of collected data, and the formulation may lead users to incorrectly assume that refusing consent prevents a free access to the website or lead to more intrusive advertisement. It was also noted that pre-ticking consent-related checkboxes was not compliant with the Recital 32 of the GDPR. It was required the list of recipients of users' data to appear immediately when consent text is displayed. In April 2019 a formal complaint[255] was filed against the IAB for a tracking wall on its own website that forces visitors to consent if they want to access the website.

9.4 Related work on consent and consent banners

In the following, we outline academic work on consent banners. All of these works focus on measuring or detecting legal violations in cookie banners from a technical point of view. We extensively discuss the legal analysis of detected violations in previous works. Table 24 displays a comparison summary between related works and ours regarding the requirements for consent banners.

Table 24 Comparison summary between related works and ours on the requirements for consent banners

Works	Tested websites	Tested high and low-level requirements	Comments
Carpineto et al.	17k Italian public administration websites	Prior (R1 prior to storing an identifier)	Violations found on websites where a banner is not displayed
Traverso et al.	100 Italian websites	Prior (R1 prior to storing an identifier), unambiguous (R15 correct consent registration)	Correct consent registration was tested via confronting the number of trackers after acceptance and after refusal of consent.
Trevisan et al.	36k EU websites	Prior (R1 prior to storing an identifier)	
Van Eijk et al.	1500 websites from 18 countries (EU, USA, Canada)	-	They test the presence of a banner on EU websites and simulating user's visit from different EU, USA and Canada.
Degeling et al.	6,500 EU websites	Specific, informed, unambiguous (R12 configurable banner)	
Sanchez-Rola et al.	2,000 websites	Prior, unambiguous (R11 affirmative action design, R12 configurable banner, R15 correct consent registration), readable and accessible (R19 using clear and plain language, R20 no consent wall), revocable (R22 delete consent cookie)	
Libert et al.	180 news websites	Prior (R1 prior to storing an identifier)	All third-party cookies were considered, independently whether they require consent or not
Utz et al.	5,000 websites	Free (R4 no tracking wall), unambiguous (R12 configurable banner, R13 balanced choice, R11 affirmative action design)	Also studies the influence of design on users choice.
Matte et al.	23k EU websites	Unambiguous (R11 affirmative action design, R12 configurable banner, R14 post-consent registration, R15 correct consent registration)	No pre-ticked boxes
Nouwens et al.	10k websites	Unambiguous (R13 balanced choices, R11 affirmative action design), free (R4 no tracking wall), informed (high level requirement)	No pre-ticked boxes. Also studies the influence of design on users' choice.
Leenes and Kosta	100 Dutch websites	Free (R4 no tracking walls), unambiguous (R12 configurable banner)	
Matte, Santos et al.	575 advertisers registered in IAB Europe TCF	-	Study of purposes used in cookie banners of IAB Europe TCF

Carpineto et al.[256] developed a tool to automatically check the legal compliance of cookie banners in Italian Public Administration websites in 2016. The authors used language-dependent text analysis methods and detected cookies based on lists of known trackers. In this study, the only criteria for non-compliance is whether the website uses tracking cookies however does not display a cookie banner. By automatically analyzing Italian Public Administration websites, authors identified 1,140 non-compliant websites placing tracking

[254] French DPA, "Decision n° MED 2018-042 of October 30th, 2018 enforcement notice against the company Vectaury" (2018) www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000037594451 accessed 7 May 2020.

[255] Brave, "Complaint against IAB Europe's "cookie wall"(2019)https://brave.com/wp-content/uploads/2019/04/3-April-2019-complaint-to-Data-Protection-Commission-of-Ireland-regarding-IAB-Europe-cookie-wall-and-consent-guidance.pdf accessed on 19 June 2020.

[256] Claudio Carpineto, Davide Lo Re, Giovanni Romano, "Automatic assessment of website compliance to the European cookie law with CooLCheck" (*Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society*, ACM, Vienna, Austria, 2016) 135-138.

cookies in the user's browser.

Traverso et al.[257] measured the impact of the ePD's cookie policy on web tracking on 100 Italian websites. Visiting the same website before and after giving consent by clicking on the accept button of the cookie banner, they measured the difference in the number of included trackers. Their results are alarming: there were few differences between both scenarios. In the no-consent-given scenario, they found an average of 29.5 trackers per webpage, none of them containing 0 tracker, and half of them containing more than 16.

Trevisan et al.[258] built an automatic tool "CookieCheck" to check violations of the ePD in 36 000 popular websites popular in the European Union (plus 4 extra-EU countries) in early 2017. Using a list-and heuristic-based tracking cookie detection method, they tested whether websites requested consent before installing cookies. They found that 49% of websites installed profiling cookies before user consent, a number raising to 74% when considering any third-party cookie. On a smaller set of 241 websites from 3 European countries, they observed that 80.5% of those installing tracking cookies did not regard the user's consent. Interestingly, they observed no significant difference in the number of installed tracking cookies between desktop and mobile browsers.

Van Eijk et al.[259] studied cookie banners after the GDPR came in force in 2018. Leveraging a crowd-sourced list of known banners, they automatically detected cookie banners on 40.2% of European Union websites. Accessing websites from different countries using VPNs, they found that the provenance of the user has not so much impact as the expected audience of a website regarding the prevalence of banners. They also observed important variations between websites of different top-level domains.

Degeling et al.[260] performed a study comparing the information presented to users of 6,500 EU websites before and after the GDPR, focusing on the changes in privacy policies and information presented to users. In particular, the authors studied characteristics of 31 cookie banner libraries by installing them locally. They observed a 6% increase in cookie banners adoption by website pre- and post-GDPR. They have identified the following categories within existing implementations of consent notices:

"*No option notices*" to simply inform the user that the website uses cookies and if the user continues to use the website, they agree to this use;

"*Confirmation-only banners*" displays button with an affirmative text, such as "OK", or "I agree", through which, by clicking on it expresses the user's consent;

"*Binary notices*" provide users with a button to accept and another to reject the use of all cookies on the website;

"*Category-based notices*" assembles the cookies used by the website into categories. Users can allow or disallow cookies of each category individually by (un)checking a settings menu or toggling an "on–off" switch;

"*Vendor-based notices*" allow visitors to accept or decline cookies for each third-party service used by the website (conceding more fine-grained control). They originate from third-party libraries, as the IAB Europe's Transparency and Consent Framework, which refers to its advertising partners as "vendors".

Sanchez Rola et al.[261] performed a wide manual evaluation of tracking in 2,000 websites, inside and outside of the EU. The aim was to measure how easy it is to opt-out from tracking if the user desires to do so and assessing whether it is possible at all. Their results show that tracking is prevalent, happens mostly without user's consent, and opt-out is difficult. They note whether banners respect many requirements relevant to our work: whether they offer a way to refuse tracking, whether consent is set automatically when visiting the website, whether tracking happens despite a refusal of consent, and whether the consent cookie is deleted upon refusal.

Concerning tracking, Libert et al.,[262] in a factsheet for the press, studied the impact of the GDPR on the amount of third-party content and cookies on news websites. On about 180 European news sites, they observe a 22% drop in the number of third-party cookies before (April 2018) and after (July 2018) the GDPR, but only 2% drop in third-party content.

Another prominent work related to ours is the research from Utz et al.[263] The authors ran a number of studies, gathering ~5,000 of cookie notices from leading websites to compile a snapshot (derived from a random sub-sample of 1,000) of the different cookie consent mechanisms. They also worked with a German ecommerce website over a period of four months to study how more than 82,000 unique visitors to the site interacted with various cookie consent designs. The authors reached the following findings significant to our paper:

Cookie consent notices do not offer a choice to the users; they are placed at the bottom of the screen (58%); not blocking the interaction with the website (93%); and offering no options other than a confirmation button that does not do anything (86%);

The more choices offered in a cookie notice, the more likely visitors were to decline the use of cookies;

A majority also try to nudge users towards consenting (57%) — such as by using "dark pattern" techniques like using a color to highlight the "agree" button (which if clicked accepts privacy-unfriendly defaults) vs displaying a much less visible link to "more options" so that pro-privacy choices are buried off screen;

Mentioning cookies in a consent notice decreases the chance that users allow cookie use.

Matte et al.[264] found several plausible violations of both the GDPR and the ePD in the implementations of cookie banners by actors using IAB Europe's Transparency and Consent Framework (TCF). They automatically and semi-automatically detected four suspected GDPR and ePD violations on more than 1400 websites using this framework (found among 23k websites) to display cookie banners

[257] Stefano Traverso, Martino Trevisan, Leonardo Giannantoni, Marco Mellia, Hassan Metwalley less, "Benchmark and comparison of tracker-blockers: Should you trust them?" (*Network Traffic Measurement and Analysis Conference*, Dublin, Ireland, 2017) 1-9.

[258] Martino Trevisan, Stefano Traverso, Eleonora Bassi and Marco Mellia, "4 Years of EU Cookie Law: Results and Lessons Learned" (*Proceedings on Privacy Enhancing Technologies*, Issue 2, 2019) 126-145.

[259] Rob Van Eijk, H. Asghari, Philipp Winter, Arvind Narayanan, "The Impact of User Location on Cookie Notices (Inside and Outside of the European Union)" (*Workshop on Technology and Consumer Protection* (ConPro '19), San Francisco, CA, 2019).

[260] Martin Degeling et al. (n 127).

[261] Sanchez-Rola et al (n 213).

[262] Timothy Libert, Lucas Graves and Rasmus Kleis Nielsen, "Changes in third-party content on European news websites after GDPR" (Reuters Institute for the Study of Journalism Reports: Factsheet, Reuters Institute for the Study of Journalism, 2018).

[263] cf. Christine Utz et al. (n 223).

[264] cf. Matte et al. (n 143).

and found at least one violation in more than half of them. Considered violations are positive consent registered before any user action, no option to refuse consent, registered consent not respecting user's decision, and pre-ticked boxes. Their work includes an analysis by a co-author which is an expert in law as to why considered violations can be considered legal violations.

Nouwens et al.[265] detected dark patterns in about 700 websites using IAB Europe's TCF, and found that only 11.8% of banners meet minimal legal requirements: unambiguous consent, accepting being as easy as rejecting consent, no pre-ticked boxes. They also measure how some design choices in banners affect users' decision on consent.

From a legal perspective, both studies by Kosta[266] and Leenes[267] on a regulatory approach towards cookies were prominent to our analysis of the legal and technical side of consent requirements for BTT. Of particular relevance to our work is the study performed by Leenes and Kosta,[268] in which the authors examined manually the practices of 100 Dutch websites with regard to cookie consent mechanisms. They found that most of these websites do not respect the ePD. Those researchers defined a four-tier classification of consent implementation from the analyzed banners:

- explicit agreement to all cookies used on the site, without possibility to opt out;
- implicit agreement to all cookies used on the site, i.e. banners whose button's text is not a response to a question regarding the user's consent;
- coerced agreements to all cookies, i.e. "cookiewalls", when users cannot access the website without accepting tracking cookies;
- detailed choice/consent of cookies, i.e. banners containing a "settings" button.

Among the 100 sites studied, they found 25 banners of the 1st type, 54 of the 2nd one, none of the 3rd one and 6 of the last one. 87% of visited websites installed cookies "of various type" on first page load, i.e. irrespective of the choice of the user.

Finally, Matte, Santos et al.[269] analyzed the purposes of data processing defined in IAB Europe's Transparency and Consent Framework (TCF). They derive from their legal analysis that most of the purposes defined in this framework cannot be exempted of consent. Measuring purposes that advertisers registered in the TCF declare to use, they observe that hundreds of advertisers rely on a legal basis that could not be considered compliant under the GDPR. While their paper focused on the legal requirements for purposes, our work analyzes design requirement in banners.

10. Conclusion

In this paper, we have performed an interdisciplinary analysis of how consent banners are supposed to be implemented to be fully compliant with the European data protection rules on consent. As a result, we defined **22 operational** fine-grained requirements for consent banners to level-set current practices of websites. For each requirement, we assessed if verification by technical means (with computer science tools), manual means (by a human operator) or based on user studies (with surveys and interviews) is needed to assess compliance with valid consent (see Table 6). We identified hurdles related to the assessment of requirements with technical means, described advances made so far in the corresponding technical areas, and identified the need for standardization of consent interfaces and of consent storage and sharing in web applications. Because of the absence of standards for consent, many requirements can be fully assessed manually by a human operator, which is time-consuming and not scalable.

Additionally, our analysis of requirements applies only to cases when browser-based tracking technology (BTT) is used for the purposes that require consent (see Table 5). However, even if computer scientists can detect the presence of a BTT on a given website, in general it is not possible to identify what purpose BTT is used for on a given website (see the discussion on page 102) . We therefore believe that legislators should propose standardized and machine-readable means to specify purposes for each BTT that can be further analyzed automatically with technical tools at scale.

Notably, the number of website audits and sweeps, complaints and fines – in response to alleged ePrivacy and GDPR violations when using BBT – have increased over the last year. Very recent guidelines by the EDPB and DPAs have been issued to surpass rogue website practices while requesting and registering consent, as reflected in section 8.

However, we observed regulatory discrepancies regarding the guidelines for a valid consent. Some DPAs move a step further in several issues compared to corresponding guidelines of other DPAs of other Member States of the European Union, thus signifying an increasing trend towards stricter rules concerning online trackers. Thus, the *EBDP may be the most appropriate institution to assure cooperation*, streamlined guidance and consistency of DPA guidelines, leveraging from the collective knowledge of other DPAs. More legal certainty can possibly be given by the upcoming ePrivacy Regulation. On the other hand, the European Commission needs to map inconsistencies and gaps of national laws and work with Member States to further harmonize the implementation of the GDPR.

Some low-level requirements described in this paper (see a full list in Table 6) were not yet discussed at EDPB level (and only some DPAs propose in their guidelines), while few of them have never been considered and result from a cooperation of the authors of this work – legal and computer science experts. We hope that all the low-level requirements presented in this work raise discussion and are included in further updates of the ePrivacy Regulation, and if is not possible, then at least explicitly recommended by the European Data Protection Board.

Acknowledgements

We are grateful to Fabiano Dalpiaz who has given feedback on the section related to natural language processing (NLP), as well as to Frederik Borgesius and Gaëtan Goldberg who have validated some of the requirements during the development of this paper. This work has been partially supported by the ANR JCJC project PrivaWeb (ANR-18-CE39-0008 and by the Inria DATA4US Exploratory Action program.

[265] cf. Nouwens et al. (n 178).
[266] cf. Kosta (n 27).
[267] cf. Leenes (n 82).
[268] Ronald Leenes, Eleni Kosta, "Taming the Cookie Monster with Dutch Law - A Tale of Regulatory Failure" (2015) *Computer Law & Security Review*, Volume 31, Issue 3, 317-335.
[269] Matte C, Santos C, Bielova N. (2020), Measuring the usage of purposes and their legal basis by advertisers in the IAB Europe's Transparency and Consent Framework, Annual Privacy Forum (APF) (forthcoming).

10

Technology and Regulation

Researching with Data Rights

Jef Ausloos* & Michael Veale**

right of access, GDPR, data protection, transparency, research methods

j.ausloos@uva.nl

m.veale@ucl.ac.uk

The concentration and privatization of data infrastructures has a deep impact on independent research. This article positions data rights as a useful tool in researchers' toolbox to obtain access to enclosed datasets. It does so by providing an overview of relevant data rights in the EU's General Data Protection Regulation, and describing different use cases in which they might be particularly valuable. While we believe in their potential, researching with data rights is still very much in its infancy. A number of legal, ethical and methodological issues are identified and explored. Overall, this article aims both to explain the potential utility of data rights to researchers, as well as to provide appropriate initial conceptual scaffolding for important discussions around the approach to occur.

1. The GDPR: Research Curse or Blessing?

Data protection legislation, in particular the EU General Data Protection Regulation (GDPR),[1] has been seen by some researchers as creating frustrating barriers to their work.[2] Data minimization and storage limitation restrict the extent to which large databases can be amassed for future consultation. Information requirements can limit covert or subtle collection, and sit at tension with web-scraping and research on social media. Uncertainty and anxiety in risk-averse organizations can stifle data-driven research initiatives, leaving researchers dissuaded or simply encouraging them to disregard the rules.[3] The GDPR does not appear to improve the situation much: while it recognizes the importance of scientific research through derogations from the default data protection rules, it also leaves a lot unsaid, and outsources crucial interpretative guidance to Member States.[4] If Member States do not clarify the issue in national law or regulatory guidance, the interpretative burden falls to the organizations who benefit from these exemptions. Increased public scrutiny in light of the Cambridge Analytica scandal, which involved online data collection infrastructure established by the University of Cambridge,[5] has only increased fear of infringement.

At the same time, the societal stakes for investigating the online world have never been higher, or the need more urgent.[6] Networked systems have changed individuals' experiences of the world, their enhanced and more pervasive mediating roles 'affecting the ways in which we understand our own capabilities, our relative boundedness, and the properties of the surrounding world'.[7] Many readers will not need much introduction into the 'algorithmic war-stories' unearthed in recent years that focus on the impact of these mediating systems,[8] particularly through the work of journalists, civil society and activist-minded research groups. Work by journalists such as Julia Angwin, Lauren Kirchner and Kashmir Hill has explored the way that technol-

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 (hereafter 'GDPR').

2 See, e.g.: Edward S Dove, 'The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era' (2018) 46 *J Law Med Ethics* 1013; Wouter Van Atteveldt, 'Toward Open Computational Communication Science: A Practical Road Map for Reusable Data and Code' [2019] 20; Rossana Ducato, 'Data Protection, Scientific Research, and the Role of Information' (2020) 37 *Computer Law & Security Review* 105412.

3 e.g., the extensive work by David Erdos, 'Stuck in the Thicket? Social Research under the First Data Protection Principle' (2011) 19 *Int J Law Info Tech* 133; David Erdos, 'Systematically Handicapped? Social Research in the Data Protection Framework' (2011) 20 *Information & Communications Technology Law* 83; David Erdos, 'Constructing the Labyrinth: The Impact of Data Protection on the Development of "Ethical" Regulation in Social Science' (2012) 15 *Information, Communication & Society* 104.

4 GDPR, art 89(2–3).

5 https://www.theguadian.com/news/series/cambridge-analytica-files

6 Also emphasised in European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust' (19.2.2020); European Commission, 'A European Strategy for Data' (Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions, 19.2.2020).

7 Julie E Cohen, *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice* (Yale University Press 2012).

8 Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 *Duke Law & Technology Review* 18. See also Malte Ziewitz, 'Governing Algorithms: Myth, Mess, and Methods' (2016) 41 *Science, Technology, & Human Values* 3 (on 'algorithmic drama').

* Jef Ausloos is a Postdoctoral Research Fellow at the Institute for Information Law, University of Amsterdam

** Michael Veale is a Lecturer in Digital Rights and Regulation at the Faculty of Laws, University College London

Received 3 Dec 2020, Accepted 28 Dec 2020, Published 4 Jan 2021.

ogy firms and society interact in the context of discrimination and manipulation. Research groups such as the Data Justice Lab at Cardiff University have mapped the use of technology in the public sector,[9] and research teams have made use of freedom of information rights to discover more about the way that data-driven systems are being procured.[10] Researchers from teams such as the Algorithm Auditing Research Group at Northeastern University[11] have been using bots and online scraping and analysis tools to better understand discrimination and inequality in AdTech systems,[12] and data leakage from apps to third party trackers.[13] Meanwhile, workers' collectives are mobilizing in an attempt to reclaim data from platform companies to prove their eligibility for basic employment rights, and to use as evidence in tribunals and other proceedings.[14] These efforts often run into a range of methodological, practical and legal hurdles which in many cases are easily arguable to have been preserved by powerful forces seeking to retain the secrecy that allows them to work with limited scrutiny and accountability.

We present one flipside of this (deliberately) dismal picture. Could the GDPR *enable*, rather than stifle, data-rich research? In particular, in a world where private and influential data infrastructures are coordinated by a limited number of powerful actors, might the GDPR's provisions be used as a *source* of data, rather than applying constraints on collection? We believe it can be. Researching with data rights can provide data for a range of 'digital methods', which include applying and adapting existing methods such as surveying, ethnography or text analytics to new, digitized sources of information, as well as fueling new 'natively digital' methods aimed at building understanding based on features of digital spaces (such as hyperlinking, wireless sensing, recommender systems or browsing histories) which have no clear offline analogue.[15] Potential access to such data sources is made possible through the GDPR's strengthened information provision measures, found predominantly in Articles 12 through 15, and underpinned by the overarching transparency principle in Article 5(1)(a).

2. Existing means to access enclosed data and their limits

Digital methods are plagued by the problem of 'special access', which is 'required for the study of certain natively digital objects'.[16] While much of our life is entwined with sensors and actuators,[17] this data gathered on us is not, generally, stored on our own devices. Even though the average user 'has in their pocket a device with vastly more resource than a mainframe of the 1970s by any measure', they usually end up 'using [their] devices as vastly over-specified dumb terminals.'[18] Instead, computation and data storage generally happens in rented 'cloud' infrastructure. This move to the 'cloud' is value-laden in nature, coming with a natural tendency to concentrate the power that comes from data and the constant experimental decisions made around its use in the hands of central, proprietary nodes.[19] Despite the fact that data is not, generally, considered a form of property, platforms in the informational economy have established 'de facto property arrangements' by enclosing such data using legal strategies such as terms-of-use agreements to heavily structure interactions.[20] These entities only rarely release data entirely and/or unconditionally, whether for legal, economic or technical reasons, and appear willing to fight against initiatives that would force them to do so more readily.

Lack of access has made private entities the gatekeepers of the data or infrastructure necessary for utilizing digital methods. Consequently, research that happens inside or with the blessing of these entities tends to be limited to that in the private entity's interests (notably profit and reputation), rendering it hard to impossible for outside actors (notably academia, journalists, and civil society more broadly) to perform critical parallel inquiry. Internal research undertaken for the genuine purpose of discovery, but which might impugn the firm's legitimacy, is unlikely to see the light of day.[21] Sealing off societally important data processing operations has rendered it very hard to scrutinize the practices of these entities.[22]

We identify roughly four main groups of approaches through which researchers external to these entities attempt to study them with digital methods:

- voluntary data sharing agreements (ad hoc arrangements);
- programmatic access (technical tools offered by data controllers);
- scraping and interception (independent technical tools); and

9 Lina Dencik and others, 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services' (Data Justice Lab, Cardiff University, 2018) https://perma.cc/39CY-H8L7 (accessed 21 August 2020); Lina Dencik and others, 'The "Golden View": Data-Driven Governance in the Scoring Society' (2019) 8 *Internet Policy Review*.

10 Marion Oswald and Jamie Grace, 'Intelligence, Policing and the Use of Algorithmic Analysis: A Freedom of Information-Based Study' (2016) 1 *Journal of Information Rights, Policy and Practice*; Robert Brauneis and Ellen P Goodman, 'Algorithmic Transparency for the Smart City' (2018) 20 *Yale Journal of Law & Technology* 103.

11 https://personalization.ccs.neu.edu

12 Michael Carl Tschantz and Anupam Datta, 'Automated Experiments on Ad Privacy Settings' (2015) *2015 Proceedings on Privacy Enhancing Technologies* 92.

13 Reuben Binns and others, 'Third Party Tracking in the Mobile Ecosystem' in *Proceedings of the 10th ACM Conference on Web Science* (WebSci '18, New York, NY, USA, ACM 2018).

14 James Farrar, 'Why Uber Must Give Its Drivers the Right to All Their Data', (*New Statesman*, 2 April 2019) https://www.newstatesman.com/america/2019/04/why-uber-must-give-its-drivers-right-all-their-data accessed 22 July 2019; 'Uber drivers demand access to their personal data' (Ekker Advocatuur, 19 July 2020) https://ekker.legal/2020/07/19/uber-drivers-demand-access-to-their-personal-data (accessed 17 August 2020).

15 Richard Rogers, *Digital Methods* (The MIT Press 2013).

16 Rogers (n 15) 15.

17 See generally Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar Publishing 2015).

18 Jon Crowcroft and others, 'Unclouded Vision' in Marcos K Aguilera and others eds, *Distributed Computing and Networking* (Springer Berlin Heidelberg 2011) 29.

19 Seda Gürses and Joris van Hoboken, 'Privacy after the Agile Turn' in Evan Selinger and others (eds), *The Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018).

20 See Julie E Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019) 44–45.

21 See e.g. Karen Hao, 'We read the paper that forced Timnit Gebru out of Google. Here's what it says' (*MIT Technology Review*. 4 December 2020) https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru (on the concerns with independence of the process surrounding the scholarly publication of a paper on bias and environmental issues in large language models co-authored by fired Google researcher Timnit Gebru).

22 See references in Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015); Nicholas Diakopoulos, 'Algorithmic Accountability: Journalistic Investigation of Computational Power Structures' (2015) 3 *Digital Journalism* 398; Muhammad Ali and others, 'Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes' [2019] arXiv:190402095 [cs], 5; European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (6 January 2020).

- data disclosure requirements (legal transparency requirements).

These approaches all have their benefits and shortcomings – discussed below – and can further be categorised along two axes, depending on (a) the relationship between researcher and data holder (collaborative v adversarial) and (b) the point of access (top-down v bottom-up) (Table 1). This last qualification is based on whether data is obtained through the entity holding the data directly, or via its users. Put briefly, *top-down* data access enables a helicopter view or overarching insights (e.g. internet platform content moderation or ad archives), but the respective data will often be very high-level, notably to safeguard users' privacy. *Bottom-up* data access enables granular insights into individuals' data (e.g. reactions to personalized media-diets), but may fail to give a global picture, may require significant technical expertise and raises legal concerns. *Collaborative* data access arrangements may be very advantageous if they work, but can create undesirable dependencies and solidify power dynamics. *Adversarial approaches* – ie independent of data holders' goodwill to release data – are therefore often the only way for researchers to obtain access to data, but come with their own set of (legal, technical, economical) challenges.

Table 1 Current approaches to data access

	Collaborative	Adversarial
Top-down	Voluntary data sharing	Data disclosure requirements
Bottom-up	Programmatic access	Scraping and interception

Against this backdrop, we believe there to be an important role for the law – democratically designed and enforceable – in framing the scope and limits of adversarial data access approaches. GDPR transparency rights show particular promise as such an adversarial, bottom-up tool for research data access (notably considering the drawbacks of scraping and interception). In order to better appreciate this, let us briefly zoom into the different approaches to data access.

2.1 Voluntary data sharing agreements

Some researchers/institutions obtain access to privately held data via 'data philanthropy' initiatives[23] and/or through amicable relationships they might entertain with the relevant actors (e.g. Facebook's *Social Science One* initiative;[24] or the UK's *Consumer Data Research Centre*[25]). Beneficial as these may be to the respective researchers, such an approach risks further solidifying existing power dynamics in academia (and the private sector). Efforts like these have also been characterized as 'corporate data philanthropy', designed to generate positive publicity rather than critical research.[26] Moreover, researchers/institutions may have several reasons for not wanting to associate with private entities as a precondition for doing research, such as fear of real or perceived loss of independence that may result from, for example, an obligatory sign-off procedure on produced findings.[27]

Threats from researchers to pull out of Facebook's *Social Science One* initiative after they were denied the data promised have only stoked scepticism about the feasibility of this ad hoc style of data access to form a basis for future digital methods.[28] Indeed, recent efforts aim to introduce more of a formal structure and regulatory oversight to data sharing arrangements, through the development of data protection codes of conduct in this area.[29]

2.2 Programmatic access

Researchers and institutions may find creative ways to re-purpose entities' existing programmatic tools, such as application programming interfaces (APIs) in order to get access to data. These allow users to access the data of themselves, others, or the environment through programmatic querying which will return machine readable data according to a given specification. There are several challenges with this approach.

APIs are generally designed with developers, not researchers, in mind, and can consequently fail to return research-grade data. API access to a stream of content may only provide a limited, non-random sample. Twitter's public APIs showed at most 1% of public tweets, and systematic biases compared to the full data-stream has cast the representativeness of reliant studies into question.[30] Such APIs in general only show public information — even then, only data that developers consider important — with available sampling and filtering commands lacking the necessary expressiveness for research.[31]

API use for research may also go against applicable Terms of Service, and researchers may therefore risk retaliatory action, such as being kicked off the platform.[32] In some jurisdictions, contract law and computer misuse law has been blurred, creating heightened legal risk as well.[33]

APIs have more recently become political tools used by platforms to exclude certain business or functionality from integration, and the interaction between developers and the changing nature of APIs has been described as 'risky territory', an 'ongoing battle' and 'hostile'.[34] Strategic changes to an API may break an entire set of business

23 See e.g., https://www.mastercardcenter.org/action/call-action-data-philanthropy.

24 Facebook, 'Facebook Launches New Initiative to Help Scholars Assess Social Media's Impact on Elections' (*Facebook Newsroom*, 9 April 2018) https://newsroom.fb.com/news/2018/04/new-elections-initiative (accessed 23 April 2019).

25 https://www.cdrc.ac.uk/about-cdrc

26 Axel Bruns, 'After the "APIcalypse": Social Media Platforms and Their Fight against Critical Scholarly Research' (2019) 22 *Information, Communication & Society* 1544, 1551.

27 Bruns (n 26) 1553. See generally the open letter regarding corporate support of research into technology and justice at https://fundingmatters.tech.

28 Camilla Hodgson, 'Facebook given Deadline to Share Data for Research', (*Financial Times*, 28 August 2019) https://www.ft.com/content/147eddec-c916-11e9-af46-b09e8bfe60c0 (accessed 11 September 2019); Social Science One, 'Public Statement from the Co-Chairs and European Advisory Committee of Social Science One' (11 December 2019) https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one (accessed 5 January 2020).

29 See arts. 40–41 GDPR. See also Mathias Vermeulen, 'The Keys to the Kingdom. Overcoming GDPR-Concerns to Unlock Access to Platform Data for Independent Researchers' (OSF Preprints 27 November 2020); 'Call for Comment on GDPR Article 40 Working Group' (EDMO, 24 Nov 2020) https://edmo.eu/2020/11/24/call-for-comment-on-gdpr-article-40-working-group (accessed 23 December 2020).

30 Andrew Yates and others, 'Effects of Sampling on Twitter Trend Detection' (2016) *Proceedings of the International Conference on Language Resources and Evaluation*.

31 Alexandra Olteanu and others, 'Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries' (2019) 2 *Front Big Data*.

32 Olteanu and others (n 31).

33 e.g., the arguments in the US case *Sandvig et al. v. Sessions*, No. 1:16-cv-01368 (D.D.C. June 29, 2016), See generally Annie Lee, 'Algorithmic Auditing and Competition Under the CFAA: The Revocation Paradigm of Interpreting Access and Authorization' (2019) 33 *Berkeley Tech L J* 1307. But see *hiQ Labs, Inc v LinkedIn Corporation* 2019 WL 4251889 (United States, Ninth Circuit).

34 Tania Bucher, 'Objects of Intense Feeling: The Case of the Twitter API : Computational Culture' (2013) 3 *Computational Culture: A Journal of Software Studies*; Paddy Leerssen and others, 'Platform Ad Archives: Promises and Pitfalls' (30 April 2019).

models, while privileged access can create economic advantage. Social media platform Facebook was accused by the UK House of Commons Digital, Culture, Media and Sport Committee of using API access to take 'aggressive positions' against competitor apps, taking actions leading to the failure of businesses.[35] In this context, APIs cannot be easily relied on by researchers, who may find their software rendered dysfunctional by external business logics or even a shift in functionality aimed at breaking their efforts to rigorously interrogate a system.[36] Unless many streams of their work rely on this software, researchers rarely have the time or resource to engage in this 'arms race' and maintain software in the face of sudden, unexpected and often ill-documented changes.[37] API-based research with inconvenient findings for private entities is unlikely to be sustainable.

Connectedly, and perhaps more problematically, is the fact that from a privacy and data protection point of view the use of APIs does not preclude bad faith (or at least ethically questionable) actors obtaining access to personal, or even sensitive, data. The quintessential example of this is Aleksander Kogan and Cambridge Analytica, whose Facebook add-on 'thisisyourdigitallife' harvested millions of Facebook profiles of both the users of the add-on and those users whose data they in turn had access to. The ensuing mixture of *bona fide* research and data privacy scandal has challenged the field of researchers using digital methods.[38] As a result, several APIs do not or no longer allow access to users who are not somehow connected to the requester, limiting data access to those users within the requester's 'social graph'.[39] While the dropping of access has been framed in terms of privacy and security, sceptics see it also as 'a convenient step towards disabling and evading unwanted independent, critical, public-interest scholarly scrutiny.'[40]

2.3 Scraping and interception

Researchers also rely on independent technical or methodological tools to obtain useful data otherwise sealed-off by private entities without their blessing.

Scraping tools or bots are common sources of data where APIs are restrictive or unavailable. Such an approach has some legal support in several jurisdictions with *text and data mining* exemptions in copyright laws. In some cases, such as in Japanese law, these exemptions are not restricted to actors or purposes,[41] while in other laws, such as in the UK since 2014 and in the new 2019 EU Copyright Directive, there are limitations of scope for 'non-commercial' and/or 'research' purposes.[42]

In some cases, what is of interest is how the platform, its users and its non-users behave in interaction with it. At scale, this is likely to require the use of bots or crowd workers. However, the use of both bots and crowd workers for research, particularly when bots impersonate a 'real' user or crowd workers make use of their own social profiles is, in at least some cases, legally and ethically contentious.[43] However, it also brings opportunities for co-creation of research, potentially seeing participants as co-researchers rather than research subjects.[44]

In other cases, data is trickier to obtain due to advanced enclosure techniques by firms.[45] Researchers wishing to understand what data mobile apps send and to where they send it often have to resort to monitoring users' internet traffic using a virtual private network (VPN), requiring invasive device access.[46] Approaches on the Web, which is a little more open in this regard, include browser plugins to monitor social media (WhoTargetsMe,[47] Algorithms Exposed,[48] FB-Forschung[49]), search engine (e.g. DatenSpende)[50] or general browsing activity (e.g. Robin).[51]

These approaches are more resistant to retaliatory action by the respective entities[52] or misuse by bad actors, and the active recruit-

35 See generally documents presented and published under privilege by Damian Collins MP to the Commons DCMS Committee. These documents were a selection of emails that were obtained through discovery in the US Courts in a lawsuit involving developer Six4Three and Facebook. Despite being held under seal by the San Mateo Superior Court, they were given to the UK Parliament which published them under privilege, and are available at https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Note-by-Chair-and-selected-documents-ordered-from-Six4Three.pdf. See generally Digital, Culture, Media and Sport Committee, 'Disinformation and "Fake News"' (18 February 2019).

36 e.g., Ariana Tobin Jeremy B. Merrill, 'Facebook Moves to Block Ad Transparency Tools — Including Ours' (*ProPublica*, 28 January 2019) https://www.propublica.org/article/facebook-blocks-ad-transparency-tools (accessed 17 April 2019).

37 See, on the death of Netvizz, a popular research tool for those studying Facebook, Bruns (n 26) 1549.

38 Tommaso Venturini and Richard Rogers, '"API-Based Research" or How Can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach' (2019) 0 *Digital Journalism* 1.

39 Dietmar Janetzko, 'The Role of APIs in Data Sampling from Social Media' in Luke Sloan and Anabel Quan-Haase (eds), *The SAGE Handbook of Social Media Research Methods* (SAGE Publications Ltd 2016).

40 Bruns (n 26) 1550.

41 On the Japanese text and data mining exemptions, See FutureTDM, D3.3+ Baseline Report of Policies and Barriers of TDM in Europe (Horizon 2020 Project 665940 2016) 75. This law has been recently clarified and extended by the Act of Partial Revision of the Copyright Act (Japan) 2018, which clarifies the use of copyrighted works in relation to machine learning. See generally European Alliance for Research Excellence, 'Japan Amends Its Copyright Legislation to Meet Future Demands in AI' (European Alliance for Research Excellence, 9 March 2018) http://eare.eu/japan-amends-tdm-exception-copyright (accessed 24 June 2019).

42 For the recently passed European provision, See Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance), arts 3–4; for the 2014 UK provision, See Copyright, Designs and Patents Act 1988 s 29A.

43 See generally (n 21).

44 See Alexander Halavais, 'Overcoming Terms of Service: A Proposal for Ethical Distributed Research' (2019) 22 *Information, Communication & Society* 1567, 1578.

45 See generally on data enclosure Julie E Cohen, 'Property and the Construction of the Information Economy: A Neo-Polanyian Ontology' in Leah A Lievrouw and Brian D Loader (eds), *Handbook of Digital Media and Communication* (Routledge forthcoming).

46 e.g., Abbas Razaghpanah and others, 'Haystack: In Situ Mobile Traffic Analysis in User Space' [2015] 14; Jingjing Ren and others, 'ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic' in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (MobiSys '16, New York, NY, USA, ACM 2016); Yihang Song and Urs Hengartner, 'PrivacyGuard: A VPN-Based Platform to Detect Information Leakage on Android Devices' in *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices* (SPSM '15, New York, NY, USA, ACM 2015); Anastasia Shuba and others, 'AntMonitor: A System for On-Device Mobile Network Monitoring and Its Applications' [2016] arXiv:161104268 [cs].

47 https://whotargets.me/en

48 https://algorithms.exposed

49 https://fbforschung.de/. This tool combines a data-gathering plugin with occasional surveys with participants, enabling more in-depth information than what can merely be observed.

50 https://datenspende.algorithmwatch.org

51 Balázs Bodó and others, 'Tackling the Algorithmic Control Crisis – the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents' (2017) 19 *Yale JL & Tech* 133.

52 Though certainly not immune, as illustrated by plugins of ProPublica and WhoTargetsMe slightly being blocked by Facebook changing some of its HTML code. See generally Jeremy B Merrill (n 36); Digital, Culture, Media and Sport Committee (n 35) 64.

ment of research subjects gives an opportunity to inform them about the study and its consequences. Nevertheless, this 'reverse engineering' approach is fragile and labor intensive. Infrastructure such as operating systems or web browsers can change and be changed, disrupting these tools in the process. Because of the predominance of vertically integrated companies in the digital economy,[53] firms often control both this infrastructure and the data of research interest (e.g. Alphabet, Google and Chrome), creating issues similar to that of APIs.[54]

2.4 Data disclosure requirements

Researchers may also put their hopes in regulatory interventions (or threats of legislation) forcing more transparency. So far, legal instruments primarily focus on transparency of public sector information (i.e. freedom of information acts or the EU's Public Sector Information and Open Data Directives[55]), but new initiatives are underway to open up privately held data as well.[56] Targeted transparency and disclosure policies are familiar policy instruments in many policy areas such as the environment, health and safety.[57] Such instruments are commonly used by individuals, civil society and journalists[58] — and often designed with them in mind — but also have surprisingly high usage by commercial entities for profitable ends.[59] Some disclosures, such as curated datasets of information on disinformation, have been forced from platforms more-or-less at threat of legislation in times of political contestation.[60] Freedom of information (FoI) laws have been used to study data-driven systems already,[61] but their scope is generally limited to the public sector, and in some jurisdictions, contractors thereof.[62] Transparency obligations also exist in many safety-critical sectors, such as electronics, food, pharmaceuticals and the like, although in practice these are rarely triggered by citizens, and usually relate to access to documents through regulators, or policies concerning labelling. Data protection impact assessments (DPIAs), which may contain useful information about processing practices for researchers, are not obliged to be made public under EU data protection law, and therefore do not count amongst transparency measures covered here.[63]

A spate of new and proposed digital regulation *does*, however, include transparency reporting on digital phenomena applicable to private entities. The proposed EU Terrorist Content Regulation would have hosting service providers set out 'a meaningful explanation of the functioning of proactive measures including the use of automated tools',[64] and published annual transparency reports containing information on detection measures and statistics on takedown information.[65] The recently adopted Regulation on promoting B2B fairness and transparency, covering platforms which intermediate trade such as online e-commerce marketplaces and 'app' stores, requires providers to reveal 'the main parameters determining ranking and the reasons for the relative importance of those main parameters as opposed to other parameters', and require search engines to provide such information in 'an easily and publicly available description, drafted in plain and intelligible language' and to 'keep that description up to date'.[66] In line with its ambitious 'strategy for data',[67] the European Commission also put forward three major policy proposals at the tail end of 2020. All three – the Data Governance Act, Digital Services Act, and Digital Markets Act – place strong emphasis on transparency obligations for digital services.[68] Obligations under the proposed Digital Services Act would mandate influential 'gatekeepers' to provide data to vetted researchers investigating systemic societal risks.[69] In the run-up to the 2019 EU elections, the European Commission also managed to make a number of powerful platforms issue monthly transparency reports on a voluntary basis.[70] Inspiration might also be drawn from gender pay gap disclosure legislation increasingly common throughout the world.[71] Finally, it is also worth

53 Ian Brown and Christopher T Marsden, *Regulating Code: Good Governance and Better Regulation in the Information Age* (MIT Press 2013) xii.

54 See also Thomas Claburn, 'Google Nukes Ad-Blocker AdNauseam, Sweeps Remains out of Chrome Web Store', (*The Register*, 5 January 2017) https://www.theregister.co.uk/2017/01/05/adnauseam_expelled_from_chrome_web_store (accessed 18 June 2019).

55 Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information OJ L 345 ('PSI Directive'); from June 2021 repealed and replaced by Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information OJ L 172/56 ('Open Data Directive').

56 cf. European Commission, 'Building a European Data Economy' (10 January 2017) https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy (accessed 28 April 2018); European Commission, 'A European Strategy for Data' (Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions, 19.2.2020). For proposed regulations, see Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) COM/2020/767 final; Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final; Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) COM/2020/842 final.

57 For two case studies drawing from the environmental and health context respectively, See Jef Ausloos and others, 'Operationalizing Research Access in Platform Governance What to Learn from Other Industries?' (25 June 2020). See generally: Archon Fung and others, *Full Disclosure: The Perils and Promise of Transparency* (Cambridge University Press 2007).

58 See Matt Burgess, Freedom of Information: A Practical Guide for UK Journalists (Routledge 2015).

59 See Margaret B Kwoka, 'FOIA, Inc.' (2016) 65 *Duke Law Journal* 1361.

60 See generally Amelia Acker and Joan Donovan, 'Data Craft: A Theory/Methods Package for Critical Internet Studies' (2019) 22 *Information, Communication & Society* 1590.

61 e.g., Oswald and Grace (n 10); Brauneis and Goodman (n 10).

62 The UK Information Commissioner has been active in her attempts to try to argue for contractors to fall under freedom of information law. See generally Information Commissioner's Office, Outsourcing Oversight? The Case for Reforming Access to Information Law (ICO 2019).

63 Reuben Binns, 'Data Protection Impact Assessments: A Meta-Regulatory Approach' (2017) 7 *International Data Privacy Law* 22.

64 Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online (COM/2018/640 final) (hereafter Proposed Terrorist Content Regulation), art 8(1).

65 Proposed Terrorist Content Regulation art 8(3).

66 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (European Union 2019) Article 5.

67 European Commission, 'A European strategy for data' (n 6).

68 See references in (n 56).

69 Proposed Digital Services Act (n 56), art 31(2).

70 European Commission - DG Connect, 'Code of Practice on Disinformation' (Text, Digital Single Market, 26 September 2018) https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation (accessed 19 July 2019). Relatedly, see European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European Democracy Action Plan' (3 December 2020).

71 The Equality Act 2010 (Gender Pay Gap Information) Regulations 2017, SI 2017/172. But note that The Financial Times caused a stir when it noted that many companies reporting their gender pay gaps under new UK legislation reported an identical mean and median: something so statistically improbably it was effectively indicative or an error, a cover-up, or both. See Billy Ehrenberg-Shannon and others, 'Cluster of UK Companies Reports Highly Improbable Gender Pay Gap', (*Financial Times*, 12 July 2017) https://www.ft.com/content/ad74ba76-d9cb-11e7-a039-c64b1c09b482 (accessed 17 June 2019).

pointing to the European Commission's ambitious data strategy, which includes the tabling of an 'enabling legislative framework for the governance of common European data spaces' by the end of 2020.[72]

As it stands under current legislation, the scope of these disclosure obligations is patchy at best. In Europe, tensions exist between FoI and privacy law,[73] which in turn limit the extent to which even public agencies can make disclosures of individual level data. Recent tensions between ICANN and European data protection regulators around the WHOIS database for website registrars have further illustrated these tensions.[74] This stands in contrast to several US cases, such as the famed *COMPAS* study into recidivism systems by *ProPublica*, where journalists used public records access to analyze a proprietary software system they accused of racial bias.[75] Replicating this method in Europe would likely run into difficulties as authorities would be unlikely to release identifiable data of convicts or ex-convicts as they did to *ProPublica* for reasons of data protection and privacy.[76]

***

These four approaches to data for digital methods all have their benefits and shortfalls. This paper does not seek to present a panacea, but it does seek to add a tool to the ever-changing toolkit. That tool is data protection transparency, in particular, the use of data rights. The rest of this paper considers legal, social, technical and ethical aspects of this proposed data source in research contexts.

3 Transparency Provisions in the GDPR

Data protection is characterized in large part by its transparency provisions. These started off as a form of general oversight over the primarily state-affiliated 'databanks' motivating early data protection law, and now are best known as tools for coping with information asymmetries that in many cases originate today's predominantly private-sector information economy.[77]

This article focusses primarily on European data protection law, and in particular the GDPR. This legal framework contains a panoply of tools, ranging from individual rights to more collectively and collaboratively-flavored provisions. Amidst this panoply, the *right to access* is explicitly highlighted in the EU Charter of Fundamental Rights. Not only should data be processed *fairly*,[78] but the Charter's Article 8(2) proclaims that everyone 'has the right of access to data which has been collected concerning him or her.'

More recently, ensuring transparency of automated processing and profiling in particular has also become a considerable public and legislative concern. Developments in the Council of Europe illustrate this well in the (recently modernized) Convention 108[79] and earlier recommendations.[80] The modernized convention provides that each individual shall have a right 'to obtain, on request, knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her'.[81] Related provisions are found in EU data protection law and French administrative law,[82] as well as in national adaptations to data protection legislation in EU member states.[83]

3.1 Flavors of data protection transparency

Transparency provisions come in many different shapes and flavors in the GDPR. Firstly, transparency provisions in the GDPR range from **overarching to concrete**. Transparency as an overarching principle informs the interpretation and application of all of the GDPR.[84] Indeed, it is listed in the first substantive provision in the GDPR, requiring any data processing operation to be lawful, fair and transparent.[85] Throughout the GDPR, more specific, concrete rights and obligations formalize how transparency should be routinely carried out.[86]

Transparency provisions have both **intrinsic and instrumental aims**.[87] The most explicit transparency provisions have a strong flavor of transparency as intrinsically important: meta-data about processing must be provided to data subjects (and often the public more broadly) upon collection,[88] upon receipt of data from a third party,[89] or upon request.[90] In other provisions, the instrumental component is more prominent, such as concerning establishing a lawful basis for processing or automated decision-making through consent;[91] in data breach notifications to data subjects;[92] in moving data to another controller;[93] and in certification mechanisms.[94]

Transparency provisions can have **different target audiences**: individual data subjects are generally considered to be the intended users of the rights to access or portability;[95] while the public at large, including

72 And which would be specifically designed to 'facilitate decisions on which data can be used, how and by whom for scientific research purposes in a manner compliant with the GDPR.' European Commission, 'A European strategy for data' (n 6) 12–13.
73 See generally Ivan Szekely, 'Freedom of Information Versus Privacy: Friends or Foes?' in Serge Gutwirth and others (eds), *Reinventing Data Protection?* (Springer Netherlands 2009).
74 See generally Stephanie E Perrin, 'The Struggle for WHOIS Privacy: Understanding the Standoff Between ICANN and the World's Data Protection Authorities' (PhD Thesis, University of Toronto 2018).
75 Julia Angwin and others, 'Machine Bias', (*ProPublica*, 23 May 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Jeff Larson and Julia Angwin, 'How We Analyzed the COMPAS Recidivism Algorithm', (*ProPublica*, 23 May 2016) https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed 28 September 2018).
76 However, the EU set-up does provide a defence against recent reported uses of freedom of information law for harassment of e.g. scientists. See e.g., Claudia Polsky, 'Open Records, Shuttered Labs: Ending Political Harassment of Public University Researchers' (2019) 66 *UCLA L Rev.*
77 Jef Ausloos and Pierre Dewitte, 'Shattering One-Way Mirrors – Data Subject Access Rights in Practice' (2018) 8 *International Data Privacy Law* 4, 5–7.
78 cf. Damian Clifford and Jef Ausloos, 'Data Protection and the Role of Fairness' (2018) 37 *Yearbook of European Law* 130.
79 Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (opened for signature 10 October 2018) 228 CETS (hereafter Convention 108+), art 9(c).
80 See e.g., Council of Europe, 'Recommendation on the Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling CM/Rec(2010)13' (23 November 2010).
81 Convention 108+, art 9(c).
82 Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 *IEEE Security & Privacy* 46.
83 Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) 35 *Computer Law & Security Review*.
84 GDPR, art 5(1)(a). See generally Clifford and Ausloos (n 78).
85 GDPR, art 5(1)(a).
86 GDPR, arts 13–15.
87 See generally Ausloos and Dewitte (n 77).
88 GDPR, art 13.
89 GDPR, art 14.
90 GDPR, art 15(2–3).
91 In general for consent, GDPR, arts 4(11), 7; for automated decision-making, See GDPR, art 22(2)(c) and recital 71.
92 GDPR, art 33.
93 GDPR, art 20.
94 GDPR, art 42.
95 GDPR, arts 15, 20. But See René LP Mahieu and others, 'Collectively Exercising the Right of Access: Individual Effort, Societal Effect' (2018) 7

civil society watchdogs, often benefit through transparency obligations (often fulfilled through privacy policies or public signage).[96] Supervisory authorities are important beneficiaries of transparency, which they can obtain through a range of data controller obligations[97] as well as their own information retrieval powers.[98] Transparency provisions can also treat sensitivity with nuance and blend target audiences in doing so. For example, in the case of sensitive data in the policing context which cannot be directly released, the data subject has a right to exercise transparency provisions *through* a supervisory authority, who must verify the legality of the processing illuminated by the data they receive.[99]

Transparency provisions can **kick in either before or after data is first processed**, a topic which we will return to further below **(*ex ante* and *ex post* transparency)**. A final, related distinction distinguishes **push and pull** transparency provisions, differentiating whether the controller[100] or the target audience[101] must take the initiative before information is released. This distinction largely corresponds to transparency *obligations* versus transparency *rights*.

While these ways of categorizing GDPR transparency overlap, they help better situate the twofold goal of transparency measures in the GDPR. Transparency provisions have a protective dimension, ensuring demonstrable accountability. Yet some measures also bring an important empowerment dimension, putting control in the hands of different stakeholders, and data subjects in particular, to be more informed. Both dimensions can be considered to contribute to a common goal: redistributing power stemming from information/data asymmetries.

3.2 *Ex ante* transparency

The epicenter of transparency measures in the GDPR, as well as the most well-known and explicit, is found within Articles 13–15. The first two of these list the information that controllers—those determining the means and purposes of data processing—need to provide proactively, at their own initiative and *before* they start processing personal data.[102] In substance, Article 13 (focused on situations where personal data was obtained from individuals *directly*) and Article 14 (personal data was obtained *indirectly*) differ very little. These provisions can first and foremost be qualified as *protective* measures, forcing controllers to give proper thought to, and be upfront about, their processing operations and enabling to hold them to account later on. As such, they also serve as a useful compliance–testing tool for data protection authorities and/or other interest-groups.

Articles 13–14 also have an empowering facet to them. After all, they make data subjects — those to whom the personal data being processed relates — aware of processing taking place and as such can be seen as a *sine qua non* for empowering individuals to invoke one or more of their rights (e.g. object, erasure, portability).[103] The most important components of *ex ante* transparency relate to the scope, purposes and the lawful bases for processing, the risks involved, the retention period and how to exercise data subject rights.

3.3 *Ex post transparency*

There are two main sources of *ex post* transparency in the GDPR that can be triggered by data subjects — the right of access, commonly known as the data subject access right and the right to data portability.

3.3.1 Subject access rights

Article 15 complements *ex ante* information obligations by granting data subjects an explicit, *user-triggered right* to obtain additional information (cf. Table 1, page 15). There are two main components to this right. The first largely replicates the information that was, or should have been, provided under Articles 13–14, which is useful when the information was missed at the time or spread across multiple sources, incomplete, or not specific to the data subject's situation. In this regard, Article 15 can be qualified as an *ex post* empowerment measure and essentially gives individuals the ability to force more timely and specific transparency.[104]

The second component is more radical, at least compared to regimes that in general lack it. It demands that data controllers 'shall provide a copy of the personal data undergoing processing',[105] which explains why the right has become known as a subject access request (SAR). It is worth noting that 'processing' is an extremely broad term, meaning 'any operation or set of operations' performed on personal data.[106] Consequently, data undergoing processing is not just data actively being used, but also includes data that is being stored. Furthermore, the wide scope of personal data[107] means that opinions or comments, including those undertaken computationally or those which may be incorrect, are, *prima facie*, often going to be within the remit of the right of access.[108]

Table 2 Information Requirements under Article 15, GDPR.

Information Requirement	Art 15
Confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the **personal data**	1
Purposes of the processing	1(a)
Categories of personal data concerned	1(b)
Recipients or categories of recipients to whom the personal data **have been or will be disclosed**, in particular recipients in third countries or international organisations	1(c)
Retention period, or if that is not possible, the criteria used to determine that period	1(d)
Existence of the data subject rights to rectification, erasure, restriction of processing, and to object	1(e)
Right to lodge a complaint with a supervisory authority	1(f)
Where personal data are not collected from the data subject, any information on the **source**	1(g)

Internet Policy Review.
96 GDPR, arts 13–14.
97 e.g., GDPR, art 30(4).
98 GDPR, art 47(1).
99 Law Enforcement Directive, art 17.
100 e.g., GDPR, art 13.
101 e.g., GDPR, art 15.
102 Processing in data protection law includes collection. GDPR, art 4(2).
103 See further Ausloos and Dewitte (n 77).
104 Ausloos and Dewitte (n 77).
105 GDPR, art 15(3).
106 GDPR, art 4(2).
107 See generally Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' (2018) 10 *Law, Innovation and Technology* 40. For a view tempering the wide scope argued in that paper, See Lorenzo Dalla Corte, 'Scoping Personal Data: Towards a Nuanced Interpretation of the Material Scope of EU Data Protection Law' (2019) 10 *European Journal of Law and Technology*.
108 Case C-434/16 *Peter Nowak v Data Protection Commissioner* ECLI:EU:C:2017:994 [34].

Existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful **information about the logic involved, as well as the significance** and the envisaged consequences of such processing for the data subject	1(h)
In case of transfer to third country, information about the appropriate safeguards	2

3.3.2 Data portability

The new right to data portability offers some further promise for use in order to obtain research data. Article 20 grants data subjects the right to receive their personal data, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance.[109] Moreover, data subjects can request their personal data to be directly transferred from one controller to another where technically feasible. It is not hard to see how this provision may make the process of sharing personal data with researchers a lot smoother.[110] Indeed, in contrast to the right of access, the right to data portability actively recognizes the value and ability for data subjects to move their personal data between entities, and thus has provisions and wording that facilitate such sharing.[111] The version of the Digital Markets Act proposed by the Commission, if passed, would further strengthen data portability rights against large 'gatekeepers' by enabling them to be used continuously and in real-time.[112]

Unlike the right of access in Article 15(3), which applies to all data being processed, three important constraints limit the scope of the right to data portability:

1. It only applies to personal data that the data subject has provided to the controller, excluding for example 'inferred' and 'derived' data.[113]
2. It only applies where processing is based on 'consent' or 'necessity for the performance of a contract' as a lawful ground. This effectively exempts data processed only with one or a mixture of the four other grounds.[114] Crucially, this includes the important 'legitimate interests' ground, upon which data is gathered on an 'opt-out' or objection basis, rather than an affirmative consent basis.
3. Although not particularly restrictive for our purposes, the right to data portability only applies in situations where the respective personal data is processed 'by automated means'. Data protection also applies to physical records that meet the definition of personal data and 'which form part of a filing system or are intended to form part of a filing system'.[115] Data controllers have no obligation to digitize such data in a machine-readable format for the purposes of the right to portability, although such data remains within scope of the right of access.

3.3.3 Transparency modalities

The GDPR also lists a number of modalities to ensure transparency is effective. The key provision for this is Article 12, but some specific modalities can also be found within the respective provisions discussed above. Importantly, individuals cannot be charged a fee for claiming transparency[116] and there are strict timing requirements as well as broader conditions for the way in which transparency is provided.[117] The European Data Protection Board (EDPB)[118] has further specified that controllers should actively consider the audience's 'likely level of understanding' when accommodating transparency (e.g. appropriate level of detail, prioritizing information, format, etc.).[119] This means the controller will need to consider the context of data processing, the product/service experience, device used, nature of interactions, and so on.[120] As a result, the information obligation may also differ throughout time.[121]

Finally, it is worth keeping in mind that controllers have a duty to facilitate the exercise of data subject rights by 'implementing appropriate technical and organizational measures'[122] and only work with processors who can guarantee doing the same.[123] While the GDPR seems to imagine standard-setting and/or APIs, collaborations in complex ecosystems that facilitate data subjects' rights remain easier

109 The format should be interoperable and machine-readable, both notions being defined in EU law, cited in: Article 29 Working Party, 'Guidelines on the Right to Data Portability' (wp242, 13 December 2016)16–18. It is further specified that '[w]here no formats are in common use [...], data controllers should provide personal data using commonly used open formats (e.g.XML, JSON, CSV,...) along with useful metadata at the best possible level of granularity'.

110 Even the European Data Protection Board (previously known as Article 29 Working Party) explained how the right might be useful to learn more about music consumption by using the right with streaming services or assessing carbon footprint by using the right with loyalty cards. Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 4–5.

111 That being said, the Commission did recently state that 'as a result of its design to enable switching of service providers rather than enabling data reuse in digital ecosystems the right [to data portability] has practical limitations.' European Commission, 'A European strategy for data' (n 6) 10.

112 Proposed Digital Markets Act (n 56), art 6(1)(h).

113 The EDPB does however advocate for a broad interpretation, encompassing both 'data actively and knowingly provided by the data subject' as well as 'observed data provided by the data subject by virtue of the use of the service or the device'. Data such as search histories, browsing/location behaviour, 'raw data' collected through 'mhealth devices' (mobile health) therefore fall within the scope of the right to data portability. Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 9–11.

114 GDPR, art 6(1) lists six lawful grounds on the basis of which personal data may be processed: (a) consent; (b) necessary for the performance of a contract; (c) necessary for compliance with a legal obligation; (d) necessary to protect the data subject or another natural person's vital interests; (e) necessary for tasks carried out in the public interest, or exercise of official authority; (f) necessary for the purposes of the legitimate interests pursued by the controller or third parties, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject.

115 GDPR, art 2(1).

116 This also means that a controller cannot require you to be a paying customer as a condition to accommodate your rights. Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' (11 April 2018) 13. Previous empirical work has demonstrated that certain controllers effectively only enable access requests filed by people who have an account with the service and/or have bought something with the service before. See Ausloos and Dewitte (n 77) 12–13.

117 See generally Jef Ausloos and others, 'Getting Data Subject Rights Right: A Submission to the European Data Protection Board from International Data Rights Academics, to Inform Regulatory Guidance' (2020) 10 *JIPITEC*.

118 Prior to the entry into force of the GDPR, this organisation – which groups together all Member State data protection authorities – was known as the Article 29 Working Party.

119 See also Recital 60 Article 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (n 116) 11.

120 This may require running (and documenting) trials before 'going live'. See Article 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (n 116) 14.

121 cf. Article 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (n 116) 16–17.

122 GDPR, arts 12(2), 25. For a more detailed explanation on data rights modalities, See Ausloos and others (n 117).

123 GDPR, art 28.

said than done.[124]

4 Opportunities for Researching Through Data Rights

How can data rights help researchers? This will effectively depend on a variety of disciplinary, practical, legal, ethical and methodological factors. Indeed, it all starts with a research question or goal, which is to be situated in a certain (number of) discipline(s) that comes with its (their) own im-/explicit rules on valid data gathering. Next, one will need to assess what exact data is needed and what GDPR transparency measure may appropriate to capture it (cf. section 3). Researchers will also need to consider the scope of the data required, both in width (i.e. how many research subjects, if any at all, are needed to have a representative sample) and in depth (i.e. how exhaustive and/or granular does the data have to be). This scope will, in turn, inform whether research subjects are needed, and if so, how to recruit them. Researchers will also need to carefully consider an interaction strategy with data controllers (including contingency plans), which may be more or less burdensome depending on the scope, but also on the identity of the data controller.[125] Indeed, based on preliminary research (including filing access requests themselves), researchers may prepare a manual or script on how research subjects should obtain the required information and interact with data controllers.[126] Finally, researchers should also anticipate how the data they might obtain through data rights will actually be analyzed in light of the research aim. Summarized, the following seven steps may serve as a useful starting point for researchers interested in using data rights in their project:

1. Aim. What is your research goal? What purpose are you gathering data for?
2. Data. What specific data do you need to achieve said purpose?
3. Legal Approach. What GDPR transparency measure is appropriate for obtaining said data (if any)?
4. Scope. What does your (ideal) research sample look like?
5. Recruitment Strategy. Based on the scope, how to identify and recruit research participants accordingly?
6. Interaction Strategy. How will you interact with your participants and the respective data controllers?
7. Data Analysis Strategy. How will you actually gather the insights you need?

These steps remain necessarily vague, in light of the broad potential of data rights as a research method in many different disciplines. To tie it back to the many variables determining the actual usefulness of data rights for any given research project – i.e. disciplinary, practical, legal, ethical and methodological factors – the abstract workflow mentioned above will have to be given shape depending on the respective discipline(s) and research questions. There are also many *practical* factors that might influence the usefulness of data rights. Again, these will depend very much on the concrete circumstances of a given research project. Nonetheless, in order to make things more concrete, and invite readers to contemplate different use cases, this section lays out some illustrative potential and promising uses of data rights. The following section will then dig into some of the legal, ethical and methodological considerations.

For our purposes here, we identify three main categories of research (goals) as being enabled by data rights (in order of specificity):

- studying infrastructures (research into the actual infrastructures to which the respective data relates);
- studying impacts (research into how data infrastructures affect individuals, communities or society at large); and
- repurposing digital traces (research into broader questions that might be far from issues of digital rights).

4.1 Understanding infrastructures

Researchers can use data transparency rights to study digital infrastructures and practices in today's economies and society.

Studies examining data protection law in practice are one example of this. Researchers have, for example, studied the privacy policies of cloud service providers to identify common industry approaches and legal mismatches.[127] These privacy policies exist in the form they do in large part due to the GDPR's transparency provisions in Articles 13–14.[128] Other research has taken the form of exploring *how* rights are responded to by controllers, the quality of which might say something about enforcement more generally.[129]

Yet *ex post* transparency measures offer wider potential as research tools beyond studying the way the law is being interpreted and adhered to. Many use cases can be envisaged which would use specific *ex post* transparency measures to uncover substantive issues. We consider a number of them below.

4.1.1 Tracking

The state of online tracking and advertising today has been both lauded for supporting online services that do not directly cost consumers money, as well as lambasted for undermining democracy, journalism and a range of fundamental rights. One thing is certain: it is a challenging area to study. Data rights provide a useful set of tools to shine further light on issues of concern.

For example, both users and researchers know little about how effective privacy protective browsers and extensions really are. While it is relatively simple to secure a device from explicitly saving tracking cookies (although that may damage Web functionality), it is very hard to disguise the unique fingerprint of a browser, particularly in the presence of advanced fingerprinting tactics utilized in modern advertising technologies.[130] Because fingerprinting does not always query

124 cf. Chris Norval and others, 'Reclaiming Data: Overcoming App Identification Barriers for Exercising Data Protection Rights' [2018] arXiv:180905369 [cs], 4.

125 As research has shown, many data controllers are often unwilling to comply in full with data access requests, unless they are repeatedly contacted. See, e.g. Ausloos and Dewitte (n 77); Jef Ausloos, 'Paul-Olivier Dehaye and the Raiders of the Lost Data' (*CITIP blog*, 10 April 2018) https://www.law.kuleuven.be/citip/blog/paul-olivier-dehaye-and-the-raiders-of-the-lost-data (accessed 23 April 2018); Mahieu and others (n 95).

126 For example, one could envisage a website or an app that makes it easier for research subjects to file access requests, follow up on them, and/or filter the personal data obtained, before it is sent to the researchers. See also section 5.3.2

127 Dimitra Kamarinou and others, 'Cloud Privacy: An Empirical Study of 20 Cloud Providers' Terms and Privacy Policies—Part I' (2016) 6 *International Data Privacy Law* 23; Jamila Venturini and others, *Terms of Service and Human Rights: An Analysis of Online Platform Contracts* (Revan 2016).

128 See section 3.2, '*Ex ante* transparency'.

129 See e.g., Ausloos and Dewitte (n 77); Mahieu and others (n 95); Janis Wong and Tristan Henderson, 'How Portable is Portable?: Exercising the GDPR's Right to Data Portability' in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (UbiComp '18, New York, NY, USA, ACM 2018); Clive Norris and others (eds), *The Unaccountable State of Surveillance: Exercising Access Rights in Europe* (Springer 2016).

130 Nick Nikiforakis and others, 'Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting' (May 2013) *2013 IEEE Symposium on Security and Privacy* 541.

a user's device directly, but observes it passively, it is unclear to what extent real protection is provided. Technologies such as re-spawning cookies, 'evercookies',[131] font and battery level fingerprinting all present methodological challenges to detect, understand and effectively and provably block.[132] The use of data rights to ascertain the data a firm actually holds on users through a separate channel could be used as means of assessing the efficacy or tracking prevention, or understanding the true nature and purpose of certain tracking practices online.[133]

The number of actors in the tracking business and the nature of their interactions with each other also considerably restricts understanding. Online advertising increasingly functions through a complex 'real-time bidding' system whereby an individual's browser, generally unbeknownst to the user, sends out personal data about them to an advertising exchange, which in turn forwards it to thousands of potential bidders. These thousands of bidders utilize the services of *data management platforms* to enrich the data received: to effectively see if your eyes are worth bidding for in relation to the adverts they are attempting to place. The UK and Belgian regulators have noted that such a system is likely not legally compliant on a number of fronts.[134] Detailed evidence is, however, scarce, due to the secrecy and complexity of these practices.

The fact that these actors often share data *server-to-server* has created a blind spot for current studies—a blind spot that data rights might help remedy. While it is possible for researchers to monitor a user's browser to observe the destination of the traffic, for example by using a VPN (local or remote) with the consent of the user,[135] data that is transmitted around the user from server-to-server cannot be observed. Researchers working in this space have to come to an unhappy compromise of either simulating these server-to-server transmissions with almost no evidence on how they actually occur in practice,[136] or to try and guess at data practices by experimenting on how users are differentially targeted further downstream.[137] If researchers were to use data rights—and, if these firms were forced to answer them truthfully and fully—information on the data, the source of the data, and potentially on the recipients could be obtained, which would help both modelling assumptions as well as enable further research questions to be answered.

Related to this, data rights may also benefit studying the intersection of user preferences and tracking infrastructures. Some researchers have been presenting users with information about tracking activities (e.g. types of data, data flows), attempting to understand the effects on decision-making by users, as well as any impacts on their ongoing formation of privacy and data control preferences.[138] To do this, they have relied on indirect methods to understand these flows, such as running an app in a virtual environment and monitoring and classifying the entities data directly flow to.[139] Yet this data is still a step removed from the tracking that has occurred to particular participants. To reflect on their own information, users would typically have to rely on tools to collect and reflect on this data,[140] such as local logging of information. Tools to give users a 'history' function on their digital activities do exist,[141] but are unwieldy to force participants to use day-to-day, and may not even log as invasively as third-party trackers currently do.[142] Insofar as these tracking infrastructures *already exist*, data rights provide an alternate means to get access to them, enabling research that takes advantage of users seeing and reflecting on tracking data that truly was captured about and relates to them.

Data rights might help economic studies too. Despite considerable interest in how online content should be funded, 'the conventional wisdom that publishers benefit too from behaviorally targeted advertising has rarely been scrutinized in academic studies'. Recent studies have indicated that when a user's cookie is available, publishers' revenue increases by only about 4%.[143] This adds to anecdotal evidence from publishers such as the New York Times that reducing tracking has increased profits in their European markets, suspected to be related to the market structure of advertising technology and the proliferation of intermediaries.[144] Data rights might help to gather datasets on which publishers, advertising technology firms, ad exchanges and other actors[145] are active in this area, and use that data to create and validate economic models which can shine light on market functioning.

In a similar vein, there has been considerable recent interest in

131 Evercookies use practices found in malware more broadly to re-establish cookies even when users or browsers attempt to purge them.

132 Gunes Acar and others, 'The Web Never Forgets: Persistent Tracking Mechanisms in the Wild' in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (CCS '14, New York, NY, USA, ACM 2014); David Fifield and Serge Egelman, 'Fingerprinting Web Users Through Font Metrics' in Rainer Böhme and Tatsuaki Okamoto eds, *Financial Cryptography and Data Security* (Lecture Notes in Computer Science, Springer Berlin Heidelberg 2015); Łukasz Olejnik and others, 'The Leaking Battery' in Joaquin Garcia-Alfaro and others eds, *Data Privacy Management, and Security Assurance* (Lecture Notes in Computer Science, Springer International Publishing 2016).

133 They would not be without their challenges of course: some tracking practices may be illegal, for example, meaning that already–infringing data controllers are unlikely to readily to openly share information.

134 Information Commissioner's Office, 'Update Report into Adtech and Real Time Bidding' (Information Commissioner's Office, 20 June 2019) https://perma.cc/X7PX-EL3L (accessed 20 June 2019); Natasha Lomas, 'IAB Europe's Ad Tracking Consent Framework Found to Fail GDPR Standard' (*TechCrunch*, 16 October 2020) https://social.techcrunch.com/2020/10/16/iab-europes-ad-tracking-consent-framework-found-to-fail-gdpr-standard (accessed 16 October 2020).

135 See e.g., Razaghpanah and others (n 46); Ren and others (n 46); Song and Hengartner (n 46); Shuba and others (n 46).

136 See e.g., Muhammad Ahmad Bashir and Christo Wilson, 'Diffusion of User Tracking Data in the Online Advertising Ecosystem' (2018) *2018 Proceedings on Privacy Enhancing Technologies* 85.

137 Tschantz and Datta (n 12).

138 See e.g., Max Van Kleek and others, 'X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps' in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18, New York, NY, USA, ACM 2018); Max Van Kleek and others, 'Better the Devil You Know: Exposing the Data Sharing Practices of Smartphone Apps' in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17, New York, NY, USA, ACM 2017).

139 See generally Binns and others (n 13).

140 See generally M Janic and others, 'Transparency Enhancing Tools (TETs): An Overview' (June 2013) 2013 *Third Workshop on Socio-Technical Aspects in Security and Trust* 18; P Murmann and S Fischer-Hübner, 'Tools for Achieving Usable Ex Post Transparency: A Survey' (2017) 5 *IEEE Access* 22965.

141 See e.g., Jennifer Pybus and others, 'Hacking the Social Life of Big Data' (2015) 2 *Big Data & Society* 2053951715616649.

142 Murmann and Fischer-Hübner (n 140) 22988.

143 Veronica Marotta and others, 'Online Tracking and Publishers' Revenues: An Empirical Analysis' (June 2019) *Proceedings of the Workshop on the Economics of Information Security* (WEIS 2019), 2–4 June, Boston, MA.

144 Jessica Davies, 'After GDPR, The New York Times Cut off Ad Exchanges in Europe - and Kept Growing Ad Revenue' (*Digiday*, 16 January 2019) https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue (accessed 19 June 2019). See also David Beer and others, Landscape Summary: Online Targeting (Centre for Data Ethics and Innovation, HM Government 2019) 32.

145 See generally Bashir and Wilson (n 136).

competition issues around online tracking from both academics[146] and policy-makers.[147] Insofar as data rights can help with issues of accountability and provenance,[148] they may help to map the space of actors and data practices in ways which better shine light on structural power relations that matter for evidencing competition policy interventions in different jurisdictions.

4.1.2 Content moderation

For well over a decade, researchers have been investigating the freedom of expression and information implications of online copyright enforcement.[149] Considerable efforts have been put into forcing more transparency and accountability from both copyright-holders as well as the (user-generated) content platforms in taking down content.[150] More recently, growing concerns over platform power and regulatory initiatives on online content moderation have breathed new life into this work.[151] Indeed, a lot of important questions have been raised in relation to content moderation and platforms' potential political biases,[152] their role in facilitating cyber-bullying,[153] impact on inclusiveness and participation by vulnerable or minority groups,[154] and the increased privatization of the public sphere more broadly.

In mapping the available empirical literature on these issues, Keller and Leerssen make a similar distinction to those we made above separating disclosures from platforms and other direct stakeholders from independent research through, for example, APIs, secondary processing of released or scraped data, or surveys with users and other stakeholders.[155]

These methods can be lacking in depth to explore exactly the processes, data sources and reasoning automated takedowns involve. Subject access requests may be a valuable addition in researchers' toolbox, providing a legally enforceable mechanism to force platforms to be more open about their decision-processes that have affected the data subject(s) at stake. One reason for this is that any decision on (not) taking down content may significantly affect either the uploader,[156] or person(s) featuring in the actual content. In those situations, Article 15(1)h provides data subjects the right to obtain *meaningful information about the logic involved, as well as the significance and the envisaged consequences* of the respective decision(s).[157] In general, data about the uploaders' actions or account may also be considered personal data and subject to Article 15 (or 20) more broadly. Such personal data can in turn be examined for its sources, gaining a better understanding of the processing activities underlying content moderation today.

That being said, as with any of these cases, data rights are no panacea. While enabling deeper insights into certain content moderation practices, using subject access requests for mapping platform-wide trends may prove more challenging. They may, however, create new research questions and challenge commonly held assumptions about data processing for these purposes, and form an important part of a researcher's toolkit as a result.

4.2 Understanding impacts

Considerable recent concern has centered around the impact of data-driven systems, particularly in reinforcing structural disadvantage affecting marginalized communities.[158] Such systems create data infrastructures, often focused on optimization, which disregard subsets of individuals (such as those considered 'low value') or contextual and environmental factors,[159] and which may use seemingly non-sensitive data to deliberately or inadvertently make decisions based on legally protected characteristics.[160] They may perform more poorly on certain demographics, such as facial recognition or analysis systems disproportionately misclassifying or misrepresenting Black women.[161] Such systems have also been accused of using micro-targeting in an electoral context in ways unsuited for demo-

146 See e.g., Elettra Bietti and Reuben Binns, 'Acquisitions in the Third Party Tracking Industry: Competition and Data Protection Aspects' [2019] *Computer Law & Security Review*; Reuben Binns and others, 'Measuring Third-Party Tracker Power Across Web and Mobile' (2018) 18 *ACM Trans Internet Technol* 52:1.

147 See e.g., Jacques Crémer and others, 'Competition Policy for the Digital Era' (European Commission, 2019) http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf (accessed 4 April 2019).

148 See generally David Eyers and others, 'Towards Accountable Systems' (Dagstuhl Seminar 18181) [2018].

149 For a comprehensive overview, See Aleksandra Kuczerawy, *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Safeguards* (Intersentia 2018); Daphne Keller and Paddy Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation' in N Persily and J Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (CUP 2019).

150 Most notably perhaps, the early work of Wendy Seltzer and in particular the Lumen database (formerly 'Chilling Effects Clearinghouse') https://lumendatabase.org, collecting and analysing removal requests of online materials.

151 See generally Robert Gorwa, 'What is Platform Governance?' (2019) 22 *Information, Communication & Society* 854.

152 Oscar Schwartz, 'Are Google and Facebook Really Suppressing Conservative Politics?', (*The Guardian*, 4 December 2018) https://www.theguardian.com/technology/2018/dec/04/google-facebook-anti-conservative-bias-claims (accessed 1 December 2019).

153 Tijana Milosevic, 'Social Media Companies' Cyberbullying Policies' [2016] 22; Pat Strickland and Jack Dent, *Online harassment and cyber bullying* House of Commons Rep 07967 (UK House of Commons 2017).

154 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018); Stefanie Duguay and others, 'Queer Women's Experiences of Patchwork Platform Governance on Tinder, Instagram, and Vine' [2018] Convergence 1354856518781530; Jillian C. York and Karen Gullo, 'Offline/Online Project Highlights How the Oppression Marginalized Communities Face in the Real World Follows Them Online' (*Electronic Frontier Foundation*, 3 June 2018) https://www.eff.org/deeplinks/2018/03/offlineonline-project-highlights-how-oppression-marginalized-communities-face-real (accessed 19 July 2019).

155 Keller and Leerssen (n 149) 13–32.

156 Relatedly, it is worth referring to FairTube, an initiative set up by (semi-) professional youtubers aimed at forcing fairer and more transparent decision-making on de-monetization of YouTube content. Subject access rights played a role in this effort. René Mahieu and Jef Ausloos, 'Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access' (Preprint, 2 July 2020) 29, DOI:10.31228/osf.io/b5dwm.

157 Some scholars have argued that, according to grammar found in the recitals, information will not relate to specific decisions, e.g., Sandra Wachter and others, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 76. Others have instead examined the GDPR in light of its overarching principles, arguing that specific information may, under some circumstances, be provided Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 *International Data Privacy Law* 233. No case law has definitively determined one way or another.

158 See generally Oscar H Gandy, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (Routledge 2009); Tawana Petty and others, 'Our Data Bodies: Reclaiming Our Data' (*Our Data Bodies Project*, June 2018); Seeta Peña Gangadharan and J drzej Niklas, 'Decentering Technology in Discourse on Discrimination' (2019) 22 *Information, Communication & Society* 882.

159 Rebekah Overdorf and others, 'POTs: Protective Optimization Technologies' [2018] arXiv:180602711 [cs].

160 Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 *Calif L Rev* 671.

161 Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in *Conference on Fairness, Accountability and Transparency* (FAT* 2018) (2018).

cratic society,[162] as well as manipulating individuals more generally and pervasively.[163] Concerns exist that individuals lose their ability to reflect on morally challenging tasks from pervasive use of affective (emotional) predictive systems in their ambient environments.[164] Policy-makers are also concerned about 'addiction' to devices, 'dark patterns' attempting to foster profitable but undesirable habits,[165] underpinned by systems designed to predict individuals who might be easily swayed into, for example, spending more on an app.[166]

4.2.1 Discriminatory decision-systems

The transparency provisions around machine learning in the GDPR,[167] such as Article 15(1)(h) (see Table 1) as well as access rights more generally, might be directly and indirectly useful in achieving transparency over complex, automated systems.[168] While the utility of individualized transparency has been questioned,[169] data rights could play a role in creating aggregate, societal-level transparency and accountability. Data from access rights might be used to seek inferences, data and meta-data about prediction and training data which can inform researchers around how systems function. Algorithmic 'explanations', where provided, might be compiled to shine light on the functioning of a model,[170] or compared across individuals, demographics or applications. Data rights could also help understand where models come from, which actors were involved in training and building them, and when, which is particularly salient given the rise in business models involving the trading of trained machine learning models.[171]

One example of an attempt to do just this with data protection rights can be found in the German credit scoring context. OpenSCHUFA was a campaign in Germany run by AlgorithmWatch and the Open Knowledge Foundation Deutschland attempting to reverse-engineer the main system used to determine creditworthiness of German residents. It built a data donation platform that was used by over 4,000 people to collate SCHUFA access information on the basis of data rights, in particular, asking for copies of data under the right to access that could later be analyzed. While such a campaign was a logistical success, and placed pressure on the SCHUFA, it also revealed an array of challenges in using data rights in this way, such as sampling bias, which we discuss later below.[172]

If sampling challenges can be overcome, the information that can be gathered through data rights of diverse populations might shine a light on some discrimination concerns. Studies that have tried to understand discrimination in job adverts, for example, have relied on different methods. The challenges of one of these, web scraping, have already been described.[173] Add to these the challenges of creating a credible 'data exhaust' which can be mistaken as that of a real person — a challenge which flummoxes even the intelligence services[174] — and it becomes clear that the bot approach might fast drift from the lived experience of individuals online. Others have relied on self-reported performance data from the platform itself:[175] whether such data can be trusted when there are strong incentives to make adverts look well-performing and non-discriminatory are unclear.

4.2.2 Recommenders and media exposure

A considerable deal of concern has centered around the creation of digital 'echo chambers' or 'filter bubbles' in relation to content viewed online.[176] There is limited empirical evidence to support their existence in many cases, particularly within traditional news source.[177] but the field is still poorly understood, particularly in the context of platforms working to enclose content within walled gardens.[178] Indeed, empirical work on the power of media recommender algorithms in radicalizing viewers would greatly benefit from more granular insights that access rights enable.[179] Where data about content shown, clicked on and/or viewed is stored or retained, it might prove useful for independent analysis and comparison to understand the extent of this tracking.[180]

4.3 Repurposing digital traces

Data rights can also provide data for other scientific and humanistic questions. The sensing infrastructure provided by mobile phones or 'smart' home devices have already been considered for 'citizen science' or 'community science' and 'participatory sensing'. However, these applications have typically focused on environmental factors, such as air, noise and water pollution,[181] and rely on the user send-

162 Information Commissioner's Office, Democracy Disrupted? Personal Information and Political Influence (ICO 2018).
163 Karen Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 *Information, Communication & Society* 118.
164 Sylvie Delacroix and Michael Veale, 'Smart Technologies and Our Sense of Self: Going Beyond Epistemic Counter-Profiling' in Mireille Hildebrandt and Kieron O'Hara (eds), *Life and the Law in the Era of Data-Driven Agency* (Edward Elgar 2020).
165 Forbruker Rådet, 'Deceived by Design' (27 June 2018).
166 Ronan Fahy and others, 'Data Privacy, Transparency and the Data-Driven Transformation of Games to Services' (August 2018) *2018 IEEE Games, Entertainment, Media Conference* (GEM) 1; Digital, Culture, Media and Sport Committee, 'Immersive and Addictive Technologies' (House of Commons, HC 1846, 12 September 2019).
167 See generally Edwards and Veale (n 8).
168 Edwards and Veale (n 8).
169 Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 *New Media & Society* 973.
170 Some work has recently shown that model reconstruction attacks can be heightened by the use of model explanations. e.g., Smitha Milli and others, 'Model Reconstruction from Model Explanations' in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT* '19, New York, NY, USA, ACM 2019). Work is ongoing to understand what explanations can be used in relation to models, See further Martin Strobel, 'Aspects of Transparency in Machine Learning' in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (AAMAS '19, Richland, SC, International Foundation for Autonomous Agents and Multiagent Systems 2019).
171 Michael Veale and others, 'Algorithms that Remember: Model Inversion Attacks and Data Protection Law' (2018) 376 *Phil Trans R Soc A* 20180083.
172 See infra section 5.2.
173 See supra section 2.3.
174 Sam Jones, 'The Spy Who Liked Me: Britain's Changing Secret Service', (*Financial Times,* 29 September 2016) https://www.ft.com/content/b239dc22-855c-11e6-a29c-6e7d9515ad15 (accessed 29 April 2019).
175 Anja Lambrecht and Catherine Tucker, 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' (2019) 65 *Management Science* 2966; Ali and others (n 22).
176 Frederik J Zuiderveen Borgesius and others, 'Should We Worry about Filter Bubbles?' (2016) 5 *Internet Policy Review*.
177 Zuiderveen Borgesius and others (n 176); Judith Möller and others, 'Do Not Blame It on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity' [2018] *Information, Communication & Society* 1; Mario Haim and others, 'Burst of the Filter Bubble?' (2018) 6 *Digital Journalism* 330.
178 See generally Angela M Lee and Hsiang Iris Chyi, 'The Rise of Online News Aggregators: Consumption and Competition' (2015) 17 *International Journal on Media Management* 3; Paddy Leerssen, 'The Soap Box is a Black Box: Regulating Transparency in Social Media Recommender Systems' (2020) 11 *EJLT*.
179 cf. Kevin Munger and Joseph Phillips, 'A Supply and Demand Framework for YouTube Politics' [2019] 38; Rebecca Lewis, 'Alternative Influence: Broadcasting the Reactionary Right on YouTube' (18 September 2018).
180 See in this regard, Leerssen (n 178) 2.
181 Stacey Kuznetsov and Eric Paulos, 'Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes' in *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (DIS '10, New York, NY, USA, ACM 2010); Prabal Dutta and others, 'Common Sense:

ing data directly rather than repurposing data collected for another purpose. Data rights could widen the scope of citizen/community science — we highlight some potential directions below.

4.3.1 Location data

Location data is one of the richest forms of data, and the rise in location aware applications has long attracted privacy concerns.[182] Because mobile phones connect so regularly to base stations, they leave a long trace of location. As users increasingly rarely turn phones off,[183] such location traces effectively extend to all times when the phone is in contact with telecoms infrastructure. As a result, telecoms data has been used by national statistical agencies and humanitarian groups alike—at times attracting considerable ethical controversy.[184] Mobile phone location data might, for example, be used to infer the type of transport someone is using,[185] socioeconomic information about them,[186] or places that they consider important,[187] among many other potential applications. But equally, with consent and with proper ethical consideration, it might be that a research subject would be happy to pass over parts of their location history to better understand some intervention or experiment they have been part of.

4.3.2 Biosensors

Commercial devices with self-monitoring sensing capabilities are becoming increasingly popular,[188] and there has been increasing interest in the medical domain in validating these consumer-grade devices to understand if their data collection has the required validity for scientific use.[189] Many devices and software are tracking physical and social characteristics of individuals, from the number of steps taken to the use of houses, vehicles, software, and even clothing. Researchers have highlighted that

> the [device] companies could allow access to more data that are collected. At present, the trackers provide users with only a subset of data that is actually collected. The companies control the output available, making the day-level summary variables the easiest to obtain. For example, despite capturing GPS and heart rate on two trackers, Fitbit currently limits the export of these full datasets. Furthermore, the resulting output is derived through proprietary algorithms that may change over time and with new features. [..] At a minimum, it would be helpful for companies to reveal what pieces of data are being used by the trackers to calculate each output measure.[190]

The role of trade secrets in this area is particularly pertinent. For example, many people use 'smart watches' to measure features of their circulation. Many research fields utilize photoplethysmography data, also known as blood volume pulse. It can be used to measure oxygen saturation, blood pressure and cardiac output, to assess autonomic function and to detect peripheral vascular disease.[191] Smart watches do not measure this directly, however: they infer it from a series of sensed measurements, often using proprietary and changing machine learning systems.[192] For researchers, this (whether in commercial or research grade) products can present challenges, as changing algorithmic systems introduce features which can be difficult to control for. For users, it might not be an issue however: they likely want the most robust and accurate measure of their heartbeat, step-count, sleep patterns or the like over time, and do not care about internal validity over the months and years of device usage.

Depending on the structure of processing, researchers interested in utilizing these sensors may be able to use transparency rights to obtain additional datasets. This might be particularly useful if and when a time comes where users are *already* using high-grade sensors in their daily lives, and research studies would work better by co-opting existing infrastructure rather than adding a further device which is not part of a user's existing routine, or may be redundant to something they already are familiar with.

4.3.3 Labor patterns

Between 1% and 5% of the EU population is estimated to have taken place in some form of paid platform work, with some countries exhibiting significantly higher rates of participation than that.[193] The growth of these markets for informal labor, such as through taxi services provided through Uber or Lyft, or workers on computers using platforms such as Amazon Mechanical Turk, has led to serious concerns, culminating in high profile legal fights, over the employment status of such individuals and the rights they possess. For example, informal work can necessarily bring a considerable amount of overhead, such as sifting through jobs online to find those which are legitimate, and being 'hypervigilant' in order to secure desirable or profitable jobs.[194] In this context, there are important factual questions, with legal ramifications, around the timings and behavior of 'gig economy' workers, such as the amount of time they are active on the app waiting for

Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors' in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, (ACM 11 April 2009) http://dl.acm.org/citation.cfm?id=1644038.1644095 (accessed 18 June 2019); Nicolas Maisonneuve and others, 'NoiseTube: Measuring and Mapping Noise Pollution with Mobile Phones' in Ioannis N Athanasiadis and others eds, *Information Technologies in Environmental Engineering* (Springer Berlin Heidelberg 2009).

182 AR Beresford and F Stajano, 'Location Privacy in Pervasive Computing' (2003) 2 *IEEE Pervasive Computing* 46.

183 UK regulator Ofcom report that 71% of adults claim they never turn their phones off. See Ofcom, 'A Decade of Digital Dependency' (3 May 2019) https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/decade-of-digital-dependency (accessed 24 July 2019).

184 Linnet Taylor, 'No Place to Hide? The Ethics and Analytics of Tracking Mobility Using Mobile Phone Data' (2016) 34 *Environ Plan D* 319; Linnet Taylor and Dennis Broeders, 'In the Name of Development: Power, Profit and the Datafication of the Global South' (2015) 64 *Geoforum* 229.

185 Donald J Patterson and others, 'Inferring High-Level Behavior from Low-Level Sensors' in Anind K Dey and others eds, *Ubiquitous Computing 2003,* UbiComp 2003 (Lecture Notes in Computer Science, Springer Berlin Heidelberg 2003).

186 Christopher Smith-Clarke and others, 'Poverty on the Cheap: Estimating Poverty Maps Using Aggregated Mobile Communication Networks' in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14, New York, NY, USA, ACM 2014).

187 Sibren Isaacman and others, 'Identifying Important Places in People's Lives from Cellular Network Data' in Kent Lyons and others eds, *Pervasive Computing* (Lecture Notes in Computer Science, Springer Berlin Heidelberg 2011).

188 See generally Gina Neff and Dawn Nafus, *Self-Tracking* (MIT Press 2016); Deborah Lupton, 'Self-Tracking Cultures: Towards a Sociology of Personal Informatics' in *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (OzCHI '14, New York, NY, USA, ACM 2014).

189 e.g., Kelly R Evenson and others, 'Systematic Review of the Validity and Reliability of Consumer-Wearable Activity Trackers' (2015) 12 *International Journal of Behavioral Nutrition and Physical Activity* 159.

190 Evenson and others (n 189) 19.

191 John Allen, 'Photoplethysmography and Its Application in Clinical Physiological Measurement' (2007) 28 *Physiol Meas R1*.

192 See e.g., Empathica, 'Utilizing the PPG/BVP Signal' (*Empatica Support*, 31 March 2016) http://support.empatica.com/hc/en-us/articles/204954639-Utilizing-the-PPG-BVP-signal (accessed 19 June 2019).

193 Chris Forde and others, 'The Social Protection of Workers in the Platform Economy' (Study for the European Parliament's EMPL Committee, IP/A/EMPL/2016-11, November 2017) 38.

194 Mary L Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Houghton Mifflin Harcourt 2019).

jobs,[195] which may require data access and analysis in order to assess compensation, fairness, and even potentially an individual's legal status with regards to e.g. holiday pay, breaks workers are entitled to, or other legal rights. Data rights are already central to civil society groups, such as Worker Info Exchange,[196] but if the informal economy continues to increase in density and complexity, more advanced, collective use of digital rights to gather data to understand exploitation, labor patterns, and the changing nature of work may be required.[197]

5 Considerations for Researching with Data Rights

While the opportunities seem promising, the research use of data rights is made difficult by several nuanced limitations. In this section we delineate some of the most important limitations, categorized as legal, methodological and ethical considerations.

5.1 Legal considerations

Even though the right of access is grounded in both the principles of the GDPR and Article 8 of the Charter, there are still legal questions as to its utility in the research context. Some of these issues have clearer answers in guidance and case law than others do. In this section, we group and tackle some of the major issues, misconceptions and open questions around the use of access rights in the contexts discussed earlier.

5.1.1 Motivation of the request

Prima facie, it might appear that a data controller could seek to refuse a request because enabling research was not a stated purpose of the GDPR. Yet case law and regulatory guidance falls behind the view that GDPR rights are *intent-agnostic*. Access rights have commonly been used in relation to highly specific pieces of information, often as part of disputes that might be related to issues of criminal,[198] employment,[199] financial,[200] fiscal,[201] immigration,[202] trust[203] or defamation proceedings.[204] These types of cases can create, in the words of AG Bobek, 'certain intellectual unease as to the reasonable use and function of data protection rules'.[205]

As European data protection has traditionally had a close connection with the right to privacy, one might argue that it is especially aimed at safeguarding the respective individual's interests. If such a view were taken, data protection transparency measures to gather research data might then appear to misuse/retrofit a legal device for unintended purposes, calling its legal enforceability into question. Yet there is an argument to be made that this type of usage is aligned *extremely well* with data protection's primary, historical purpose of regulating data infrastructures underlying society (from large, centralized data mainframes to the complex ecosystem today) rather than (just) supporting individually-focused privacy. The GDPR's legal toolbox that gives some level of control over personal data and/or how it is processed, and the use of these tools is arguably envisaged to be used by a range of stakeholders, including regulators, academics, journalists, artists and civil society organizations, not just by individual data subjects for purely individualistic purposes. As such, the GDPR's transparency measures, as a general tool with many potential uses for promoting oversight and agency, can only be intent agnostic: it is up to these stakeholders to use them flexibly as part of governance, self-determination and oversight.[206]

Indeed, the right of access is an explicit part of the fundamental right to data protection in the Charter, and courts and regulators have held that a 'privacy' motive is not required for its use. In *YS and others* for example, the Court of Justice made no reference to fact that the claimants were seeking to use the right of access in order to support litigation as evidence that their use of rights should fail. National case law has been supportive of this approach too. For example, both English[207] and Dutch[208] courts in recent years have reached a clear consensus that access requests are purpose-blind, and the guidance of the Information Commissioner's Office[209] and Autoriteit Persoonsgegevens[210] is in alignment with this. It is worth noting that some restrictions on motivation of access rights exist at national level to prevent data subjects from being coerced into making them.[211]

Especially insofar as research aims to shed light on the use of personal data in contemporary infrastructures, research uses of data rights seem not just possible within this intent-agnostic regime, but a prime example of an empowerment mechanism working on the side of data subjects.

5.1.2 Infringement of the rights and freedoms of others

Controllers might (partially) fend off access and portability requests when they can establish that accommodating them would 'adversely

195 *Uber BV v Aslam* [2018] EWCA Civ 2748 at [100]; 'Uber drivers demand access to their personal data' (n 14).

196 Farrar (n 14). See further https://workerinfoexchange.org and Mahieu and Ausloos (n 157) 8–10; 29; 'Uber drivers demand access to their personal data' (n 14).

197 The European legislator already took a step in this direction with: Directive (EU) 2019/1152 of the European Parliament and of the Council of 20 June 2019 on transparent and predictable working conditions in the European Union of 2019 EP, CONSIL 32019L1152, EP, CONSIL (European Union EP, CONSIL 2019).

198 *Kololo v Commissioner of Police for the Metropolis* [2015] EWHC 600 (QB). *Lin & Anor v Commissioner of Police for the Metropolis* [2015] EWHC 2484 (QB).

199 *Ittihadieh v 5-11 Cheyne Gardens RTM Company Ltd & Ors* [2017] EWCA Civ 121.

200 *Rechtbank Zwolle-Lelystad 103434 / HA RK 04-215 9 maart 2005*; Parket bij de Hoge Raad 9 Nov 2018.

201 Amélie Lachapelle and Elise Degrave, 'Le Droit d'accès Du Contribuable à Ses Données à Caractère Personnel et La Lutte Contre La Fraude Fiscale, Revue Générale Du Contentieux Fiscal, 2014, 5, p. 322-335' [2014].

202 Joined Cases C-141/12 and C-372/12 *YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S* ECLI:EU:C:2014:2081.

203 *Dawson-Damer & Ors v Taylor Wessing LLP* [2017] EWCA Civ 74.

204 *Rudd v Bridle & Anor* [2019] EWHC 893 (QB).

205 Case C-13/16 *Valsts policijas R gas re iona p rvaldes K rt bas policijas p rvalde v R gas pašvald bas SIA 'R gas satiksme'* ECLI:EU:C:2017:43, Opinion of AG Bobek, para 93.

206 See also: René Mahieu and Jef Ausloos, 'Harnessing the Collective Potential of GDPR Access Rights: Towards an Ecology of Transparency' [2020] *Internet Policy Review*.

207 e.g., *Dawson-Damer & Ors v Taylor Wessing LLP* [2017] EWCA Civ 74 at [104]–[108]; *B v The General Medical Council* [2019] EWCA Civ 1497 at [79] ('the general position is that the rights of subject access to personal data [...] are not dependent on appropriate motivation on the part of the requester') and case law cited therein.

208 *Parket bij de Hoge Raad, 9 November 2018* ECLI:NL:PHR:2018:1273, at para 3.37.

209 See generally Information Commissioner's Office, 'Subject Access Code of Practice' (9 June 2017) 47.

210 Autoriteit Persoonsgegevens, *Recht op inzage* (Netherlands Autoriteit Persoonsgegevens).

211 e.g., Data Protection Act 2018 (United Kingdom) s 184, which, albeit not relevant to researching through data rights, creates offences for employers or providers of contracts to make arrangements conditional on the production of 'relevant records' obtained by use of a SAR. See generally (in relation to the previous regime) Alexander de Gaye and Sabba Mahmood, 'Enforced Subject Access—is It Finally the End?' (2014) 15 *Privacy and Data Protection* 10; Information Commissioner's Office, Enforced Subject Access (Section 56) (ICO 2015).

affect the rights and freedoms of others'.[212] Understanding of this clause by the EDPB has centered on two rights which might be balanced against information rights in the GDPR — the right to privacy and trade secrets/intellectual property rights.[213]

Privacy of third parties. The rights to privacy and data protection of third parties is one of the most important roles of this provision, and likely its most common use. It is common that personal data relates to more than one natural person — messages, notes about one person made by another, ratings and reputation systems, or shared 'smart' devices, for example. This is not a *carte blanche* to refuse data provision, however. The European Court of Human Rights (ECtHR) has held that an access rights regime would be in breach of Article 8 of the Convention if there was no independent authority to determine if access had to be granted if an individual to whom data also relate failed to provide or withheld consent.[214]

The European Court of Human Rights has also weighed in on the argument that the inclusion of some personal data in a document renders it ineligible for release. In *Társaság a Szabadságjogokért v Hungary*,[215] an NGO attempted to access a complaint to the Constitutional Court submitted by a member of parliament. The Government of Hungary denied this request, arguing that the complaint contained the personal data of the member, and consequently was ineligible for release. The Court found it 'quite implausible that any reference to the private life of the MP [...] could be discerned from his constitutional complaint', and noted that it would be 'fatal for freedom of expression in the sphere of politics if public figures could censor the press and public debate in the name of their personality rights, alleging that their opinions on public matters are related to their person and therefore constitute private data which cannot be disclosed without consent'.[216] It found a violation of Article 10 (freedom of expression), in relation to the freedom to 'receive and impart information and ideas without interference by public authority and regardless of frontiers'.

This emerging regime appears favorable to the use of data rights in research, particularly if ethical reviews are undertaken to carefully consider third party privacy interests.[217]

Intellectual property of the controller. The EDPB anticipated that controllers will invoke this clause in relation to an adverse effect on *their* rights and freedoms. [218] A clear example would be where a trade secret or IP argument is forwarded by the controller.[219] Yet, as counselled in the recitals to the GDPR, 'the result of those considerations should not be a refusal to provide all information to the data subject.'[220] How this would play out in the situation where access requests *en masse* might threaten intellectual property in a different way is unclear. It is worth noting however that it would be very difficult for a data controller to accurately pre-empt the fact that data rights were being used in that way. Indeed, from the CJEU's case law on copyright protection, it can be derived that the mere potentiality of an IP breach will not generally be sufficient to impinge on the right to data protection in Article 8 of the Charter (which includes a right of access as mentioned before).[221]

Freedom to conduct a business. It could also be envisaged that a company claims that its 'freedom to conduct a business'[222] has been adversely affected. Yet the freedom to conduct a business is not an absolute right, but must be considered in relation to its societal function.[223] Restrictions of and interferences with this freedom are possible in cases where they correspond to an objective of general interest pursued by the Union, and respect the 'actual substance' of the freedom.[224] Furthermore, the Court has upheld that the tentative wording of Article 16,[225] which differs from that of other rights and freedoms in Title II of the Charter, reflects a broader leeway to restrict this freedom than they would have otherwise.[226]

Indeed, cases where the Court has held a measure in breach of Article 16 are rare, and even in these cases have only been in breach when read closely with EU secondary legislation.[227] In *Scarlet Extended*, the Court held that the installation of 'a complicated, costly, permanent computer system at [the company's] own expense' (to monitor internet traffic) would be a 'serious infringement' of the freedom to conduct a business in Article 16 of the Charter.[228] This was upheld in *SABAM v Netlog*.[229] However, it is important to consider the broader context in both cases, where the freedom to conduct a business aligned with the respective service providers' rights to data protection and freedom of expression (resp Articles 8 and 11 Charter). Moreover, in the latter case the Court relied specifically on the explicit language of the IPR Enforcement Directive to this effect, which forbids intellectual property enforcement measures that are 'unnecessarily complicated or costly'.[230] No comparable language or provision exists

212 GDPR, arts 15(4), 20(4).
213 See Article 29 Working Party, 'Guidelines on the Right to Data Portability (n 109)9–10 (mentioning only these two areas as examples of an issue to be considered as part of Article 20[4]).
214 *Gaskin v United Kingdom* [1990] EHRR 36. The United Kingdom for example entered a balancing test into the law as a result, stating there is no obligation to provide data 'to the extent that doing so would involve disclosing information relating to another individual who can be identified from the information' unless that third individual has consented to the release or it is reasonable to do so (determined with regard to the type of information, duties of confidentiality which might exist, attempts to seek consent, capabilities to give consent or express refusal of consent). See Data Protection Act 2018 (United Kingdom) sch 2 para 16.
215 *Társaság a Szabadságjogokért v Hungary App* no 37374/05 (ECtHR 2009).
216 *Társaság a Szabadságjogokért v Hungary* (n 215) para 37.
217 Research purposes benefit from a more lenient approach in the GDPR, See GDPR, art 89. See generally Miranda Mourby and others, 'Governance of Academic Research Data under the GDPR—Lessons from the UK' (2019) 9 *International Data Privacy Law* 192. See also Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 12.
218 See Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 9–10 (mentioning only these two areas as examples of an issue to be considered as part of Article 20[4]).
219 Gianclaudio Malgieri, 'Trade Secrets v Personal Data: A Possible Solution for Balancing Rights' (2016) 6 *International Data Privacy Law* 102; Jef Ausloos, *The Right to Erasure in EU Data Protection Law. From Individual Right to Effective Protection* (Oxford University Press 2020) 343–48.
220 GDPR, recital 63. Similarly, See Caroline Colin and Yves Poullet, 'Du Consommateur et de Sa Protection Face à de Nouvelles Applications Des Technologies de l'information: Risques et Opportunités' (2010) 2010/3 *DCCR* 94, 117; Malgieri (n 219) 103.
221 See case law references in Ausloos (n 219) 345.
222 Charter, art 16.
223 Joined Cases C-184/02 and C-223/02 *Spain and Finland v Parliament and Council* ECLI:EU:C:2004:497; Ausloos (n 219) 335–43.
224 Case C-554/10 *Deutsches Weintor eG v Land Rheinland-Pfalz* ECLI:EU:C:2012:526, para 54.
225 'The freedom to conduct a business in accordance with Union law and national laws and practices is recognised.'
226 Case C-283/11 *Sky Österreich GmbH v Österreichischer Rundfunk* ECLI:EU:C:2013:28, para 47.
227 See generally Peter Oliver, 'What Purpose Does Article 16 of the Charter Serve?' in Ulf Bernitz and others (eds), *General Principles of EU Law and European Private Law* (Kluwer Law International 2013).
228 Case C-70/10 *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL* (SABAM) ECLI:EU:C:2011:771.
229 Case C-360/10 *Belgische Vereniging van Auteurs Componisten en Uitgevers CVBA (SABAM) v Netlog NV* ECLI:EU:C:2012:85.
230 Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights (Text with EEA relevance) OJ L 157/45 ('IPR Enforcement Directive'), art 3(1).

in the GDPR. Even in cases where significant costs are placed upon a business, such as in *Denise McDonagh v Ryanair Ltd*, where the airline's duty to provide care after the eruption of the Icelandic volcano Eyjafjallajökull left passengers stranded, the existence of articles in secondary legislation that could be understood to reconcile fundamental rights (in this case, freedom to conduct a business and the right to property with the right to consumer protection) led the Court to rule no breach of the right to conduct a business had occurred.[231]

The GDPR has many provisions designed to respect (or enable Member States to navigate) the balancing between, among other fundamental rights and freedoms, Articles 8 and 16 of the Charter, such as Article 12 (on transparency modalities), Article 14(5) (on situations where information obligations can be avoided or relaxed) and Article 23(1) ('Restrictions'). Furthermore, for information-intensive companies, the marginal cost of providing information to each individual once a compliant infrastructure is established is very low (compared to, for example, flight compensation). Indeed, it does not generally require the establishing of any new modalities of communication, as information-intensive companies already have data and computational infrastructures, as well as log-in accounts and/or email, which can be used to this end.[232] Consequently, in agreement with many scholars,[233] we do not see much chance of the freedom to conduct a business as standing in the way of the use of data rights, including in research situations.

5.1.3 Abuse of rights?

The idea that an access right could, in certain situations, construe an abuse of rights was considered by Advocate General (AG) Kokott in her opinion in *Nowak*.[234] Abuse of rights is, however, 'rarely used, or at least not successfully',[235] usually implicated in politically charged, high level issues concerning freedom of expression or freedom of association, often when pitted against values of the defense of democracy. Yet, as AG Kokott noted, the risk of abuse of rights which was present in the 1995 Data Protection Directive is 'resolved' in the GDPR by the considerations of the rights and freedoms of others (see section 5.1.2).[236] We agree, noting further that the intent-agnostic nature of the right to access under the GDPR makes abuse more difficult to construe.[237]

5.1.4 Data, not documents

It is important to note that accommodating the right of access does not necessarily require sharing an exact copy of the data on the servers (or in the manual filing system) of the data controller in question. Both the CJEU and national courts have affirmed that a SAR is not a right to access whole documents, for example to provide context, but the right to the personal information contained within.[238] Such data could be extracted and provided in a variety of forms, and need not be in the original format. Indeed, there are times where that original format might actually be undesirable, such as if it is proprietary in nature, requiring the data subject to have specific software or expertise to examine it. No cases have been ruled on or are pending in the CJEU relating to the new right to data portability, but we can safely assume that that right, too, does not provide access to documents. As a result, there will be research designs that are better suited to freedom of information legislation,[239] or access to environmental information legislation,[240] which both can provide documentation for matters within their respective scopes. In many cases however, data controllers may find it easier to provide documents, and as such while it cannot be relied upon, data rights may be useful in studies where the original context is crucial for understanding.

5.1.5 Lack of consistency and machine readability

The 2012 GDPR proposal had a role for the European Commission, through implementing acts, of specifying a standard for the format of SAR responses in different sectors.[241] This aspect of the GDPR was a casualty of the intense, half-decade political battle over the text. The result is that access (and portability) rights do not have a common standard or format which data subjects can expect. This, in turn, makes it hard to build tools which are data controller agnostic, and which are reliable enough not to break if a data controller decides to switch the form of response they provide.[242] While codes of conduct and certification mechanisms under the GDPR may yet provide a means to help standardize this area,[243] we are still to see one on access or portability rights take concrete shape[244] — although a plethora of third parties seeking to sell back-end software to data controllers with the promise of consolidation and automation have emerged.[245]

Obtaining machine-readable data is crucial for research.[246]

231 Case C-12/11 *Denise McDonagh v Ryanair Ltd* ECLI:EU:C:2013:43 at [59]–[65].

232 ccf. Case C-649/17 *Bundesverband der Verbraucherzentralen und Verbraucherverbände—Verbraucherzentrale Bundesverband eV v Amazon EU Sàrl* ECLI:EU:C:2019:576, where the Court emphasised with reference to the freedom to conduct a business that a firm should not be obliged to establish a phone line for the purposes of communication with consumers where alternative means of direct and effective communication have been established.

233 See Hielke Hijmans, 'The European Union as a Constitutional Guardian of Internet Privacy and Data Protection' (PhD Thesis, University of Amsterdam 2016) 196, 216–17, 258; Ausloos (n 219) 333–49.

234 Case C-434/16 *Peter Nowak v Data Protection Commissioner* ECLI:EU:C:2017:582, Opinion of AG Kokott, paras 42–50.

235 Lorna Woods, 'Abuse of Rights' in Steve Peers and others (eds), *The EU Charter of Fundamental Rights : A Commentary* (Hart Publishing 2014) 1545.

236 Case C-434/16 *Peter Nowak v Data Protection Commissioner* ECLI:EU:C:2017:582, Opinion of AG Kokott, para 48.

237 See section 5.1.1.

238 Joined Cases Joined Cases C-141/12 and C-372/12 *YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S* [2019] ECLI:EU:C:2014:2081 [48].; *Dunn v Durham County Council* [2012] EWCA Civ 1654, [2013] 2 All ER 213 at 16; *Ittihadieh v 5-11 Cheyne Gardens RTM Company Ltd & Ors* [2017] EWCA Civ 121 at 93; *Rudd v Bridle* [2019] EWHC 893 (QB); *Rechtbank Noord-Holland (24 May 2019)* ECLI:NL:RBNHO:2019:4283; Parket bij de Hoge Raad 9 Nov 2018.

239 e.g., Freedom of Information Act 2000 (United Kingdom).

240 e.g., implementations of the UN/ECE Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters ('Aarhus Convention') such as transpositions of Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC OJ L 41/56.

241 Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM/2012/011 final - 2012/0011 (COD)), Art 15.

242 This is comparable to the politics of APIs and programmatic access, See generally section 2.2 above.

243 GDPR, arts 40, 42.

244 Though it is worth pointing to a recent initiative by the European Digital Media Observatory (EDMO), which is setting up a working group in order to develop a code of conduct on 'Access to Data Held by Digital Platforms for the Purposes of Social Scientific Research'. See notably: 'Call for Comment on GDPR Article 40 Working Group' (n 29); Vermeulen (n 29). There are also some self-regulatory initiatives, none of which really seem to have gained a lot of traction, most notably the 'Data Transfer Project' (with among its contributors: Apple, Facebook, Google, Microsoft and Twitter). See https://datatransferproject.dev

245 The IAPP compiled a list of such providers, accessible at https://iapp.org/resources/article/privacy-tech-vendor-report.

246 cf. European Commission, 'A European strategy for data' (n 6) 10.

Machine-readable data is not the same as digital data. For example, a PDF containing tabular data is designed to be printed rather than read and processed by a computer, and as such are not generally marked-up in such a way which makes automatic processing easy.[247] Machine-readable has been defined in EU law as 'a file format structured so that software applications can easily identify, recognize and extract specific data, including individual statements of fact, and their internal structure.'[248]

A teleological reading of the GDPR would require controllers to share personal data from the right to access in a consistent, machine-readable format unless great effort was involved. This effort is unlikely to be required in the context of online services, which given their automated nature already hold the data in such forms, as both consistency and machine readability are key to their business models — and to take more effort to obstructively destroy such properties would be highly questionable in light of the overarching data protection principle of fairness.[249] Even more so considering the European Commission's more recent push for 'stricter requirements on interfaces for real-time data access and making machine-readable formats compulsory for data from certain products and services'.[250]

5.1.6 (Re)identifying data subjects

Article 11(1) explains that controllers do not have to retain personal data *only* for the ability to potentially accommodate data subject rights at a later stage. Put differently, the requirement to accommodate data subject rights does not prevent them from anonymizing their datasets. Be that as it may, data subjects still have the possibility to provide the controller with additional information so as to (re-) identify their data in anonymized data-sets.[251] In practice however, this may lead to a frustrating back-and-forth between data subject and controller, where the data controller appears to have designed systems that are deliberately challenging to reidentify data subjects within.[252] In particular, the data controller may argue that the data, while clearly falling within the GDPR's scope (with high re-identification potential and in practice used to target or single out data subjects), may not be re-identifiable to the very high reliability needed to ensure that data not relating to an individual is delivered to them by mistake.[253] This is an argument Apple makes to refuse accommodating access requests with regard to the voice-data gathered in relation to its Siri-service.[254] Such arguments will generally be insufficient to block access requests entirely however.[255]

5.1.7 'Disproportionate effort?'

Some data controllers have read into data protection law the existence of a 'disproportionate effort' exemption which would exempt them from fulfilling an access request.[256] Such an exemption does not appear to exist in the GDPR, although it did in some transpositions of the now defunct 1995 Data Protection Directive.[257] Complaints around this are ongoing and it seems likely that more clarity will be forthcoming. Indeed, even if the increasing complexity of data processing ecosystems may render it hard to accommodate the core transparency requirements,[258] it does not exonerate controllers. To the contrary, Recital 58 highlights transparency is even more important in complex situations involving many actors.[259] When the controller processes a large quantity of personal data, Recital 63 does permit the controller to request the data subject to specify the information or processing activities to which the request relates.

One related provision that *does* exist in the GDPR is the ability to refuse a request if the nature of that request is 'manifestly unfounded or excessive, in particular because of their repetitive character'. Where this is done, 'the controller shall bear the burden of demonstrating the manifestly unfounded or excessive character of the request.'[260] This provision relates to the character of the request itself, rather than the character of the burden of fulfilling that request.

The EDPB have noted that for information society services such as large social media firms which specialize in automated data processing, 'there should be very few cases where the data controller would be able to justify a refusal to deliver the requested information, even regarding multiple data portability requests.'[261] They also note that the cost of building the infrastructure to comply with these requests is irrelevant to the notion of 'excessive' requests. In particular, they state that 'the overall system implementation costs should neither be charged to the data subjects, nor be used to justify a refusal to answer portability requests.'[262] Under these conditions, it appears that there are limited general reasons to refuse a data subject access request or portability request on effort grounds.[263]

5.1.8 National exemptions

It should also be noted that Article 23 grants Member States (and the EU legislator) the ability to install specific exemptions to the rights of access/portability in their national and/or sector-specific laws.[264] While most of the situations in which such exemptions can be pre-

247 The EDPB has stated that PDFs are unlikely to meet portability requirements, also noting that the requirements of portability must be interpreted in the context of the intention of the portability requirement, which the recitals (68) note is to promote interoperability. See Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 18. See also Ausloos and others (n 117) 286–87.

248 Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information [2019] OJ L 172/56, art 2(13).

249 See generally Clifford and Ausloos (n 78).

250 European Commission, 'A European strategy for data' (n 6) 20.

251 GDPR, art 11(2).

252 This is further detailed in: Michael Veale and others, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8 *International Data Privacy Law* 4; Ausloos (n 125). For mobile app-specific considerations, See Norval and others (n 124).

253 On the security implications of data rights, See Andrew Cormack, 'Is the Subject Access Right Now Too Great a Threat to Privacy?' (2016) 2 *European Data Protection Law Review* 15; Coline Boniface and others, 'Security Analysis of Subject Access Request Procedures How to Authenticate Data Subjects Safely When They Request for Their Data' (*2019 - Annual Privacy Forum*, 13 June 2019) https://hal.inria.fr/hal-02072302/document (accessed 4 April 2019).

254 Veale and others (n 252).

255 Ausloos and others (n 117) 308–09.

256 Ausloos (n 125).

257 e.g., Data Protection Acts 1998, 2003 (Ireland) s 4(9) (repealed).

258 René Mahieu and others, 'Responsibility for Data Protection in a Networked World: On the Question of the Controller, "Effective and Complete Protection" and Its Application to Data Access Rights in Europe' (2019) 10 *JIPITEC*.

259 Article 29 Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (Guidelines, 6 February 2018) 25.

260 GDPR, art 12(5).

261 Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 15.

262 Article 29 Working Party, 'Guidelines on the right to data portability' (n 109) 15.

263 See generally Ausloos and others (n 117).

264 For a review, See Access Now, One Year Under the EU GDPR: An Implementation Progress Report (Access Now 2019).

scribed relate to specific contexts[265] and are subject to conditions,[266] there is a catch-all included that makes it hard to anticipate the level of derogations to access/portability rights. Especially because this catch-all—enabling EU or Member State laws to restrict data subject rights in order to safeguard 'the rights and freedoms of others'—may be deployed in any kind of legislation (so not just the GDPR implementation laws). For example, while the seminal *Nowak* case in 2017 highlighted that data protection subject access rights applied to exam scripts,[267] this jurisprudence had limited direct applicability within the United Kingdom, which had an exemption for exam scripts being subject to access requests since 1998, replicated in the new law of 2018.[268]

In any case, such exemptions or derogations ought to be formulated and interpreted restrictively and narrowly. Hence, it is fair to say that the default position should be that data subject rights *are* applicable, *unless* the controller can clearly establish the applicability of a (national) exemption or derogation.[269] Such derogations are subject to potential challenge on the grounds of data protection principles and the fundamental right to data protection more generally.

5.2 Ethical considerations

While not easily split from other concerns, there are several ethical challenges that are distinctly applicable to data rights in research.

5.2.1 Who are the research subjects?

One ethical argument against the use of data rights in research is that it places a heavy burden on infrastructures that can prevent them from carrying out their normal function. A relevant question is whether data controllers (and their staff) would then be research subjects in the context of such a study.

Useful analogies can be found in studies of the peer review system. A 1982 study considered the rejection of duplicate papers by fictitious less-prestigious authors by selective American psychology journals.[270] They submitted 12 papers that journals had already accepted, authored by researchers from prestigious American psychology departments, but changed the names on the papers to fictious ones to see whether the prestige of the authors biased the reviewers' responses. A different 1987 study investigated whether social work journals' editorial processes were biased in favor of studies showing interventions to be effective, sending 146 submissions to test this hypothesis.[271] Both works were published by journals only trepidatiously and in an unusual manner. Despite referees' reservations about both the rigor of both studies, the journals that published these pieces (*Behavioral and Brain Sciences* and *Science, Technology and Human Values* respectively) did so only alongside commentaries (5 and 55 (short form) commentaries respectively) on relevant methodological and ethical issues.[272] In later years, the issue of studies into peer review was reignited by the 'Sokal affair', where a paper designed to be non-sense was submitted by Alan Sokal, a physics professor into a post-modern cultural studies journal and accepted, and follow-up events that have become known as 'Sokal Squared'.[273]

Scholars considering the ethical implications of these types of studies have questioned the 'social overhead of social research',[274] asking whether the 'costs of studying and correcting an injustice consume so many resources that they create new injustices, or create a net social loss [..] if too many people designed [peer review bias testing experiments], they would simply clog the peer review machinery altogether and bring the system to its knees.'[275] Parallel concerns have been raised in relation to issues of 'survey fatigue',[276] that 'indiscriminate use of surveys may be undercutting their effectiveness as a data collection approach by creating survey fatigue and lowering response rates',[277] particularly among student populations.[278] Others have considered that perhaps the journal editors and peer reviewers should have consented in line with widely accepted norms of research ethics. '[S]cientists do have rights,' one commentator noted, 'and [those] rights are not less than those guaranteed other human subjects'.[279] Others yet consider it important to weigh the stress on the system with the need to scrutinize gatekeepers of power and prestige.[280]

A key question to take away and analyze from this is whether formal processes of research ethics should be engaged simply because individuals are burdened as a result of the research. It is not clear in the case of data rights that simply because a human is involved in the fulfilment of a statutory obligation that the research should be treated as 'human subject' research.

There are some jurisdictions that have exempted studies concerning data rights from ethical review on the basis that disclosures mandated by legislation already have processes of custodianship associated with them and built into their respective regimes. Canada's three

265 For example, national security; defence; public security; prevention, investigation, detection or prosecution of criminal offences. See GDPR, art 23(1).

266 GDPR, art 23(2).

267 Case C-434/16 *Peter Nowak v Data Protection Commissioner* ECLI:EU:C:2017:994 (Nowak).

268 Data Protection Act 1998 (United Kingdom) sch 7 para 9 (repealed); Data Protection Act 2018 (United Kingdom) sch 2 para 25.

269 While certainly an interesting and much needed exercise, mapping the different implementations of Article 23 across EU Member States, even when only focusing on GDPR implementation laws, far reaches beyond the scope of this paper.

270 Douglas P Peters and Stephen J Ceci, 'Peer-Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again' (1982) 5 *Behavioral and Brain Sciences* 187.

271 William M Epstein, 'Confirmational Response Bias Among Social Work Journals' (1990) 15 *Science, Technology, & Human Values* 9.

272 Susan E Cozzens, 'Editorial' (1990) 15 *Science, Technology, & Human Values* 5.

273 Issues that arose in the Peters and Ceci and the Epstein studies also returned in subsequent peer-review 'hoax' studies, such as the so-called Sokal Affair, where the mathematician Alan Sokal sought to test his belief that the journal Social Text would accept an article that did not make sense, but supported the editors' ideological views. Despite the Sokal Affair reaching higher peaks of notoriety than either Peters and Ceci's or Epstein's controversies, Sokal submitted only a single paper, and therefore it is the parallel with the two studies above that is the most interesting for our purposes. See generally Stephen Hilgartner, 'The Sokal Affair in Context' (1997) 22 *Science, Technology, & Human Values* 506. On the later hoaxes, See Yascha Mounk, 'What an Audacious Hoax Reveals About Academia', (*The Atlantic*, 10 May 2018) https://www.theatlantic.com/ideas/archive/2018/10/new-sokal-hoax/572212 (accessed 30 November 2019).

274 See generally the special issue commencing with Joan E Sieber, 'Whose Ethics? On the Perils and Dilemmas of Studying Powerful Persons' (1983) 9 *SASP Newsletter* 1.

275 Mary Clark, 'Comments from the Side Lines' (1983) 9 *SASP Newsletter* 10, 11.

276 Stephen R Porter and others, 'Multiple Surveys of Students and Survey Fatigue' (2004) 2004 *New Directions for Institutional Research* 63.

277 Curtis A Olson, 'Survey Burden, Response Rates, and the Tragedy of the Commons' (2014) 34 *Journal of Continuing Education in the Health Professions* 93, 93.

278 Stephen R Porter, 'Survey Research Policies: An Emerging Issue for Higher Education' (2005) 2005 *New Directions for Institutional Research* 5, 8.

279 Michael J Mahoney, 'Bias, Controversy, and Abuse in the Study of the Scientific Publication System' (1990) 15 *Science, Technology, & Human Values* 50, 53.

280 Rachelle D Hollander, 'Journals Have Obligations, Too: Commentary on "Confirmational Response Bias"' (1990) 15 *Science, Technology, & Human Values* 46; Mahoney (n 279) 53.

federal research agencies note in their statement on ethical conduct for human-subject research that

> [r]esearch that relies exclusively on information that is publicly available, or made accessible through legislation or regulation, does not require REB [Research Ethics Board] review. Exemption from REB review for research involving information that is legally accessible to the public is based on the presence of a legally designated custodian/steward who protects its privacy and proprietary interests (e.g., an access to information and privacy coordinator or a guardian of Canadian census data).281

It is worth considering freedom of information (FoI) rights as a parallel case. A recent paper by Walby and Luscombe makes three core arguments in favor of not subjecting FoI-based research to ethical review.[282] Firstly, they claim that FoI already involves a bureaucratic vetting process, and only results in data being officially published by governments and redacted as appropriate with respect to national legislation. To extend ethical review to FoI-based research could be considered a form of unwarranted 'ethics creep',[283] where researchers become subject to restrictions on the use of *secondary* data. Data protection rights too have such built-in exemptions. Secondly, they use an analogy to the legal notion of *double jeopardy* to argue that researchers should not be subject to both the process of the 'quasi-ethical' exemptions in FoI law *and* university procedure. Thirdly, they argue that research ethics processes cannot infringe on a citizenship right: universities should not block a researcher's right to know, which in some cases (like New Zealand) is even constitutional in nature. They note a university refusing to push a right to know to its limit additionally could be accused of not carrying out its duty as a knowledge-seeking institution. The fundamental rights nature of access rights in EU law make this additionally convincing in the case of data protection.

Yet there is a significant difference between freedom of information and data protection transparency rights: the former are supposedly subject independent, whereas the latter are most certainly not. This creates ethical challenges that are more unique to data subject rights, to which we now turn.

5.2.2 Privacy of research subjects

Unlike the case argued above for data controllers, in many cases, those undertaking transparency requests — the data subjects themselves — should be treated as human subjects.

If the researcher themselves is gathering information (e.g. that relates to them) with data rights, fewer ethical considerations around the data subject are relevant. For example, a researcher may only need a single response per data controller to answer their research question. They may also be fabricating data subjects, such as through simulating web or app behavior to study tracking. Yet the researcher undertaking data requests alone does not mean that there are no ethical issues relating to data subjects. A helpful parallel is autoethnography — a qualitative research method that uses a researcher's autobiographical experiences as primary data to analyze and interpret the sociocultural meanings of such experiences.[284] In autoethnography, while the subject of the study is ostensibly the researcher, many other individuals are implicated through the stories being told and analyzed.[285] Where data relates to more than one person, these privacy issues may require ethical considerations that cannot be resolved by the data subject–researcher alone .[286]

However, in many of the scenarios illustrated above,[287] we have envisaged recruiting participants to carry out data rights where one of the aims is to contribute to the research project in question. This raises several issues.

While one of the tenets of research ethics is informed consent, information asymmetries in data rights use cases make this challenging.[288] The research team will not always be able to foresee the content or categories of personal data returned to the data subject, posing two main challenges.

The first is that the data subject might discover something that distresses them. There seems little need to pre-emptively protect subjects from dismal revelations about, for example, the sheer extent of online tracking, or reflection on their own experiences through data more generally.[289] Indeed, a call for participation could be structured to make the aim of triggering such experiences clear. However, data often inadvertently relate to more than one person,[290] and may reveal sensitive information that, for example, could create rifts and divisions between families and friends.

The second challenge is that the returned data might be so complex, or rich with potential inferences, that the individual themselves is unable to accurately appraise the sensitivity of what it is they are handing over to researchers. Individuals participating in citizen or participatory science projects do express privacy concerns, but a tendency to focus on 'openness, sharing, and the personal and collective benefits that motivate and accompany participation' can mask these and limit the attention paid to them by coordinating researchers.[291] This is problematic because even 'dull' seeming data framed as part of a significant collective good, such as smart meters in the context of climate change, can be extremely revealing of individuals' lifestyle and preferences.[292] Practices around genetic research indicate some of the challenges when individuals provide extremely potent data about themselves to third parties.[293] Yet in these cases, what genetic

281 Canadian Institutes of Health Research and others, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (2014) 16.
282 Kevin Walby and Alex Luscombe, 'Ethics Review and Freedom of Information Requests in Qualitative Research' (2018) 14 *Research Ethics* 1.
283 Kevin D Haggerty, 'Ethics Creep: Governing Social Science Research in the Name of Ethics' (2004) 27 *Qualitative Sociology* 391.
284 Heewon Chang, 'Autoethnography in Health Research: Growing Pains?' (2016) 26 *Qual Health Res* 443, 444.
285 See generally on the ethics of autoethnography Martin Tolich, 'A Critique of Current Practice: Ten Foundational Guidelines for Autoethnographers' (2010) 20 *Qualitative Health Research* 1599; Anita Gibbs, 'Ethical Issues When Undertaking Autoethnographic Research with Families' in *The SAGE Handbook of Qualitative Research Ethics* (SAGE Publications Ltd 2018).
286 See generally on the entangled nature of privacy Solon Barocas and Karen Levy, 'Privacy Dependencies' [2019] 95 *Washington Law Review* 555.
287 See supra section 4.
288 See on the overlap with data protection law: European Data Protection Supervisor (n 22) 18 et seq.
289 See generally Petr Slovák and others, 'Reflective Practicum: A Framework of Sensitising Concepts to Design for Transformative Reflection' in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17, New York, NY, USA, ACM 2017).
290 This is referred to as a bycatch by Barocas and Levy (n 286).
291 Anne Bowser and others, 'Accounting for Privacy in Citizen Science: Ethical Research in a Context of Openness' in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17, New York, NY, USA, ACM 2017) 2134.
292 Ian Brown, 'Britain's Smart Meter Programme: A Case Study in Privacy by Design' (2014) 28 *International Review of Law, Computers & Technology* 172; Michael Veale, Data Management and Use: Case Studies of Technologies and Governance (The Royal Society and the British Academy 2017).
293 See generally, concerning the list of findings that researchers should report in the United States by way of a voluntary code, Sarah S Kalia

diagnosis can potentially do, can be communicated to research subjects better than what a (personal) dataset of unknown variables of unknown extent might reveal.

In both these cases, the genetic analogy lends an important structural finding that may ameliorate concerns. This field has emphasized 'a duty on the part of a research investigator to consider what incidental and secondary results might occur from genomic testing, to create a plan for the possible return of results to participants, and to inform research participants of that plan before the tests are conducted'.[294] In data rights, a similar plan should be made clear. In the case where only a small subset of data would ever be needed and analyzed, a strict plan should be made to discard the rest as soon as possible, either before it leaves the research subject's control, or as soon as possible after if separation is technically challenging. If the aim is for the research subject to explore the data themselves, researchers should be aware of the potential for findings about e.g. others in the datasets that may concern or alarm the researcher and prepare the data subject accordingly. Particular care should be taken if the researchers are to ask open-ended questions of potentially large and unknown datasets provided by research subjects, and situations where researchers do this on their own without supervision or guidance from research subjects may be best avoided unless there is a very clear and justified reason to do so.

Apart from this, researchers should of course also comply with relevant legal protections, including the GDPR, that are aimed at safeguarding research subjects' privacy. This holds particularly true for key data protection principles such as 'purpose limitation', 'data minimisation', 'storage limitation', 'integrity and confidentiality'.[295] As emphasized by the EDPB, 'the principles of necessity and proportionality are essential' and it will not be sufficient for researchers to simply claim that the processing of personal data is 'necessary for the purposes of scientific research'.[296] Important here is that 'informed consent' in research ethics should be distinguished from research/data subjects consenting to the processing of their personal data (consent being one out of six grounds for rendering the processing of personal data lawful).[297] Indeed, in some situations there might be a clear imbalance between data subjects and the controller/researcher (e.g. because of the scale of the research project and/or how invested the research/data subject may be), which would challenge the GDPR-requirement for consent to be *freely given*.[298] Put briefly, to the extent researchers plan on receiving personal data of their participants, they will have to give due regard to data protection law. In this regard, it is worth referring to the European Commission's plans to propose a data governance legal framework that would also include rules to 'facilitate decisions on which data can be used, how and by whom for scientific research purposes in a manner compliant with the GDPR.'[299]

5.2.3 Risk of retribution

A last risk, which does not have considerable legal support but which may nevertheless pose ethical risks for data/research subjects is the possibility of some retribution by a data controller. Prior research into data rights has highlighted the tendency of some data controllers, for example, to respond to access requests as if they were erasure requests, presumably to avoid regulatory burden of troublesome data subjects.[300] This indicates a security risk that is posed to data subjects in relation to the availability of the data in the systems they are being asked to query. Deleting data before access has been provided may be considered a violation of the GDPR (notably the fairness, lawfulness and integrity principles in Article 5(1)), subject to considerable fines and even criminal prosecution in some countries.[301] However, in some cases it is also possible that the data controller or their agents are personally known to the research subject: for example, in the case of previous employers or medical practitioners. Considerations must be given to the social repercussions of requesting research subjects to use rights against controllers such as these.

5.2.4 Relationship to enforcement action

As data controllers are often responding to data rights in ways that do not seem compliant with the law,[302] researchers may feel they should work with research subjects to author complaints to data protection authorities to ensure the law is properly upheld. Given the overburdened and under-resourced nature of many authorities,[303] we feel this move should be supported in general as researchers will often be very well placed to explain breaches in detail and clarify important technical issues to the regulators. However, this does raise a challenge when research subjects are involved, as while a complaint seems like a simple form, in many jurisdictions it can open a legal process with the research subject as a party. While the research subject should not be put under any legal liability as a result, there is a small possibility they may be asked to eventually be party or intervenor to a legal case that could occur, such as an appeal against the decision of a supervisory authority. If this is undertaken, the potential role of the research subject going forward should be made clear, and while researchers may wish to provide the means and support for a research subject to complain, they should emphasize that this aspect should be considered an activity independent of the research project.

5.2.5 Broader ethical issues

None of this is to suggest that research questions themselves cannot bring ethical issues that are not well characterised by privacy concerns. A mass data access campaign to access and utilise biometric data to create facial recognition systems, for example, can bring ethical questions regardless of individual data subjects' consent. These are out of scope of this paper, which focusses on issues more specific to researching through data rights.

and others, 'Recommendations for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing, 2016 Update (ACMG SF v2.0): A Policy Statement of the American College of Medical Genetics and Genomics' (02 2017) 19 *Genet Med* 249.

294 Christine Weiner, 'Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts (December 2013 Report of the Presidential Commission for the Study of Bioethical Issues)' (2014) 180 *Am J Epidemiol* 562.

295 Article 5 GDPR.

296 European Data Protection Supervisor (n 22) 11–12, 16.

297 Article 6(1) GDPR

298 European Data Protection Supervisor (n 22) 18.

299 European Commission, 'A European strategy for data' (n 6) 12–13.

300 Ausloos and Dewitte (n 77).

301 In the UK, for example, it is considered a criminal offence 'to alter, deface, block, erase, destroy or conceal information with the intention of preventing disclosure of all or part of the information that the person making the request would have been entitled to receive'. Data Protection Act 2018 s 173(3).

302 Ausloos and Dewitte (n 77); Mahieu and others (n 95); Janis Wong and Tristan Henderson, 'The Right to Data Portability in Practice: Exploring the Implications of the Technologically Neutral GDPR' [2019] *International Data Privacy Law*.

303 See generally European Data Protection Board, 'First Overview on the Implementation of the GDPR and the Roles and Means of the National Supervisory Authorities' (Report presented to the European Parliament's Civil Liberties, Justice and Home Affairs Committee (LIBE), 26 February 2019).

5.3 Methodological considerations

5.3.1 Integrity of research

Certain uses of data rights might struggle for methodological validity when assessed in a strictly quantitative frame. In particular, some scholars advance a quantitative approach as a general template for conducting research with inferential, empirical validity in both quantitative and qualitative projects.[304] One characteristic result of this logic is the advice that increasing the number of records (assuming they are sampled in a random manner) will increase inferential leverage. For *ex post* data rights especially, this can be challenging, as uptake of the use of rights in a particular study might be limited, both in a general sense and among specific subgroups. According to a classic quantitative view, this might mean that the sample may be insufficiently large or representative to draw generalizable statistical conclusions from.

These problems mainly arise, however, when we confuse data rights and their potential with 'Big Data' research. The logic of research over large datasets made available through the digital economy, such as scraped Web data or global search patterns,[305] is that even data not collected for a particular purpose might reveal important societal phenomena due to the number of subjects and the richness of collected data. As *ex post* data rights require manual effort, they are not akin to this type of research, but more akin to citizen or participatory science. This field has well-known effort and participation biases, such as oversampling on weekends[306] or in certain areas[307] which researchers actively work to compensate.[308]

This indicates that data rights are more useful when certain characteristics of a research program are met. Studies that are considering small, well defined populations are apt for data rights. If participants were always going to be enlisted and worked with directly, and perhaps compensated for their time, then many of the biases simply reduce down to the classic representativeness challenges in fields such as psychology. If an attempt is made to generalize from a small sample to the world, significant challenges exist, such as capturing phenomenon as they manifest in easily accessed 'convenience samples' of participants,[309] such as students on campus,[310] which may differ from the world more generally. However, if the aim is to study exactly that type of student, this poses little problem.

If there is pre-existing reason to believe that a phenomenon will be homogeneous across populations, then data rights may also be appropriate. If the aim is, for example, to study how web tracking systems work online, these remain the same between individuals, although the websites sampled and technologies (such as tracker blockers) used may differ. In this situation, researchers are a gateway into a homogenous phenomenon, such as policy or infrastructure. Where this becomes challenging is where the aspect of infrastructure observed is heavily contingent on the data subject, as German credit scoring reverse-engineering effort OpenSCHUFA discussed above[311] found when it was unable to study issues such as discrimination due to a bias in white, male volunteers. OpenSCHUFA reflected that they 'were not able to get the attention of demographic groups that are probably most affected by poor SCHUFA scores' and as a result it was difficult to make generalizable conclusions, or understand all parts of the system.[312]

Statistical and methodological challenges around data rights must also be seen in the context of the pitfalls and biases in 'Big Data' research about the digital economy[313] — and data rights can potentially help provide alternative datasets as a check on these biases for the same types of phenomena — for example, for focusing on obtaining data about certain difficult to identify populations and communities that may be underserved or underrepresented in 'Big Data' held either by firms or obtained through other methods by external researchers.

5.3.2 Interactional considerations

Data requests can be made directly by the data subject or indirectly by an individual or organization mandated by a data subject. The latter option, however, can present difficulties as data controllers are concerned around releasing data to individuals pretending to be the data subject.[314] Individuals having been given demonstrable power of attorney are unlikely in practice to see problems of authentication,[315] but other agents, such as researchers, may be refused or requested for specific information to aid verification which only the data subject can provide. The data may also be provided to the data subject for sending on further to the third party again, necessitating a significant back-and-forth. We leave detailed legal analysis of mandating data rights to third parties to future work, but note that this is a challenging area, and in the absence of clear judicial clarification, it seems unlikely that controllers will adopt a consistent approach broadly necessary for research.

If rights are not to be delegated to a third party, it will be up to data subjects to interact with the data controller and obtain the necessary data, and to make all or relevant portions of that data available for research. This is easier said than done, as interaction with these controllers can take many different forms along a spectrum of collaborative to adversarial. In some cases, adversarial approaches

304 e.g., Gary King and others, *Designing Social Inquiry* (Princeton University Press 1994).

305 e.g., Shihao Yang and others, 'Accurate Estimation of Influenza Epidemics Using Google Search Data via ARGO' (2015) 112 *Proceedings of the National Academy of Sciences* 14473.

306 Jason R Courter and others, 'Weekend Bias in Citizen Science Data Reporting: Implications for Phenology Studies' (2013) 57 *Int J Biometeorol* 715.

307 Yexiang Xue and others, 'Avicaching: A Two Stage Game for Bias Reduction in Citizen Science' in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (AAMAS '16, Richland, SC, International Foundation for Autonomous Agents and Multiagent Systems 2016).

308 e.g., Chankyung Pak and others, 'Auditing Algorithms With Donated Data: Methods for Poor Scientists?' (*ICA*, Virtual, 20–[26]May 2020).

309 Robert A Peterson and Dwight R Merunka, 'Convenience Samples of College Students and Research Reproducibility' (2014) 67 *Journal of Business Research* 1035.

310 e.g., Patricia M Greenfield, 'Sociodemographic Differences Within Countries Produce Variable Cultural Values' (2014) 45 *Journal of Cross-Cultural Psychology* 37 (arguing that the difference between student populations from different socioeconomic backgrounds can be larger than cultural differences between countries); Paul HP Hanel and Katia C Vione, 'Do Student Samples Provide an Accurate Estimate of the General Public?' (2016) 11 *PLoS One* (arguing that different student populations significantly differ from the general public in ways that are difficult to explain).

311 See supra section 4.2.1.

312 'OpenSCHUFA' (OpenSchufa, no date) https://openschufa.de/english (accessed 24 June 2019).

313 Olteanu and others (n 31).

314 See generally Coline Boniface and others, 'Security Analysis of Subject Access Request Procedures How to Authenticate Data Subjects Safely When They Request for Their Data' [2019] *Annual Privacy Forum*, Jun 2019, Rome, Italy; Cormack (n 255).

315 See eg Information Commissioner's Office, 'Right of Access' (Guide to the GDPR, 12 August 2019) https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access (accessed 1 December 2019).

may be required as data controllers are unwilling to provide data they are required to by law. It may be possible for the research team to pre-empt and avoid these adversarial encounters by testing the process for relevant controllers before the research begins, allowing identification of any hurdles, the enlistment of the local data protection authority if required,[316] and the creation of both a tailored request and pre-built responses that are suitable for the particular issues and views of the controller in question. Researchers will have to consider participants' skills when crowd-sourcing data-gathering using the GDPR. This can be dealt with to some extent, by providing explanations, personal or technical assistance and tools.[317]

In some cases however, the research project may have to be postponed while enforcement or legal action can be carried out.[318] On the more collaborative side of the spectrum, one could imagine company and researchers agreeing to include a specific tag in participants' access requests so that they are prioritized and/or responded to in a predefined format. Researchers may also simply rely on available 'download my data' functionalities already offered by many online services, which currently only generally provide a fraction of eligible data, but which may be suitable for the research question.

Information may be provided in a variety of ways, such as files through secure drop facilities, as email attachments with or without passwords, or on physical media (particularly for data outside of the digital economy such as CCTV footage). The research team must prepare for these different formats and create a secure, suitable and ideally easy-to-use way for data subjects to grant access to this data. There may be an important role for the researchers to carry out an initial request to create more bespoke guidance of what to expect from a data controller. Relatedly, the research team should also make efforts to ensure that data subjects are not storing this data in insecure ways, and advise them on the correct storage or disposal if appropriate.

6 Conclusion

The concentration and privatization of data infrastructures, turns (mainly big technology) companies into *de facto* gatekeepers of research agendas. Independent researchers have developed a wide variety of approaches in order to pierce through enclosed datasets, each with their benefits and drawbacks. This article outlines a fairly new approach to add to researchers' toolset for obtaining relevant research data (Section 3). Compared to other tools, data rights under the GDPR have the advantage of being potent (legally enforceable) and enabling access to very fine-grained data (Section 4). That being said, they also raise a number of (legal, ethical and methodological) issues whose significance will vary depending on the actual research projects (Section 5).

Given the multi-faceted nature of using data rights outlined throughout this paper, it is not possible to outline a detailed procedure or plan that would fit each potential research project. That said, the seven steps identified at the start of section 4 may serve as a useful starting point for researchers interested in using data rights in their project.[319] The research team should reflect upon the process in the context of methodological, ethical, legal and data security and protection challenges described in Section 5. Such analysis will depend in large part on national and local processes specific to different countries, university systems or funders. Methodological issues will be largely discipline-specific, and cross-cutting guidance cannot be easily provided linking this specific data collection approach to the broad and welcome array of potential analysis techniques.

In conclusion, using data rights requires a triangle of expertise – domain, technical and legal – the constellation of which may vary from one research project to another. Any research project will of course rely on adequate *domain expertise* relating to the actual research questions. Data rights in particular require a minimum level of *legal expertise* to properly identify the opportunities and limitations, as well as manage the interaction strategy. Finally, *technical expertise* may be necessary in order to understand and process the data received.

***

Researching with data rights is still at a very early stage. Our aim with this article was both to explain the potential utility of data rights to researchers, as well as to provide appropriate initial conceptual scaffolding for important discussions around the approach to occur. We do not claim to have exhausted either the possibilities or the challenges of using the transparency provisions in data protection law for research, and offer only a non-exhaustive tour through some of the issues and questions that might arise. Data rights may not be the right tool for every job, but there are many investigations of data and power in particular that remain open. Data protection is a flexible instrument designed to address asymmetries of informational power, and we believe researchers should be at the forefront of finding new ways to use that flexibility for societally critical knowledge generation.

Acknowledgements

This article underwent countless alterations and is the product of many discussions. We would like to thank everyone that has provided us with their feedback in many different forms. We are particularly grateful for the engaged audience at TILTing Perspectives (Tilburg, 2019), as well as the rich conversation, led by Joshua Kroll, at the the Privacy Law Scholars Conference where an earlier version of this paper was workshopped (Berkeley, June 2019). We also want to thank the TechgReg reviewers for their incredibly fast and thoughtful feedback and suggestions.

316 e.g., Johnny Ryan, 'Regulatory Complaint Concerning Massive, Web-Wide Data Breach by Google and Other "Ad Tech" Companies under Europe's GDPR' (*Brave Browser*, 9 December 2018) https://www.brave.com/blog/adtech-data-breach-complaint (accessed 1 May 2019); 'Our Complaints against Acxiom, Criteo, Equifax, Experian, Oracle, Quantcast, Tapad' (*Privacy International*, no date) http://privacyinternational.org/advocacy-briefing/2426/our-complaints-against-acxiom-criteo-equifax-experian-oracle-quantcast-tapad (accessed 8 April 2019).

317 One of the authors has undertaken several types of research set-ups, interacting with subjects in different ways. Unsurprisingly, the project with only a limited number (3) of law students, with bi-weekly follow-up calls, appeared the most successful.

318 See the Uber references in n 14.

319 Aim > Data > Legal Approach > Scope > Recruitment Strategy > Interaction Strategy > Data Analysis Strategy.