

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

Stefania Milan and Emiliano Treré

Sociology and Digital Media

*Edited by Deana A. Rohlinger and Sarah Sobieraj*

Subject: Sociology, Social Theory, Economic Sociology Online Publication Date: Sep 2021

DOI: 10.1093/oxfordhb/9780197510636.013.7

### Abstract and Keywords

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,

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

---

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 (Suzi-

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

---

na 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,

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

---

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

---

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

---

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” (Rodríguez 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” (Rodríguez 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 1. Building Blocks to Analyze Data at the Margins and Illustrative Examples

Data infrastructure	<ul style="list-style-type: none"><li>• Data infrastructures in the Global South (e.g., citizen scoring systems)</li><li>• Creation of autonomous data infrastructure (e.g., cell networks, data sets)</li><li>• Experiments with local data sovereignty</li></ul>
Data imaginaries	<ul style="list-style-type: none"><li>• Fear and resignation associated with surveillance</li><li>• Facial recognition technology as dehumanizing</li><li>• #AbolishBigData2009, connecting datafication to abolitionism</li></ul>
Data practices	<ul style="list-style-type: none"><li>• Data activism</li><li>• Citizen sensing of environmental degradation</li><li>• Rituals associated with everyday encounters with data (e.g., self-tracking)</li></ul>

### 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

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

---

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, Lomborg, 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-Másmela 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

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

---

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



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

---

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 (Bhambra 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 en-

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

---

courage 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 Sindors (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 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.

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

Table 2. Analytical Matrix

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

## 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

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

---

with an anti-authoritarian ethos, articulating also the urgency of a dialogue between different national sociologies (Hanafi 2019).

### **References**

Akrich, Madeleine. 1992. "The De-Description of Technical Objects." In *Shaping Technology/ Building Society. Studies in Sociotechnical Change*, edited by W. E. Bijker and J. Law, 205–224. Cambridge, MA: MIT Press.

Ali, Syed Mustafa. 2016. "Algorithmic Racism: A Decolonial Critique." *Proceedings of the 10th International Society for the Study of Religion*. [http://www.academia.edu/14124452/Algorithmic\\_Racism\\_-\\_A\\_Decolonial\\_Critique](http://www.academia.edu/14124452/Algorithmic_Racism_-_A_Decolonial_Critique).

Andrejevic, Mark. 2014. "Big Data, Big Questions: The Big Data Divide." *International Journal of Communication* 8: 1673–1689.

Aouragh, Miriyam, Seda Gürses, Jara Rocha, and Femke Snelting. 2015. "Let's First Get Things Done! On Division of Labour and Techno-Political Practices of Delegation in Times of Crisis." *Fiberculture* 26: 208–235. <https://doi.org/10.15307/fcj.26.196.2015>.

Aradau, Claudia, and Tobias Blanke. 2017. "Politics of Prediction: Security and the Time/Space of Governmentality in the Age of Big Data." *European Journal of Social Theory* 20, no. 3: 373–391.

Arcila Calderón, Carlos, Alejandro Barranquero, and Eva González Tanco. 2018. "From Media to Buen Vivir: Latin American Approaches to Indigenous Communication." *Communication Theory* 28, no. 2: 180–201.

Arora, Payal. 2016. "The Bottom of the Data Pyramid: Big Data and the Global South." *International Journal of Communication* 10: 1681–1699. <https://ijoc.org/index.php/ijoc/article/viewFile/4297/1616>

Awori, Kagonya, Nicola J. Bidwell, Tigist Sherwaga Hussan, Satinder Gill, and Silvia Lindtner. 2016. "Decolonising Technology Design." In *Proceedings of the First African Conference on Human Computer Interaction (AfriCHI'16)*, 226–228. New York: ACM. <https://doi.org/10.1145/2998581.2998622>.

Baack, Stefan. 2015. "Datafication and Empowerment: How the Open Data Movement Re-articulates Notions of Democracy, Participation, and Journalism." *Big Data & Society* 2, no. 2. <https://doi.org/10.1177/2053951715594634>.

Baca-Feldman, Carlos F., Erick Huerta Velázquez, María Álvarez Malvido, Daniela Parra Hinojosa, and Karla Velasco Ramos. 2019. "Weaving Technological Autonomy in Indigenous Peoples: Community Cellular Telephony in Oaxaca, Mexico." In *Internet Governance and Regulations in Latin America*, edited by L. Belli and O. Cavalli, 275–288. Rio de Janeiro: FGV Direito Rio.

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

---

- Balch, Oliver. 2013. "Buen Vivir: The Social Philosophy Inspiring Movements in South America." *The Guardian*, February 4. <https://www.theguardian.com/sustainable-business/blog/buen-vivir-philosophy-south-america-eduardo-gudynas>.
- Barranquero, Alejandro Carretero, and Chiara Saez Baeza. 2017. "Latin American Critical Epistemologies toward a Biocentric Turn in Communication for Social Change: Communication from a Good Living Perspective." *Latin American Research Review* 52, no. 3: 431–445. <https://doi.org/10.25222/larr.59>.
- Beraldo, Davide, and Stefania Milan. 2019. "From Data Politics to the Contentious Politics of Data." *Big Data & Society* 6, no. 2. <https://doi.org/10.1177/2053951719885967>.
- Berti Suman, Anna, and Marina van Geenhuizen. 2020. "Not Just Noise Monitoring: Rethinking Citizen Sensing for Risk-Related Problem-Solving." *Journal of Environmental Planning and Management* 63, no. 3: 546–567. <https://doi.org/10.1080/09640568.2019.1598852>.
- Bhambra, Gurminder K. 2007. *Rethinking Modernity: Postcolonialism and the Sociological Imagination*. Basingstoke, UK, and New York: Palgrave MacMillan.
- Bigo, Didier. 2002. "Security and Immigration: Toward a Critique of the Governmentality of Unease." *Alternatives* 27, no. 1: 63–92. <https://doi.org/10.1177/03043754020270S105>.
- boyd, dana, and Kate Crawford. 2012. "Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5: 662–679.
- Castro-Gómez, Santiago, and Ramón Grosfoguel, eds. 2007. *El giro decolonial. Reflexiones para una diversidad epistémica más allá del capitalismo global*. Bogota, Colombia: Siglo del Hombre Editores.
- Chenou, Jean-Marie, and Carolina Cepeda-Másmela. 2019. "#NiUnaMenos: Data Activism from the Global South." *Television & New Media* 20, no. 4: 396–411. <https://doi.org/10.1177/1527476419828995>.
- Cheruiyot, David, and Raul Ferrer-Conill. 2018. "Fact-Checking Africa. Epistemologies, Data and the Expansion of Journalistic Discourse." *Digital Journalism* 6, no. 8: 964–975.
- Connell, R. 2014. "Using Southern Theory: Decolonizing Social Thought in Theory, Research and Application." *Planning Theory* 13, no. 2: 210–223.
- Cortesi, Arianna, Claudia Magnani, Roberto Romero, Paula C.P. Silva, Sueli Maxakali, Isael Maxakali, and Ana Maria R. Gomes. 2021. "Under other Skies: Astronomy as a Tool to Face COVID-19-Induced Isolation in the Indigenous Village of Aldeia Verde, Brazil." In *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the*

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

---

*Datified Society*, edited by S. Milan, E. Treré, & S. Masiero, 259–262. Amsterdam: Institute of Network Cultures.

Couldry, Nick. 2014. "A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World." *Sociological Review* 62, no. 4: 880–897.

Crawford, Kate. 2015. "Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics." *Science, Technology & Human Values* 41, no. 1: 77–92. <https://doi.org/10.1177/0162243915589635>.

Criado Perez, Caroline. 2019. *Invisible Women. Exposing Data Bias in a World Designed for Men*. London: Vintage.

Crooks, Roderic. 2019. "What We Mean When We Say #AbolishBigData2019." Medium, March 22. <https://medium.com/@rncrooks/what-we-mean-when-we-say-abolish-big-data2019-d030799ab22e>.

Currie, Morgan, Brittany Paris, I. Pasquetto, and J. Pierre. 2016. "The Conundrum of Police Officer-Involved Homicides: Counter-data in Los Angeles County." *Big Data & Society* 3, no. 2. <https://doi.org/10.1177/2053951716663566>.

Dalton, Craig M., Linnet Taylor, and Jim Thatcher. 2016. "Critical Data Studies: A Dialog on Data and Space." *Big Data & Society* 3, no. 1. <https://doi.org/10.1177/2053951716648346>.

Daskal, Efrat. 2018. "Let's Be Careful out There ...: How Digital Rights Advocates Educate Citizens in the Digital Age." *Information, Communication & Society* 21, no. 2: 241–256. <https://doi.org/10.1080/1369118X.2016.1271903>.

Data for Black Lives. n.d. "About Us." <http://d4bl.org/about.html>.

Dauvergne, Peter. 2020. "Is Artificial Intelligence Greening Global Supply Chains? Exposing the Political Economy of Environmental Costs." *Review of International Political Economy*. Published ahead of print September 3, 2020. <https://doi.org/10.1080/09692290.2020.1814381>.

Dencik, Lina. 2018. "Surveillance Realism and the Politics of Imagination: Is There no Alternative?" *Krisis: Journal for Contemporary Philosophy* 1: 31–43.

Dencik, Lina, Arne Hintz, Joanna Redden, and Harry Warne. 2018. *Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services*. Cardiff, UK: Data Justice Lab. <http://orca.cf.ac.uk/117517/1/data-scores-as-governance-project-report2.pdf>.

Deutch, Jeff, and Hadi Habal. 2018. "The Syrian Archive: A Methodological Case Study of Open-Source Investigation of State Crime Using Video Evidence from Social Media Platforms." *State Crime* 7, no. 1: 46–76.

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

---

D'Ignazio, Catherine, and Lauren Klein. 2020. *Data Feminism*. Cambridge, MA: MIT Press.

Duffy, Brooke Erin, and Ngai Keung Chan. 2018. " 'You Never Really Know Who's Looking': Imagined Surveillance across Social Media Platforms." *New Media & Society* 21, no. 1: 119–138. <https://doi.org/10.1177/1461444818791318>.

Dunbar-Hester, Christina. 2012. "Soldering toward Media Democracy. Technical Practice as Symbolic Value in Radio Activism." *Journal of Communication Inquiry* 36, no. 2: 149–169.

Escobar, Arturo. 2001. "Culture Sits in Places: Reflections on Globalism and Subaltern Strategies of Localization." *Political Geography* 20, no. 2: 139–174. [https://doi.org/10.1016/S0962-6298\(00\)00064-0](https://doi.org/10.1016/S0962-6298(00)00064-0).

Escobar, Arturo. 2018. *Designs for the Pluriverse. Radical Interdependence, Autonomy, and the Making of Worlds*. Durham, NC: Duke University Press.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Felizi, Natasha, and Joana Varon. n.d. "MENSTRUAPPS—How to Turn Your Period into Money (for Others)." Coding Rights. <https://chupadados.codingrights.org/en/menstruapps-come-transformar-sua-menstruacao-em-dinheiro-para-os-outros/>.

Flichy, Patrice. 2007. *The Internet Imaginaire*. Cambridge, MA: MIT Press.

Gamson, William A. 1992. *Talking Politics*. Cambridge: Cambridge University Press.

Gangadharan, Seeta P., and Niklas Jędrzej. 2018. *Between Antidiscrimination and Data: Understanding Human Rights Discourse on Automated Discrimination in Europe*. London: London School of Economics. [http://eprints.lse.ac.uk/88053/13/Gangadharan\\_Between-antidiscrimination\\_Published.pdf](http://eprints.lse.ac.uk/88053/13/Gangadharan_Between-antidiscrimination_Published.pdf).

Gudynas, Eduardo. 2011. "Buen Vivir: Today's Tomorrow." *Development* 54, no. 4: 441–447. <https://link.springer.com/article/10.1057/dev.2011.86>.

Gupta, Anushri, Panos Panagiotopoulos, and Frances Bowen. 2020. "An Orchestration Approach to Smart City Data Ecosystems." *Technological Forecasting and Social Change* 153: 119929. <https://doi.org/10.1016/j.techfore.2020.119929>.

Gutierrez, Miren. 2018. *Data Activism and Social Change*. London: Palgrave MacMillan.

Haggerty, Kevin D., and Richard V. Ericson. 2000. "The Surveillant Assemblage." *British Journal of Sociology* 51, no. 4: 605–622.

Hanafi, Sari. 2019. "Global Sociology Revisited: Toward New Directions." *Current Sociology* 68, no. 1: 3–21. <https://doi.org/10.1177/0011392119869051>.



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

---

Heeks, Richard, and Jaco Renken. 2018. "Data Justice for Development: What Would It Mean?" *Information Development* 34, no. 1: 90–102. <https://doi.org/10.1177/0266666916678282>.

Helles, Rasmus, Stine Lomborg, and Signe Sophus Lai. 2020. "Infrastructures of Tracking: Mapping the Ecology of Third-Party Services across Top Sites in the EU." *New Media & Society* 22, no. 11: 1957–1975. <https://doi.org/10.1177/1461444820932868>.

Herrera Huérfano, Eliana, Francisco Sierra Caballero, and Carlos Del Valle Rojas. 2016. "Hacia una epistemología del sur. Decolonialidad del saberpoder informativo y nueva comunicología Latinoamericana. Una lectura crítica de la mediación desde las culturas indígenas." *Chasqui. Revista Latinoamericana de Comunicación* 131: 77–105.

Hintz, Arne, Lina Dencik, and Karin Wahl-Jorgensen. 2018. *Digital Citizenship in a Datafied Society*. Cambridge: Polity.

International Telecommunication Union. 2019. "Statistics." <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>.

Kaun, Anne, Maria Kyriakidou, and Julie Uldam. 2016. "Political Agency at the Digital Crossroads?" *Media and Communication* 4, no. 4: 1–7.

Kazansky, Becky, and Stefania Milan. 2021. "'Bodies not Templates': Contesting Mainstream Algorithmic Imaginaries." *New Media & Society* 23, no. 2: 363–381. <https://doi.org/10.1177/1461444820929316>.

Kazansky, Becky, Guillén Torres, Lonneke van der Velden, Kersti R. Wissenbach, and Stefania Milan. 2019. "Data for the Social Good: Toward a Data-Activist Research Agenda." In *Good Data*, edited by A. Daly and M. Mann, 244–259. Amsterdam: Institute of Network Cultures.

Kennedy, Helen, Rosemary Lucy Hill, William Allen, and Andy Kirk. 2016. "Engaging with (Big) Data Visualizations: Factors That Affect Engagement and Resulting New Definitions of Effectiveness." *First Monday* 21, no. 11. <https://firstmonday.org/article/view/6389/5652>

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Kubitschko, Sebastian. 2017. "Acting on Media Technologies and Infrastructures: Expanding the Media as Practice Approach." *Media, Culture & Society* 40, no. 4: 629–635. <https://doi.org/10.1177/0163443717706068>.

Leff, Enrique. 1994. *Ecología y capital: Racionalidad ambiental, democracia participativa y desarrollo sustentable*. Mexico City: Siglo XXI.

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

---

Lehtiniemi, Tuukka, and Minna Ruckenstein. 2018. "The Social Imaginaries of Data Activism." *Big Data & Society* 6, no. 1. <https://journals.sagepub.com/doi/full/10.1177/2053951718821146>.

López, Joan. 2020. *Experimentando con la pobreza: El sisbén y los proyectos de analítica de datos*. Bogota, Colombia: Fundación Karisma. <https://web.karisma.org.co/wp-content/uploads/download-manager-files/Experimentando%20con%20la%20pobreza.pdf>.

Lupton, Deborah. 2016. *The Quantified Self*. Cambridge: Polity.

Lynch, Casey R. 2020. "Contesting Digital Futures: Urban Politics, Alternative Economies, and the Movement for Technological Sovereignty in Barcelona." *Antipode* 52, no. 3: 660–680. <https://doi.org/10.1111/anti.12522>.

Martin-Barbero, Jesus. 1993. *Communication, Culture and Hegemony: From the Media to Mediations*. London: Sage.

Masiero, Silvia. 2020. "Biometric Infrastructures and the Indian Public Distribution System." *South Asia Multidisciplinary Academic Journal* 23. <https://doi.org/10.4000/samaj.6459>.

Mattoni, Alice, and Emiliano Treré. 2014. "Media Practices, Mediation Processes, and Mediatization in the Study of Social Movements." *Communication Theory* 24, no. 3: 252–271.

Medina, Eden, Ivan da Costa Marques, and Christina Holmes, eds. 2014. *Beyond Imported Magic. Essays on Science, Technology, and Society in Latin America*. Cambridge, MA: MIT Press.

Mignolo, Walter D. 2000. *Local Histories/Global Designs: Essays on the Coloniality of Power, Subaltern Knowledges and Border Thinking*. Princeton, NJ: Princeton University Press.

Mignolo, Walter D. 2011. *The Darker Side of Western Modernity: Global Futures, Decolonial Options*. Durham, NC: Duke University Press.

Milan, Stefania. 2013. *Social Movements and Their Technologies: Wiring Social Change*. London: Palgrave Macmillan.

Milan, Stefania. 2017. "Data Activism as the New Frontier of Media Activism." In *Media Activism in the Digital Age*, edited by G. Yang and V. Pickard, 151–163. London and New York: Routledge.

Milan, Stefania. 2019. "Acting on Data(fication)." In *Citizen Media and Practice: Currents, Connections, Challenges*, edited by H. Stephansen and E. Treré, 212–226. London and New York: Routledge.

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

---

Milan, Stefania. 2020. "Techno-solutionism and the Standard Human in the Making of the COVID-19 Pandemic." *Big Data & Society* 7, no. 2. <https://doi.org/10.1177/2053951720966781>.

Milan, Stefania, and Niels ten Oever. 2017. "Coding and Encoding Rights in Internet Infrastructure." *Internet Policy Review* 6, no. 1. <https://doi.org/10.14763/2017.1.442>.

Milan, Stefania, and Emiliano Treré. 2019. "Big Data from the South(s): Beyond Data Universalism." *Television & New Media* 20, no. 4: 319–335. <https://doi.org/10.1177/1527476419837739>.

Milan, Stefania, and Emiliano Treré. 2020. "The Rise of the Data Poor: The COVID-19 Pandemic Seen from the Margins." *Social Media + Society* 6, no. 3. <https://doi.org/10.1177/2056305120948233>.

Milan, Stefania, and Emiliano Treré. 2021. "Latin American Visions for a Digital New Deal: Learning from Critical Ecology, Liberation Pedagogy and Autonomous Design." In *A Digital New Deal. Visions of Justice in a Post-Covid World*, edited by S. Sarkar and A. Korjan, 101–111. Bangalore, India: IT for Change.

Milan, Stefania, Emiliano Treré, and Silvia Masiero. 2021. *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society*. Amsterdam: Institute of Network Cultures.

Milan, Stefania, and Lonneke van der Velden. 2016. "The Alternative Epistemologies of Data Activism." *Digital Culture & Society* 2, 57–74. <https://doi.org/10.14361/dcs-2016-0205>.

Mohamed, Shakir, Marie-Therese Png, and William Isaac. 2020. "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence." *Philosophy & Technology* 33, no. 4: 659–684. <https://doi.org/10.1007/s13347-020-00405-8>.

Mosco, Vincent. 2004. *The Digital Sublime: Myth, Power, and Cyberspace*. Cambridge, MA: MIT Press.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.

Pelizza, Annalisa. 2020. "'No Disease for the Others': How COVID-19 Data Can Enact New and Old Alterities." *Big Data & Society* 7, no. 2: 2053951720942542. <https://doi.org/10.1177/2053951720942542>.

Pelizza, Annalisa, Stefania Milan, and Yoren Lausberg. Forthcoming. "Undocumented Migrants in COVID-19 Counting: Rethinking the Data-(In)Visibility Nexus." *Data & Policy*.

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

---

Quijano, Anibal. 2007. "Coloniality and Modernity/Rationality." *Cultural Studies* 21, no. 2-3: 168-178.

ReclaimYourFace. 2020. "Secretive. Unlawful. Inhumane." <https://reclaimyourface.eu/the-problem/>.

Rodrigues, Gemma F., Christopher Csikszentmihalyi, Daniel Mwesigwa, Jude Mukundane, and Michelle Kasprzak. 2018. *Social Tech Ecosystems in Sub-Saharan Africa*. Funchal, Madeira: Madeira Interactive Technologies Institute. <https://doi.org/10.5281/zenodo.1244086>.

Rodriguez, Clemencia. 2017. "Studying Media at the Margins: Learning from the Field." In *Media Activism in the Digital Age*, edited by V. Pickard and G. Yang, 49-60. London and New York: Routledge.

Sadowski, Jathan. 2019. "When Data Is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6, no. 1. <https://doi.org/10.1177/2053951718820549>.

Schoemaker, Emrys, Dina Baslan, Bryan Pon, and Nicola Dell. 2020. "Identity at the Margins: Data Justice and Refugee Experiences with Digital Identity Systems in Lebanon, Jordan, and Uganda." *Information Technology for Development* 27, no. 1: 13-36. <https://doi.org/10.1080/02681102.2020.1785826>.

Sinders, Caroline. 2017. *Feminist Data Set*. <https://carolinesinders.com/feminist-data-set/>.

Sun, Y., and W. Yan. 2020. "The Power of Data from the Global South: Environmental Civic Tech and Data Activism in China." *International Journal of Communication* 14: 2144-2162.

Suzina, Ana Cristina. 2020. "English as *Lingua Franca*. Or the Sterilisation of Scientific Work." *Media, Culture & Society* 43, no. 1: 171-179. <https://doi.org/10.1177/0163443720957906>.

Taylor, Linnet, and Dennis Broeders. 2015. "In the Name of Development: Power, Profit and the Datafication of the Global South." *Geoforum* 64: 229-237. <https://doi.org/10.1016/j.geoforum.2015.07.002>.

Thatcher, Jim, David O'Sullivan, and Dillon Mahmoudi. 2016. "Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data." *Environment and Planning D: Society and Space* 34, no. 6: 990-1006. <https://doi.org/10.1177/0263775816633195>.

Thorat, Dhanashree. 2021. "Modalities of Data Colonialism and South Asian Hashtag Publics." *Feminist Media Studies* 21, no. 1: 151-153. <https://doi.org/10.1080/14680777.2021.1864874>.

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

---

Torres, Guillén. 2019. "Institutional Resistance to Transparency: The Quest for Public Sector Information in Mexico." *Journal of Resistance Studies* 5, no. 2: 68–96.

Treré, Emiliano. 2019. *Hybrid Media Activism*. London and New York: Routledge.

van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance and Society* 12, no. 3: 197–208.

Van Dijk, Jan. 2020. *The Digital Divide*. Cambridge: Polity Press.

van Doorn, Niels. 2017. "Platform Labor: On the Gendered and Racialized Exploitation of Low-Income Service Work in the 'On-Demand' Economy." *Information, Communication & Society* 20, no. 6: 898–914. <https://doi.org/10.1080/1369118X.2017.1294194>.

van Schie, Gerwin, Alex Smit, and Nicolás López Coombs. 2020. "Racing through the Dutch Governmental Data Assemblage: A Postcolonial Data Studies Approach." *Global Perspectives* 1, no. 1: 12779. <https://doi.org/10.1525/gp.2020.12779>.

Vera, Lourdes A., Lindsey Dillon, Sara Wylie, Jennifer Liss Ohayon, Aaron Lemelin, Phil Brown, et al. 2018. "Data Resistance: A Social Movement Organizational Autoethnography of the Environmental Data and Governance Initiative." *Mobilization: An International Quarterly* 23, no. 4: 511–529. <https://doi.org/10.17813/1086-671X-24-4-511>.

Voo, Teck Chuan, Hannah Clapham, and Clarence C Tam. 2020. "Ethical Implementation of Immunity Passports during the COVID-19 Pandemic." *Journal of Infectious Diseases* 222, no. 5: 715–718. <https://doi.org/10.1093/infdis/jiaa352>.

Walker, Maggie, Tahu Kukutai, Stephanie Russo Carroll, and Desi Rodriguez-Lonebear, eds. 2021. *Indigenous Data Sovereignty and Policy*. London and New York: Routledge.

Winner, Langdon. 1999. "Do Artefacts Have Politics?" In *The Social Shaping of Technology*, edited by D. MacKenzie and J. Wajcman, 28–40. Buckingham, UK, and Philadelphia: Open University Press.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. London: Profile Books.

### **Stefania Milan**

Department of Media Studies, University of Amsterdam, The Netherlands

### **Emiliano Treré**

School of Journalism, Media and Culture (JOMEC), Cardiff University