Big Data from the South(s): An Analytical Matrix to Investigate Data at the Margins

Milan, S.; Treré, E.

DOI
10.1093/oxfordhb/9780197510636.013.7

Publication date
2022

Document Version
Author accepted manuscript

Published in
The Oxford Handbook of Digital Media Sociology

Citation for published version (APA):
Big Data from the South(s). An Analytical Matrix to Investigate Data at the Margins
Stefania Milan (University of Amsterdam) and Emiliano Treré (Cardiff University)

Datafication—through which many aspects of social life are transformed into data—is usually equated with a more efficient use of resources and improved state–citizen relations. But it can have negative consequences on those at the margins of society, such as refugees, racialized individuals, gig workers, and citizens of countries with scant respect for human rights. How can we understand the ways in which the disempowered are impacted by and resist datafication? This essay presents an analytical matrix to study data at the margins. The matrix identifies three components of data at the margins: (1) infrastructure, emphasizing the material dimension; (2) practices, pinpointing agency in people’s encounters with datafication; and (3) imaginaries, that is, the cultural and symbolic facets of data at the margins. The matrix offers also three lenses of interpretation through which to observe the components: (1) decoloniality and race, (2) intersectionality and feminism, and (3) the “pluriverse.” Together, they help in questioning datafication and why certain social groups are oppressed while uncovering pathways toward justice and equality.

Keywords: datafication, artificial intelligence, data infrastructure, data imaginary, data practice, political agency, decoloniality, intersectionality, feminism, pluriverse

The computational turn witnessed since the 1960s has spectacularly accelerated since the 2010s, thanks to the increase in computing power and the advances in automation and artificial intelligence (AI). Through this omnipresent digital infrastructure, more and more aspects of human existence and social life are transmuted into data points, in a process known as “datafication.” Cities become “smart” (Gupta, Panagiotopoulos, and Bowen 2020), service work moves to the “platform” (van Doorn 2017), and citizens are increasingly “datafied” (Hintz, Dencik, and Wahl-Jorgensen 2018). These developments are typically associated with positive outcomes and accompanied by flamboyant narratives of empowerment. They yield the promise of flexible income generation, media content targeted to personal preferences, and increased efficiency in the use of public resources. However, there exists an “asymmetric relationship between those who collect, store, and mine large quantities of data, and those whom data collection targets” (Andrejevic 2014, 1673). This data trade, often invasive of people’s privacy, is at the core of “surveillance capitalism,” an exploitative economic system that benefits a handful of tech corporations at the expense of the majority of users (Zuboff 2019).
While these exploitative dynamics affect societies across the globe regardless of geography, census, or education, communities in the fringes of the neoliberal system appear to suffer the worst consequences—in the so-called Global South as much as in the north of the hemisphere. Migrants and refugees, racialized groups, indigenous peoples, subjects with disabilities, children and minors, women and non-gender-conforming individuals, and impoverished households are increasingly at risk of discrimination. Take the example of data-driven policymaking, whereby public policies are based on data automatically generated by dashboards and devices. Ethnic prejudices are reproduced in racially biased decision-making software adopted by judiciary systems worldwide, targeting especially minority communities (Gangadharan and Jędrzej 2018). Also, citizens of countries with weak rule of law, under authoritarian governments, or simply with poor sovereignty over digital infrastructure are exposed to data exploitation, which might result in human rights violations and the (re-)production of inequality. Think of the pervasive data infrastructure run by the state for the management of digital identity like Aadhaar in India, the world’s largest biometric identity system (Masiero 2020). “Citizen scoring” schemes such as the Social Credit System in China (Dencik et al. 2018) are another case in point: algorithms amenable to false positives are used to segment population groups according to their “risk profile.” Both systems might eventually generate social exclusion. Yet, individuals and communities at the margins are not merely paying the social costs of datafication. On the contrary, they increasingly explore the potential of “counter data action” (Currie et al. 2016) and “data resistance” (Vera et al. 2018) to defy and subvert the pervasive datafication of social life.

How can we understand the multiple ways in which the disempowered are impacted by and seek to confront data exploitation? To date, research tends to rely on a limited set of Western conceptual tools to make sense of the specificities of these novel types of social vulnerability and resistance. Frameworks, epistemologies, and ontologies used to decode what large-scale data collection does to people at the margins or to determine what data infrastructure—impoverished communities might need to “develop” (Taylor and Broeders 2015) emerge almost entirely from “a world economy of knowledge structured by the history of colonialism and current north–south global inequalities” (Connell 2014, 210). They reproduce the “digital sublime” of Silicon Valley narratives, meaning the allure exerted by digital technologies that often obfuscates their drawbacks (Mosco 2004). This fascination is often uncritically extended to include also the dispossessed, ignoring the relevance of situated knowledge and contextual differences. In other words, we tend to extend theories and categorizations developed in a handful of centers of epistemic power in liberal democracies and wealthy countries to make sense of the datafied society in the fringes of the neoliberal system. We thus risk “universalizing” both the interpretation of the problem and the proposed solutions (Milan and Treré 2019), instead of trying to “decolonize” our approach (Ali 2016). In addition, most scholarship on these subject matters circulates merely in English, which acts as lingua franca for critical thinking on these matters (Suzina 2020). Significant interventions from academia and nonprofits within Latin America (e.g., Castro-Gómez...
and Grosfoguel 2007; Herrera Huérfano, Sierra Caballero, and Del Valle Rojas 2016), Africa (e.g., Cheruiyot and Ferrer-Conill 2018; Rodrigues et al. 2018), or Asia (e.g., Thorat 2021) circulate poorly or pay the price of linguistic diversity.

This chapter argues that we need to critically interrogate and redefine our conceptual toolbox if we are to understand non-mainstream data vulnerabilities and practices. To explore this claim, the essay promotes an interdisciplinary dialogue between critical data studies, sociology, science and technology studies, and decolonial thinking. The result is an analytical matrix which allows us to approach non-mainstream engagement with data from a sociological standpoint. More specifically, the matrix takes into consideration three key aspects of the datafied society: firstly, data infrastructure, meaning the structural dimension of datafication (e.g., the technical and governance arrangements imposed on data subjects, their governmentality consequences, and any potential self-organized alternatives); secondly, data practices, which allow us to understand how agency unfolds when people relate to datafication (e.g., how people and communities incorporate data in their action repertoires); and thirdly, data imaginaries, which stand in for the cultural and symbolic dimension of data at the margins (e.g., meaning-making processes as they are mediated by data and data infrastructure). Furthermore, the matrix mobilizes three lenses of interpretation which may help in overcoming the blind spots in current “Western” analyses of the datafied society, namely decoloniality, or the decolonial “thinking and doing” (Mignolo 2011, xxiv); race, intersectionality, and feminism; and the “pluriverse,” which stands in for the myriad of “alternative” epistemologies that emerge outside the mainstream (Escobar 2018).

The chapter is organized as follows. Firstly, we review useful literature to approach and further define the margins, also evoking the concept of a “plural” south. Secondly, we present our three-pronged approach which accounts for three fundamental aspects of the datafied society, namely infrastructure, practices, and imaginaries. Finally, we sketch three valuable lenses of interpretations to read data at the margins, questioning power dynamics and making room for distinct epistemologies.

**Thinking from the Margins: A Critical Literature Review**

Only 53% of the world population is connected to the internet today (International Telecommunication Union 2019). But in our increasingly datafied society, reducing the problem to the dichotomy between the “big data rich” and the “have nots” (boyd and Crawford 2012) would be a mistake. It is also more than simply lacking access to digital infrastructure, as the classical literature on digital divide seems to suggest (Van Dijk 2020)—for three reasons. Firstly, policymakers increasingly rely on “calculated publics,” that is to say publics evoked by automatized quantification exercises (Crawford 2015), to make decisions and allocate public resources. As a result, people’s existence is more and more tied to data, as the COVID-19 global health crisis has made evident (Milan, Treré, and Masiero 2021). Being “datafied” during a pandemic (e.g., visible to the state, gaining access to social welfare and vaccines) has turned into a necessary condition for
survival and care (Milan and Treré 2020). For example, undocumented migrants in many EU countries are invisible to government data sets, which prevents them from accessing healthcare or unemployment benefits (Pelizza, Milan, and Lausberg forthcoming). While this new type of “data poverty” (Milan and Treré 2020) has to do with fundamental forms of inequality that predate the datafied society, it is aggravated by the limited citizen agency in the complex ecosystem of data extractivism and commodification (Sadowski 2019). Secondly, many individuals and communities lack the data literacy and sociocultural capital to situate their voices in the public sphere. Thus, they struggle to engage with the challenges of an increasingly datafied policymaking process where it is often impossible to seek redress. Finally, technology innovation itself may paradoxically contribute to worsen the problem. To name just one example, the emergence of digital technology breaking the boundaries between the networking infrastructure and the application data, such as the celebrated fifth generation of cellular networks (5G), risks limiting the possibilities of end users to run and/or use the infrastructure as they want, deepening the gulf between the “haves” and the “have nots.”

Scholars of disciplines as diverse as media studies, law, sociology, and informatics are increasingly devoting their attention to the implications of datafication from a critical perspective. They have denounced the novel forms of exploitation of those “at the bottom of the data pyramid” (Arora 2016) and exposed how these mechanisms contribute to reproduce colonial power relations (Thatcher, O’Sullivan, and Mahmoudi 2016). They have brought under the spotlight a variety of bottom-up data justice projects which emerged in the Global South as well as in the plurality of “Souths”—in other words, pockets of marginality—that survive in the fringes of Western democracies. These grassroots projects expand the space of possibilities of citizen action (Sun and Yan 2020), promote social justice through data (Heeks and Renken 2018), reclaim citizenship through transparency activism (Torres 2019), and seek ways of decolonizing data and technology (Awori et al. 2016). Instead of replicating the mainstream, they generate novel data epistemologies and alternative ways of participating in the datafied society (Milan and van der Velden 2016). Regrettably, in the literature these disparate case studies seldom result in wide-ranging theory development. They remain siloed in distinct disciplines which rarely converse with each other, with sociologists often overlooking the social affordances of technology and media scholars disregarding fundamental questions of power.

This chapter offers an analytical grid that combines insights from four disciplines—sociology, science and technology studies, critical data studies, and postcolonial and decolonial studies. An interdisciplinary approach is required because investigating the impact of datafication on people at the margins means taking a deep dive into complex processes at the intersection of several dimensions, including the infrastructural, cultural, political, and legal. Sociology, especially political sociology and social movement studies, allows us to center human agency with respect to data, infrastructure, and software (Couldry 2014). Science and technology studies reminds us that data infrastructure is not merely the outcome of decisions of technical nature but
embodies the values and preferences of its designers, producers, and shareholders (Winner 1999). Critical data studies, at the crossroads of the humanities, social sciences, and informatics, brings under the spotlight the potential exclusion, discrimination, and unfairness embedded in the datafied society (Dalton, Taylor, and Thatcher 2016). Finally, scholarship on colonialism and decoloniality, especially its Latin American strand, invites us to question the narratives of technology as “imported magic” (Medina, Marques, and Holmes 2014) pushed onto the Global South. It forces us to consider the colonial rationalities (Quijano 2007) embedded also in the datafied society and to give voice to diversity and multiplicity (Mignolo 2000).

To situate these “non-mainstream” forms of engagement with data, in our earlier work we proposed taking the plurality of the South as a frame of reference. South is “not merely a geographical or geopolitical marker (as in ‘Global South’). Rather, it is a plural entity subsuming also the different, the underprivileged, the alternative, the resistant, the invisible, and the subversive” (Milan and Treré 2019, 321). This flexible and expansive definition of the South(s) identifies “a place of (and a proxy for) alterity, resistance, subversion, and creativity” (2019, 325, original italics). It empowers us to take into consideration inequality as “it transcends boundaries and known geographies” (2019, 321). However, to avoid the potential reductionism harbored by a spatial metaphor, here we summon a second, broader frame of reference: the margins. “[T]he margin,” claims Colombian citizen media scholar Clemencia Rodríguez, is “a shortcut to speak of complex dynamics of power inequality. Processes of asymmetrical access to material and symbolic resources shape differentiated and unequal access to the public sphere” (Rodriguez 2017, 49). The margins are understood as complex sites of struggle, where the challenges of datafication unfold in distinctive ways but also where particular data ecologies divergent from the mainstream emerge and thrive.

Using the media analogy introduced by Rodríguez, we can identify a first distinction between “data at the center” and “data at the margins,” whereby the latter questions technological and data universalism, or the tendency to gloss over diversity and impose Western epistemologies (Milan and Treré 2019). Nonetheless, current interpretation of forms of resistance and inequality associated with data tend to flatten the intrinsic multiplicity of data at the margins. They overgeneralize the features of people’s engagement with data and overlook key differences between distinct types of data practices, imaginaries, and infrastructure. Data at the margins are grounded in local time, social geography, values, and agendas that might be diametrically opposed to “global” trends and market imperatives. At the margins, data generally exist out of sight, rendered invisible by the glitter, hype, and excitement that characterize “data at the center.” Data at the margins hardly ever looks like one might expect. Instead, it is frequently used in unpredictable ways that differ from its originally intended purpose. It might even be absent because people lack the infrastructure or skills needed to produce, share, or make sense of it. Journeying into data at the margins is “stepping into the land of otherwise” (Rodriguez 2017, 49). Thus, how can we understand datafication as it unfolds at the margins of our increasingly interconnected world?
Three Building Blocks to Investigate the Land of Otherwise

To explore peripheral ways of engaging with data, we propose a three-pronged approach that allows for locating three fundamental aspects of the datafied society. These can be seen as the building blocks for a comprehensive analysis of the interplay between datafication and inequality—one that is able to knit together the research loci of the four disciplines inspiring our work. These building blocks can be studied both as a whole and in isolation. In this section we introduce each building block, reviewing useful concepts for its analysis and identifying potential areas of investigation. Table 1 provides an overview of the building blocks, offering illustrative examples as they relate specifically to data at the margins.

<table>
<thead>
<tr>
<th>Data infrastructure</th>
<th>Data infrastructures in the Global South (e.g., citizen scoring systems)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Creation of autonomous data infrastructure (e.g., cell networks, data sets)</td>
</tr>
<tr>
<td></td>
<td>Experiments with local data sovereignty</td>
</tr>
<tr>
<td>Data imaginaries</td>
<td>Fear and resignation associated with surveillance</td>
</tr>
<tr>
<td></td>
<td>Facial recognition technology as dehumanizing</td>
</tr>
<tr>
<td></td>
<td>#AbolishBigData2009, connecting datafication to abolitionism</td>
</tr>
<tr>
<td>Data practices</td>
<td>Data activism</td>
</tr>
<tr>
<td></td>
<td>Citizen sensing of environmental degradation</td>
</tr>
<tr>
<td></td>
<td>Rituals associated with everyday encounters with data (e.g., self-tracking)</td>
</tr>
</tbody>
</table>

Table 1. Building Blocks to Analyze Data at the Margins and Illustrative Examples

Data Infrastructure: Investigating Structure in the Datafied Society

The first building block concerns the material dimension of data at the margins. We can think of physical infrastructure from large (e.g., corporate data centers) to small (e.g., personal devices like smartphones) but also of the mechanisms generating, collecting, and processing data and the related governance arrangements. The material dimension of the datafied society is also made visible in a myriad of software-based “sociotechnical assemblages” (Akrich 1992) like social media platforms and their “infrastructures of tracking” (Helles, Lombok, and Lai 2020) or COVID-19 data sets and dashboards (Milan 2020). These data assemblages are “composed of many apparatuses and elements that are thoroughly entwined, and develop and mutate over time and space,” shaping “what is possible, desirable and expected of data” (Kitchin 2014, 24–25). Data infrastructure, rising out of and existing in a complex web of relationships, can be sociologically understood as the recurrent yet mutable structure of the datafied society determining the environment, choices, and opportunities available to social actors.

How can we understand infrastructure at the margins? Useful concepts include the popular notions of “surveillant assemblage,” indicating how individuals are profiled from information collected in a variety of digital places such as social media (Haggerty and Ericson 2000); “dataveillance,” or surveillance through data infrastructure (van Dijck
2014); and “governmentality” as applied to big data (Aradau and Blanke 2017). Alternative perspectives include the Latin American critical ecology approach (Barranquero and Baeza 2017), which connects resistance to the Western-led, one-way vision of development with sustainability and “environmental rationality” concerns (Leff 1994). The latter directions are particularly promising given the growing environmental footprint of the datafied society, and AI in particular (Dauvergne 2020).

In the field, we may focus our attention on the features of data infrastructure in the Global South, investigating, for instance, citizen scoring systems to regulate access to welfare like Sisbén in Colombia (López 2020) and the emergence of digital identity systems in the African continent (Schoemaker et al. 2020). One could analyze the discriminatory effects of the governance by data infrastructure by means of “immunity passports” (Voo, Clapham, and Tam 2020), the governmentality consequences of the securitization of migration (Bigo 2002), or local experiments in data ownership and technological sovereignty seeking to empower individuals to control their data as experimented in Barcelona, Spain (Lynch 2020). But we can also investigate the emergence of alternative data infrastructure, ranging from the self-organized cellular networks operated by indigenous communities in Oaxaca, Mexico (Baca-Feldman et al. 2019) to the citizen-led curation of a feminist index of domestic violence in Argentina (Chenou and Cepeda-Másmeila 2019) or a data set of human rights violations in the Syrian conflict (Deutch and Habal 2018).

**Data Imaginaries: Meaning-Making in the Datafied Society**

Social actors, both individually and collectively, seek to make sense of their datafied environment. To mobilize people, sense-making (i.e., the process of interpretation of the complex reality of datafication) must be paired with meaning attribution (i.e., the generation of shared insights and beliefs). In this process, the perception of some form of injustice plays a key role (Gamson 1992). Movement frames might also help in attributing mobilizing value to otherwise technical or expert issues (Milan 2013). Examples of narratives associated with datafication include empowerment (Baack 2015), unfairness (O’Neil 2017), and fear or resignation (Dencik 2018). Popular mobilizing frames include state resistance and data appropriation and transparency (Torres 2019) and seek to counteract the mainstream narratives of securitization and self-empowerment. Recent examples include characterizing period-tracking mobile applications as “unpaid work” that “must be considered in light of the historic lack of recognition for women’s sexual, reproductive and relational labor” (Felizi and Varon, n.d.) and biometric surveillance in public space as objectifying, commodifying, and dehumanizing people rather than increasing their safety (ReclaimYourFace 2020).

Suitable notions to capture how social actors engaging in resistant data practices make sense of datafication include “alternative epistemologies” as applied to datafication (Milan and van der Velden 2016), “social imaginaries” (Lehtiniemi and Ruckenstein 2018), and “counter imaginaries” (Kazansky and Milan 2021). Despite illuminating distinct ways of “feeling out” datafication, these notions similarly capture collective
visions that connect “intentions and projects as well as utopias and ideologies” (Flichy 2007, 4). Knitting together the “technological” and the “social,” they embody a normative dimension which often has mobilizing potential (Milan and ten Oever 2017). Venues to study emerging data imaginaries include initiatives to come to terms with surveillance (e.g., Duffy and Chan 2018), tools and platforms designed by activists to support other activists (e.g., Aouragh et al. 2015), and events like digital rights festivals and security training workshops (Daskal 2018).

Examples of alternative imaginaries of datafication at the margins include interventions like #AbolishBigData2009, promoted by grassroots organizations representing or working with racialized and minoritized communities in the United States. The initiative aims at changing the way big data and AI are conceived in society by connecting the mainstream discourse to abolitionism (Crooks 2019; see also Mohamed, Png, and Isaac 2020). Groups like Data 4 Black Lives, mobilizing data as “protest” and as “accountability,” contribute to change the perception of data itself by positing data tools as statistical modeling as “powerful instruments for fighting bias, building progressive movements, and promoting civic engagement” (Data for Black Lives n.d.). Indigenous perspectives, too, provide alternative points of entry able to promote the decolonization of mainstream approaches to data, arguing for indigenous data sovereignty (Walker et al. 2021). Among others, the indigenous social philosophy of Andean origin known as el buen vivir (“good living”)—a way of doing things that is community-centric, ecologically-balanced and culturally-sensitive (Balch 2013) can be fruitfully adapted to reconcile humanism, sustainability, and the ecological imperative (Gudynas 2011; Arcila Calderón, Barranquero, and González Tanco 2018) in the datafied society.

Data Practices: Reclaiming Agency in the Datafied Society
But far from passively losing out to the sweeping structure of datafication, social actors may engage in resistant, subversive, and creative practices that reclaim their political agency. Political agency has to do with the ability of social actors in “making sense of the world so as to act within it” (Couldry 2014, 891). It is “transformative of the structures within which it is embedded” as it makes “use of knowledge and resources in creative and often radical ways” (Kaun, Kyriakidou, and Uldam 2016, 2). The dynamics of datafication forces us to rethink the very same conditions of political agency. In particular, it encourages us to focus on the interstitial spaces between institutionalized politics and the datafied public sphere (including social media platforms). Looking at practices of engagement with data offers a point of entry to track the manifestations of agency in the datafied society as social actors renegotiate their possibilities of intervention. Data practices point to routinized and creative sociotechnical practices of engagement with data, understood both in their guise of objects (e.g., data sets, data visualizations) and in their processes (e.g., surveillance, self-quantification) (see also Mattoni and Treré 2014 on media practices). Practices emerge within hybrid informational ecologies (Treré 2019) and can be individual or collective or both.
In the search for ways of exercising agency in the datafied society, we can explore how variably skilled users engage with data. Examples include data activism, data journalism, citizen sensing, and rituals associated with everyday encounters with data. Data activism appropriates or generates data to exert social change (Gutierrez 2018) or seeks to resist surveillance often by means of technical solutions such as communication encryption (Milan 2017). Data journalism points to the use of data for journalistic storytelling, often with an investigative goal (Baack 2015). Citizen sensing concerns the engagement of citizens in the collection of evidence to support, for example, environmental preservation campaigns (Berti Suman and van Geenhuizen 2020). Finally, everyday encounters with data include the rituals of the quantified self, whereby individuals engage in self-tracking by means of wearable devices (Lupton 2016), but also the efforts of making sense of the many data visualizations that populate pop culture today (Kennedy et al. 2016).

“Contentious data politics,” pointing to the sociocultural mobilizations that critically interrogate datafication from the bottom up (Beraldo and Milan 2019), offers a useful framework to understand data practices. Furthermore, a range of notions allows us to zoom in on what people do with information and technology. These include the idea of “acting on” data and data infrastructures as a politics of intervention (Kubitschko 2017; Milan 2019), as well as “emancipatory communication practices” (Milan 2013) and “critical technical practice” (Dunbar-Hester 2012) foregrounding hands-on, do-it-yourself modes of engagement with data and data infrastructures. Martin-Barbero’s (1993) analysis of “mediations” as opposed to “media,” pointing to autonomous ways of appropriating media products, can help us in thinking about the grammar of grassroots appropriation and subversion when applied to datafication. Finally, Latin American movement praxis that foregrounds community understood as “common-unit (común-unidad)” (Barranquero and Baeza 2017) can help us to rethink collective agency in the datafied society.

Interpretative Lenses to Understand Data at the Margins

The horizontal axis of our matrix identifies three fruitful lenses of interpretation to explore emerging peripheral ways of making sense of data, namely decoloniality and race, intersectionality and feminism, and the pluriverse. Mobilizing these lenses of interpretation pursues two main goals. Firstly, it helps to overcome the blind spots in current “Western” analyses of the datafied society, zeroing in on specific genealogies of dispossession by means of a sociohistorical approach. Secondly, and most importantly, asking why and how certain social groups are oppressed through data and data infrastructure nurtures ongoing efforts to uncover viable pathways toward justice and equality in the datafied society. In other words, by evoking these perspectives to interrogate datafication, we want to make room for thinking along the lines of the “activist work that is required to turn that belief [of equality and justice] into reality” (D'Ignazio and Klein 2020, 3). It is worth noting, however, that there are other possible interpretative lenses that one might adopt to investigate data at the margins: our other selection is situated and by no means conclusive.
The first lens of interpretation we acknowledge is the broad domain of decoloniality and race. Decoloniality and race identify two interconnected perspectives that give voice to those “options confronting and delinking from... the colonial matrix of power” (Mignolo 2011, xxvii). They represent a concrete call to think from and at the margins and peripheries of the world system (Bhabha 2007). They also point to a process of “epistemic reconstruction” (Quijano 2007, 176) that outlines reparative measures and invites redress. Inspired by decolonial and critical race theorists, we can question the persistence of (European and Western) colonial structures in the contemporary datafied society. These colonial structures are visible, for example, in the reproduction of racial bias in search engine algorithms (Noble 2018) or in the use of high-tech tools in predictive policing that profile and punish the poor (Eubanks 2018). Concretely, adopting a decoloniality and race perspective entails interrogating how human characteristics like ethnicity and class are encoded in web applications designed to mediate the relation between the state and its citizens that mirror the colonial lineages of a country like The Netherlands (van Schie, Smit, and López Coombs 2020). It means critically approaching the data infrastructure designed to curb the COVID-19 pandemic by asking in what ways it renders migrants and minorities invisible (Pelizza 2020).

The second lens of interpretation is offered by intersectional and feminist scholarship and movement praxis. Intersectionality and feminism point to traditions that forefront the situated and contextual nature of datafication, grounding it on an analysis of social change (D’Ignazio and Klein 2020). Feminism, often referred to with the plural “feminisms” to emphasize the vast diversity of critical feminist thinking, upholds the idea of equality between the sexes, while challenging sexism and other forces of oppression. Intersectionality broadens the scope of the critique, considering also other aspects of privilege and dispossession that characterize any individual identity. Taken together, these approaches encourage us to “center embodiment and situatedness in relation to disempowerment” (Milan and Treré 2019, 327). Grounded on the feminist imperative of equality, data feminism is “a way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought” (D’Ignazio and Klein 2020, 8). The art project Feminist Data Set by Caroline Sinders (2017) offers a practical example of an intersectional approach to datafication: it purports to collect cultural material about feminism to train AI systems to locate feminist and other intersectional ways of thinking across online media content. Adopting a data feminist perspective might also entail interrogating the gender data gap, exposing how women’s contributions to society have been historically silenced and investigating how said gender data gap concretely impacts their lives today (Criado Perez 2019).

Finally, the third lens of interpretation we put forward addresses the need to make room for distinct ways of seeing and interpreting the world we inhabit, interrogating emerging “southern” (as in the plurality of Souths) epistemologies in our analysis of datafication from the bottom up. Rather than “a fixed set of propositions,” giving voice to southern epistemologies represents “a challenge to develop new knowledge projects

Electronic copy available at: https://ssrn.com/abstract=3933711
and new ways of learning with globally expanded resources” (Connell 2014, 210). The notion of the pluriverse, which foregrounds ontological difference, comes in handy. Advanced by anthropologist Arturo Escobar, the pluriverse is a “world where many worlds fit” and “a tool for reimagining and reconstructing local worlds” (Escobar 2018, xvi). Engaging with the pluriverse paves the way for “an ethical and political practice of alterity that involves a deep concern for social justice, the radical equality of all beings, and nonhierarchy” (p. xvi). And because “culture sits in places” (Escobar 2001, 139), localizing data knowledge and data uses is a first step toward recognizing place as a sociocultural object of struggle also in the datafied society. Concretely, adopting a pluriverse lens to interpret datafication asks, for instance, what buen vivir with data might mean in practice, in an effort to reconcile distinct cosmogonies and local specificities with datafication as a Global North project with a high environmental footprint (Milan and Treré 2021). But a pluriversal approach represents also an invitation to question the epistemology of research itself, paying close attention to relation-building with communities and methods that empower research subjects (Kazansky et al. 2019). It may mean, for example, bridging the agenda of astronomy researchers in rural Brazil with the needs of the local indigenous groups during a global pandemic (Cortesi et al. 2021).

Table 2 presents our analytical matrix populated with the illustrative example discussed in the chapter. The vertical axis details the three components of data at the margins, whereas the horizontal axis identifies the three lenses of interpretation we detected.

<table>
<thead>
<tr>
<th>Components of Data at the Margins</th>
<th>Lenses of Interpretation</th>
<th>Intersectionality and Feminism</th>
<th>The Pluriverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data infrastructure</td>
<td>Decoloniality and Race</td>
<td>Feminist index of domestic violence in Argentina</td>
<td>Servers for local data sovereignty (e.g., Barcelona)</td>
</tr>
<tr>
<td>Data imaginaries</td>
<td>Algorithmic racism in facial recognition technology</td>
<td>Period tracking apps as data exploitation</td>
<td>Buen vivir with data</td>
</tr>
<tr>
<td>Data practices</td>
<td>Data as accountability for Black lives</td>
<td>Feminist Data Set by Caroline Sinders (2017)</td>
<td>Indigenous data sovereignty initiatives</td>
</tr>
<tr>
<td></td>
<td>Creation of “southern” autonomous data infrastructure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Analytical Matrix

Conclusion
Far from being merely an economic resource of global interest, big data and associated technological innovations, including AI applications, might alter citizen agency, jeopardize human rights, and reproduce or create new forms of discrimination. They might also flatten the richness, plurality, and diversity of data cultures, visions, and infrastructure emerging in the fringes of society. This chapter introduced the notion of
data at the margins as a starting point to redefine our conceptual toolbox to address non-mainstream data vulnerabilities. It offered an analytical matrix to study the complex entanglements of infrastructures, imaginaries, and practices. It argued that for a sociological understanding of data at the margins we ought to dialogue with distinct scholarly fields, including critical data, algorithms, and AI studies but also decolonial, feminist, critical race, and critical design studies. In so doing, it sketched a research agenda able to future-proof our understanding of the complex relation between people and technology.

Much work remains to be done. Sociology is particularly well placed to bring valuable insights in this timely debate, in virtue of its long-standing engagement with power, poverty, inequality, and social justice and its tradition of investigating social movements, forms of collective solidarity, and resistance. Particularly promising are recent developments in global sociology, where scholars increasingly complement their postcolonial approach with an anti-authoritarian ethos, articulating also the urgency of a dialogue between different national sociologies (Hanafi 2019).

References


http://eprints.lse.ac.uk/88053/13/Gangadharan_Between-antidiscrimination_Published.pdf.


