7

Big Data Analytics and Human Rights

Privacy Considerations in Context

*Mark Latonero**

The technology industry has made lucrative use of big data to assess markets, predict consumer behavior, identify trends, and train machine-learning algorithms. This success has led many to ask whether the same techniques should be applied in other social contexts. It is unquestionable that new information and communication technologies bring both benefits and costs to any given domain. And so, when it comes to the human rights context, with higher stakes due to vulnerable populations, the potential risks in applying the technologies associated with big data analytics deserve greater consideration.

This chapter argues that the use of big data analytics in human rights work creates inherent risks and tensions around privacy. The techniques that comprise big data collection and analysis can be applied without the knowledge, consent, or understanding of data subjects. Thus, the use of big data analytics to advance or protect human rights risks violating privacy rights and norms and may lead to individual harms. Indeed, data analytics in the human rights monitoring context has the potential to produce the same ethical dilemmas and anxieties as inappropriate state or corporate surveillance. Therefore, its use may be difficult to justify without sufficient safeguards. The chapter concludes with a call to develop guidelines for the use of big data analytics in human rights that can help preserve the integrity of human rights monitoring and advocacy.

"Big data" and "big data analytics" are catchphrases for a wide range of interrelated sociotechnical techniques, tools, and practices. Big data involves the collection of large amounts of data from an array of digital sources and sensors. Collection often occurs unbeknownst to those who are data subjects. In big data, the subjects are the individuals creating content or emitting data as part of their everyday lives (e.g., posting pictures on social media, navigating websites, or using a smartphone

* Zachary Gold, JD, a research analyst at the Data & Society Research Institute, contributed research to an early version of this work that was presented as a conference paper.

https://doi.org/10.1017/9781316838952.007 Published online by Cambridge University Press

with GPS tracking operating in the background). This data can be collected, processed, analyzed, and visualized in order to glean social insights and patterns. Behavioral indicators at either the aggregate or individual level can be used for observation, decision-making, and direct action.

Privacy is a fundamental human right.[1] As those in the human rights field increasingly address the potential impact of new information and communication technologies,[2] privacy is of particular concern. Indeed, as G. Alex Sinha states in Chapter 12, since the Edward Snowden revelations in 2013, "perhaps no human rights issue has received as much sustained attention as the right to privacy." Digital technologies have called into question the traditional expectations of privacy, including the right to be free from interference with one's privacy and control over one's personal information, the ability to be left alone, and the right to not be watched without permission. As the technology and ethics scholar Helen Nissenbaum states, "information technology is considered a major threat to privacy because it enables pervasive surveillance, massive databases, and lightning-speed distribution of information across the globe."[3]

The emergence of details about the use of pervasive surveillance technology in the post-Snowden era has only heightened anxieties about the loss of privacy. According to a 2014 report from a UN Human Rights Council (UNHRC) meeting on the right to privacy in the digital age, the deputy high commissioner noted that

> digital platforms were vulnerable to surveillance, interception and data collection. . . [S]urveillance practices could have a very real impact on peoples' human rights, including their rights to privacy, to freedom of expression and opinion, to freedom of assembly, to family life and to health. In particular, information collected through digital surveillance had been used to target dissidents and there were credible reports suggesting that digital technologies had been used to gather information that led to torture and other forms of ill-treatment.[4]

The UNHRC report gives examples of the risks of surveillance technologies, which expose sensitive information that can produce harms to political freedoms and

[1] *See* Universal Declaration of Human Rights, December 10, 1948, U.N. G.A. Res. 217 A (III), art. 12; International Covenant on Civil and Political Rights, December 16, 1966, S. Treaty Doc. No. 95–20, 6 I.L.M. 368 (1967), 999 U.N.T.S. 171, art. 17.

[2] M. Land et al. demonstrate how the impact of emerging information and communication technologies can be examined from a rights-based framework and international human rights law. M. K. Land et al., "#ICT4HR: Information and Communication Technologies for Human Rights," World Bank Institute, Nordic Trust Fund, Open Development Technology Alliance, and ICT4Gov, November 2012, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2178484.

[3] H. Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Palo Alto, CA: Stanford University Press, 2010).

[4] United Nations Human Rights Council, "Summary of the Human Rights Council panel discussion on the right to privacy in the digital age," December 19, 2014, www.un.org/en/ga/search/view_doc.asp?symbol=A/HRC/28/39.

physical security. The role of big data analytics in perpetuating anxieties over surveillance will be discussed later in this chapter, after highlighting the importance of understanding the human rights contexts in which big data analytics might transgress privacy norms. The chapter will first take a closer look at what is meant by privacy in relation to the technologies that comprise big data analytics in the human rights context.

I BIG DATA ANALYTICS FOR MONITORING HUMAN RIGHTS: COLLECTION AND USE

Assessing the legitimate application of big data analytics in human rights work depends on understanding what the right to privacy protects. Nissenbaum's framework of "contextual integrity" can provide a way to understand the value of privacy within the human rights context and the situations in which the use of big data might infringe this right. Contextual integrity ties "adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it."[5] Nissenbaum does not go down the perilous path of trying to define privacy in absolute terms or finding a precise legal definition against the thicket of competing legal regimes, and neither will this chapter. Instead, she discusses privacy in terms of an individual's right to determine the flow of information about him- or herself.[6] What individuals care about, Nissenbaum argues, is that their personal information flows appropriately depending on social context. This chapter will employ similar terminology, such that privacy is violated when an individual's information is collected, analyzed, stored, or shared in a way that he or she judges to be inappropriate.

One challenge with examining whether a specific technology may violate privacy is that technology is not a single artifact that exists by itself. Technology is a combination of sociotechnical processes. Thus, it is useful to divide the technologies that comprise big data broadly into two categories: collection and use. As Alvaro Bedoya notes, most questions dealing with data privacy and vulnerable populations focus on how data will be used rather than how it will be collected.[7] Yet collection and use are intrinsically tied together. Disaggregating these categories into their individual processes provides us with a better understanding of potential privacy concerns related to human rights monitoring.

[5] H. Nissenbaum, "Privacy as Contextual Integrity" (2004) 79(1) *Washington Law Review* 119–58.
[6] Nissenbaum, *Privacy in Context*.
[7] A. Bedoya, "Big Data and the Underground Railroad," *Slate*, November 7, 2014, www.slate.com/articles/technology/future_tense/2014/11/big_data_underground_railroad_history_says_unfettered_collection_of_data.html.

https://doi.org/10.1017/9781316838952.007 Published online by Cambridge University Press

A *Collection*

Discovery, search, and crawling are activities that involve finding data sources that may contain information relevant to purpose, domain, or population. Data sources can be publicly available; for example, Twitter tweets, articles on online news sites, or images shared freely on social media.[8] Other sources, such as Facebook posts, are quasi-public in that they contain data that may be intended to be accessible only to specific members of an online community with appropriate login credentials and permissions. Other data sources, such as the e-mail messages of private accounts, are not publicly searchable. Even the collection of data that is publicly available can violate the privacy expectations of Internet users whose data is being collected. These users have their own expectations of privacy even when posting on sites that are easily accessible to the public. Users may feel that their posts are private, intended only for their friends and other users of an online community.[9]

Scraping involves the actual collection of data from online sources to be copied and stored for future retrieval. The practice of scraping can have an impact on individual Internet users as well. With scraping, users' data may be collected by an entity unbeknownst to them, breaching their privacy expectations or community/social norms.

Classification and indexing involve categorizing the collected data in a structured way so it can be searched and referenced. The data can be classified according to social categories created by the data collector or holder, such as name, gender, religion, or political affiliation. Classification extends the privacy risk to individual Internet users whose data has been collected. The subjects' data, put in a database, is now organized in a way that may not correctly represent those subjects or may expose them if the data were inadvertently released. Placing subjects' personally identifiably data into categories that may be incorrect may cast those in the dataset in a false light.

Storing and retention of large quantities of data is becoming more prevalent as storage become less expensive. This situation means that more data can be kept for longer periods of time by more entities. A post someone may have thought was fleeting or deleted can persist in numerous unseen databases effectively in perpetuity. Storing data for long periods of time exposes users to unforeseen privacy risks. Weak information security can lead to leaks or breaches that reveal personal data to others whom either the collectors or users did not intend to inform. This could expose individuals to embarrassment, extortion, physical violence, or other harms.

[8] See Chapter 6. *See also* J. Aronson, "Mobile Phones, Social Media, and Big Data in Human Rights Fact-Finding: Possibilities, Challenges, and Limitations," in P. Alston and S. Knuckey (eds.), *The Transformation of Human Rights Fact-Finding* (Oxford: Oxford University Press, 2015).

[9] In Chapter 12, G. Alex Sinha discusses the need to determine when revealing information to others constitutes a waiver of the human right to privacy.

B *Use*

Big data analytics involves deploying a number of techniques and tools designed to find patterns, behavioral indicators, or identities of individuals, groups, or populations. Structuring data, performing statistical modeling, and creating visualizations transform otherwise incomprehensible datasets into actionable information.

The threat to privacy from the use of big data analytics is clear. The entity performing the analysis could learn more about a person's life than would be anticipated by a typical citizen, thereby violating the right to determine the flow and use of one's personal information. By combining disparate data sources, these entities may be able to link online identities to real-world identities or find out about a person's habits or personal information.[10] There is also a risk is that the analysis is wrong. Datasets, and the analyses carried out on them, always carry some form of bias, and such analyses can lead to false positives and negatives that decision-makers may later act on. Deploying resources to the wrong place or at the wrong time can cause significant harm to individuals. Even if the analysis is correct in identifying a human rights violation, the victims may be put at greater risk if publicly identified due to community stigma or retaliation by perpetrators.

Access and sharing applies to both data use and collection. The way the data is indexed or classified or the type of data collected can already reveal information considered private. For some human rights issues, personally identifiable information is needed to monitor, assess, and intervene in real time. Unauthorized sharing and access presents a major breach of privacy norms in the human rights context.

II PRIVACY TRADE-OFFS, URGENCY, AND TEMPORALITY

Privacy considerations associated with applying big data analytics in the human rights context has received considerably less attention in the literature than issues like representativeness and validity. A recent volume edited by Philip Alston and Sarah Knuckey discusses new advances in technology and big data that are poised to transform the longstanding activity of human rights fact-finding.[11] In his contribution to that volume, Patrick Ball raises major reservations about using big data for human rights. Ball argues that sampling procedures and the data itself in many big datasets are riddled with biases.[12] Regardless of its "big-ness," data that is not

[10] "The 'mosaic theory' describes a basic precept of intelligence gathering: Disparate items of information, though individually of limited or no utility to their possessor, can take on added significance when combined with other items of information." D. E. Pozen, "The Mosaic Theory, National Security, and the Freedom of Information Act" (Note) (2005) 115 *Yale Law Journal* 628–79.

[11] Alston and Knuckey (eds.), *The Transformation of Human Rights Fact-Finding*.

[12] P. Ball, "The Bigness of Big Data," in Alston and Knuckey (eds.), *The Transformation of Human Rights Fact-Finding*, p. 428.

https://doi.org/10.1017/9781316838952.007 Published online by Cambridge University Press

representative leaves out key populations. Incomplete data can hide the very populations most affected by human rights violations. For example, when estimating the number of victims in a conflict based on available data, those who were secretly killed by paramilitary groups and never reported in the media or government records are left out. The models and analyses resulting from such flawed data, Ball argues, can lead to faulty conclusions that could imperil or discredit human rights fact-finding and evidence gathering.

Patrick Meier, on the other hand, suggests an alternative perspective, buoyed by the high-profile applications of big data analytics in humanitarian crises.[13] From crowdsourced maps after the Haiti earthquake in 2009 to millions of tweets related to Hurricane Sandy in New York, such data is important and can help save lives. Meier concedes that big data can be biased and incomplete. Yet data and information on vulnerable populations are almost always lacking in completeness, even more so in the immediate aftermath of a crisis. Thus, big data, for all its flaws, can serve to inform decision-making in real time (i.e., during a crisis event) where comprehensive information does not exist.

One of the core questions that needs to be answered when using big data for human rights purposes is the extent to which the urgency of the need being addressed impacts the decision to use imperfect data and risk privacy violations. Consider, on the one hand, a fact-finding mission to ascertain whether a human rights violation took place in the past. Sometimes the data collection can take months or years in order to produce evidence for use in justice and accountability proceedings or for historical clarification. On the other hand, in a humanitarian response to a disaster or crisis, data collection seeks to intervene in the present, to find those in immediate need. Of course, temporality is not an absolute rule separating human rights and humanitarian domains. Data can be collected both in protracted humanitarian situations and when human rights violations are happening in the present. The issue at hand is whether an urgent situation places unique demands on the flow of personal information, which impacts one's dignity, relationships with others, and right to privacy.

During humanitarian responses to time-bound crises like natural disasters, decisions must often be made between maintaining privacy and responding quickly by using available data that is often deeply personal. For example, in the response to the Ebola crisis, humanitarian organizations deliberated on how big data could be used legitimately.[14] Consider a responder requesting the mobile phone contact list of a person who tested positive for Ebola in an attempt to stop the spread of the disease.

[13] P. Meier, *Digital Humanitarians* (Boca Raton, FL: CRC Press, 2015); P. Meier, "Big (Crisis) Data: Humanitarian Fact-Finding with Advanced Computing," in Alston and Knuckey (eds.), *The Transformation of Human Rights Fact-Finding*.

[14] B. Campbell and S. Blair, "How 'big data' could help stop the spread of Ebola," *PRI's The World*, October 24, 2014, www.pri.org/stories/2014-10-24/how-big-data-could-help-stop-spread-ebola.

Further, consider a response organization asking a mobile phone company for the phone numbers and records of all the users in the country in order to trace the network of individuals who may have become infected. That data would need to be analyzed to locate those at risk, and personal information might be shared with other responders without the consent of the data subjects.

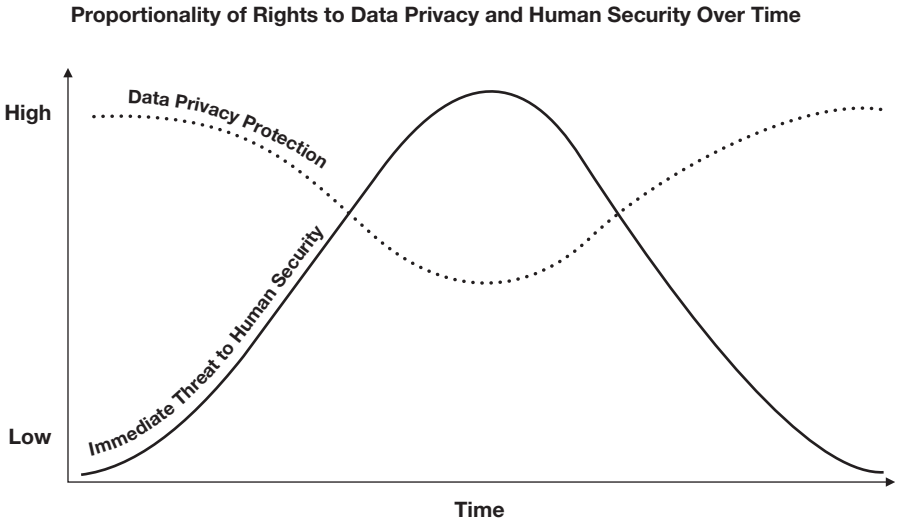
The assumption in such an operational decision is that saving the lives of the person's contacts and protecting the health of the broader public outweigh the privacy concerns over that person's personal information. Yet such decisions about trade-offs, particularly in the case of the Ebola response, remain highly controversial due to the potential privacy violations inherent in collecting the mobile phone records of individuals at a national level without the consent of individuals.[15] Even in an urgent humanitarian response context, there is little agreement about when it is appropriate and legitimate to limit privacy rights. Applying real-time data collection and analytics to the investigation of historical human rights violations could raise even more concerns.

When, if ever, is it appropriate to limit privacy rights in order to achieve human rights objectives? Big data can support human rights fact-finding by providing a basis for estimating the probability that an individual falls within a group that has been subjected to past rights violations. Data analytics can also personally identify an individual whose rights are being violated in the present. For example, access to millions of mobile phone records in a large city may reveal patterns of calls by individuals that suggest human trafficking for commercial sexual exploitation. An analysis of these "call detail records" can indicate the movement of people to buildings where exploitation is known to take place at unique times and pinpoint calls with known exploiters. Yet to establish whether a specific individual's rights have been violated, personal information like cell phone numbers, names, and building ownership would need to be identified. Such data can reveal sensitive information (e.g., religion, political affiliation, or sexual orientation) that could lead to physical harm and retribution if shared with the wrong party.

Ultimately, permissible limitations on the right to privacy will vary depending on the urgency and importance of the objective to be achieved. The accompanying graph gives a visual sketch of the potential trade-offs involved in privacy. The right to privacy is not an absolute right, and it may be limited in order to achieve a legitimate objective. The extent to which the right is infringed, however, must be proportional to the objective. Thus, the graph illustrates a scenario involving a decision about the proportional relationship between the urgency of the objective of human security and permissible limits on privacy. In a human rights context, protecting the right to privacy is of high concern since safeguarding an individual's data also protects the right to freedom of expression, association, and related rights. Yet, when there is an

[15] S. McDonald, "Ebola: A Big Data Disaster," The Centre for Internet and Society, March 1, 2016, http://cis-india.org/papers/ebola-a-big-data-disaster.

increasing threat to human security and safety, such as the imminent danger of an individual's death due to violent conflict, the concern over privacy may start to decrease. Perhaps a decision must be made about whether to obtain or release an individual's personal mobile phone's GPS coordinates without their consent or permission in order to attempt to locate and rescue that person to save his or her life. There may come an inflection point in the situation where the immediate danger to human security is higher than the protection of an individual's privacy. When there is no longer an immediate threat to human security, there is every reason to uphold the integrity of the individual's privacy right at a high level. Of course, there are a number of additional factors that might influence this analysis and decision. In some cases the release of an individual's data will expose that person to the very forces threatening his or her safety and security or may put another individual directly at risk. Clearly, more foundational work is needed to begin to understand how these trade-offs may be operationalize in practice when making a real-time decision.



Does this mean that we cannot use big data collection for long-term and continuous human rights monitoring? Long-term monitoring of digital data sources for human rights violations holds the promise of providing unprecedented insight into hidden or obscured rights abuses such as labor violations, sexual abuse/exploitation, or human trafficking. The use of big data analytics for monitoring human rights, even by human rights organizations, carries both known and unknown risks for privacy. Indeed, the very nature of data collection and analysis can conflict with normative expectations of privacy in the human right context. It is unclear whether

https://doi.org/10.1017/9781316838952.007 Published online by Cambridge University Press

the uncertain benefits of long-term human rights monitoring and advocacy can outweigh the very concrete risks to privacy that accompany the use of big data.

III HUMAN RIGHTS MONITORING AND SURVEILLANT ANXIETIES

The human rights field has a long tradition of employing methods and techniques for protecting privacy while monitoring rights abuses and violations. Determining "who did what to whom" involves identifying victims and alleged violators, doing an accounting of the facts, and collecting relevant information.[16] Ensuring the privacy and confidentiality of sources and victims is a critical concern in human rights fact-finding. A training manual published by the UN Office of the High Commissioner for Human Rights (OHCHR), for example, urges the use of monitoring practices that "keep in mind the safety of the people who provide information," seeking consultation in difficult cases and maintaining confidentiality and security.[17] At the local level, Ontario's Human Rights Commission states that data collection should include informing the public, consulting with the communities that will be affected, using the least intrusive means possible, assuring anonymity where appropriate, and protecting privacy.[18]

In reality, though, ethical data collection protocols that protect privacy in the human rights context, such as obtaining informed consent, are extraordinarily difficult to utilize when deploying big data analytics. Big data analytics often relies on "found" data, which is collected without a user's consent or even knowledge. It also necessarily involves using this information in ways not intended by the individual to whom it relates. Thus, with respect to big data, privacy harms are likely unavoidable.

The use of big data also harms privacy by adding to a growing, although ambiguous, sense of "surveillant anxiety," in which we fear surveillance to the point that it affects our thoughts, behaviors, and sense of self. Kate Crawford describes this anxiety as "the fear that all the data we are shedding every day is too revealing of our intimate selves but may also misrepresent us... [N]o matter how much data they have, it is always incomplete, and the sheer volume can overwhelm the critical signals in a fog of possible correlations."[19] A fact-finding mission that creates privacy

[16] UN Human Rights Office of the High Commissioner, *Monitoring Economic, Social and Cultural Rights*, HR/P/PT/7/Rev. 1, 2011, www.ohchr.org/Documents/Publications/Chapter20-48pp.pdf.

[17] UN Human Rights Office of the High Commissioner, *Training Manual on Human Rights Monitoring: The Monitoring Function*, March 21, 1999, www.ohchr.org/Documents/Publications/training7part59en.pdf, § A–C, G, J–K.

[18] Ontario Human Rights Commission, "Count me in! Collecting human rights based data – Summary (fact sheet)," www.ohrc.on.ca/en/count-me-collecting-human-rights-based-data-summary-fact-sheet.

[19] K. Crawford, "The Anxieties of Big Data," *The New Inquiry*, May 30, 2014, https://thenewinquiry.com/the-anxieties-of-big-data/.

https://doi.org/10.1017/9781316838952.007 Published online by Cambridge University Press

risks is not equivalent to surveillance. And certainly not all surveillance technologies violate privacy in a legal sense. Yet any use of big data analytics to monitor human rights creates concerns and anxieties about surveillance, whether or not the surveillance is intended for "good."

Thus, the issue that must be addressed is whether human rights researchers using big data analytics would themselves produce this kind of surveillant anxiety in their data subjects in ways that feel similar to traditional government or corporate surveillance. According to the Special Rapporteur on the right to Privacy, surveillance creates privacy risks and harms such that "increasingly, personal data ends up in the same 'bucket' of data which can be used and re-used for all kinds of known and unknown purposes."[20] Although surveillance for illegitimate purposes necessarily violates privacy, even surveillance for legitimate purposes will do so if the associated privacy harms are not proportional to that purpose. For example, a public health surveillance program that continuously monitors and identifies disease in a population might constitute an appropriate use for a common good shared by many, but could still violate privacy if it produces undue harms and risks. As Jeremy Youde's study of biosurveillance contends:

> [T]he individual human right to privacy had the potential to be eroded thought the increased use of biosurveillance technology by governments and international organizations, such as WHO. This technology requires an almost inevitable intrusion into the behaviours, habits, and interests of individuals – collecting data on individual entries into search engines, Facebook entries in individual travel history and purchases.[21]

In their work on disease surveillance in the United States, Amy L. Fairchild, Ronald Bayer, and James Colgrove document the conflict around public health surveillance during the early days of the AIDS crisis and the struggle to balance the need to collect and share medical records against charges of institutional discrimination against marginalized groups.[22]

Big data collection and analysis by technology companies, sometimes called corporate surveillance, can produce the same kinds of anxieties. And big data collection by either governments or human rights organizations often rely on technologies that serve as intermediaries to the digital life of the public. Both a government agency and a human rights organization may collect data on the lives of millions of individuals from major social media platforms like Facebook. The very same tools, techniques, and processes in the collection and use of big data can be

[20] UN Human Rights Council, *Report of the Special Rapporteur on the Right to Privacy, Joseph Cannataci*, U.N. Doc. A/HRC/34/60 (February 24, 2017), p. 9 ("Cannataci Report").

[21] S. Davies and J. Youde (eds.), *The Politics of Surveillance and Response to Disease Outbreaks: The New Frontier for States and Non-State Actors* (London: Routledge, 2016), p. 3.

[22] A. Fairchild, R. Bayer, J. Colgrove, *Searching Eyes: Privacy, the State, and Disease Surveillance in America* (Berkeley: University of California, 2007).

employed to both violate and protect human rights. The collection of one's personal data by governments, corporations, or any number of other organizations using big data analytics may contribute to the constant feeling of being watched and curtail privacy and freedom of expression.

The pressing question is whether the use of big data by human rights organizations is permissible, given the risks to privacy involved. It is fair to say that human rights organizations should be held to the same standards that privacy advocates require of government. Yet the best intentions of human rights organizations using big data are not enough to protect privacy rights or automatically justify privacy violations. Furthermore, any organization collecting, classifying, and storing sensitive human rights data needs to address issues like data protection, secure storage, safe sharing, and access controls. If a human rights organization deploys data collection and analytic tools, how can they incorporate safeguards that responsibly address the inherent risks to privacy and minimize the potential harms?

IV TOWARD GUIDELINES FOR BIG DATA APPLICATIONS IN HUMAN RIGHTS

Because of concerns about privacy, value trade-offs, surveillance, and other potential risks and harms that may befall vulnerable populations, a rigorous assessment of the legitimacy and impact of the use of any big data analytics by organizations in the human rights context is vital. This begs the question of what type of guidelines could help steer such an assessment, particularly given the global proliferation of technologies and the plethora of context-specific harms. As the UN Special Rapporteur on Privacy states, "The nature of trans-border data flows and modern information technology requires a global approach to the protection and promotion of human rights and particularly the right to privacy."[23] Human rights monitoring organizations may need to update their standards for data collection and analysis to take new technologies like big data into account.

New technological applications do not necessarily require entirely new high level principles. For example, the principles of safety, privacy, and confidentiality outlined in the OHCHR Training Manual would still apply to big data, but these principles may need further elaboration and development when they are applied to new information collection regimes. At the same time, new technological challenges need new solutions. For example, confidentiality policies may not be achieved simply through anonymization techniques alone, since the currently accepted fact in computer science is that no dataset can be fully anonymized.[24]

[23] Cannataci Report, p. 9.

[24] *See* P. Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2009) 57 *UCLA Law Review* 1701–88; A. Narayanan and E. Felton, "No silver bullet: De-identification still doesn't work," July 29, 2014, http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf.

A major advance in addressing the ethical and responsible use of data is the Harvard Humanitarian Initiative's Signal Code, subtitled "A Human Rights Approach to Information during Crisis," which focuses on the right to information in times of crises. Bridging the humanitarian and human rights contexts, the code states that the right to privacy is fundamental in crisis situations and that "[a]ny exception to data privacy and protection during crises exercised by humanitarian actors must be applied in ways consistent with international human rights and humanitarian law and standards."[25] Other standards in the code include giving individuals the right to be informed about information collection and use as well as the right to have incorrect information about themselves rectified.

Policy work from other fields, such as international development, will also be relevant to creating guidelines for human rights. UN Global Pulse, the data innovation initiative of the UN Secretary General, has published its own "privacy and data protection principles"[26] for using data in development work. These principles recommend that actors "adhere to the basic principles of privacy," "maintain the purpose for data," and ensure "transparency [as] an ongoing commitment."[27] Human rights organizations may also learn lessons from civil society's demands of governments engaging in surveillance. The International Principles on the Application of Human Rights to Communications Surveillance suggests a number of guidelines designed to ensure that government surveillance does not infringe on the right to privacy, including notifying people when they are being watched.[28]

This brings us to the issue of how to ensure the accountability of human rights organizations using big data for interventions or monitoring. Any actor that uses big data to intervene in or monitor human rights, whether a small domestic NGO or a large international organization, should be responsible for the potential risks and harms to the very populations it seeks to help. Yet since human rights organizations often hold governments to account, asking governments to regulate the use of big data by those very organizations will likely provide an avenue for state suppression of human rights monitoring. As such, traditional regulatory mechanisms are unlikely to be effective in this context.

One emerging framework that seeks to regulate any entity engaged in data collection is the EU's General Data Protection Regulation, which comes into

[25] F. Greenwood, et al., "The Signal Code: A Human Rights Approach to Information during Crisis," Harvard Humanitarian Initiative, January 2017, http://hhi.harvard.edu/publications/signal-code-human-rights-approach-information-during-crisis.

[26] UN Global Pulse, "Privacy and Data Protection Principles," www.unglobalpulse.org/privacy-and-data-protection; *see also* UN Global Pulse, "Workshop on ICT4D Principle 8: Address Privacy & Security in Development Programs," www.unglobalpulse.org/events/workshop-ict4d-principle-8-address-privacy-security-development-programs.

[27] UN Global Pulse, Unpublished report on data privacy and data security for ICT4D (2015).

[28] Necessary and Proportionate, "International Principles on the Application of Human Rights to Communications Surveillance," May 2014, https://necessaryandproportionate.org/.

force in 2018.[29] This regulation requires all organizations collecting data on EU residents to follow privacy directives about data collection, protection, and consent. For example, individuals in the EU would have the right to withdraw consent given to organizations that collect or process their data. Such organizations would be required to alert the authorities if a data breach of personal information occurred. Any organization, including both companies and nonprofits, could incur heavy fines for noncompliance, such as 4 percent of global revenue.

Another possible avenue may lie in encouraging human rights organizations to engage with technology privacy professionals to assess the possible use of new technologies like big data. Professionals with the appropriate technical, legal, and ethical expertise may be capable of conducting privacy impact assessments of risks that may harm vulnerable populations *before* a new technology is deployed.[30] Or perhaps large donors can require grantees to follow data privacy protections as a condition of ongoing funding.

At the end of the day, though, the lack of certainty about the effectiveness of these approaches speaks to the pressing need for both more foundational research and context-specific assessments in this area. Addressing questions about context would require researchers to include more direct input and participation of the vulnerable communities themselves; for example, through field research. As Nissenbaum suggests, protecting privacy is about upholding the contextual integrity of the underlying norms governing information flow. Understanding norms around data privacy and consent, for example, should necessarily involve the communities where those norms exist.

Applying big data analytics in human rights work reveals tensions around privacy that need to be resolved in order to guide current thinking and future decisions. Unfortunately, a sustained knowledge gap in this area puts vulnerable populations at greater risk. And the anxieties and trade-offs around norms and interventions will be compounded as the human rights field addresses "newer" technologies such as artificial intelligence (AI).[31] Since AI is fueled by big data collection and analytics the same privacy concerns discussed above would apply. Furthermore, AI entails some form of automated decision making, which creates dilemmas over whether only a human rights expert, rather than a computer algorithm, can decide about proportionality and trade-offs between rights in real time. A research agenda should work toward guidelines, principles, and practices that *anticipate* the risks, costs, and benefits inherent in each process involved in emerging technological interventions in the human rights context.

[29] General Data Protection Regulation (Regulation [EU] 2016/679), 2–17, http://data.consilium.europa.eu/doc/document/ST-5419-2016-INIT/en/pdf.

[30] See Lea Shaver's chapter (Chapter 2) for an argument that the right to science also requires such an assessment.

[31] See Enrique Piracés's chapter (Chapter 13).

https://doi.org/10.1017/9781316838952.007 Published online by Cambridge University Press