Technology and Regulation

The Double Helix of Data Extraction
Radicalizing Reflexivity in Critical Data Studies

Dr Sebastián Lehuedé

sebastian.lehuede@kcl.ac.uk

The increasing value of digital data is rendering data extraction ubiquitous. Looking at recent research, I argue that critical examinations of such extractive dynamics can give rise to a 'double-helix of data extraction' that exacerbates existing asymmetries by appropriating the affected populations' means of critique, effectively disempowering them. I show how this dynamic might play out in practice by turning to my research on astronomy data in Chile, which highlighted the situation of Indigenous activists through a decolonial lens. I advocate a radical embrace of reflexivity attentive to both positionality and political economy as a condition for conducting properly critical data studies (CDS), especially in the case of research relying on decolonial approaches. Although reflexivity cannot solve the double helix, it can expose the conditions underpinning research and acknowledge the limitations of research on its own.

1. Introduction

Data extraction has become ubiquitous. Some of the world's wealthiest companies, such as Meta/Facebook, rely to a great extent on an aggressive and unconsented appropriation of user data.[1] Other types of data, such as pharmaceutical data that are derived from biodiverse regions, have also started to get produced and commodified at a large scale, in some cases without the consent of the Indigenous communities who have safeguarded the environment for centuries.[2] Against this backdrop, scholars have denounced how the increasing political and economic value of data is exacerbating already existing power asymmetries, giving rise to a field of study known as *critical data studies (CDS)*.[3] This field has adopted an increasingly critical bent, building upon decolonial,[4] critical race,[5] feminist[6] and other jus-tice-oriented frameworks to explore how data extraction connects with longstanding histories of dispossession, discrimination and social control in modernity.

Despite the valuable contribution made by CDS, my research experience shows that merely denouncing data extraction practices does not suffice for bringing about social change. This is because, as Indigenous women from different regions have expressed, critical theory and research are not immune to reproducing dynamics of extraction themselves, even when data production and interpretation is made in the name of social justice.[7] In the context of my PhD research, I interviewed Lickan Antay Indigenous activists who saw the operation of the international astronomical observatories in the Atacama Desert, Chile, as involving a form of data extraction.[8] For these activists, decisions such as the construction of the Atacama Large Millimeter/submillimeter Array (ALMA) on the sacred Chajnantor mountain speak to the lack of dialogue proper of the extractivist ethos.[9] However, a critical analysis of my own assumptions and practices also reveals extractive patterns in my approach to these activists. For example,

1   Joseph Turow and Nick Couldry, 'Media as Data Extraction: Towards a New Map of a Transformed Communications Field' (2018) 68 Journal of Communication 415.
2   Shane Greene, 'Indigenous People Incorporated? Culture as Politics, Culture as Property in Pharmaceutical Bioprospecting' (2004) 45 Current Anthropology 211.
3   Andrew Iliadis and Federica Russo, 'Critical Data Studies: An Introduction' (2016) 3 Big Data & Society 1.
4   Nick Couldry and Ulises A Mejias, *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism* (Stanford University Press 2019).
5   Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity 2019).
6   Catherine D'Ignazio and Lauren F Klein, *Data Feminism* (MIT Press 2020).

Lecturer in Ethics, AI and Society, Department of Digital Humanities, King's College London
Research Associate, Centre of Governance and Human Rights, University of Cambridge

7   Silvia Rivera Cusicanqui, 'Ch'ixinakax Utxiwa: A Reflection on the Practices and Discourses of Decolonization' (2012) 111 South Atlantic Quarterly 95; Leanne Betasamosake Simpson and Naomi Klein, 'Dancing the World into Being: A Conversation with Idle No More's Leanne Simpson On Extractivism' [2013] *Yes! Magazine*; Linda Tuhiwai Smith, *Decolonizing Methodologies: Research and Indigenous Peoples* (3rd Ed., Zed 2021).
8   Sebastián Lehuedé, 'Territories of Data: Ontological Divergences in the Growth of Data Infrastructure' (2022) 5 Tapuya: Latin American Science, Technology and Society 1.
9   It is important to note that, although astronomy data does not have a commercial value, the public and the private sector in Chile regard this data as a potential tool for scientific and economic development.

members of Lickan Antay Indigenous group explained to me that the operations of the observatories were not their priority since the local communities were more concerned with the boom of lithium extraction, a mineral employed in the development of rechargeable batteries.[10] This situation triggered relevant questions for me, which ranged from power dynamics regarding whose concerns come to matter in the definition of research themes to how the different parties involved, such as the Indigenous activists and myself, would benefit from the exchange.

More broadly, conducting this research made me question whether the people I talked to were being subject to not just one form of data extraction, namely the one conducted by the observatories, but another form as well: my research project. If that were the case, and despite my intentions, a phenomenon that I call 'double helix of data extraction' could have exacerbated existing hierarchies by exposing local communities to two interrelated forms of data extraction – one pertaining to the environment where they live and another one concerning their means of resistance to extractive dynamics. My research on astronomy data made me realize that relevant questions need to be asked by CDS scholarship, especially in the case of research drawing on decolonial and other justice-oriented approaches.

Looking ahead, in this article I advocate *radical reflexivity* as a means to identify, expose and address usually overlooked power dynamics taking place in critical data research. Reflexivity calls for self-scrutiny in the research process, inviting researchers to explore how their positionality has shaped the different stages of research, from the production of the data to their presentation in the form of findings.[11] The idea of *radical* reflexivity points to the fact that the current context requires putting reflexivity at the centre of CDS and carrying it out in a way that, following Rivera Cusicanqui,[12] takes into consideration issues of both positionality and political economy. In this article I also discuss participatory-action research (PAR) as an approach that is particularly well suited for addressing the gaps identified through reflexivity.

Certainly, reflexivity cannot 'solve' the double helix of data extraction; however, it constitutes a necessary step to advance research on data extraction that is aware of the struggles underpinning the theoretical and empirical sources employed and of the limitations of academic research for advancing social justice on its own. In my view, radical reflexivity constitutes a condition for conducting properly 'critical' data research that does not reproduce the very dynamics it aims to denounce. As Medrado and Verdegem explore in their work on Artificial Intelligence (AI), reflexivity can inform participatory research frameworks in order to advance a more horizontal research process in collaboration with participants.[13]

This article proceeds as follows. The first section describes the double helix of data extraction, introducing the 'mainstream' and 'critical' forms of data extraction and explaining how they can come to operate together. After that, I narrate how the double helix of data extraction could have come to matter in my research on astronomy data in Chile. In particular, I analyze three extractive dynamics that came to the fore during different stages of my project: (1) an undermining of the relevance of Indigenous voices in the research design, (2) the contrast between my and the Lickan Antay communities' priorities and (3) the unequal opportunities of the different parties involved to take advantage of my research. After that, I pinpoint radical reflexivity and unpack its relevance for conducting properly critical data research. Finally, in the conclusion I concur with some critiques with regard to reflexivity but argue that the double helix of data extraction makes radical reflexivity a necessary step for challenging the asymmetries denounced by CDS.

2. The Double Helix of Data Extraction

One of the most common definitions of data portrays them as self-standing forms of evidence employed to sustain truth claims.[14] However, a critical lens reveals that data have been crucial in facilitating diverse forms of extraction. As an influential strand of work has argued, the current stage of capitalism is marked by the large-scale extraction of personal data by digital platform companies. For Shoshana Zuboff, new forms of data extraction are giving rise to a novel economic order where powerful technology companies expropriate the data from users and sell it to buyers interested in predicting and manipulating human behavior. As she writes: 'Ford's inventions revolutionized *production*. Google's invention revolutionized *extraction* and established surveillance capitalism's first economic imperative: the *extraction imperative*'.[15] In a similar vein, although adopting a broader framework that brings into view historical patterns of domination, Couldry and Mejias[16] argue that *data colonialism* represents a new phase of both capitalism and colonialism. Just like the European empires took the land, resources and bodies of the colonies during formal colonialism, at present powerful transnational companies are appropriating the very life of individuals through data extraction.

As Zuboff, Couldry and Mejias, other authors have also pointed out the relevance of extractive dynamics in relation to platform capitalism[17] and so-called smart technologies[18] as well as used decolonial frameworks to approach activist data,[19] humanitarian initiatives[20] and other applications of data. Unfortunately, this strand of research has not delved into the extent to which critical research can also involve forms of extraction. This lack of reflexivity, I argue later, reflects a deficit present in the field of critical data studies more generally.

Besides digital platforms, it is also possible to identify patterns of colonialist data extraction across modernity and in relation to other types of data. One of the first manifestations of the idea of 'data

10 Danae Tapia and Paz Peña, 'White Gold, Digital Destruction: Research and Awareness on the Human Rights Implications of the Extraction of Lithium Perpetrated by the Tech Industry in Latin American Ecosystems', *Technology, the environment and a sustainable world* (Global Information Society Watch 2020).

11 Wanda S Pillow, 'Confession, Catharsis, or Cure? Rethinking the Uses of Reflexivity as Methodological Power in Qualitative Research' (2003) 16 International Journal of Qualitative Studies in Education 175.

12 Rivera Cusicanqui (n 7).

13 Andrea Medrado and Pieter Verdegem, 'Participatory Action Research (PAR) in Critical Data Studies: Questioning AI for Social Good from a South-North Approach' (forthcoming).

14 Daniel Rosenberg, 'Data Before the Fact' in Lisa Gitelman (ed), *'Raw Data' Is an Oxymoron* (MIT Press 2013).

15 *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs 2019) 87, emphases in the original.

16 Couldry and Mejias (n 4).

17 Nick Srnicek, *Platform Capitalism* (Polity 2017).

18 Jathan Sadowski, *Too Smart: How Digital Capitalism in Extracting Data, Controlling Our Lives, and Taking Over the World* (The MIT Press 2020).

19 Paola Ricaurte, 'Data Epistemologies, The Coloniality of Power, and Resistance' (2019) 20 Television & New Media 350.

20 Mirca Madianou, 'Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises' (2019) 5 Social Media and Society.

deluge' took shape in the Renaissance, when European travellers 'collected' a range of specimens in the so-called New World that challenged existing taxonomies and categories.[21] Just like AI aims at making sense of vast volumes of data, new systems of classification and standardization were developed at that time and improved throughout modernity. In many cases, the production and processing of this data encompassed blatant acts of extraction. A perfect case in point stems from European botany in early modernity. While the knowledge of Indigenous and Black populations on plants and new specimens was crucial for elaborating drugs and increasing the power and wealth of Europe, quite often this information was obtained through coercion and without any sense of reciprocity.[22] This story might sound like a thing of the past, but the current extraction of genetic data by transnational pharmaceutical companies from plants and animals in places such as Peru's high forest, which is being safeguarded by Indigenous communities, speaks to a similar dynamic.[23]

To the above form of data extraction, which I call 'mainstream', it is necessary to add a 'critical' one. Critical data extraction emerges as research and activism turn to theoretical or empirical sources stemming from dissenting groups[24] to denounce data extraction. Critical data extraction encompasses a form of appropriation even if it strives for social justice and embraces horizons such as that of decoloniality. The key to identifying critical data extraction concerns the type of relationship established by critical researchers with the sources inspiring such critiques. As long as these sources represent dissenting forms of knowledges that emerged in relation to specific struggles, an extraction can occur when those experiences are utilized without acknowledging the context in which they emerged. In such cases, it is possible to speak of a co-optation of the means of critique: while dissenting groups undertook the labour required to articulate relevant reflections about the working of power, critics of data extraction obtain symbolic and material benefits without attending to the struggle of those who inspired their critique in the first place. While in this article I explain how this dynamic plays out in relation to Indigenous activists, other dissenting groups, such as Black feminist authors, can also become subject to critical forms of data extraction.

The ideas put forward by Indigenous women from different regions of the world are particularly helpful for understanding how seemingly critical research can also encompass a form of extraction. Reflecting on her experience in the Idle No More movement in Canada, Michi Saagiig Nishnaabeg thinker Leanne Betasamosake Simpson considers that the logic of extraction consists of transforming Indigenous lands, plants, animals and bodies into mere resources.[25] For Simpson, extraction 'removes all of the relationships that give whatever is being extracted meaning ... [Extracting] is taking without consent, without thought, care or even knowledge of the impacts that extraction has on the other living things'.[26] For example, Métis/otipemisiw scholar Zoë

Todd[27] has denounced how renowned European social scientists are relying on ancestral knowledges as a means to face the climate crisis, but without engaging in dialogue with Indigenous thinkers.[28] Similarly, Ngāti Awa, Ngāti Porou, Tūhourangi intellectual Linda Tuhiwai Smith opens her influential book 'Decolonising Methodologies' by arguing that '[i]t appalls us that the West can desire, extract and claim ownership of our ways of knowing ... and then simultaneously reject the people who created and developed those'.[29] From this perspective, not only material resources and labour but also critical knowledges can be subject to extraction when there is no sense of responsibility and reciprocity.

Similar ideas have been expressed by Aymara/Bolivian thinker Silvia Rivera Cusicanqui in relation to the Latin American context.[30] For Rivera Cusicanqui, some decolonial thinkers of Latin American origins have managed to thrive in Global North universities by undertaking a form of extraction. These authors have increased their intellectual authority in the North and the South and moved up in the academic ladder by relying on local knowledges without proper acknowledgement of those who first formulated such ideas or by omitting the struggles that inspired them. This is how Rivera Cusicanqui presents the overall picture emerging from this form of extraction:

> Ideas run, like rivers, from the south to the north and are transformed into tributaries in major waves of thought. But just as in the global market for material goods, ideas leave the country converted into raw material, which become regurgitated and jumbled in the final product.[31]

The reflections by Simpson, Todd, Smith and Rivera Cusicanqui help delineate what I call the 'critical' form of data extraction, which takes place when research on data extraction relies on Indigenous or other dissenting theoretical and empirical sources but does not engage with the context that prompted such reflections in the first place. As the case discussed by Rivera Cusicanqui shows, quite often 'critical' data extraction can reinforce North/South asymmetries and take place even when research embraces radical political horizons such as that of decoloniality.

Mainstream and critical data extraction might appear as separate phenomena, but in practice they act in a concerted way. I speak of a *double-helix of data extraction* because, similarly to the physical structure of DNA, these two forms of data extraction can wind around each other and establish different contact points. While mainstream data extraction appropriates the value of environments, bodies and the life of individuals in different ways, critical data extraction appropriates the means of critique by relying on the vocabularies and experience of dissenting groups without engaging with the contexts that inspired their struggles. Because of this, these two forms of data extraction can encompass a double strike that affects the chances of resisting data extraction. As companies, scientific organizations and other data extraction actors appropriate the value from the resources and labour of a broad range of groups (mainstream data extraction), researchers formulate sophisticated critiques of data extraction in

21   Bruno J Strasser, 'Data-Driven Sciences: From Wonder Cabinets to Electronic Databases' (2012) 43 Studies in History and Philosophy of Science 281.
22   Londa Schiebinger, 'Prospecting for Drugs: European Naturalists in the West Indies' in Sandra Harding (ed), *The Postcolonial Science and Technology Studies Reader* (Duke University Press 2011).
23   Greene (n 2).
24   By dissenting groups, I refer to groups engaged in the construction of worlds that are not based on racist, patriarchal and capitalist precepts.
25   Simpson and Klein (n 7).
26   Simpson and Klein (n 7) para 11.
27   Zoe Todd, 'An Indigenous Feminist's Take On The Ontological Turn: "Ontology" Is Just Another Word For Colonialism' (2016) 29 Journal of Historical Sociology 4.
28   Todd (n 27).
29   Smith (n 7) 1.
30   Rivera Cusicanqui (n 7).
31   Rivera Cusicanqui (n 7) 104.

a way that improves their prestige and status but does not advance the struggle that inspired their sources (critical data extraction). Even if this is not always the case, it is noteworthy that mainstream data extraction usually involves *large-scale* data extraction (as in the case of digital platforms) and critical data extraction constitutes a more selective and therefore *small-scale* appropriation of data emerging from the experience of dissenting groups.

The above does not mean that research in itself is wrong or undesirable. One reason for this is because research is not only conducted by academics but also by communities and groups in their struggle for social justice. A tradition of praxis research in Latin America sees binary opposition between theory (or research, in this case) and practice as a modern artificial construct. As Catherine Walsh and Walter Mignolo ask: 'Are you not doing something when you theorize or analyze concepts? Isn't doing something praxis? And from praxis ... do we not also construct theory and theorize thought?'.[32] The question for academic research, thus, is how to ensure horizontal dialogue with practitioners, activists, communities and a range of actors working with, or affected by, datafication. In this article I highlight participatory-action research as particularly well suited for establishing such horizontal collaborations.

In the next section I turn to my study on astronomy data in Chile to explain how I came up with the idea of the double helix of data extraction and provide an empirical example of how this dynamic might play out in practice. In so doing, I first discuss the type of mainstream data extraction at stake in the Chilean case and then identify three extractive patterns that made me wary of my own role in the setting I was studying.

3.   Extractivist Patterns in my Research on Astronomy Data in Chile

One of the few things I knew when I started my PhD in 2017 at the Department of Media and Communications at the London School of Economics (LSE), a prestigious university according to global rankings, was that I wanted to look into global data exchanges. I was particularly interested in the power dynamics playing out in such exchanges, including issues of epistemology and political economy as they were being explored by CDS. The case of astronomy data in Chile fit perfectly with that goal. As someone who grew up in Chile, I was aware that the Atacama Desert was the host of sophisticated and futuristic astronomical observatories constructed by predominantly US and European scientific organizations. Besides the unique visibility of the area, international astronomy projects had been attracted thanks to a series of tax, labour and environmental exceptions put forward by the Chilean state.[33]

In 2017 I learnt that these observatories were producing increasingly vast volumes of data — around 16.5 petabytes per year in 2021[34] — and that ambitious initiatives were emerging in order to take advantage of this situation for national development. For example, the Data Observatory public-private partnership aimed at taking advantage

of astronomy data for industry development and had Amazon Web Services as one of its founding members. In terms of theory, I decided to analyze these initiatives from a Latin American decolonial lens, situating them in the context of a capitalist modern/colonial world system.[35] As for the methodology, I conducted 34 interviews with astronomers, astroinformaticians, members of the public and the private sector and Indigenous activists affected by the construction of one of such observatories in 2018 and 2019.

A crucial source in my research were the Lickan Antay Indigenous activists. The Lickan Antay people inhabit oases, valleys and streams around the Atacama salt flat and the Loa River in the Antofagasta region. They were part of the Incan empire in the fifteenth century, the Spanish empire almost a century later, and their territory was claimed by Bolivia in 1883 and annexed by Chile in that same year. The Chilean state has held contradictory approaches to the Lickan Antay people — while in the twentieth century it repressed their culture, causing the extinction of the Kunza language, some policies advanced since the nineties have contributed to strengthen the Lickan Antay identity.[36] As other Indigenous communities, the Lickan Antay people have an intimate connection with the territories they inhabit, considering the mountains active agents shaping their lives and conducting rituals in order to keep the streams of water, a scarce resource, thriving.

As per my personal experience, I did not know any Lickan Antay person before conducting this research. I grew up in an urban context (Santiago de Chile) and do not hold any known Indigenous ancestry. Even though in legal terms the Lickan Antay people and I share the same nationality, some of them would call me 'Chilean'.

The Lickan Antay activists I talked to expressed to me that their neighbor, the Atacama Large Millimeter/submillimeter Array (ALMA) astronomical project, could be considered an extractivist actor. As one of them explained to me: 'ALMA's would be a more superficial form of extractivism. An extractivism of the knowledge of the territory ... Because the thing about extractivism is that it does not give back. It only takes'. Such a definition of extractivism holds striking resemblances with the one by Leanne Betasamosake Simpson I mentioned in the previous section. In the case of astronomy in Chile, the 'only taking' approach of ALMA reflects on at least two conflicts that this observatory has had with the local communities. In 2002, the local communities opposed the concession of the Chajnantor mountain by the state to ALMA. This mountain constitutes Lickan Antay ancestral lands, and some members of the local community carried out protests before reaching an agreement with the observatory. More recently, ALMA decided to install natural gas pipelines, which required undertaking excavations and other works in the area. Concerned about this situation, Toconao villagers undertook a field visit and noted that the pipelines were threatening the life of 22 chululo colonies, a small rodent that only lives in the Atacama Desert. Besides such conflicts, Lickan Antay activists expressed to me that ALMA had not made efforts to build any type of sustained dialogue with the local communities.

32   Catherine E Walsh and Walter D Mignolo, 'Introduction' in Catherine E Walsh and Walter D Mignolo (eds), *On Decoloniality: Concepts, Analytics, Praxis* (Duke University Press 2018) 7.

33   Javiera Barandiarán, 'Reaching for the Stars? Astronomy and Growth in Chile' (2015) 53 Minerva 141.

34   Ministry of the Economy, 'Data Observatory: Llamado a Propuestas de Valor [Data Observatory: Call for Partners]' (2019) https://drive.google.com/file/d/1DXCjoAcF7lrDiNI2GcpiOv7RJooyU3-b/view accessed 20 June 2022.

35   Ramón Grosfoguel, 'The Epistemic Decolonial Turn: Beyond Political-Economy Paradigms' (2007) 21 Cultural Studies 211.

36   Hans Gundermann, 'Los Atacameños y Sus Relaciones Interétnicas [The Atacamenos and Their Interethnic Relations]' in Sergio Missana (ed), *Pueblos originarios y sociedad nacional en Chile: La interculturalidad en las prácticas sociales* (PNUD 2012).

Going back to the double helix, ALMA's extractive operations identified by Lickan Antay activists could be considered a 'mainstream' form of extraction. But in addition to this, my research also reproduced dynamics of extraction that could be considered a 'critical' one. In hindsight, I have identified at least three extractivist patterns at different stages of my work.

This extraction pertained to both the relationship I established with Lickan Antay Indigenous activists and my use of decolonial theory. In relation to the latter, I employed the modernity/coloniality framework proposed by a group of scholars who, inspired by Peruvian sociologist Aníbal Quijano, relied on knowledges and experiences of resistance in Latin America to come up with a macro-historical vision of the current world system. The modernity/coloniality group, in turn, has been accused of appropriating Indigenous[37] and Chicana[38] knowledges, which reveals that the extraction patterns I reproduced are part of a broader compound of appropriation facilitated by global academia.

The first of these patterns concerns the very research design I envisioned while preparing my fieldwork. Although I was relying on Latin American decolonial thinking, which to a large extent draws on Indigenous knowledges,[39] I initially considered the interviews I would conduct with Lickan Antay activists as mere background information for my study rather than a relevant source of thought on data governance, the subject matter of my study. This approach is not a coincidence. Modernity has presented Indigenous knowledges as opposed to scientific ones and approached these knowledges as potentially relevant for understanding the past but not the future.[40] However, after talking to Lickan Antay activists I did identify relevant points of discussion related to data governance. In particular, I ended up dedicating an entire chapter of my thesis to unpacking the Lickan Antay vision of territory and its relevance for the design of data infrastructure.[41] In this way, and just as large-scale mining and other extractivist actors, I only considered the effects of my object of study (astronomy data) on Indigenous communities and the environment at a second stage. Furthermore, it did not occur to me to examine how local forms of resistance and knowledge could provide an alternative vision to the one held by the observatories and the Chilean state.

The second extractivist pattern I reproduced in my research, which I mentioned in the introduction, concerns the contrast between my and the Lickan Antay activists' priorities and concerns. During my fieldwork, Lickan Antay activists conceded me their valuable time and agreed to share their views and experience on the astronomical observatories. However, they made it clear to me that neither astronomy nor astronomy data were among their main concerns. At the time we talked, they were channelling their energies towards another, and potentially more harmful, technology-related issue. Since around 2015 the Atacama Desert has been facing a boom in the extraction of lithium, a mineral used in the development of rechargeable batteries. While for the Global North lithium is employed to build 'green technologies' such as electric cars, the Lickan Antay communities inhabiting the Atacama Desert have seen rivers and meadows go drier

and Algarrobo trees and flamingo colonies decrease.[42] This is because lithium is extracted in Chile through a water-intensive process despite the Atacama Desert being one of the driest areas in the world. As of today, 45% of lithium mining projects worldwide are located in Chile, Argentina and Bolivia.

Considering this worrying situation, the more I talked to Lickan Antay activists the more I wondered whether I was merely chasing astronomy data due to the interest of Global North academia in 'all things data' without having considered looking at other technological developments, such as rechargeable batteries, that were wreaking havoc in the setting I was studying. Just as the 'only take' approach criticized by Simpson and Lickan Antay activists, my research design did not consider how my focus on astronomy data could represent a distraction from more profound technology-related issues. It could be said that my priorities and concerns as a researcher spoke to the agenda of the Global North academia rather than to the priorities and concerns of the local communities in the Atacama Desert.

The third and final extractivist pattern concerns the unequal capacity of the different actors involved to take advantage of my research for advancing their goals. Like other doctoral dissertations written at the LSE, it is possible to find mine on the 'LSE Theses Online' repository.[43] Some of the available theses look at case studies in places such as China, the United States, South Sudan and Brazil. It is certainly great to see that academia is interested in studying such a varied range of geographies and contexts, and that in some cases this is accompanied by frameworks that challenge Eurocentric epistemologies. However, the diversity of the research conducted by PhD students can also be used in order to reinforce the LSE's, and at the same time the United Kingdom's and the English language's, dominant position in world academia by collecting 'raw' data from distant locations and 'processing' them at home. This strategic form of approaching research can be explicitly mobilized in marketing and branding material. For example, in 2019 the LSE launched its 2030 strategy, which has as an overarching slogan 'Shape the world'.[44] When I saw this slogan, I started to understand with more clarity how the research conducted by PhD researchers could be used by the LSE to reach communities and places all over the world, including Lickan Antay people and the Atacama Desert, for the benefit of the LSE itself.

In contrast, it is not certain whether and how the Lickan Antay people would be able to draw on my research in order to advance their causes. The Council of Atacameno Peoples does have an archive of publications and outputs stemming from academics and researchers. However, issues such as the English language and the use of academic jargon, in addition to the fact that astronomy data does not constitute their main concern, makes it difficult, if not completely impossible, for the local communities to engage in dialogue with research produced on the basis of their experiences and knowledges. Certainly, there are still things I could do to improve this situation; in fact, in 2022 I had the chance to go back to San Pedro and meet with the local communities.[45] Still, Western academia does not consider

37   Rivera Cusicanqui (n 7).
38   Mariana Ortega, 'Woes and Practices of Un-Knowing' (2017) 31 The Journal of Speculative Philosophy 504.
39   Arturo Escobar, *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds* (Duke University Press 2018).
40   Escobar (n 40).
41   Sebastián Lehuedé (n 8).

42   Tapia and Peña (n 10).
43   Sebastián Lehuedé, 'Governing Data in Modernity/Coloniality: Astronomy Data in the Atacama Desert and the Struggle for Collective Autonomy (PhD thesis, LSE 2021) <http://etheses.lse.ac.uk/4321/ accessed 1 August 2023.
44   Minouche Shafik, 'LSE 2030' <https://www.lse.ac.uk/2030 accessed 10 June 2022.
45   For consent reasons I am not able to reproduce these conversations in this article.

engagement with local communities a relevant criterion of research excellence, and instead tends to render it a secondary aspect.

The three extractivist patterns I identified in my research pertain to different research stages: research design (initial omission of the Lickan Antay people), fieldwork (different priorities) and outputs (who can take advantage of research). Still, they all speak to the type of appropriation outlined by Simpson and Rivera Cusicanqui, where Indigenous knowledges are either ignored or transformed into research data without a proper and serious acknowledgment of the needs, visions and struggles of the local context. These extractivist patterns were not necessarily helpful for developing a non-extractive relationship between the Lickan Antay people and ALMA. Also, these patterns of extraction relied on the local communities' ancestral knowledges and experience of struggle in a way that could help reinforce existing asymmetries — with me and my academic institutional scaffolding being able to take advantage of this exchange — whereas the same exchange would not allow Lickan Antay communities to address their main preoccupations. As the metaphor of the double-helix suggests, both the 'mainstream' and 'critical' forms of data extraction intertwined in the Atacama Desert in order to maintain existing power imbalances.

The three extractivist patterns I discussed in this section shine a light on relevant aspects to be considered when conducting critical data research. First, they imply that it is crucial to map all the relevant actors involved in or affected by the production and management of data. A particular emphasis should be put on how research can challenge whose voices are deemed legitimate sources in matters related to data. Second, critical data research should encompass a critical assessment of the extent to which researching data could represent the interests and views of powerful actors, such as academic institutions, rather than the participants'. In some cases a focus on data can obscure issues that are more relevant and urgent to the eyes of the communities at stake. Finally, critical data scholars should be aware of how the different parties would be able to take advantage of the outputs of the research. This can be especially critical when research involves actors from the Global North and the South.

Before moving to the next section, I would like to highlight that the above points do not reflect the whole range of relationships I established with the Lickan Antay activists. In fact, the extractivist patterns I just described coexisted with some degree of complicity. I am borrowing the term 'complicity' from anthropologist George E. Marcus, for whom fieldwork should involve questioning how power operates in a given setting rather than assuming pre-constructed understandings of the position of the actors involved. For Marcus, this stance can enable a complicity between researcher and participants, understanding complicity as 'having a sense of being *here* where major transformations are under way that are tied to things happening simultaneously *elsewhere*, but not having a certainty or authoritative representation of what those connections are'.[46] The dialogue I sustained with Lickan Antay activists generated complicity, inasmuch as it involved figuring out what was going on with the astronomy data in the Atacama Desert (here) and how it could connect with broader changes taking place in the world economy (elsewhere). Our exchange revealed actors and relationships previ-

ously unknown to the local communities, such as the Data Observatory public-private partnership involving Amazon Web Services, and in turn I was able to comprehend how data extraction also relies on a notion of territory as a flat surface offering easily extractable information. This form of collaboration, of mutual figuring out, does not make up for the extractivist patterns reproduced in my research, but does provide a more holistic picture of the politics involved in my research on astronomy data in Chile.

The extractivist patterns I identified represent relevant concerns pertaining to the design, conduction and dissemination of research. However, they by no means cover all the power dynamics that can take place when doing critical data studies. In the next section I propose that the main takeaway from my experience concerns the need to embracing self-scrutiny in the entire research cycle, a practice that anthropologists, feminist thinkers and social scientists more broadly call *reflexivity*.

4.   Radicalizing Reflexivity for a Properly 'Critical' Data Studies

In the previous section I critically examined how my practices and positionality, as well as the institutional scaffolding supporting my research, could have reproduced the same extractive patterns I was seeking to identify and denounce. This form of self-scrutiny, known in the social sciences as *reflexivity*, is crucial when it comes to formulating critiques of data extraction. So far, reflexivity has not been central in discussions about data extraction nor in the field of CDS as a whole, making critical data research impervious to the double helix of data extraction. Against this backdrop, in this section I call for radicalizing reflexivity in CDS, by which I mean (1) granting reflexivity a central place in critical reflections about the role of data and society, and (2) doing so in a way that takes into consideration power dynamics involving both positionality and political economy.

Almost a decade ago, the increasing significance of data in social life gave rise to a new field known as CDS. A foundational article for this field is boyd and Crawford's *Critical Questions for Big Data*. As terms such as *big data* gained currency, these authors argued that 'it is still necessary to ask critical questions about what all this data means, who gets access to what data, how data analysis is deployed, and to what ends'.[47] Along these lines, different authors advocated the creation of a common space of critique in order to bring together discussions stemming from different disciplines, as well as to collectively situate the emergence of this phenomenon in history and in relation to broader sociotechnical associations.[48] Ultimately, CDS came to represent 'a formal attempt at naming the types of research that investigate all forms of potentially depoliticized data science and to track the ways in which data are generated, curated, and how they permeate and exert power on all manner of forms of life'.[49] Works looking at data extraction from decolonial,[50] feminist[51] and other justice-oriented frameworks I mentioned in the first section, as well as my research on astronomy data in Chile, fit with CDS' aim.

46   Geroge E Marcus, 'The Uses of Complicity in the Changing Mise-En-Scène of Anthropological Fieldwork' [1997] Representations 85, 96 emphases in the original.

47   danah boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15 Information, Communication & Society 662, 664.

48   Jim Dalton and Jim Thatcher, 'What Does a Critical Data Studies Look like, and Why Do We Care? Seven Points for a Critical Approach to "Big Data"' [2014] Society and Space 1; Rob Kitchin and Tracey P Lauriault, 'Towards Critical Data Studies : Charting and Unpacking Data Assemblages and Their Work' [2014] Geoweb and Big Data 1.

49   Iliadis and Russo (n 3) 2.

50   Couldry and Mejias (n 4); Ricaurte (n 19).

51   D'Ignazio and Klein (n 6).

CDS has been identified as 'open to self-critique and dialogue',[52] but existing proposals have not elaborated in depth how reflexivity could help interrogate the data practices of CDS research itself. The etymology of 'data' refers to what is given, a form of transparent evidence that does not require further discussion.[53] Challenging this idea, one of the main streams of critique advanced by CDS has focused on revealing the socially constructed nature of data and the power dynamics that come to bear in this process. From the very definition of what comes to count as 'data'[54] to the broader role of data in colonialism and capitalism,[55] relevant research has made it increasingly difficult to assume data as synonymous with objectivity and neutrality. However, and despite few exceptions,[56] CDS in general has not applied that critique to its own data practices, keeping this nascent field from asking uncomfortable questions such as how CDS itself constructs the theoretical and empirical data it uses to advance its arguments. While a study showed that CDS might be igniting reflexivity among data scientists,[57] little has been discussed on the need for reflexivity among critical data scholars themselves.

Reflexivity has a long tradition in the social sciences, emerging with force in disciplines and fields such as anthropology[58] and feminist studies.[59] There are different definitions of reflexivity, and even more ways of conducting reflexivity in practice.[60] For some, reflexivity constitutes a tool that can help develop less 'biased', and therefore more 'truthful', research.[61] However, a definition more attuned to CDS's epistemological standpoint would understand reflexivity as an inquiry into

> how one is inserted in grids of power relations and how that influences methods, interpretations, and knowledge production ... [and] how one relates to research participants and what can/cannot be done vis-á-vis the research within the context of institutional, social and political realities'.[62]

What is required, thus, is that CDS grant a central role to reflexivity, i.e., a *radical* embrace of reflexivity. In fact, some authors have argued that one of the aspects that distinguishes 'critical' from traditional theory is precisely the former's departure from the idea that there is such thing as neutral social research, making it necessary to interrogate the social factors shaping knowledge generation and the way this knowledge interacts with society.[63] Because of this, a properly *critical* data studies would require us to pay attention to the positionality of

CDS researchers, the interests underlying the institutions supporting their research and their relationship with theoretical and empirical data. CDS scholars have shown a great capacity to identify the usually hidden politics of data, and the current context requires a turn towards reflecting on their own practices.

My call for radicalizing reflexivity not only involves granting reflexivity a central position within CDS but also conducting a particular type of reflexivity that acknowledges the role of both positionality and what Silvia Rivera Cusicanqui calls the 'political economy of knowledge'.[64] One of the most common ways of conducting reflexivity in the social sciences has been through a focus on how the researcher's positionality in the social world—including aspects such as gender, race and class—could have informed the research process. As Wendy Pillow argues, '[o]ne of the most noticeable trends to come out of a use of reflexivity is increased attention to researcher subjectivity in the research process – a focus on how does who I am, who I have been, who I think I am, and how I feel affect data collection and analysis'.[65] While some authors have criticized the transformation of positionality into either 'apologies' or 'badges',[66] this form of reflexivity can still be deployed in a way that generates productive discomfort as long as it rejects pre-constructed identities and allows for more creative forms of granting voice to research participants.[67]

Representation is certainly crucial when it comes to critical data studies, but it would not suffice for unearthing the power dynamics involved in the double helix of data extraction. Following Silvia Rivera Cusicanqui, an excessive focus on representation runs the risk of over-emphasizing discourse at the expense of a scrutiny of the gap between discourse and practice, i.e., between what is being said and what is being done. For Rivera Cusicanqui, research can underpin truly generative and reflective texts but nonetheless still comprise problematic extractive practices. To address this issue, she advocates for a 'political economy of knowledge' approach that brings to the fore 'the economic strategies and material mechanisms that operate behind discourses'.[68]

Rivera Cusicanqui's concern revolves around how academics based in the North can obtain individual benefits from the alleged support to decolonization, generating 'an economy of salaries, perks, and privileges'[69] through the appropriation of ideas stemming from Indigenous people and intellectuals in the South. Rivera Cusicanqui uses the example of Félix Patzi Paco, an author who discusses communal systems by referring to 'decolonial' established academics such as Walter Mignolo but ignoring the origins of this idea in the Indianist and Katarist struggles in the sixties and eighties in Bolivia.[70] In such cases, including an attentiveness to the political economy of knowledge makes it possible to bring to the surface the extractive dynamics of 'critical' research by delving into how decolonial, feminist or justice-oriented critical data research, which might look transgressive at a first glance, can still reinforce already existing asymmetries.

52 Iliadis and Russo (n 3) 2.
53 Rosenberg (n 14).
54 Lisa Gitelman (ed), *'Raw Data' Is an Oxymoron* (MIT Press 2013).
55 Couldry and Mejias (n 4).
56 e.g., D'Ignazio and Klein (n 6) 217.
57 Gina Neff and others, 'Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science' (2017) 5 Big Data 85.
58 James Clifford and George E Marcus (eds), *Writing Culture: The Poetics and Politics of Ethnography* (University of California Press 1986).
59 Mary Margaret Fonow and Judith A Cook, 'Back to the Future: A Look at the Second Wave of Feminist Epistemology and Methodology' in MM Fonow and J Cook (eds), *Beyond Methodology: Feminist Scholarship as Lived Research* (Indiana University Press 1991).
60 Pillow (n 11).
61 Farhana Sultana, 'Reflexivity, Positionality and Participatory Ethics: Negotiating Fieldwork Dilemmas in International Research' (2007) 6 Acme 374.
62 Sultana (n 62) 376.
63 Thomas McCarthy, 'On the Idea of a Critical Theory and Its Relation to Philosophy' in D Couzens Hoy and T McCarthy (eds), *Critical Theory* (Blackwell 1994).
64 Rivera Cusicanqui (n 7) 102.
65 Pillow (n 11) 176.
66 Daphne Patai, 'U.S. Academics and Third-World Women: Is Ethical Research Possible?' in SB Gluck and D Patai (eds), *Women's Words: The Feminist Practice of Oral History* (Routledge and Kegan Paul 1991) 36.
67 Pillow (n 11).
68 Rivera Cusicanqui (n 7) 102.
69 Rivera Cusicanqui (n 7) 102–103.
70 *Una Alternativa al Sistema Liberal [The Communal System: An Alternative to the Liberal System]* (CEEA 2004).

Based on the previous paragraphs, by *radicalizing* reflexivity in CDS I mean both a radical embrace of reflexivity in research and an attentiveness to both representation and political economy. Still, it is important to clarify that I am not advocating for radical reflexivity as a 'solution' to the double helix of data extraction. There are serious reasons why reflexivity would not suffice to overcome the double helix of data extraction. Behind the idea of reflexivity as a 'solution' lies the fantasy that problems can be solved by merely exposing them. Instead, overcoming the double helix of data extraction would require not only *speaking* about certain things but undertaking a profound transformation of power structures involving academia, especially North-South gaps. Furthermore, as a narrow notion of reflexivity becomes routine practice, it is increasingly difficult to carry out a truly original and honest self-scrutiny of the researcher's positionality and the political economy involving their work.

Despite the above, and as this article has shown, radical reflexivity can still facilitate two relevant and necessary moves in CDS. The first obvious move is to expose what has so far remained unexamined, making it possible to provide a more complex and nuanced account of how data extraction works, the actors at stake and the role of critical research in sustaining it. Such an exposition can certainly generate discomfort, since allegedly critical work might end up sharing some of the practices of mainstream big data and data science research. A second move facilitated by radical reflexivity would be to generate awareness about the limitations of research on its own for achieving social justice. If critical research is shown to participate in some way in the reinforcement of injustices, then a more humble and grounded type of work would emerge.

Reflexivity should not be understood as synonymous with lack of action, either. In this regard, PAR constitutes one of the most assertive ways of proactively addressing the power dynamics involved in critical data research. Instead of extractive relations, participatory action-research (PAR) values the knowledge of communities and people by fostering horizontal collaborations.[71] In PAR, participants themselves define the research questions, methods and outcomes of research. This approach also involves the search of justice as it grants agency to oppressed groups who have been deemed as mere 'objects' of research.[72] As its name suggests, PAR undertakes *action* in order to advance social change; however, this is not an excuse for avoiding reflexivity. In fact, actions are conducted in combination with cycles of reflexivity to scrutinize power dynamics taking place within the research process and examine their connections with broader structures.[73]

Some studies have shown the productivity of participatory methods for thinking critically about data without reproducing extractive patterns. For example, Ramesh Srinivasan[74] undertook collaborative ethnographic research with Native American tribes dispersed across San Diego County in Southern California in the development of access networks and digital environments. Srinivasan's stories are rich in descriptions of how he sought to incorporate the concerns and priorities of the groups he collaborated with, as well as with a self-scrutiny of his own anxieties and biases. Aspects such as the respect for local leaderships and a different understanding of information sharing emerged as valuable take-aways of the processes. Along similar lines, PAR allowed Andrea Medrado and Pieter Verdegem to unearth relevant questions regarding autonomy, empathy and dialogue that speak to the daily life of marginalized groups but that have not been prominent in discussions regarding AI systems. Importantly, and in line with this article, they note that PAR cannot be undertaken without radical reflexivity as academic institutional constraints in the Global North makes it difficult to approach participants with a truly open agenda.

However, PAR can also constraint reflexivity when tied to an inclination to 'action'. More broadly, an impatience with reflexivity has become widespread in academia. In a conference where I presented these ideas, a colleague commented that reflexivity was necessary and urgent in order to 'move forward' quickly and enable even more radical critiques of data extraction. In my view, though, radical reflexivity might invite CDS researchers to reconsider the practices underpinning their work, adopt a more humble attitude and encompass a form of retracting, reassembling and moving backwards. Furthermore, while PAR takes the community as its starting point (with whom the research design is developed), reflexivity might indicate that the groups we deem the powerful and the oppressed, or even what groups are relevant in the first place, might not be totally clear from the outset and might well change during the research process. Research on citizen science[75] and geographic crowdsourcing initiatives[76] have shown the need for reflexivity, since data participatory initiatives are not absent of power dynamics such as co-optation by powerful actors.

In 2017, when I undertook my research on astronomy data in Chile, CDS was a relatively new field, which can explain why reflexivity was not a central issue discussed by researchers unearthing the previously obscure politics of data. Some trends in the field, however, are calling for a radical embrace of reflexivity. Two of such trends are the embrace of critical and justice-oriented theoretical sources and an increasing attentiveness to dissenting communities.

CDS' interest in decolonial frameworks and an attentiveness to North-South dynamics makes it the more urgent to approach reflexivity as not only discussions about the positionality of the researcher but also broader patterns of intellectual extraction. Because of this, the radical type of reflexivity I am advocating require us to pay attention to Rivera Cusicanqui's political economy of knowledge to explore how historical and global-scale dynamics can generate new forms of extraction even when research is conducted under the banner of decoloniality. Only such a bold embrace and multidimensional approach to reflexivity can help address the complex dynamics involved in the double helix of data extraction and give way to properly 'critical' research on data.

71   Sean A Kidd and Michael J Kral, 'Practicing Participatory Action Research' (2005) 52 187.
72   Orlando Fals-Borda and Muhammad Anisur Rahman (eds), *Action and Knowledge: Breaking the Monopoly with Participatory Action-Research* (The Apex Press 1991).
73   Carri Hand and others, 'Initiating Participatory Action Research with Older Adults: Lessons Learned through Reflexivity*' (2019) 1 512.
74   *Whose Global Village? Rethinking How Technology Shapes Our World* (NYU Press 2017).
75   Jennifer Gabrys, 'Programming Environments : Environmentality and Citizen Sensing in the Smart City' (2014) Environment and Planning D: Society and Space 32 30.
76   Tracey P Lauriault and Peter Mooney, 'Crowdsourcing: A Geographic Approach to Public Engagement Keywords' [2014] The Programmable City 1.

5.   Conclusion

In this article I have outlined the double helix of data extraction, a pattern in which both mainstream and critical forms of data extraction become complicit in affecting dissenting groups. On the one hand, mainstream data extraction is conducted by social media platforms, scientific organizations, states and other actors for the sake of profits, the accumulation of intellectual capital or the exercise of social control. On the other hand, critical data extraction takes place when researchers turn to dissenting empirical and theoretical sources to denounce mainstream data extraction without an attentiveness to the struggles, needs and visions of the context where such critical sources emerged. The double helix of data extraction is becoming increasingly worrisome as the critique of data extraction becomes mainstream and CDS scholars adopt decolonial, feminist and other justice-oriented frameworks.

In this article I looked at how the double helix of data extraction might have played out in my research on astronomy data in Chile. In particular, I identified three extractivist patterns that I reproduced in my exchange with Lickan Antay communities affected by astronomy data extraction: (1) a disregard for the struggles and knowledges held by local communities affected by data extraction, (2) a disconnect between the concerns and priorities of CDS research with those of the dissenting groups it studies, and (3) the different capacity of academic institutions and research subjects to appropriate and take advantage of CDS research. These three patterns coexisted with dynamics of complicity, in which the Lickan Antay activists and I as researcher collaborated on figuring out what was going on, but they nonetheless attest to the presence of the same dynamics I was seeking to denounce in my research project.

Finally, I proposed that the rise of the double helix of data extraction relates to a lack of reflexivity within CDS. While CDS research has succeeded in bringing to the fore some of the previously obscure politics of data, this field has not been able to do the same in relation to its own data practices. Against this backdrop, radicalizing reflexivity can expose the way CDS itself can reproduce extractive dynamics and provide a more grounded understanding of the possibilities of research on its own for advancing social justice. I employed the term 'radical' to call for providing reflexivity a central role in CDS and privileging a type of reflexivity that focuses on both positionality and political economy. As Silvia Rivera Cusicanqui has shown, an attentiveness to the flow of material and intellectual capital is especially relevant in the case of research adopting a decolonial lens and involving North/South power dynamics. Participatory methods might offer a way of addressing such power dynamics, but their inclination to action should be critically assessed so as to not stifle reflexive processes.

As some scholars have noted, radical reflexivity would not suffice for overcoming the double helix. For Daphne Patai, reflexivity is prone to become a matter of privileged academics concerned with issues of representation and language, and might not even deliver 'better research.'[77] In addition to this, the lack of action that some attribute to reflexivity could be tackled by adopting participatory approaches in which research subjects have more capacity to shape the questions, methods and outputs of research. While I agree with these two points, I would add that reflexivity constitutes a necessary step for becoming aware of how seemingly critical research can reproduce mainstream extraction patterns. Without reflexivity, radical critiques

and horizontal methodologies run the risk of shooting themselves in the foot by skipping the relevant question that would allow CDS to challenge, rather than reproduce, existing asymmetries.

The insights offered in this article can also provide an orientation for organizations and groups seeking to advance 'ethical' or 'humane' approaches to data and data-intensive developments such as AI. In these cases, reflexivity should not only be tolerated but proactively promoted by education entities, private companies and the state. When it comes to education entities, ethical assessments tend to privilege normative principles or checkboxes instead of the facilitation of instances of deep and open self-scrutiny. Some big technology companies have adopted ethical guidelines and frameworks; however, such initiatives have shown intolerance to reflexivity. Employees drawing attention to the structural problems of these companies run the risk of being removed from their jobs — the case of Timnit Gebru being one of the most cited ones.[78] Against this backdrop, a big challenge for institutions consists of promoting environments where it is possible to question the institutional scaffolding underlying the production and management of data.

77   Patai (n 67) 69.

78   Nitasha Tiku, 'Google Hired Timnit Gebru to Be an Outspoken Critic of Unethical AI. Then She Was Fired for It.' (2020) <https://www.washingtonpost.com/technology/2020/12/23/google-timnit-gebru-ai-ethics/> accessed 12 May 2021.