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.

[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.

* 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.

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]

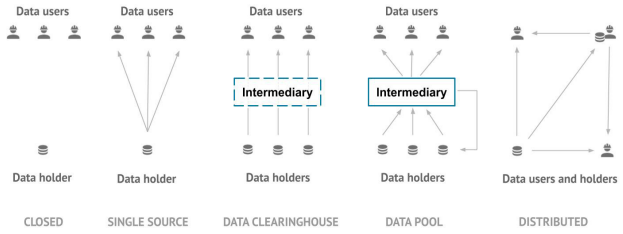

Fig. 1: Data Governance Models

[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-projekt.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.

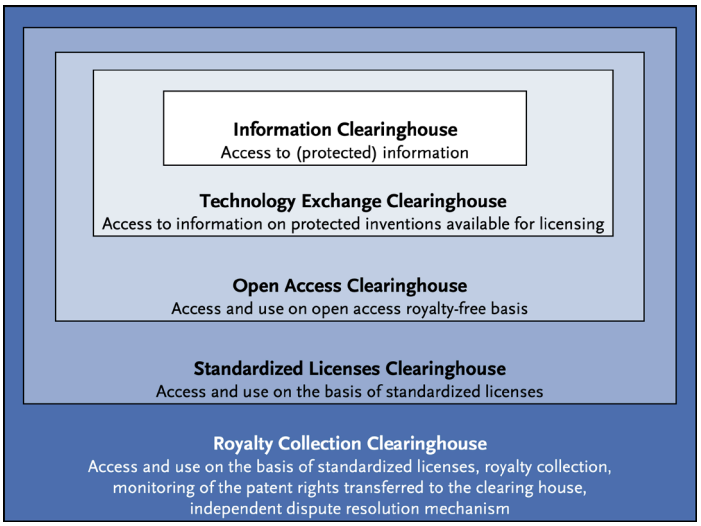

Fig. 2: Categorisation of clearinghouses following van Zimmeren et al (n 94) 354.

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.

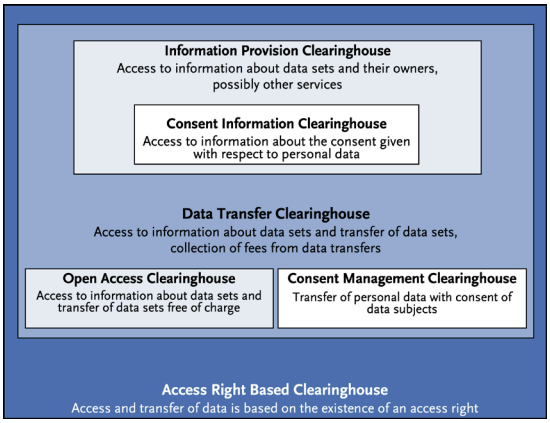

Fig. 3: Data Clearinghouses

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.