

Resilient nations.

DISCUSSION PAPER

Data Philanthropy, International Organizations and Development Policy: Ethical Issues to Consider

April 2020



Contents

PART ONE: DATA PHILANTHROPY EXPLORED	5
Unpacking Data Philanthropy	5
What is "Good"?	6
The Benefits of Data Sharing	7
Modes of Data Sharing	8
PART TWO: CHALLENGES IN THE CONTEXT OF DEVELOPMENT POLICY	9
Value of the Data	9
Privacy Concerns	
Sustainability and Capacity Building	
PART THREE: MOVING FORWARD	15
A Trusted Data Sharing Platform	15
Safe Sharing Sites	15
CONCLUSION	17

This discussion paper was prepared by Charlotte M. Westbrook, LL.M. Candidate (International Legal Studies), May 2019, New York University School of Law, as part of the NYU Law International Finance and Development Fellowship program.

The fellowship assignment was supervised by Serge Kapto, Policy Specialist, Data for Development, SDG Integration Team, Bureau for Policy and Programme Support, UNDP

© United Nations Development Programme

The analysis and recommendations expressed in this discussion paper do not necessarily reflect the official views and positions of the United Nations, its agencies or its Member States. Textual material may be freely reproduced with proper citation and/or attribution to the authoring agency.

Introduction

In an increasingly datafied society, possessing big data means being able to save lives.¹ In emergency situations, big data can help organizations pinpoint the location of impacted individuals and allocate aid resources and response teams accordingly.² Most of this information can be gleaned from where people check their phones following disasters or which terms are used when communicating on social media.³ However, big data can also assist in the handling of non-emergency situations. Companies' data can tell us not only whether a given policy intervention works or doesn't work, but also how it can be fixed.⁴ As a result, a

As beneficiaries of "data philanthropy", international development organizations must consider the legal and ethical risks inherent in data sharing agreements, beyond privacy risks. growing number of international development, humanitarian and public sector organizations have called for private digital data to be shared for the public good.⁵ As a result, a "responsible data" or "data for good" movement has emerged, which calls on companies to share their data for philanthropic purposes.

The "data for good" movement promotes data-driven projects that can increase the efficiency of social initiatives, and has highlighted the significant role that the private sector can play by providing valuable data to further social action. Data generated via platforms such as telecom operators, satellite companies, and social media networks has the potential to provide a range of insights into development issues that can assist immeasurably with social research.⁶ To this end, "data philanthropy" is a term that has been used to describe the giving of private sector data, access to it, or the production of data-driven insights for a socially beneficial purpose.⁷ Reportedly coined by World Economic Forum Chief Technology Officer Brian Behlendorf at the World Economic Forum in 2011, the definition of "data philanthropy" remains unsettled.⁸ However, this paper chooses to operate within the parameters of the definition developed by Professor Yafit Lev-Aretz, whose project is the first to engage in a privacy-focused legal analysis of data philanthropy.⁹ Her project describes data philanthropy as a practice in which (1) privately-held data or proprietary data-driven insights, (2) are shared or given access to, (3) for the public good.¹⁰

⁴ Alberto Alemanno, Big Data for Good: Unlocking Privately-Held Data to the Benefit of the Many, 9 EUR. J. OF RISK & REG. 183, 184 (2018).

⁷ Id.

3

¹ Alberto Alemanno, Big Data for Good: Unlocking Privately-Held Data to the Benefit of the Many, 9 EUR. J. OF RISK & REG. 183, 183 (2018).

² Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1493 (2019).

³ Annika Richterich, THE BIG DATA AGENDA: DATA ETHICS AND CRITICAL DATA STUDIES (London, University of Westminster Press, 2018).

⁵ *Id*, 186.

⁶ Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1493 (2019).

⁸ Presently, it includes variations such as (1) the "donation of privately-held commercial data towards beneficial causes": Jane Wu, *Big Data Philanthropy: The Social Impact of Donating Data*, LINKEDIN: PULSE (Jul. 1, 2015), https://www.linkedin.com/pulse/data-philanthropy-social-impact-donating-june-wu/; (2) a "partnership in which private sector companies share data for public benefit": Andreas Pawelke & Anoush Rima Tatevossian, *Data Philanthropy: Where Are We Now*? UNITED NATIONS GLOBAL PULSE: BLOG (May 8, 2013), https://www.unglobalpulse.org/data-philanthropy-where-are-we-now; and (3) "companies sharing proprietary datasets for social good": Patrick Meier, *Big Data for Humanitarian Response*, IREVOLUTIONS (June 4, 2012).

⁹ See, Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491 (2019).

¹⁰ Id, 1498.

Organizations in both the public and private sector increasingly view data sharing as key to the success of data-driven innovation. Thus, reducing the complexities involved in data sharing is seen as crucial in enabling further collaboration.¹¹ However, scholars from a range of disciplines have identified a need for legal guidance on various aspects of the data philanthropy practice, especially insofar as privacy risks are concerned.¹² While guidance concerning privacy-related legal risks has been provided, ¹³ this paper seeks to highlight the normative risks that should be considered before entering into certain data philanthropy arrangements. In order to do this, the paper first unpacks the notion of data philanthropy and evaluates some of the key assumptions underlying the practice. The paper then discusses the normative risks inherent in sharing arrangements, including the value of the data itself, the sustainability of data philanthropy arrangements, and the privacy harms that are not presently addressed by legal regimes. As regards privacy specifically, the paper examines issues such as the risk of re-identification, the concept of collective privacy and the need for informed consent. Finally, this paper seeks to provide an initial discussion of two infrastructural modalities that address some, but not all, of the issues discussed in this paper, including the use of the United Nations (UN) Global Platform governed by the UN Global Working Group on Big Data for Official Statistics,¹⁴ and the Safe Sharing Site proposal.

In order to provide a unique contribution to the literature, this paper narrows its focus in three ways. First, the paper concentrates particularly on the risks inherent in the provision of *personal* data, such as data collected by social media companies. This is to be contrasted with, for example, the provision of spatial mapping data. Second, it discusses the unique ethical risks involved when data is aimed at furthering *policy* development, as distinct from ad hoc emergency or disaster-response arrangements. Third, it frames its discussion of these risks from the perspective of international organizations, NGOs or civil society groups or, in other words, the *receivers* of the data. The scope of this paper is deliberately limited to what international organizations, as recipients of data philanthropy, need to consider before entering into such arrangements.¹⁵ In so doing, the paper seeks to assist such organizations in making informed decisions about entering into data philanthropy arrangements, and does not provide a discussion of the legal or other risks that such arrangements pose to private actors providing data.

¹¹ Lisa M. Austin & David Lie, Safe Sharing Sites, 94 N.Y.U. LAW. REV. 581, 585 (2019).

¹² See, e.g. Robert Kirkpatrick, Big Data for Development, 1 Big DATA, 3 (2013); Jeffrey P. Kahn, Effy Vayena & Anna C. Mastoianni, Opinion: Learning as We Go: Lessons from the Publication of Facebook's Social-Computing Research, 111 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (2014); Nir Kshetri, The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges and Concerns, 1 Big DATA & SOCIETY (2014).

¹³ See Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491 (2019).

¹⁴ See UN Global Working Group on Big Data for Official Statistics, United Nations Global Platform: Data for the world (2019), http://publications.officialstatistics.org/assets/pdf/UNGlobalPlatform_Brochure.pdf.

¹⁵ As a separate matter, it would be useful to conduct research on broader uses of data in, for example, crisis contexts, particularly considering the Do No Harm principles, and what guidance the United Nations system could provide in this regard.

PART ONE: DATA PHILANTHROPY EXPLORED

Unpacking Data Philanthropy

The notion of "data for good" has its roots in the open data movement, which has advocated for the release of

Stakeholders in data philanthropys, problem holders, data holders and skill holders, are driven by different interests. The assumption that relationships between these stakeholders serve the public good must be questioned. governmental data in downloadable, useable and distributable formats.¹⁶ However, given that private actors harbor the vast majority of this data, the open data movement has expanded its target to also include private sector entities.¹⁷ Lee-Aretz outlines a variety of different ways that organizations can engage in data philanthropy, including: (1) *Data Analytics Services* such as DataKind, which pairs high impact organizations with leading data scientists, and advocates meaningful utilization of data to solve humanitarian issues around poverty, health, human rights, education and the environment;¹⁸ (2) *Data Storage and Data-Based Utilities* such as Microsoft Corporation's Microsoft Cloud Services which offers nonprofits and university researchers useful tools for the collection, storage and management of

data, but does not involve the giving of actual data; (3) *Monetary Donations for Data Science Education and Development*; and (4) *Individual Data Sharing* which can involve (i) giving personal data about the sharing individual, such as the Personal Genome Project, which lets participants share their genome sequence and health data for the use of researchers, ¹⁹ and (ii) citizen scientists, being amateur individuals who help professional scientists speed up discoveries and innovation.²⁰

Further, data philanthropy involves various stakeholders, each guided by individual interests. Mikel Niño and others provide a useful taxonomy to explain the different roles stakeholders can play, distinguishing between (1) problem holders, (2) data holders and (3) skill holders.²¹ *Problem holders* are the individuals or institutions closest to the target population affected by the social issue to which the data relates. They may include representatives of public administration, such as governmental agencies and non-profit organizations, who work to identify the social problem, needs of affected groups, and potential solutions. *Data holders* include the individuals or institutions that hold the data, whose interest in it may arise because they collected the data themselves or acquired intellectual property or contractual rights in it. Relevantly, from a legal perspective, this category

¹⁶ Yafit Lev-Aretz, *Data Philanthropy*, 70 HASTINGS L.J. 1491, 1501 (2019); Tim Berners-Lee, *The Year Open Data Went Worldwide*, TED (Feb. 2010), https://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.

¹⁷ Rajesh Chandy, Magda Hassan & Prokriti Mukherji, Big Data for Good: Insights from Emerging Markets, 34 J. PROD. & INNOVATION MGMT. 703 (2017); Frederika Welle Donker, Bastiaan van Loenen, & Arnold K. Bregt, Open Data and Beyond, 5 INT'L J. GEO-INFO. 48 (2016) (arguing that private organizations that are mandated to perform a public task and generate data in the process should not be exempted from open government data policies); Beth Simone Noveck, Collaboratives: Sharing Public Data in Private Hands for Social Good, FORBES (Sept. Data 24. 2015, 9:30 AM), https://www.forbes.com/sites/bethsimonenoveck/2015/09/24/private-data-sharing-for-public-good/#397107b351cd.

¹⁸ See Jake Porway, Using Collaboration to Harness Big Data for Social Good, STAN. SOC. INNOV. REV. (Jun. 14, 2017).

¹⁹ THE PERSONAL GENOME PROJECT, http://personalgenomes.org/ (last visited Mar. 12, 2020). *See also*, AMERICAN GUT, http://americangut.org/ (last visited Mar. 12, 2020). 2020).

²⁰ Yafit Lev-Aretz, *Data Philanthropy*, 70 HASTINGS L.J. 1491, 1501-1503 (2019).

²¹ Mikel Niño et. al., *Data Projects for "Social Good": Challenges and Opportunities*, 11 INT'L J. HUMAN & Soc. Sci. 1073 (2017), http://www.waset.org/publications/10006923.

does *not* include the data subjects themselves, i.e. those people to whom the data personally relates.²² Finally, *skill holders* are those possessing the technical expertise to analyze and extract insights from the data and translate the results to the problem holders for the purpose of developing policy solutions. In the social media context, companies oftentimes operate as skill holders as well as data holders. For example, in 2016, Facebook analyzed its own data extract from the posts of Brazilian users regarding the Zika virus.²³ It then passed on these insights to UNICEF for the purposes of developing an ad campaign raising awareness about the virus.²⁴

What is "Good"?

Having outlined the more structural aspects of what makes a data philanthropy arrangement, it is useful to question its key normative assumption: that these relationships function to serve the public good. The literature is currently lacking a comprehensive description of "public good".²⁵ Rather, scholars have pointed to discrete and ad hoc examples of collaborations that are widely accepted as being in the public interest. However, as Lee-Aretz posits, this approach is flawed. It results in a selfdefining notion of "good" that is largely left to the various stakeholders, with their own interests and incentives, to describe.²⁶ This approach is particularly problematic considering two factors: (1) the non-rivalry nature of data and (2) the reputational gains it offers for the private sector. First, the mere notion that the giving of data is philanthropic is uncomfortable. Traditional corporate philanthropy involves a cost for the corporation (such as the giving of money) and a practical limit on the amount that can be given without harmfully interfering with the corporation's business. However, in the absence of corporations also acting as skill holders, data is offered for reuse, meaning that there is no cost to the corporation and no practical limit on the amount that can be given. In other words, by reusing data, the corporation will never run out of it. Second, data philanthropy offers various reputational benefits for corporations, which for social media companies in the present political climate may be extremely valuable.²⁷ Thus, in circumstances where the giving of data is done at no cost and provides data holders with a number of incentives, the ability of corporations to self-define what is "good" must be resisted. Accordingly, the data philanthropy practice calls for a comprehensive definition of social good and concrete guidelines to instruct collaborations, such that controversial characterizations of social good are thoroughly scrutinized.²⁸

²² Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1509 (2019). US law does not recognize data subjects as having any proprietary interest in the data of which they are subjects. See, e.g., Daniel J. Solove, Privacy Self-Management and the Consent Dilemma, 126 HARV. L. REV. 1880 (2013).

²³ Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1509 (2019).

²⁴ Catherine Cheney, How Facebook Statuses Informed the Zika Response in Brazil, DEVEX (Dec. 13, 2016), https://www.devex.com/news/how-facebook-statusesinformed-thezika-response-in-brazil-89290.

²⁵ Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1513 (2019).

²⁶ Id.

²⁷ These incentives are discussed in section II.c.ii below.

²⁸ Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1513 (2019).

The Benefits of Data Sharing

(i) Data as a Valuable Commodity

Effective public policy rests on good information about problems and the efficacy of the solution. While in the 19th and 20th centuries, governments obtained this information through statistics offices, a modality often touted currently is the analysis of big data, through which far more nuanced understandings of social problems can be gained.²⁹ Several qualities of big data have accelerated the growing demand of its use for the public good. First, there is an increasing demand for evidence-based social action³⁰ and the ability to capture and analyse real-time measurements that big data affords is especially valuable for social action that must quickly respond to unforeseen events.³¹ Second, big data can reveal less conspicuous social action and public sentiment that may not be visible through traditional metrics, such as national statistics.³² This is particularly the case for social media data, where individuals "team up" with other users to respond to social events, and express opinions on issues that they may not be so willing to share in an official forum. Finally, mobile data (while not the focus of this paper) is of particular value for humanitarian organizations because of its ability to reveal population movements in real time.³³

(ii) **Private Sector Incentives**

As alluded to above, data philanthropy arrangements offer a variety of incentives for private sector providers of data. While this paper presents a discussion of the risks inherent in data philanthropy to the organizations receiving the data, it is useful for such organizations to be aware of these private sector incentives so that they can better scrutinize proposals for data sharing. Klein and Verhulst suggest six categories of sharing incentives for private sector entities: reciprocity; research, recruitment and insights; reputation and public relations; increasing revenue; regulatory compliance; and responsibility and corporate philanthropy.³⁴ Of these, reciprocity, reputational advantage and corporate philanthropy are the most relevant to this discussion. As regards reciprocity, Klein and Verhulst outline one type that takes the form of compensation. Companies that collect personal information, particularly those in the social media industry, may act under the impression that they are giving back to society in order to counterbalance what they have taken from individuals, by way of their personal data.³⁵ As Lev-Aretz elaborates, this kind of data philanthropy serves as a form of non-mandatory tax, akin to the carbon tax, in that a company that pollutes society with surveillance "pays" by donating its data back for socially beneficial causes.³⁶ In relation to reputational advantage, it seems clear

³⁵ *Id*, 12.

²⁹ Rohan Samarajiva & Srinivas Lokanathan, Using Behavioural Big Data for Public Purposes: Exploring Frontier Issues of an Emerging Policy Arena 1, 3 (Feb. 1, 2016), https://lirneasia.net/wp-content/uploads/2013/09/NVF-LIRNEasia-report-v8-160201.pdf (last visited Nov. 19, 2019).

³⁰ Claudia J. Coulton et al., *Harnessing Big Data for Social Good: A Grand Challenge for Social Work* 4 (Am. Acad. of Soc. Work & Soc. Welfare, Working Paper No. 11, 2015), http://aaswsw.org/wp-content/uploads/2015/12/ WP11-with-cover.pdf.

³¹ Rajesh Chandy, Magda Hassan & Prokriti Mukherji, *Big Data for Good: Insights from Emerging Markets*, 34 J. PROD. & INNOVATION MGMT. 703 (2017); Yafit Lev-Aretz, *Data Philanthropy*, 70 HASTINGS L.J. 1491, 1507 (2019).

³² Peter Baeck (ed.), *Data For Good: How Big and Open Data Can Be Used for the Common Good*, NESTA 1, 25 (Feb. 2015), https://www.nesta.org.uk/sites/default/files/dataforgood.pdf.

³³ Yafit Lev-Aretz, Data Philanthropy, 70 HASTINGS L.J. 1491, 1508 (2019).

³⁴ Thilo Klein & Stefaan Verhulst, Access to New Data Sources for Statistics: Business Models and Incentives for the Corporate Sector 17 (Paris21 Partnership in Stat. for Dev. in the 21st Century, Discussion Paper No. 10, 2017), http://www.thegovlab.org/static/files/publications/paris-21.pdf.

³⁶ Yafit Lev-Aretz, *Data Philanthropy*, 70 HASTINGS L.J. 1491, 1511 (2019).

that sharing data for socially beneficial purposes can improve a business' reputation. In the wake of controversies concerning misuse of data, such as Facebook's Cambridge Analytical scandal, it seems that, now more than ever, companies wish to demonstrate that their collection and use of data practices align with the public interest.³⁷

Modes of Data Sharing

In addition to mapping out different corporate incentive structures, Klein and Verhulst have also identified five different categories of data sharing: in-house production of statistics, transferring copies of data sets, remote access model, trusted third parties, and the algorithmic sharing model.³⁶ (1) *In-house production of statistics*: In the most common form of data collaboration, data holders process the data in-house and share the resulting statistics with problem holders, thus retaining control over the raw data. (2) *Transferring copies of data sets*: Data holders transfer copies of data sets that have been de-identified and aggregated, on which users (most likely skill holders) can run their algorithms. This method is most effective for research purposes because the aggregated data sets contain detailed information and a merger of different data sets and sources. However, producers relinquish control over the data and increase risks of security breaches and privacy harms.³⁹ (3) *Remote access model*: In this model, end users securely access data remotely while data holders maintain control over the final aggregated metrics.⁴⁰ (4) *Trusted third parties*: Data holders and users rely on a trusted third party who possess reliable technical infrastructure, including large data storage capacity and secure connections, to facilitate secure access to the data.⁴¹ Users do not have access to the raw data and instead request reports from the facilitator.⁴² (5) *Algorithmic sharing model*: Finally, this model allows for data to remain in the data holders possession, and the data holder runs an algorithm of the data user's choice, with the data holder producing the final report.

预览已结束, 完整报告链接和二维码如下:

https://www.yunbaogao.cn/report/index/report?reportId=5 11630

