

Data for Development: Exploring Connections between Open Data, Big Data, and Data Privacy in the Global South

TERESA SCASSA AND FERNANDO PERINI

Abstract

This chapter considers the interrelated themes of open and big data in the Global South, with a particular emphasis on privacy concerns. Open data and big data are examined together, in part because of the emerging view that open government data alone cannot meet the data needs of developing countries. We examine the potential uses and applications of big data analytics to inform evidence-based policy-making, mostly around sustainable development, as well as some of the associated challenges. We also identify and address data-protection and privacy concerns that arise from the use of open and big data, with a particular focus on the experience of countries of the Global South. The chapter concludes with some suggestions about the kind of research needed to shape the future of open data in the Global South.

Acknowledgements

This chapter was prepared with funding support from the International Development Research Centre. An earlier version

was presented as part of a workshop at the International Open Data Conference in Buenos Aires, Argentina, in September 2018.

The so-called data revolution has profoundly changed the role of data in society and the economy. Data, data analytics, and data-fuelled technologies offer new ways to identify and address problems, and promise greater efficiency in processes and decision-making for the public and private sector alike. Data-related innovation is also driving economic prosperity. The data revolution has, in turn, motivated research about data and associated technologies, their promise, benefits, and risks. Most of this research comes from the Global North and focuses on these benefits, risks, and challenges in that context.

This chapter considers the interrelated themes of big data and open data in the Global South, with a particular emphasis on certain key concerns, such as privacy. Open data and big data are considered together since they are, in many ways, becoming inextricably linked as the data revolution continues to unfold. We review a body of research primarily produced in the Global South that has been funded entirely or in part by the International Development Research Centre (IDRC), a Canadian Crown corporation. In total, we considered thirty publications wholly or partially funded by the IDRC. This literature explores the relevance of data for development and the dimensions of data opportunities, particularly in relation to sustainable development goals (SDGs). At the same time, it identifies and explores data harms and the need for better data governance.

The chapter features three broad themes: open data, big data, and data protection/privacy. Although there are issues that cut across all three, each theme reveals a core set of concerns. The first part of the chapter examines each of the themes and identifies these core issues. Under the first theme, open data, we address the work of the global community that has pushed for the disclosure of primarily government data as a means to increase transparency, accountability, and innovation. Given that resource and capacity issues sometimes limit the scope of available government data, we also explore the potential for other sources of open data, including private-sector and research data. For big data, we examine the potential uses and applications of big data analytics to inform evidence-based policy-making, mostly around sustainable development, as well as some of the associated challenges. Materials under the third theme, data protection

and privacy, explore the use and misuse of personal data and related challenges for privacy protection.

There are clearly important overlaps between the three themes. In Part 2 of this chapter, we identify and elaborate upon a series of cross-cutting themes. In our conclusion, we identify gaps in the research and sketch out some of the work needed to shape the future of open data in the Global South.

1. Open Data, Big Data, and Data Privacy

1.1 *Open Data*

Open data generally requires “that a dataset be accessible (usually by being online) at no cost, and with no technical restrictions to prevent its re-use” (Davies & Perini, 2013, p. 3). According to commitments made through the Open Government Partnership, a majority of countries worldwide are now publishing “some or all of their data” (Davies & Perini, 2016, p. 149). Open government data are generally seen to have the potential to support three goals: transparency and accountability; innovation and economic development; and inclusion and empowerment (Open Data for Development, 2016), although Davies and Perini (2016) argue that a fourth area—the use of open data to support internal government reforms—is also important in the development context.

In the Global North, “open data” has generally been used interchangeably with “open government data.” Davies and Perini (2016) note that most open data research focuses on government as the primary data source. However, they observe that “in developing countries a wide range of government, NGOs, international agency and private actors may be involved [in] creating and holding relevant data” (Davies & Perini, 2016, p. 153). This point is also emphasized in the African context, where a report suggests that open data communities look to “data from non-governmental actors, such as oil, mining and gas companies (open extractives), aid agencies (open aid), government procurement (open contracting), and scientific publications (open access)” (UNECA, 2017, p. 27). Crowdsourced data can also be another source of open data (UNECA, 2017).

According to Davies and Perini, research is also required in order to better understand open data used for decision making. They identified four priority areas for open data research that looked at the flow

of open data to users through a variety of intermediaries; understanding how the broader context affects open data initiatives; understanding how “global standards, platforms, infrastructures and ‘eco-systems’ of open data affect local contexts”; and understanding how benefits of open data initiatives are distributed (Davies & Perini, 2013, p. 7). Davies and Perini later (2016) refined their research framework. They argued that open data research should consider not only the kinds of decisions being made using open data but the governance settings in which those decisions are made. They suggest that research should focus on “emerging outcomes,” and should consider not only the different outcomes sought through open data but the relationship between how data are supplied and the realization of outcomes. They propose a research framework for open data case studies, with a view to grouping research into key areas and supporting cross-case comparisons.

Overall, there are a number of open data challenges in the Global South. One is the sustainability of open data initiatives in some countries. In some cases, there is a need to broaden efforts to build open data capacity, including bringing in and expanding the expertise of national statistical offices (UNECA, 2017). At the same time, the literature identifies developing regional hubs and creating links across countries as a way of providing coherence and coordination to national efforts to develop open data programs (Open Data Institute, 2016). Drawing on the experiences and expertise of leaders in open data in the Global South is another way to support such movements, and the Open Data Leaders Network, convened by the Open Data Initiative, offers different stories of open data success (Open Data Institute, 2016). It emphasizes the role of open data leaders in championing open data projects, and bringing them to fruition, but recognizes as well the need for strong peer networks within and across jurisdictions (Open Data Institute, 2016, p. 8).

There is a growing awareness that open data may create new privacy challenges, as well as concerns that privacy may become an excuse not to open and share data (Open Data Research Network, 2013), even though privacy and openness are not antithetical concepts (Gurumurthy & Chami, 2016). There is a need for engagement around the balance between open data and privacy, recognizing that there may be national and culture-specific views of how best to achieve this balance (Open Data Research Network, 2013).

Open data is generally recognized as important for providing a supply of useful data for many purposes, for building capacity within

governments, and for encouraging innovation in government. However, the *Africa Data Revolution Report 2016* (UNECA, 2017), prepared by the African Centre for Statistics at the United Nations Economic Commission for Africa, questions whether its potential is over-promised, suggesting that evidence of the utility of open data is mostly anecdotal. There is also a possibility that in some countries the focus may be on providing open data as an end goal rather than on the potential applications for open data. Two phenomena—“open washing” (where data appear to be open but are not readily available for unrestricted reuse) and “open wishing” (where the benefits of open data are over-promised)—are described as pervasive (Open Data for Development, 2016). There are also concerns that unless attention is also paid to capacity building among potential user communities for open data, greater use of open data will be made, rather, by those who are privileged, reinforcing social and economic divides (UNECA, 2017, p. 30). Open Data for Development (2016), in particular, recognizes a need to address gender issues in open data but acknowledges that this remains challenging.

1.2 Big Data for Development: Leveraging the Private Sector

Although government open data programs involve governments making their own data more broadly available, many governments in the Global South face data deficits—where they lack the data they need for their own planning and decision-making processes. There is, therefore, a considerable interest in the potential of big data analytics to meet informational and data gaps in countries of the Global South. Gaps may arise where national statistical agencies lack the capacity to collect sufficient data (Gurumurthy & Chami, 2018; LIRNEasia, 2017b; UNECA, 2017). Such data are usually only collected at intervals of up to 10 years, which makes them less useful than big data for assessing progress toward SDGs or for identifying and responding quickly to new trends or situations (LIRNEasia, 2017a; Surendra et al., 2017; Lokanathan et al., 2017). There may, therefore, be a particular need to supplement state statistical data, particularly with respect to SDGs (LIRNEasia, 2017a, 2017b; UNECA, 2017). Nevertheless, the capacity of the public sector to access, use, store, and secure big data is an important issue (LIRNEasia, 2017a). Some problems relate to the lack of interoperability between available data sources, as well as the lack of country-level plans and processes for big data (Manoj, 2017). Further,

some administrative data may either be lacking or difficult to use (UNECA, 2017; Manoj, 2017).

Big data analytics require large volumes of data, and the quantity of available government data may not be sufficient—certainly not for all purposes (Gurumurthy & Chami, 2018; Lokanathan et al., 2017). As a result, there is a general recognition of a need to look to other sources. The external big data source most often referenced in the literature is mobile-communications data from telecommunications service providers (UNECA, 2017; Lokanathan et al., 2017; LIRNEasia, 2018, 2017a, 2017b). Other important sources are satellite data and social media data (LIRNEasia, 2017b; Lokanathan et al., 2017; UNECA, 2017). The *Africa Data Revolution Report* (UNECA, 2017, p. 22) identified a need for systematic efforts “to harness data held by private-sector communities for sustainable development in Africa.” Some mention is also made of using data from civil-society actors as well as citizen-generated data (UNECA, 2017; Lokanathan et al., 2017).

Sourcing big data from the private sector can raise challenges around obtaining access and permission to use these data (Lokanathan et al., 2017; Gurumurthy & Chami 2016, 2018; IT for Change, 2017). The use of intermediaries is a possible solution. Intermediaries could “gain access to (privately held) data, conduct the analyses and share insights with government institutions” (LIRNEasia, 2017a, p. 10). Such a solution might also help overcome the lack of capacity for in-house public-sector big data analytics (Lokanathan et al., 2017). By contrast, a “data as a public good” approach might see states mandate the disclosure of “critical data” by private-sector companies for use by government (Gurumurthy & Chami, 2018; IT for Change, 2017).

A number of studies demonstrate how big data analytics have been used to address issues in countries within the Global South. Some studies looked at specific applications, including national identity cards, predictive policing, credit scoring, and smart meters (Hickok et al., 2017). One case study looked at using call detail records or “telephony metadata” to predict socio-economic characteristics of the Sri Lankan population as a means of supplementing census data (Surendra et al., 2017). Big data have also been used to predict outbreaks of disease (LIRNEasia, 2017a), in intelligent transportation systems (Hickok et al., 2017), or to monitor the performance of new programs or initiatives (UNECA, 2017). One report considered how big data could be specifically used to support SDGs (Lokanathan et al., 2017).

Each of the case studies identifies both benefits and harms flowing from big data analytics. While the benefits are often quite specific to the problems sought to be addressed by the adoption of the technology, the harms tend to be similar across applications and relate to issues such as abuse of the results of the analytics; biased results; overreliance on analyzing data for decision-making while ignoring other considerations; privacy invasion and surveillance; lack of ethical frameworks; and lack of algorithmic transparency (Hickok et al., 2017; LIRNEasia, 2018; Samarajiva & Perera-Gomez, 2018; Lokanathan et al., 2017).

Gurumurthy and Chami (2018) suggest that in the rush to embrace big data solutionism there is a “side-stepping” of ethical and privacy concerns. While some work has been done on providing normative frameworks for the use of big data (LIRNEasia, 2018), there is a general recognition of significant gaps in legal and ethical frameworks (UNECA, 2017). LIRNEasia identified a need for networking and the sharing of best practices around the use of big data analytics (LIRNEasia, 2017b). A number of authors identify gaps in governance (UNECA, 2017; Gurumurthy & Chami, 2018; Manoj, 2017). Governance can be broadly understood to include methods, procedures, data management, and analysis (Manoj, 2017), as well as governance in terms of privacy, access, and transparency concerns. In their report from a workshop on big data for development, Lokanathan and Perera-Gomez (2016) flagged a need for more research on the harms of big data. There was a concern that actual harms, and not just theoretical harms, need to be identified and studied. More concrete examples could lead to strategies to prevent such harms or to limit their impact (Lokanathan & Perera-Gomez, 2016). LIRNEasia (2017a, p. 12) argues that “regulatory frameworks need to evolve in parallel with big data in development.” Lokanathan et al. (2016) have suggested that developing codes of practice, built upon professional standards, might provide guidance to those engaged in big data analytics.

The completeness and quality of data available for big data analytics are also an issue (Manoj, 2017; Lokanathan et al., 2017; IT for Change, 2017). There are concerns about who is counted and who is excluded from big data (Lokanathan & Perera-Gomez, 2016), as well as about “the marginalisation of women’s ways of knowing and an undermining of democratic life in general,” for example, in contexts in which political decision-making becomes data-driven (Gurumurthy & Chami, 2016). Gender inclusivity in datasets is identified as a

problem, particularly given the nature of some of the data that might be used, and the “structural inequalities and entrenched prejudices in many societies” that may limit the inclusion of data about women (van der Spuy & Aavriti, 2017, p. 30). For example, social media data are collected only from those who have access to and use social media; similarly, mobile-phone data can only be collected from those with mobile phones (LIRNEasia, 2017b).

There are also concerns that the limits of big data analytics are not well understood. Big data analytics deal with correlation and not causality (Lokanathan et al., 2017; IT for Change, 2017). Not only does this mean that big data analytics cannot replace deductive reasoning, some have also suggested that they might not be well-suited to dealing with many of the complex issues facing the Global South (Lokanathan et al., 2017; IT for Change, 2017).

Large datasets are also important for uses in other emerging technologies. The use of artificial intelligence (AI) will require significant amounts of training data. On the positive side, AI could lead to “innovative, data-driven, technical innovations to help address pressing social problems” (Smith, 2018, p. 9). At the same time, many of the risks of AI in the Global South mirror those discussed in relation to big data, including surveillance and loss of privacy, as well as bias and discrimination in decision-making caused by non-representative datasets, and other biases present in AI algorithms. Smith (2018, p. 13) expresses concerns over “how our current set of institutions and cultures shapes the evolution of technologies, and how, in turn, these technologies shape these institutions and cultures.” He identifies a need for policy and regulatory frameworks, as well as further research to understand both potential applications and impacts of AI.

1.3 Data Protection and Privacy

The nature and volume of personal information being collected in the big data society are concerning. The literature on the Global South identifies a lack of adequate legal and normative infrastructure to protect privacy rights, particularly with rapidly evolving technologies (Hosein, 2011; Hosein & Nyst, 2013). Although in some cases there are simply no effective privacy laws in place, even in those instances where laws are in place the proper funding and staffing of oversight bodies must be ensured (ADC, 2014). There are also concerns that technologies rejected as privacy invasive in the Global

North are being “dumped” in the Global South (Hosein & Nyst, 2013; Hosein, 2011). Hosein (2011) notes that development funds from the Global North may even be used to purchase these technologies for countries in the Global South. National ID systems that incorporate biometrics, communications-surveillance technologies, electronic health registries, and DNA databases are examples of adopted technologies that have significant privacy implications. In some cases, the privacy issues flow in part from poor security measures around digital data.

Specific technologies adopted by governments may have significant privacy implications. One of these is biometrics, often adopted in conjunction with national identification systems (Hosein, 2011; Hosein & Nyst, 2013; ADC, 2017). The most commonly used form of biometrics in the Global South appears to be fingerprinting. There are concerns about the lack of adequate privacy frameworks for government adoption of biometrics systems (Hosein & Nyst, 2013; ADC, 2017). For example, there is considerable risk of abuse if specific legal frameworks are lacking to set the parameters for when and in what circumstances law-enforcement officials can access biometric databases (ADC, 2017). Biometric-identification systems also raise issues that go beyond privacy. For example, such systems can exclude individuals or segments of the population from programs and services where biometric data cannot be collected, or where the technologies used to collect data are not suited to local conditions (ADC, 2017).

Some technologies are specifically oriented toward law enforcement and national security. These tend to raise significant privacy concerns. Communications surveillance and predictive policing are examples (Hosein & Nyst, 2013; Cortés, 2015; Hickok et al., 2017). A number of potential harms flow from data-based surveillance, including the use of such technologies to track protestors and suppress civil liberties; the risk of data breaches; and the fact that algorithms used may lack transparency, making it difficult to understand how individuals are singled out within the bulk data (Samarajiva & Perera-Gomez, 2018; Hosein, 2011; Hosein & Nyst, 2013). The effect of constant surveillance on a population is also a concern, particularly as it may lead to behaviour modification that both limits autonomy and that may undermine the usefulness of the data analysis being carried out (Samarajiva & Perera-Gomez, 2018). Van der Spuy and Aavriti (2017) suggest that behaviour modification due to surveillance is also likely to disproportionately affect women. Predictive policing can lead to

unjustified surveillance and discrimination (Hickok et al., 2017). Its methods may also be opaque (Hickok et al., 2017). Scott-Railton et al. (2017) cite the use of spyware by state officials to infect targets' phones or computers. Not only are such technologies highly invasive, they could potentially be used to target opponents or those critical of government. This highlights that privacy issues are closely intertwined with human-rights issues; poor privacy-protective frameworks can contribute to the misuse of personal information in ways that adversely affect privacy and other human rights and civil liberties.

More general privacy concerns include the potential for state surveillance that is exacerbated where large volumes of data are collected about individuals (Hosein & Nyst, 2013; Hosein, 2011; Samarajiva & Perera-Gomez, 2018). Mobile-communications data are of particular concern. As noted above, these data are identified as being a particularly useful category of data for supplementing open data in big data analytics. From a privacy perspective, such data raise problems of bulk surveillance (Samarajiva & Perera-Gomez, 2018; Hosein & Nyst, 2013). Data-minimization principles and data-retention limits are required. Lokanathan and Perera-Gomez (2016) suggest that for big data it is important to develop privacy frameworks on the front end—in other words, some form of privacy by design, rather than relying upon concepts of notice and consent. One study considered the role of information intermediaries in ensuring the privacy of Internet users by adopting appropriate protocols and safeguards for managing subscriber information. Manoj (2017) suggests that international ethical standards may be required to address privacy issues in big data.

In a study from Argentina, the Asociación por los Derechos Civiles (ADC, 2014) looked at whether citizens' personal data was adequately protected by the state. It found problems, including excessively broad exceptions for the use of personal information by state actors. It also noted that the oversight body was under-resourced to the point of being ineffective. Another case study by the same organization looked at the impact of the introduction of a national biometric-identification system on human rights (ADC, 2017). The report noted that this system is not easily reconcilable with constitutional guarantees, raising important human-rights issues. It also identified concerns about the potential of biometric systems to be used as tools to target particular segments of the population. Another study on communications-surveillance practices in Columbia identified

significant deficiencies in the legal protections for privacy and civil liberties, and suggests that the laws are not well adapted to emerging technologies (Cortés, 2015). A study on the protection of personal information in public databases in Paraguay found problems that flowed from inconsistent application of the existing law, which itself was out of date with respect to rapidly evolving technologies (Acuña, et al., 2017).

While privacy is most often considered in terms of individuals, there was also concern about the potential of privacy-invasive technologies to facilitate the identification, profiling, and targeting of minority communities (Gurumurthy & Chami, 2018). Hosein (2011) warns of the secondary effects of national registry systems, including those for health data, noting that such systems can “reveal ethnic origin or religious affiliation in a systematic manner,” raising fears of data misuse. Collective, not just individual, approaches to privacy are therefore required.

2. Cross-Cutting Themes

There is clearly a demand for open and big data in the Global South, notwithstanding the potential risks and concerns. In this data landscape, a number of sub-themes emerge that cut across the three areas of big data, open data, and data protection/privacy. These are outlined below.

2.1 A Need for Improved Legal Frameworks for Data Governance

The need for improved legal frameworks and governance emerges from the literature. In the open data context, deficits in legal governance infrastructure were linked to the fragility of open data commitments and to concerns about the overall sustainability of open data programs. In the context of big data, there were concerns that the necessary legal and governance frameworks to ensure appropriate use of big data were absent. There are calls for better privacy laws, as well as for the reform of such laws, in response to rapidly evolving technologies. In some cases, there were recommendations that existing institutions (e.g., national statistical agencies, privacy commissions) be better supported in order to enable more effective governance. Overall, there was preoccupation that governance generally lags behind the adoption

of digital technologies, and that “the world’s poorest” are the last to be heard in conversations about digital governance (Hampson, 2018).

2.2 Shifting Public- and Private-Sector Roles Raise New Challenges

Data and their associated technologies can affect the traditional roles of public- and private-sector actors. For example, in the open data context, the private sector is emerging as a source of open data. There is also potential for open government data to stimulate private-sector innovation. In the big data context, the private sector could be an important source of data for government data analytics, particularly in relation to meeting SDGs or in assessing progress towards meeting them. Mobile-communications data, remote-sensing data, and social-media data are identified as important big data sources, although it is not always clear how governments will be able to reliably access and use such data. Notwithstanding the perceived importance of private-sector data, there seems to be a consensus that national statistical offices remain important sources of data, and require financial support, capacity building, and independence. Changing public- and private-sector roles are also evident in relation to privacy. The considerable risks to privacy of massive quantities of data collected by the private sector are exacerbated when government can access these data for investigative purposes without adequate transparency and oversight.

2.3 The Development Context Raises Different Issues than in the Global North

Some open data issues may play out differently in the Global South than they do in the Global North. This can require different strategies and approaches. For example, in the Global South, achieving greater transparency through open data might be more likely to lead to the gaming of government data (Davies & Perini, 2016). And while open data in the Global North are often touted as a vehicle for stimulating innovation in the private sector, in the development context there are suggestions that it might be more appropriate for open data programs to focus on using open data for innovation in government.

Particular issues for big data include the challenges for countries of the Global South in gaining access to important and adequate data sources. This is particularly a concern since flawed data can

produce flawed analytics. Concerns were raised about the potential that certain people and/or communities might be excluded from data-sets relied upon in analytics. While bias and exclusion are also issues in the Global North, the relative paucity of data sources, as well as conditions that might lead to considerable unevenness in the data, make these issues more acute in the Global South. Gurumurthy and Chami (2018) also note that the “complexity of development requires knowledge that is contextual, requiring conventional theory-building that uses causation-based models, rather than merely correlation-based ones most often employed in big data techniques.”

The adoption and use of technologies may also play out differently in the Global North and South. For example, it was observed that some technologies with significant privacy implications are rejected by governments in the Global North, but are purchased with aid dollars for adoption and use in the Global South (Hosein, 2011). In addition to the privacy risks posed by such practices, the technologies may also not be well-adapted to local conditions, leading to the collection of incomplete or flawed data, and exacerbating issues of exclusion and disenfranchisement.

2.4 Capacity Building is Essential

Capacity building is seen as a serious need across all areas. In the open data context, capacity building was identified as a need within national statistical offices, within government (not only in operationalizing open data programs, but in learning to make use of open data), and within civil society in order to develop the skills to use open data effectively. Capacity building for big data was identified as a need for governments in order to make effective use of big data (Lokanathan & Perera-Gomez, 2016). There is a broad need for data scientists in the public-, private- and non-state-actor sectors (Lokanathan et al., 2017). There are also gender issues in capacity building (van der Spuy & Aavriti, 2017), with a “significantly lower number of females in the big data for development space” (LIRNEasia, 2017b, p. 8). Capacity building to address privacy issues is also important. The literature identifies a need for greater education and awareness around privacy issues, a need for capacity to understand, identify, and address privacy issues arising from the use of big data and other emerging technologies, and a need for greater state resources for developing and maintaining privacy protection and oversight. The *Africa Data*

Revolution Report (UNECA, 2017, p. 20) noted that there was a need for both reforms and investment if Africa “is to harness the data revolution for accelerated sustainable development.”

Capacity issues also arise in relation to gaining access to data and research that may be unavailable because it lies behind paywalls (LIRNEasia, 2017a; UNECA, 2017; van der Spuy & Aavriti, 2017). A number of papers emphasized the need for open access to research publications and research data. In addition, there were calls for more research to be carried out across each of the three themes discussed here— open data, big data, data protection/privacy—with an emphasis on the needs of the Global South. A number of papers proposed specific research agendas or identified research priorities (UNECA, 2017; van der Spuy & Aavriti, 2017; Davies & Perini, 2013, 2016; Hampson, 2018; Lokanathan & Perera-Gomez, 2016).

2.5 Data have a Complex Relationship with Sustainable Development Goals

The SDGs are closely linked to data issues, since data will be an important tool to measure progress toward meeting them (Manoj, 2017). In the context of open data, greater transparency and increased capacity to use open data within government and civil society are seen as benefits. Open data is described as “a vital part of ensuring effective monitoring of the SDG agenda, as well as improving the achievement of targets within the goals” (Open Data for Development, 2016, p. 36). Big data are seen as a means not just of measuring progress but of identifying areas where changes could be made to policy or practices in order to help meet SDGs (Lokanathan et al., 2017). A major challenge is finding sufficient sources of big data (Manoj, 2017), as well as ensuring those data sources are suitable for the purposes to which they will be put. One source notes that because the SDGs require data for assessment, they may drive a push toward improvement of data sources and an increase in supply (UNECA, 2017). The SDGs may also help shape approaches to who and what is counted, by whom and for whom. This will be crucially important. If a goal of the SDGs is to count the uncounted, then it will be necessary to find ways to ensure that the use of big data analytics does not compound the problem. It will be necessary to ensure that datasets are sufficiently inclusive so that they do not contribute to marginalization (LIRNEasia, 2017b).

In spite of the potential for data to help meet SDGs, there are reasons to be cautious. Lokanathan et al. (2017) note that using big data to address some SDGs may come at a cost. For example, the use of big data in policing may help reduce crime but may create issues around surveillance and privacy. There is also concern that if governments turn to the private sector as a source of big data, this could create new dependencies (Gurumurthy & Chami, 2018). Van der Spuy and Aavriti (2017) suggest that pressure to adopt technologies to meet SDGs can lead to adoption without adequate protection for privacy and human rights.

2.6 Gender Dimensions are Important and Insufficiently Addressed

There are important gender dimensions to data. In the open data context, for example, it is necessary to consider who gets counted in official data, as well as what subset of that data is made available as open data (Open Data for Development, 2016). Addressing gender gaps will require significant time and resources, as there are many systemic barriers. This includes a “lack of understanding of the relationship between open data and gender, and limited capacity on gender-related programming and analysis” (Open Data for Development, 2016, p. 38). Lokanathan and Perera-Gomez (2016), as noted, also raise concerns about addressing gender in big data research where women are absent from the data. For example, mobile-communications data—an important source for big data in the Global South—may be significantly under-inclusive of data about women, particularly where households share a single phone and the phone is controlled by a patriarch. The enthusiastic embrace of technology-based research methods, such as big data analytics, also carries with it the risk of marginalizing women’s ways of knowing—substituting automated analyses for experience-based and traditional knowledge (Gurumurthy & Chami, 2016). Van der Spuy and Aavriti (2017) emphasize that information communication technologies (ICTs) are not gender neutral. They suggest that although the adoption of ICTs is often linked to increased empowerment and agency, such systems can reinforce existing hierarchies. They observe that “surveillance is never gender-neutral, and reflects asymmetries in power” (2017, p. 46).

2.7 Aspects of the Data Revolution Threaten Human Rights and Civil Liberties

The literature reveals considerable concerns about the risk that under-inclusiveness in open data and in big datasets will lead to further marginalization and exclusion of certain segments of the population, including not only women but also ethnic minorities and those living in rural and remote communities. For example, the *Africa Data Revolution Report* (UNECA, 2017, p. 15) notes: “Official statistics can be and have been used as a tool for social inclusion, integration and development, but equally for social exclusion, economic extraction, exploitation and political exclusion.” As a result, it is important to keep asking the questions: “Data for whom? Controlled by whom? And to what end?” (Gurumurthy & Chami, 2018; see also UNECA, 2017; IT for Change, 2017). Nascent concepts of data sovereignty may also lead some communities to insist upon rights to their own data (Gurumurthy & Chami, 2018; IT for Change, 2017). There was some critique of techno-solutionism, and of an evolving data context in which “algorithms determine (and even replace) deliberation and discussion” (Gurumurthy & Chami, 2016). There is also the risk that participation in an increasingly automated society becomes impossible for those without technological access.

The lack of adequate legal frameworks for technology governance was clearly a concern for the protection of privacy and other human rights. A study on a biometric-identification system in Argentina raised concerns that there was insufficient transparency and oversight frameworks to ensure that the database was not used for improper purposes that could include targeting groups and individuals for discriminatory purposes, infringing their civil liberties (ADC, 2017).

Human rights that become particularly important in the big data environment include the right to be counted; the right of access to information; the right to participate in the collection, production, and dissemination of data; the right to privacy, non-discrimination, and equality; and the right to freedom of expression (UNECA, 2017).

3. Conclusion: Moving Forward—Research for the Data Society

Although there is a growing body of quality research undertaken across a range of data-related issues in the Global South, there remains room for much more. This is an area where both technology and its local adoption are evolving rapidly. As a result, there are always new issues that require consideration. The novelty of the technology, the complexity of the challenges it raises, and the diversity of its impacts across different social and economic divides means that there is an ongoing need for research into data and the Global South.

Particularly in the early days of the development and adoption of new technologies, there is a need for broad-based research that identifies key issues, whether they relate to potential applications, advantages of adoption, impacts, or governance. Such research can lay the foundation for future work by identifying key questions or unresolved issues. Broad themes that need to be further developed in future research include data sovereignty (the right of states and/or communities to control data about them), the shifting relationship of private and public sectors with respect to data, and emerging rights of individuals to control data about them. Such rights include privacy, but may go beyond conventional understandings of privacy to include concepts such as data portability.

While broad-ranging studies can fill important knowledge gaps, there is also a distinct need for research in the form of context- or technology-specific case studies. Such case studies may be particularly important in understanding successes or failures in the use of data, in identifying unanticipated problems, and in attempting to address any challenges. Case studies from the Global South will reveal and identify challenges that are particular to that context. Case studies, however, must be more than just exercises to champion or tout the successes of particular technology adoptions; they must be both rigorous and critical, enabling the identification of benefits and harms, as well as potential solutions.

In addition to case studies that focus on particular technologies, there is need for further research on governance. It is clear that data-related technologies raise significant privacy and human rights concerns. The pressure to adopt technologies that promise solutions to pressing problems may often lead to a neglect of governance issues. Research is needed to identify and examine privacy and human rights issues in the abstract, but it must also do so in the

context of concrete case studies. Governance issues include the need for effective laws and the resources to support the administration of those laws. However, there is also a need for case studies of particular governance options. The concept of data trusts, for example, is gaining ground in the Global North as a potential data governance mechanism. This might be a fruitful area of research for data governance in the Global South. In addition, research on tools other than legislation, including strategic litigation, the use of incentives, or trust systems, to cite but a few examples, is necessary and important.

References

- Acuña, J., Fulchi, L. A., & Sequera, M. (2017). *La protección de datos personales en bases de datos públicas en Paraguay*. Asunción, Paraguay : Tecnología & Comunidad. https://www.tedic.org/wp-content/uploads/2017/09/La-protecci%C3%B3n-de-Bases-de-Datos-en-Paraguay_Documento-Final.pdf
- Asociación por los Derechos Civiles (ADC). (2017). *La identidad que no podemos cambiar: Cómo la biometría afecta nuestros derechos humanos*. <https://adcdigital.org.ar/wp-content/uploads/2017/06/La-identidad-que-no-podemos-cambiar.pdf>
- Asociación por los Derechos Civiles (ADC). (2014). *The collecting state—A study about Argentina and citizens' personal data*. <https://adcdigital.org.ar/portfolio/the-collecting-state/>
- Cortés, C. (2015). *Communications surveillance in Colombia: The chasm between technological capacity and the legal framework*. Working Paper 3. Dejusticia. https://www.dejusticia.org/wp-content/uploads/2017/04/fi_name_recurso_683.pdf
- Davies, T., & Perini, F. (2016). Researching the emerging impacts of open data: Revisiting the ODDC conceptual framework. *The Journal of Community Informatics*, 12(2), 148–178. <http://ci-journal.org/index.php/ciej/article/view/1281>
- Davies, T., & Perini, F. (2013). *Researching the emerging impacts of open data: ODDC conceptual framework* (ODDC Working Papers No. 1). <http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf>
- Gurumurthy, A., & Chami, N. (2018). *Data frameworks for a right to development*. United Nations Research Institute for Social Development. <http://www.unrisd.org/TechAndHumanRights-Gurumurthy-Chami>
- Gurumurthy, A., & Chami, N. (2016). *Data: The new four-letter word for feminism*. GenderIT. <https://www.genderit.org/articles/data-new-four-letter-word-feminism>

- Hampson, F. (2018). *Improving Internet governance: Support to the global commission on Internet governance*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/56925>
- Hickok, E., Chattapadhyay, S., & Abraham, S. (Eds.). (2017). *Big data in governance in India: Case studies*. Centre for Internet & Society. <https://cis-india.org/internet-governance/files/big-data-compilation.pdf>
- Hosein, G. (2011). *Privacy and developing countries*. Office of the Privacy Commissioner of Canada. https://priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2011/hosein_201109/
- Hosein, G., & Nyst, C. (2014). *Aiding surveillance: An exploration of how development and humanitarian aid initiatives are enabling surveillance in developing countries*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/52191>
- IT for Change. (2017). *Policy frameworks for digital platforms: Moving from openness to inclusion* [Background paper]. http://itforchange.net/sites/default/files/1377/Background%20Paper_Platform%20Policies_August%202017.pdf
- LIRNEasia. (2017a). *Big data and SDGs: The state of play in Sri Lanka and India*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/56907/IDL-56907.pdf>
- LIRNEasia. (2017b). *Mapping big data for development and the global goals*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/56905/IDL-56905.pdf>
- LIRNEasia. (2018). *Annex 19: Code of practice for the secondary use of mobile network big data*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/56918/56976.pdf>
- Lokanathan, S., & Perera-Gomez, T. (2016). *Workshop report on shaping a research and policy agenda on big data for development in the global south*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/56910>
- Lokanathan, S., Perera-Gomez, T., & Zuhyle, S. (2017). *Mapping big data solutions for the Sustainable Development Goals (Draft)*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/56906>
- Manoj, M. (2017). *Big data governance frameworks for "Data Revolution for Sustainable Development."* Centre for Internet & Society. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/56914>
- Open Data for Development. (2016). *Open data for development: Building an inclusive data revolution*. IDRC. <http://od4d.net/result/od4d-annual-report-2016/>
- Open Data Institute. (2016). *Open data leaders network digest*. Open Data Institute. Available at: <http://od4d.net/wp-content/uploads/2017/05/Open-Data-Leaders-Network-Digest.pdf>
- Open Data Research Network. (2013). *Open data and privacy: Discussion notes*. <http://www.opendataresearch.org/content/2013/501/open-data-privacy-discussion-notes.html>

- Samarajiva, R., & Perera-Gomez, T. (2018). *Bulk data: Policy implications (Draft)*. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/56922>
- Scott-Railton, J., Marczak, B., Abdul Razzak, B., Crete-Nishihata, M., & Deibert, R. (2017). *Reckless V: Director of Mexican anti-corruption group targeted with NSO group's spyware*. The Citizen Lab. <https://citizenlab.ca/2017/08/nso-spyware-mexico-corruption/>
- Smith, M. L. (2018). *Artificial intelligence and human development* [White paper]. Ottawa: IDRC. Available at: https://www.idrc.ca/sites/default/files/ai_en.pdf
- Surendra, A., Lokanathan, S., Lasantha, F., & Gomez, T. (2017). *Predicting socio-economic characteristics of the Sri Lankan population using call detail records (CDRs)* [Policy brief]. IDRC Digital Library. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/58172>
- United Nations Economic Commission for Africa (UNECA). (2017). *The Africa data revolution report 2016: Highlighting developments in African data ecosystems*. Prepared by the UNECA, the United Nations Development Programme, the International Development Research Centre of Canada, and the World Wide Web Foundation. <https://www.africa.undp.org/content/dam/rba/docs/africa-data-revolution-report-2016.pdf>
- Van der Spuy, A., & Aavriti, N. (2018). *Mapping research in gender and digital technology*. Association for Professional Communications. https://www.apc.org/sites/default/files/IDRC_Mapping_0323_o.pdf

About the Authors

Teresa Scassa is Canada Research Chair in Information Law and Policy at the University of Ottawa, Faculty of Law. She is the author or co-author of several books, including *Digital Commerce in Canada* (LexisNexis, 2020), *Canadian Trademark Law* (LexisNexis, 2015, 2nd ed.), and *Law Beyond Borders* (Irwin Law, 2014); is co-editor of *Artificial Intelligence and the Law in Canada* (LexisNexis, 2021) and *Law and the Sharing Economy* (University of Ottawa Press, 2018); and has written widely in the areas of intellectual-property law, law and technology, and privacy. Scassa is a member of the Canadian Advisory Council on Artificial Intelligence and of the Geothink research partnership. She is also Senior Fellow with the Centre for International Governance Innovation's International Law Research Program.

Fernando Perini is Senior Program Specialist at the International Development Research Centre. Part of the Democratic and Inclusive

Governance Division, he coordinates a range of efforts on data for development, including open data, feminist open government, digital rights, and artificial intelligence for development. With 12 years of experience in programs and acting as the IDRC Regional Director for Latin America and the Caribbean, he has led a large global portfolio and donor partnerships. He has written extensively in the topics of data, management, governance, technology, and development, including co-editing *The State of Open Data: Histories and Horizons* (African Minds and IDRC, 2019) and serving as technical advisor for the *World Development Report 2021: Data for Better Lives* (World Bank, 2021). Perini has extensive international experience as a researcher, consultant, and lecturer, and holds a PhD from the University of Sussex (Science Policy Research Unit) in the management of technology and innovation.

