Hastings Law Journal

Volume 70 | Issue 6 Article 3

8-2019

Data Philanthropy

Yafit Lev-Aretz

Data Philanthropy

YAFIT LEV-ARETZ[†]

The term "data philanthropy" has been used to describe the sharing of private sector data for socially beneficial purposes, such as academic research and humanitarian aid. The recent controversy over an academic researcher's alleged misuse of Facebook users' data on behalf of Cambridge Analytica has brought data philanthropy into the spotlight of public debate. Calls for data ethics and platform transparency have highlighted the urgent need for standard setting and democratic oversight in the use of corporate data for public ends. Data philanthropy has also received considerable scholarly attention in various academic disciplines but has, until now, been virtually overlooked by the legal literature. This Article explains and starts filling in the resulting research gap by providing the first legal accounting of data philanthropy. Following a detailed description of current developments and scholarly thinking, this Article homes in on a normative assessment of privacy risks that are often cited as a conceptual and practical barrier to data philanthropy.

This Article refines the scope of data philanthropy's informational risks and proposes a framework for mitigating some of these risks through the Fair Information Practice Principles ("FIPs"). Specifically, the purpose specification and use limitation principles, which limit data collection to ex-ante specified purposes, are discordant with the unanticipated, ex-post quality of data philanthropy. Adopting a new "data philanthropy exception" will account for the existence and nature of the privacy risks, the time frame for action, the social risks of using the data, and the allowed retention time following the reuse. The data philanthropy exception reinforces the values at the heart of the FIPs, provides guidance in a field that currently operates in a legal vacuum, and introduces the possibility of responsible sharing by and to smaller market participants.

[†] Postdoctoral Research Fellow, NYU Law School. I would like to thank Michael Birnhack, Kiel Brennan-Marquez, Niva Elkin-Koren, Ignacio Cofone, Nizan Geslevich-Packin, Amanda Levendowsky, Helen Nissenbaum, Gideon Parchomovsky, Jules Polonetsky, Julia Powles, Ira Rubinstein, Madelyn Sanfilippo, Jason Schultz, Andrew Selbst, Abbey Stemler, Omer Tene, Katherine Strandburg, Daniel Susser, and Kurt Wimmer. I would also like to thank the participants in the Privacy Research Group at NYU; the Minerva Center for the Rule of Law under Extreme Conditions at the University of Haifa; the Internet Law-In-Progress the Privacy Law Scholars Conference; the Northeast Privacy Scholars Workshop; the Privacy Law Scholars Conference at George Washington University, the Big Data Ethics workshop at Babson College, and the Privacy, Cyber, and Technology workshop at Tel-Aviv University. I would also like to thank the Minerva Center for the Rule of Law under Extreme Conditions at the University of Haifa for generous financial support. Remaining errors are mine.

TABLE OF CONTENTS

INTRODUCTION

April 2015 brought the worst earthquake to hit Nepal in over eighty years, killing nearly 9,000 people, injuring over 22,000, and displacing over 2.8 million individuals.[1] The resulting humanitarian crisis was massive, and it was virtually impossible to quickly locate the thousands of people who might have been hurt or trapped, or those who had escaped to safe zones.[2] But Nepal's largest mobile network operator, Ncell, offered a solution. The company partnered with Flowminder, a non-profit organization that builds population movement models,[3] to map population displacements around the country using anonymized data from 12 million Ncell subscribers in the affected areas.[4] These data-based maps helped humanitarian response organizations to pinpoint the location of impacted individuals and to allocate aid resources and response teams accordingly.[5]

The Ncell-Flowminder collaboration exemplifies a new form of private sector donation: using private sector data for the public social good. Data generated via platforms like telecom operators, satellite companies, and social media networks, has the potential to enable a range of insights into economic development, medical advances, environmental issues, and various other properties of public life that could accelerate the pace and scope of social discovery and development. This understanding has triggered a "data-for-good" movement, which promotes data-driven projects that can increase the efficiency of social initiatives, extend their reach, and better tailor them to specific communities.[6] The data-for-good movement has spotlighted the imperative role of the private sector in producing useful data for social action, sparking an active conversation about models and incentives for sharing. As part of this conversation, the term "data philanthropy" was born.[7]

The term "data philanthropy" has been used to describe the giving of private sector data, providing access to it, or the production of data-driven insights for a socially beneficial purpose.[8] The recent Facebook/Cambridge Analytica debacle, at which political consulting firm Cambridge-Analytica used

---

1. Nikhil Kumar, *Go Inside the Effort to Rebuild Nepal*, TIME (June 25, 2015), http://time.com/3928685/nepal-earthquake-recovery-donors-food-rice/.

2. Matt Petronzio, *Facebook's New 'Disaster Maps' Could Revolutionize Natural Disaster Rescue Efforts*, MASHABLE (June 7, 2017), https://mashable.com/2017/06/07/facebook-disaster-maps-humanitarian-aid/#jQ0fXOB95Oqq.

3. UN GLOBAL PULSE & GSMA, THE STATE OF MOBILE DATA FOR SOCIAL GOOD REPORT 7 (2017), http://unglobalpulse.org/sites/default/files/MobileDataforSocialGoodReport_29June.pdf [hereinafter MOBILE DATA FOR SOCIAL GOOD REPORT].

4. *Id*.

5. *Id*.

6. 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. Some of my views on the term philanthropy are articulated in Part II. For more on the criticism of the term, see Yafit Lev-Aretz, *A Case for Precision: Against the Philanthropy in Data Philanthropy* (*on file with the author*) [hereinafter Lev-Aretz, *A Case for Precision*].

8. *See supra* Part II.

information originally obtained for research purposes to target potential voters during the 2016 election,[9] has highlighted the risks of data philanthropy misuse. In 2014, Facebook permitted an academic researcher, Aleksandra Kogan, to access the private information of tens of thousands of volunteers who agreed to have their data used in an academic study.[10] But Kogan then shared all that Facebook user data—which included information about all of the volunteers as well as information about all their friends and contacts that did not agree to the sharing—with Cambridge Analytica.[11] The company then used the data from Kogan to target consumers and voters on behalf of Cambridge Analytica's customers.[12] The story about the unauthorized sharing of personal data from nearly 87 million Facebook users broke nearly four years later, in 2018, and massive public outcry ensued. The blatant misuse of academic access privileges in the Facebook/Cambridge Analytica case has spotlighted the active market for sharing corporate data for academic research and the lack of acceptable standards for these collaborations.[13]

As data philanthropy collaborations have mushroomed in recent years, scholars from a range of diverse disciplines, including computer science, social science, economics, information science, business, and philosophy, have engaged in data philanthropy conversations.[14] Many of those scholars—as well as industry players and other stakeholders—have identified a pressing need for legal guidance on various aspects of the data philanthropy practice, especially

9. Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

10. *Id*.

11. *Id*.

12. *Id.*

13. *See, e.g.*, Sheera Frenkel, *Scholars Have Data on Millions of Facebook Users. Who's Guarding It?*, N.Y. TIMES (May 6, 2018), https://www.nytimes.com/2018/05/06/technology/facebook-information-data-sets-academics.html?smprod=nytcore-ipad&smid=nytcore-ipad-share.

14. *See, e.g.*, Jean Burgess & Axel Bruns, *Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research after the Computational Turn*, *in* COMPROMISED DATA: FROM SOCIAL MEDIA TO BIG DATA 93 (Ganaele Langlois et al. eds., 2015); John Karlsrud, P*eacekeeping 4.0: Harnessing the Potential of Big Data, Social Media, and Cyber-Technology*, CYBERSPACE AND INTERNATIONAL RELATIONS: THEORY, PROSPECTS AND CHALLENGES 141 (Jan-Frederik Kremer & Benedikt Müller eds., 2014); Patrick Meier, *Human Computation for Disaster Response*, *in* HANDBOOK OF HUMAN COMPUTATION 95 (Pietro Michelucci ed., 2013); Luciano Floridi & Mariarosaria Taddeo, *What Is Data Ethics?* 374 PHIL. TRANSACTIONS ROYAL SOC'Y, Dec. 28, 2016, at 1; Jeffrey P. Kahn et al., *Opinion: Learning as We Go: Lessons from the Publication of Facebook's Social-Computing Research*, 38 PROC. NAT'L ACAD. SCI. 13677 (2014); Robert Kirkpatrick, *Big Data for Development*, 1 BIG DATA 3 (2013); Nir Kshetri, *The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges, and Concerns*, BIG DATA & SOC'Y, DEC. 22, 2014, at 1; Mariarosaria Taddeo, *Data Philanthropy and the Design of the Infraethics for Information Societies*, 374.2083 PHIL. TRANSACTIONS ROYAL SOC'Y, Dec. 28, 2016, at 1; Effy Vayena et al., *Ethical Challenges of Big Data in Public Health*, PLOS COMPUTATIONAL BIOLOGY, Feb. 9, 2015, at 1; Robert Kirkpatrick, *A New Type of Philanthropy: Donating Data*, HAR. BUS. REV. (Mar. 21, 2013), https://hbr.org/2013/03/a-new-type-of-philanthropy-don.

when the practice implicates individuals' privacy. Yet, until now, legal scholars have been conspicuously missing from this discourse.[15]

One likely reason for the relative silence of legal academics and regulators alike is that contractual agreements are often sufficient to legally facilitate reuse. By broadly defining data collection purposes and granting expensive authorizations for use, collectors of personal data have been able to safeguard themselves from legal liability for sharing the data for socially beneficial purposes.[16] Even in regulated industries, businesses that have had an interest in data sharing have found a way to incorporate privacy compliance into the sharing process.[17] Evidence suggests that corporate players who are truly interested in donating data for social good have plenty of legal options to do so, such that, in practice, the extent to which legal concerns impede data philanthropy efforts is overstated.[18]

Another reason may be that most calls for guidance have been directed towards privacy scholars, asking for a framework for balancing privacy risks

---

15. As of writing this, there are only five law review pieces that mention data philanthropy. *See* Janine S. Hiller & Jordan M. Blanke, *Smart Cities, Big Data, and the Resilience of Privacy*, 68 HASTINGS L.J. 309, 335–36 (2017) (discussing "the right to be free from government and private surveillance" through transparency and the ability to opt out, and argue that "[t]his principle speaks to 'Data Philanthropy,' a framework that gives individuals the power either to consent to being involved or to opt-out of participation in the smart city"); Luca Leone, *Addressing Big Data in Eu and US Agriculture: A Legal Focus*, 12 EUR. FOOD & FEED L. REV. 507, 508 (2017) (pointing to "'open data' and 'data philanthropy' as institutional and procedural patterns to follow to achieve more knowledgeable and sustainable agriculture"); Beth Simone Noveck, *Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency*, 19 YALE HUM. RTS. & DEV. L.J. 1, 15–16 (2017) (referring to data philanthropy as "the next wave in corporate social responsibility" and giving some examples for private sector data sharing for socially beneficial purposes); Galit A. Sarfaty, *Can Big Data Revolutionize International Human Rights Law?*, 39 U. PA. J. INT'L L. 73, 97 (2017) (proposing "to broaden the scope of big data projects in the human rights field," by "incentiviz[ing] companies to engage in data philanthropy"); Stephanie Segovia, *Privacy: An Issue of Priority*, 11 HASTINGS BUS. L.J. 193, 200 (2015) (arguing that "[l]egislative legal frameworks should be created that (1) protect the individual, and (2) require contractors to make their data public, thus honing in on the business value that data philanthropy can deliver," without further reference to or elaboration on the term). None of these works provide a material descriptive or normative contribution on data philanthropy initiatives. While Sarfaty's article discusses data philanthropy in more detail than others, even her work does not offer a comprehensive analysis of data philanthropy in the context of her proposal.

16. Viktor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data through Europe's New Data Protection Regulation*, 17 COLUM. SCI. & TECH. L. REV. 315, 322 (2016). For example, Facebook's Data Policy states:

> We provide information and content to vendors and service providers who support our business, such as by providing technical infrastructure services, analyzing how our Products are used, providing customer service, facilitating payments *or conducting surveys*. . . . We also provide information and content to research partners and academics to conduct research that advances scholarship and innovation that support our business or mission, and enhances discovery and innovation on topics of general social welfare, technological advancement, public interest, health and well-being.

Data Policy, FACEBOOK, https://www.facebook.com/policy.php (last visited July 27, 2019) (first emphasis added).

17. FUTURE OF PRIVACY FORUM, UNDERSTANDING CORPORATE DATA SHARING DECISIONS: PRACTICES, CHALLENGES, AND OPPORTUNITIES FOR SHARING CORPORATE DATA WITH RESEARCHERS 11 (2017), https://fpf.org/wp-content/uploads/2017/11/FPF_Data_Sharing_Report_FINAL.pdf [hereinafter FPF REPORT].

18. *Id*.

with the likely social benefits of data philanthropy.[19] Among the privacy community of legal scholars, however, which has dedicated itself to the vital mission of identifying and addressing data-driven privacy harms, discussing the socially beneficial aspects of the information economy is a morally challenging task. To borrow from Fourth Amendment jurisprudence, this ill-gotten data has the appearance of a fruit from a very poisonous tree: how can we launder it, portray it as a charitable way of giving, and use the historically loaded and probably inaccurate term "philanthropy"?[20] Furthermore, to gift, one must first own. How can corporations share personal data when their ownership of this data is highly contested?[21]

These concerns all have merit. Today's "big data" culture imposes numerous negative externalities that are being allocated unfairly.[22] But ignoring the social benefits of hoarded information may add more to the list of negative externalities while simultaneously preventing positive ones. The two scholarly endeavors—to limit illegitimate collection and use of personal data and, simultaneously, to promote socially beneficial reuses—must coexist without, except in extreme cases, influencing each other.[23]

This Article is the first to engage in a legal analysis of data philanthropy, with a focus on the privacy aspects of the practice. Part II defines key properties of data philanthropy and situates this Article within the broader academic discourse. This in-depth descriptive contribution provides a close look at the interaction between data philanthropy and preexisting laws, social norms, and industry practices.

Part III offers a significant normative contribution by identifying a legal research gap, redefining this gap and its roots in the context of privacy, and proposing a framework for responsible sharing of private sector data for socially beneficial purposes. In Part III, this Article also provides a much-needed legal perspective on the privacy risks of data philanthropy. Before doing so, however, this Article refines the scope of relevant risks by questioning privacy's restraining power over the practice. In reality, privacy compliance has not presented a significant barrier to corporate sharing for social good.[24] Most businesses can and have sheltered themselves from legal liability through their

19. While these calls have not identified privacy scholars as such, they require guidance on privacy, which could naturally come only from experts in the field.

20. *See infra* Part II.

21. *See infra* Part II.

22. *See, e.g.*, Julie Cohen, *The Surveillance-Innovation Complex: The Irony of the Participatory Turn*, *in* THE PARTICIPATORY CONDITION IN THE DIGITAL AGE 207 (Darin Barney et al. eds., 2015); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Frank Pasquale, *Privacy, Antitrust, and Power*, 20 GEO. MASON L. REV. 1009 (2013); Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75 (2015).

23. This approach is in line with Helen Nissenbaum's contextual integrity theory that treats privacy interests as part of a complex set of interests and considerations. HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 129–85 (2010).

24. FPF REPORT, *supra* note 17, at 6.

terms of service agreements.[25] Furthermore, sharing of corporate data for public ends has been mostly done under technical and institutional privacy safeguards.[26] Instead, this Article argues for the pressing need—particularly from non-lawyers—to provide a privacy framework group broader than data-related issues under the umbrella term "privacy," and use privacy's legal framework as a proxy for general legal acknowledgment of data philanthropy. In other words, the demand voiced is for a formal legal recognition of the practice and for a framework to address broader fair information practices in data philanthropy initiatives.

Thus, and for reasons further discussed in Part IV, this Article asserts that the Fair Information Practice Principles (FIPs) are the best-suited host for data philanthropy. The FIPs are widely accepted principles for the protection of personal information that have been adopted and proposed, in various versions, by different institutions around the world, including the Organisation for Economic Co-operation and Development (OECD), the Canadian Standards Association, and the U.S. Federal Trade Commission.[27] Specifically, two principles of the FIPs—purpose specification, which requires a concrete indication of personal data collection purposes, and use limitation, which mandates that subsequent use of the personal data does not exceed the purposes specified at collection—stand to block socially beneficial reuses of data in data philanthropy collaborations.[28] This Article concludes that an exception to the FIPs would reconcile data philanthropy with the purpose specification and use limitation principles.[29] In addition to prescribing guidelines for responsible sharing of personal information in data philanthropy initiatives, a FIPs exception would also provide legal acknowledgment of the practice and introduce it to smaller private sector actors.

The data philanthropy exception would only apply when privacy protections prevent or limit socially beneficial reuse of data. The exception provides a three-tiered review process for determining the appropriate balance between privacy protections and social benefit in any given case. The socially beneficial reuse is first classified as one of three use privileges: exigencies (emergencies at which the use of the data is required under time pressure and high risk); responses (uses of corporate data to advance solutions to social problems); or collective knowledge (corporate data is used for research and knowledge advancements). Next, the review requires a risk assessment to account for the potential informational harm involved in reuse, as well as the

25. *See, e.g.*, *Data Policy*, *supra* note 16.

26. *See infra* Subpart III.B.

27. Robert Gellman, *Fair Information Practices: A Basic History* (Apr. 10, 2017), https://bobgellman.com/rg-docs/rg-FIPshistory.pdf.

28. Aaron Fluitt, *Executive Summary: Report from the Georgetown Law Round Table on the Ethical Reuse of Data in a Machine Learning World*, TECH. & PRIVACY L. BLOG (Mar. 26, 2018), https://www.techprivacylawblog.com/executive-summary-report-from-the-georgetown-law-round-table-on-the-ethical-reuse-of-data-in-a-machine-learning-world/.

29. *See infra* Subpart IV.B.

likely harm resulting from *not* using the information. After a use has been categorized under one of the three use privileges and a risk assessment has been conducted, the exception mandates the setting of a retention time for the reused data.

A data philanthropy exception to the FIPs would confirm legal recognition of widely appreciated corporate data sharing practices and provide legal guidance on responsible sharing. The wide reach of the FIPs could bring data philanthropy to the attention of smaller market players who may have not been familiar with the practice and direct them towards privacy-mindful initiatives. Legal engagement with data philanthropy is also likely to stir discussions in other areas of law about the promise and perils of the practice well beyond privacy law. But, perhaps more importantly, data philanthropy offers an opportunity to consider how the law can protect and promote the use of private data for public good.

I. BACKGROUND AND SCOPE

The term "Data Philanthropy" was reportedly coined by World Economic Forum CTO Brian Behlendorf during a spontaneous conversation at the 2011 World Economic Forum.[30] The definition of data philanthropy is unsettled and includes several variations, such as the "donation of privately-held commercial data towards beneficial causes,"[31] a "partnership in which private sector companies share data for public benefit,"[32] "the act of sharing private data assets to serve the public good,"[33] and "companies sharing proprietary datasets for social good."[34] Nonetheless, all of these interpretations involve the same basic scenario—the use of privately collected data for socially beneficial purposes. The term data philanthropy can theoretically apply to many forms of data sharing, but this project centers on one common type of sharing, where (1) privately-held data or proprietary data-driven insights (2) are shared or given access to (3) for the public good. This form of sharing has frequently been labeled either "data philanthropy" or "data collaboratives."[35]

30. Nathan Wolfe et al., *Crunching Digital Data Can Help the World*, CNN (Feb. 2, 2011), http://www.cnn.com/2011/OPINION/02/02/wolfe.gunasekara.bogue.data/.

31. Jane Wu, *Big Data Philanthropy: The Social Impact of Donating Data*, LINKEDIN (July 1, 2015), https://www.linkedin.com/pulse/data-philanthropy-social-impact-donating-june-wu.

32. Andreas Pawelke & Anoush Rima Tatevossian, *Data Philanthropy: Where Are We Now?* UNITED NATIONS GLOBAL PULSE: BLOG (May 8, 2013), http://www.unglobalpulse.org/data-philanthropy-where-are-we-now.

33. BRICE MCKEEVER ET AL., DATA PHILANTHROPY, UNLOCKING THE POWER OF PRIVATE DATA FOR PUBLIC GOOD 1 (2018), https://www.urban.org/research/publication/data-philanthropy-unlocking-power-private-data-public-good/view/full_report.

34. Patrick Meier, *Big Data Philanthropy for Humanitarian Response*, IREVOLUTIONS (June 4, 2012), https://irevolutions.org/2012/06/04/big-data-philanthropy-for-humanitarian-response/.

35. Stefaan Verhulst and David Sangokoya propose the term "data collaboratives" to describe data exchange to help solve public problems. *See* Stefaan Verhulst & David Sangokoya, *Data Collaboratives: Exchanging Data to Improve People's Lives*, MEDIUM (Apr. 22, 2015), https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a.

Characterizing data sharing of this sort as philanthropy is not without objectors.[36] Philanthropy has been criticized as patronage, reinforcement of social forces of hegemony and control, and as a representation of the systemic failures in modern societies.[37] Traditional definitions of philanthropy also assume altruistic giving that originates in the "love of mankind."[38] The attribution of such noble motivations to profit-maximizing entities is also met with skepticism.[39]

The term "philanthropy," as applied in this context, also stands on shaky legal ground. In legal terms, while private sector actors collect personal information under contractual authorization from the information subject, it is not entirely clear that the collectors become owners of the information in the legal sense.[40] If private sector actors only have a license to use this information

36. For more on these objections, *see* Lev-Aretz, *A Case for Precision*, *supra* note 7.

37. Siobhan Daly, *Philanthropy as an Essentially Contested Concept*, INT'L J. OF VOLUNTARY & NONPROFIT ORG. 535, 542–44 (2012).

38. Marty Sulek, *On the Modern Meaning of Philanthropy*, 39 NONPROFIT & VOLUNTARY SECTOR Q. 193, 196–200 (2010).

39. *See, e.g.*, Daryl Koehn and Joe Ueng, *Is Philanthropy Being Used by Corporate Wrongdoers to Buy Good Will?*, 14 J. OF MGMT. & GOVERNANCE 1 (2010); Ming Jia and Zhe Zhang, *Donating Money to Get Money: The Role of Corporate Philanthropy in Stakeholder Reactions to IPOs*, 51 J. OF MGMT. STUD. 1118 (2014); Timothy S. Mescon and Donn J. Tilson, *Corporate Philanthropy: A Strategic Approach to the Bottom-Line*, 29.2 CAL. MGMT. REV. 49 (1987). However, claims that philanthropy interferes with the traditional model of corporations, which involves making money, distributing it to stakeholders, and allowing them to decide how to spend it, show that, at least in some cases, philanthropy is motivated by reasons other than profit-maximization. *See* Milton Friedman, *The Social Responsibility of Business Is to Increase Its Profits*, N.Y. TIMES (Sept. 13, 1970) (arguing that the corporate executive's responsibility is "to conduct the business in accordance with their desires, which generally will be to make as much money as possible while conforming to the basic rules of the society"). Responses to this argument have tended to revolve around the economic gains of philanthropy. *See* M. Todd Henderson & Anup Malani, *Corporate Philanthropy and the Market for Altruism*, 109 COLUM. L. REV. 571 (2009); Faith Stevelman Kahn, *Pandora's Box: Managerial Discretion and the Problem of Corporate Philanthropy*, 44 UCLA L. REV. 579 (1997); Michel E. Porter & Mark R. Kramer, *The Competitive Advantage of Corporate Philanthropy*, HARV. BUS. REV. (Dec. 2002), https://hbr.org/2002/12/the-competitive-advantage-of-corporate-philanthropy ("Philanthropy can often be the most cost-effective way for a company to improve its competitive context, enabling companies to leverage the efforts and infrastructure of nonprofits and other institutions.").

40. The intersection of privacy and property has been widely explored in legal scholarship. Some have called for data privacy to be viewed as a property right that grants them full control in their personal information. *See, e.g.*, *Developments in the Law: The Law of Cyberspace*, 112 HARV. L. REV. 1574, 1644–48 (1999); Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193, 1246–94 (1998). Others have rejected the property-based approach to privacy, claiming that it does not effectively protect privacy. *See, e.g.*, Rochelle Cooper Dreyfuss, *Warren and Brandeis Redux: Finding (More) Privacy Protection in Intellectual Property Lore*, STAN. TECH. L. REV. 8 (1999). Others still find that such approach might even encourage the market for personal data rather than constraining it. *See, e.g.*, Jessica Litman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283 (2000). In addition to the currently non-existing property rights of information subjects, U.S. copyright law does not protect databases absent some level of creativity in their creation. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 348 (1991). The most common rights structure in the market combines contractual confirmation of ownership of any rights related to consumer data together with trade secrecy protection. This combination allows businesses to protect themselves from intellectual property lawsuits and protect the collected data from free-riding competitors.

for specific purposes, as opposed to ownership of the data, can they legally donate it or authorize access to it?

Because of the non-rivalry nature of data, "philanthropy" seems like an odd term to attribute to the sharing of private sector data for socially beneficial purposes. Traditional corporate philanthropy commonly involves the giving of money, other tangible gifts, or services.[41] Each of these endowments involve both a cost for the corporate philanthropist and a practical limit on the amount that can be given without significantly interfering with the business' operation and existence. Data, however, is offered for reuse, meaning that it could be donated at minimal to no cost and with no practical limit. A business can not only allow access to its collected information at a very low cost, but it can also offer its entire database for socially beneficial use while capitalizing on the same database for its business interests.

The term "data collaboratives," offered by Stefaan Verhulst and David Sangokoya,[42] also fails to capture the essence of private sector sharing of data for social good. The "data collaboratives" term is useful and is not subject to some of the criticism that the term "data philanthropy" rightly receives. Nevertheless, the term "data collaborative" is both under-inclusive and over-inclusive. The emphasis on collaboration leaves many instances of data sharing outside the scope of data collaboratives. For example, open data initiatives in the private sector, where datasets are released to the public with no continuous interaction between the public and the provider of the data following the release, can hardly be described as collaborative. The data collaboratives universe also appears to be broader than that of data philanthropy, as it covers access to public sector data. Most importantly, the data collaborative definition does not underscore the sharing of privately-held data or privately-owned, data-driven insights. It fails to highlight the monetary and business value of the data and does not reflect the ecosystem in which private sector data is shared.

In the absence of an alternative definition targeting the scope of what has been termed "data philanthropy," this Article adopts the term "data philanthropy" subject to the objections expressed above. To better understand the contours of data philanthropy, it is essential to distinguish this form of "data-for-good" from other forms of data-sharing, to refine the "data," "sharing," and "public good" aspects of the definition, and to illustrate the promise and perils of data philanthropy through real-life examples.

A. DATA FOR GOOD

The idea that privately held or owned data should be used to promote the greater good has its roots in the open data movement.[43] The open data movement has advocated for the release of governmental data in machine-readable,

41. Sulek, *supra* note 38 at 200; *see also* Bruce Seifert et al., *Having, Giving, and Getting: Slack Resources, Corporate Philanthropy, and Firm Financial Performance*, 43 BUS. AND SOC'Y 135 (2004).

42. Verhulst and Sangokoya, *supra* note 35.

43. Alemanno, *supra* note 6, at 185.

downloadable, usable, and distributable formats.[44] These efforts have been celebrated as a means of transparency, accountability, and civic participation, and have led governments worldwide to open up countless datasets.[45] The open data movement also rests on the understanding that the rise of big data and advances in the capture, collection, real-time processing, analysis, sharing, and visualization of information can advance a better understanding of social problems and direct practical solutions.[46]

Because much of the data needed to address societal challenges rests in private hands, calls to make data available for socially beneficial reuses have expanded their targets from states and public entities to the private sector.[47] Unlike public sector data sharing, private sector data sharing triggers competitive and privacy risks that directly affect value for money and return on investment.[48] Yet, many private sector players have joined the cause with various "data-for-good" initiatives, outlined below: data analytics services, data storage and data-based utilities, monetary donations for data science education and development, individual data sharing, and data philanthropy.

Data Analytics Services: DataKind is a non-profit organization dedicated to using data science to address critical humanitarian issues by pairing high-impact organizations with leading data scientists.[49] DataKind, together with other NGOs like Bayes Impact and corporations like IBM and SAS,[50] is part of the Data-for-Good movement, which advocates meaningful utilization of data to solve humanitarian issues around poverty, health, human rights, education and the environment.[51] In practice, DataKind's mission and most of the Data-for-

44. 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?language=en.

45. *See, e.g.*, Jillian Raines, *The Digital Accountability and Transparency Act of 2011 (DATA): Using Open Data Principles to Revamp Spending Transparency Legislation*, 57 N.Y. L. SCH. L. REV. 313 (2013); Anneke Zuiderwijk and Marijn Janssen, *Open Data Policies, Their Implementation and Impact: A Framework for Comparison*, 31 GOV'T INFO. Q. 17 (2014).

46. STEFAAN VERHULST & ANDREW YOUNG, OPEN DATA IMPACT: WHEN DEMAND AND SUPPLY MEET 7–8 (Mar. 2016), http://odimpact.org/files/open-data-impact-key-findings.pdf.

47. *See* Rajesh Chandy et al., *Big Data for Good: Insights from Emerging Markets*, 34 J. PRODUCT INNOVATION MGMT. 703 (2017); Frederika Welle Donker et al.**,** *Open Data and Beyond*, 5 INT'L J. GEO-INFO. 48 (2016), http://www.mdpi.com/2220-9964/5/4/48/htm (arguing that private organizations that are mandated to perform a public task and generate data in the process should not be exempt from open government data policies); Beth Simone Noveck, *Data Collaboratives: Sharing Public Data in Private Hands for Social Good*, FORBES (Sept. 24, 2015), https://www.forbes.com/sites/bethsimonenoveck/2015/09/24/private-data-sharing-for-publicgood/#209e001d51cd.

48. *Open Data, Driving Growth, Ingenuity, and Innovation*, DELOITTE ANALYTICS (2012), https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf.

49. DATAKIND, http://www.datakind.org/ (last visited July 27, 2019).

50. BAYES IMPACT, http://www.bayesimpact.org/ (last visited July 2, 2019). *Data Science for a Better World*, IBM RESEARCH BLOG (Jan. 21, 2016), https://www.ibm.com/blogs/research/2016/01/data-science-for-a-better-world/. *Data for Good: Analytics Helping Humanity*, SAS, https://www.sas.com/en_us/data-for-good.html# (last visited July 27, 2019).

51. Jake Porway, *Using Collaboration to Harness Big Data for Social Good*, STAN. SOC. INNOVATION REV. (June 14, 2017), https://ssir.org/articles/entry/using_collaboration_to_harness_big_data_for_social_good.

Good movement revolve around the donation of *data analytics services* to high-impact social organizations.

Data Storage and Data-based Utilities: The Microsoft Corporation committed to donating $1 billion of Microsoft Cloud Services for the use of nonprofits and university researchers between 2016 and 2019.[52] The cloud services allow users to securely store their data, compute on the cloud platform, turn data into actionable insight using effective data management tools and advanced analytics, and capitalize on the cutting-edge technologies of the Internet of Things and artificial intelligence.[53] These services offer useful tools for the collection, storage, and management of data, but do not involve the giving of actual data. Another example from Microsoft is the language translation systems the company developed to help relief workers communicate with the local community in Haiti after the destructive 2010 earthquake.[54]

Monetary Donations for Data Science Education and Development: Tableau Software announced grants to DataKind and Bayes Impact in support of the Data-for-Good movement.[55] Dartmouth alumnus William H. Neukom has committed $10 million to his alma mater for expanding and improving the school's data science programs.[56] While efforts like these contribute both directly and indirectly to data-driven innovation for the greater good, they implicate different promises and perils than donations of data.[57]

Individual Data Sharing: Individuals, too, can share data for socially beneficial ends. Two kinds of data are shared on an individual basis: personal data about the sharing individual and data collected to which individuals own the intellectual property rights. The Personal Genome Project, which lets participants share their genome sequence and health data for the use of researchers, illustrates the first category. [58] The second category includes

52. *Microsoft Philanthropies Announces Commitment to Donate $1 Billion in Cloud Computing Resources to Serve the Public Good*, MICROSOFT (Jan. 19, 2016), https://news.microsoft.com/2016/01/19/microsoft-philanthropies-announces-commitment-to-donate-1-billion-in-cloud-computing-resources-to-serve-the-public-good/.

53. *Microsoft Enterprise*, MICROSOFT, https://onedrive.live.com/about/en-us/ (last visited July 27, 2019).

54. *See* Janie Chang, *Translator Fast-Tracks Haitian Creole*, MICROSOFT: RESEARCH BLOG (Feb. 4, 2010), https://www.microsoft.com/en-us/research/blog/translator-fast-tracks-haitian-creole/.

55. Press Release, Tableau Software, Tableau Foundation Supports Data for Good Movement with Grants to DataKind and Bayes Impact (Mar. 12, 2015), https://www.prnewswire.com/news-releases/tableau-foundation-supports-data-for-good-movement-with-grants-to-datakind-and-bayes-impact-300049382.html.

56. *Dartmouth Announces $10 Million Gift from Bill Neukom '64*, DARTMOUTH NEWS (Apr. 30, 2014), https://news.dartmouth.edu/news/2014/04/dartmouth-announces-10-million-gift-bill-neukom-64.

57. For example, privacy and security risks are only marginal in this context, and considerations around tax breaks are more crucial for monetary donations of this sort than they currently are for data donations.

58. THE PERSONAL GENOME PROJECT, http://personalgenomes.org/ (last visited July 27, 2019); *see also* AMERICAN GUT, http://americangut.org/ (last visited July 27, 2019) (facilitating participation in studies about gut microbiome); Rumi Chunara & Sofia Ahsanuddin, *The GoViral Study*, J. GLOBAL HEALTH (June 14, 2016), https://www.ghjournal.org/the-goviral-study/ (last visited July 27, 2019) (using collected specimens from people who experience the flu or flulike symptoms). There are also examples of initiatives where individual data donation is only partially central to their business model, such as the personalized health network. PATIENTSLIKEME, http://news.patientslikeme.com/ (last visited July 27, 2019) (allowing patients to learn about

"citizen scientists," amateur individuals who help professional scientists to speed up discoveries and innovation.[59] Among other voluntary supporting acts, citizen scientists share information they digitally collect by observing environments, monitoring neighborhoods, and documenting occurrences that may contribute to the advancement of knowledge.[60]

Each type of initiative yields a variety of benefits and prompts a host of challenges, some of which have already been spotlighted and analyzed in academic scholarship. This Article, however, exclusively addresses data philanthropy, loosely defined, as the sharing of privately held data for socially beneficial purposes.

B. THE DEFINITION OF DATA PHILANTHROPY

Data philanthropy is largely defined through the combination of three elements: (1) unpaid for sharing of or access to (2) privately held data or proprietary data insights for (3) the greater good. Under this framework, data philanthropy may present itself in several configurations that are directly linked to the "sharing/access" model, the unique value of data, relevant stakeholders, sharing motivations, and the definition of socially beneficial causes. After discussing these elements of the definition, I move to discuss an additional element that current definitions of data philanthropy are missing: reuse outside the business model.

1. Sharing/Access Model

Scholars have identified five generic categories of data sharing: in-house production of statistics, transfer of data sets to end users, remote access, trusted third parties, and moving the algorithms.[61]

The first model, in-house production of statistics, is the most common model of collaboration.[62] Businesses do not grant access to the data itself, but

new treatment options, meeting others with similar medical conditions, and opting to contribute data for research).

59. *See generally* CAREN COOPER, CITIZEN SCIENCE: HOW ORDINARY PEOPLE ARE CHANGING THE FACE OF DISCOVERY (2016); Jonathan Silvertown, *A New Dawn for Citizen Science*, 24 TRENDS IN ECOLOGY & EVOLUTION 467 (2009).

60. For example, volunteers for the British "Track a Tree" project record and collect information about woodland trees and flowering plants to "provide insights into the seasonal timing of woodland species, and how future changes in climate may affect the interactions between trees and flowering plants." *What is Track a Tree?*, TRACK A TREE, http://trackatree.bio.ed.ac.uk/about (last visited July 27, 2019).

61. Thilo Klein and Stefaan Verhulst, *Access to New Data Sources for Statistics: Business Models and Incentives for the Corporate Sector*, PARIS21 PARTNERSHIP IN STAT. FOR DEV. IN THE 21ST CENTURY 17 (Discussion Paper No. 10, 2017), http://www.thegovlab.org/static/files/publications/paris-21.pdf. In addition to these models, a sixth model is emerging under which deep learning models, as opposed to the data itself, are being shared and run concurrently on several data sets. This method is especially helpful in the context of medical data that is held by multiple holders, but each holder holds a small sample size that by itself cannot produce reliable results. *See* Ken Chang et al., *Distributed Deep Learning Networks among Institutions for Medical Imaging*, 28 J. AM. MED. INFORMATICS ASS'N 945 (2018).

62. Klein & Verhulst, *supra* note 61, at 17.

analyze the data in-house and release or otherwise share the resulting statistics. By analyzing the data in-house, private sector actors can retain control over the generation and use of the data, guarantee adherence to standards of users' privacy protection, and increase data security.[63] MasterCard, for example, offers what it terms "data knowledge" through its Center for Inclusive Growth.[64] After identifying a need for better data, MasterCard's in-house experts analyze the relevant company's data and release the findings for broader use.[65] The internal production of statistics is limited to utilizing the expertise of company employees only; it does not leverage external skills. This model also requires internal infrastructure and technical mastery that smaller players often cannot afford.[66] Another example for this type of sharing is Facebook's "Disaster Maps." Launched in 2017, Facebook's Disaster Maps uses aggregated, anonymized Facebook data in disaster areas to deliver crucial information to aid organizations during and after crises.[67] By offering location density maps, movement maps, and safety check maps, Facebook shares the deliverable insights from the data analyzed, without sharing the actual data.[68]

Under the second model, private sector actors share data by transferring copies of data sets directly to end-users.[69] Users then develop their algorithms to run on the often de-identified and aggregated data.[70] Access to granular data is most effective for research purposes, where analysis entails detailed information and the merger of different data sets and sources.[71] Producers relinquishing control of this data, however, increase operational risks of data leakage, security breaches, and privacy harms.[72] Telecom company Orange used this sharing model in its Data for Development ("D4D") challenge,[73] in which selected researchers are granted access to transformed mobile data in order to develop applications for socially beneficial purposes, like disease monitoring and public transport improvement for developing countries.[74] Past challenges

63. *Id*.

64. Shamina Singh, *A Call to Action on Data Philanthropy*, MASTERCARD CTR FOR INCLUSIVE GROWTH (Oct. 4, 2016), https://mastercardcenter.org/action/call-action-data-philanthropy/.

65. *Id.* ("For example, in partnership with the White House's Data Driven Justice Initiative—an effort to use data to help advance criminal justice reform—the Center was able to perform an analysis to demonstrate the impact crime has on merchant locations and local job opportunities in Baltimore.").

66. *Id*.

67. Molly Jackman, *Using Data to Help Communities Recover and Rebuild*, FACEBOOK NEWSROOM, June 7, 2017, https://newsroom.fb.com/news/2017/06/using-data-to-help-communities-recover-and-rebuild/.

68. *Id*.

69. Klein & Verhulst, *supra* note 61, at 19.

70. *Id*.

71. *Id*.

72. *Id*.

73. *Data for Development*, ORANGE, https://www.orange.com/en/Footer/Thematic-features/2013/D4D/Data-for-Development (last visited July 27, 2019).

74. *Id*.

successfully forecasted the spread of epidemics and improved evaluation of internal migration and population density.[75]

Under this model, the data is sometimes offered in the form of open data repositories. Like its governmental counterpart, private sector open data involves the release of privately-owned data in machine-readable, downloadable, usable, and distributable formats, available for anyone to access and reuse. Yahoo!, for instance, through its Webscope Program, offers a reference library of "interesting and scientifically useful datasets for non-commercial use by academics and other scientists."[76] Google has released several datasets including: data on requests to remove content due to copyright; visual databases for machine learning researchers in collaboration with Carnegie Mellon and Cornell University; and YouTube-8M, a dataset of 8 million labeled YouTube videos for video understanding research.[77] In addition, Uber introduced Uber Movement, a website where users can access some of the company's internal demand and usage data.[78]

In the third model of data sharing, end users securely access the data remotely while data controllers maintain control over the extracted information.[79] Unlike the second model, under the remote access model, data is not duplicated and does not leave the premises of the data holder. When sharing remotely, data holders activate strict monitoring of the input and output traffic on their data storage devices to safeguard the shared data, and users can only export the final aggregated metrics.[80] Commonly, data holders outsource the

75. Antonio Lima et al., *Progmosis: Evaluating Risky Individual Behavior During Epidemics Using Mobile Network Data* 1 (2015) (unpublished manuscript), https://arxiv.org/abs/1504.01316. *See* Gabriel Pestre et al., *The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates*, ANN. WORLD BANK CONF. ON DEV. ECON. 1, 1 (2016), http://pubdocs.worldbank.org/en/551311466182785065/Pestre-Letouze-Zagheni-ABCDE-May-2016.pdf. Other companies have also made their datasets available via challenges or competitions, where technologists are encouraged to conduct research or analysis on the data and share their discoveries. For example, Yelp offers a dataset challenge. *Yelp Dataset Challenge*, YELP, https://www.yelp.com/dataset/challenge (last visited July 27, 2019). Netflix initiated its famous Netflix Prize, an open competition for the best collaborative filtering algorithm to predict user ratings for films based on previous ratings only. *Netflix Prize*, WIKIPEDIA, https://en.wikipedia.org/wiki/Netflix_Prize (last visited July 27, 2019). Facebook held a Recruiting Competition where the donors of top entries received the opportunity to interview with the company. *Facebook Recruiting Competition*, KAGGLE, https://www.kaggle.com/c/FacebookRecruiting (last visited July 27, 2019).

76. *Webscope: Datasets*, YAHOO!, https://webscope.sandbox.yahoo.com/ (last visited July 27, 2019). In early 2016, Yahoo released its largest-ever machine learning dataset to the academic research community. Sarah Perez, *Yahoo Releases its Biggest-Ever Machine Learning Dataset to the Research Community*, TECHCRUNCH (Jan. 14, 2016), https://techcrunch.com/2016/01/14/yahoo-releases-its-biggest-ever-machine-learning-dataset-to-the-research-community/.

77. Sudheendra Vijayanarasimhan & Paul Natsev, *Announcing YouTube-8M: A Large and Diverse Labeled Video Dataset for Video Understanding Research*, GOOGLE: AI BLOG (Sept. 28, 2016), https://research.googleblog.com/2016/09/announcing-youtube-8m-large-and-diverse.html.

78. *Movement Cities*, UBER MOVEMENT, https://movement.uber.com/cities?lang=en-US (last visited July 27, 2019).

79. Klein and Verhulst, *supra* note 61, at 21. This model, as well as the fourth model, are also referred to as "privileged access." *See* Fluitt, *supra* note 28, at 15–16.

80. Klein and Verhulst, *supra* note 61, at 21.

operation of remote access-based collaborations to external third parties.[81] This model sets a fertile ground for data philanthropy collaborations with relatively low operational risks. Identification risks, however, still exist, and so does the potential of losing a competitive edge as competitors may attempt to extract strategic insights from the aggregates.[82] The collaboration between Flowminder and Ncell, mentioned above, is a typical example of the remote access model.[83] Flowminder's reports on largest-scale population displacements after a natural disaster were produced through remote access and were later shared with the United Nations Office for the Coordination of Humanitarian Affairs and other key agencies for pre-disaster preparations and more effective post-disaster relief.[84]

In the fourth model of data sharing, data holders and users rely on a trusted third party to facilitate secure access to the data.[85] To maximize protection, users do not have direct access to the raw data and, instead, must request reports or transitional results from a trusted facilitator.[86] Trusted third parties must possess reliable technical infrastructure, including large data storage capacity and secure connections.[87] While this model increases the operational risk level as more external parties handle the data, it lowers costs and streamlines the data sharing process.[88]

In the fifth model, the data remains in the data holder's possession, and the data holder runs an algorithm of the data user's choice.[89] With shared algorithms, different data holders can perform the same analytical functions on their data sets without merging them with data sets from their competitors.[90] The results that are shared directly with users can be merged for the sake of a more comprehensive analysis.[91] The algorithmic sharing model lessens many of the risks of the other models, as most of the analyses take place within data holder's premises.

2. The Unique Value of Data

A private sector actor acquires an interest in data through collecting the data itself or by purchasing data captured by other players. In the course of supporting a product or providing a service, businesses can actively gather information by, for example, conducting surveys or passively collecting data

81. *Id*.
82. *Id*. at 21–22.
83. *Case Study: Nepal Earthquake 2015*, FLOWMINDER, http://www.flowminder.org/case-studies/nepal-earthquake-2015 (last visited July 27, 2019).
84. *Id*.
85. Klein and Verhulst, *supra* note 61, at 22–23.
86. *Id*.
87. *Id*.
88. *Id*.
89. *Id*. at 23–24.
90. Klein and Verhulst, *supra* note 61, at 23–24.
91. *Id.*

such as by recording geo-location. The potential for targeting-based monetization, either by the collecting company or by a third party acquiring the data, motivates most passive collection.[92] While users' explicit consent is a prerequisite for lawful data collection and use, collection of personal information is assumed to be permitted even when users have not explicitly given consent (commonly through clickwrap and browsewrap agreements).[93]

Several qualities of big data have induced the growing interest in potential uses of private data for the public good. There is an increasing demand for evidence-based social action, both in the process of devising policy and in practical terms.[94] Technological tools facilitate the capturing and monitoring of social activities, and analysis of the data produces actionable insights. The ability to capture and analyze real-time measurements is especially valuable for social action that must often quickly respond to unforeseeable events while evaluating short-term policies in the course of the response.[95] Big data can also increase the exposure and visibility of less conspicuous social action that is happening "below the radar" as individuals team up to identify and address a social issue.[96] Algorithms that detect hidden trends and patterns in data and analysis that foresees future development with a relatively high level of accuracy are also immensely valuable for social action. For example, these systems can not only identify individuals in need during humanitarian crises but also forecast their movements, advancing more efficient allocation of resources.[97] Big data technologies also have potential uses that were unknown when the technology was first developed, and these potential uses later turn out to be of great worth over time.[98]

In this context, mobile data possesses a set of unique qualities and is considered especially useful for social action.[99] In addition to assuring ubiquitous connectivity, mobile technologies are virtually always switched on.[100] Consequently, users leave an uninterrupted track of records that is both

92. Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect*, 2013 U. CHI. LEGAL F. 95, 95 (2013).

93. Facebook allowing developers to scrape information from its platform until 2014 exemplifies this notion.

94. 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://grandchallengesforsocialwork.org/wp-content/uploads/2015/12/WP11-with-cover.pdf.

95. Chandy et al., *supra* note 47.

96. NESTA, DATA FOR GOOD: HOW BIG AND OPEN DATA CAN BE USED FOR THE COMMON GOOD 25 (Peter Baeck ed., 2015), https://media.nesta.org.uk/documents/dataforgood.pdf.

97. KATIE WHIPKEY & ANDREJ VERITY, GUIDANCE FOR INCORPORATING BIG DATA INTO HUMANITARIAN OPERATIONS 7 (2015), http://digitalhumanitarians.com/sites/default/files/resource-field_media/IncorporatingBigDataintoHumanitarianOps-2015.pdf.

98. Chandy et al., *supra* note 47, at 710.

99. In fact, the UN Global Pulse has dedicated significant time to studying the state of mobile data for social good. *See* Mobile Data for Social Good Report, *supra* note 3.

100. Chandy et al., *supra* note 47, at 710.

high volume and granular.[101] Because cell ranges can cover less than a thousand meters in heavily populated areas, mobile data offers a high spatial resolution.[102] Mobile data is also of particular interest for humanitarian organizations because it can reveal population movements. Tracking these movements is crucial in the course of a natural disaster or disease outbreak, but it is also constructive for urban planning purposes.[103] Mobile data can also shed light on socio-economic trends, as well as the economic health and resilience of communities.[104]

3. Stakeholders

Data philanthropy entangles various stakeholders, each of which is guided by distinctive interests and tasked with turning the data into socially beneficial insights. Effective policy proposals must account for the complex operation of data philanthropy and identify both the stakeholder groups and the interests driving them. For this discussion, the most useful and inclusive taxonomy to describe the roles of stakeholders within the data philanthropy universe, which is suggested by Mikel Niño and others, distinguishes between problem holders, data holders, and skill holders.[105] The relationships between the different stakeholders are commonly facilitated by contracts that define roles, responsibilities, rights, and duties.

Problem Holders are the individuals or institutions closest to the target population affected by the social issue that the data-driven strategy tackles.[106] Stakeholders in this group usually include representatives of public administration, like governmental agencies and non-profit organizations, who work to identify the social problem, the needs of affected groups, and potential solutions.[107] Ideally positioned to provide key knowledge to other stakeholders, the problem holders act as a resourceful intermediary. They express the interests of the target population members who do not directly interact with other stakeholders. By doing so, the problem holders not only give voice to isolated or otherwise muted collectives, but also improve the data-driven solution by sharing their perspectives with the skill and data holders.

Data Holders include the individuals or institutions that hold the data. The specifics of the interest in the data depend on the particular instance of data

101. Nicholas Robin et al., *Public-Private Partnerships for Statistics: Lessons Learned, Future Steps* 6–7 (OECD, Working Paper No. 27, 2016), https://www.oecd-ilibrary.org/docserver/5jm3nqp1g8wf-en.pdf?expires=1562909961&id=id&accname=guest&checksum=DD01771516A6A731342E12B4DE0DA910.

102. *Id*.

103. MOBILE DATA FOR SOCIAL GOOD REPORT, *supra* note 3, at 5.

104. *Id*.

105. Mikel Niño et al., *Data Projects for "Social Good": Challenges and Opportunities*, 11 INT'L J. HUMAN. & SOC. SCI. 1094 (2017). The mobile data for social good report proposes a more detailed list of stakeholders that includes roles such as the Data Producer (the individual creating the data), and Data Steward (the one entrusted with monitoring and evaluating the data to confirm its adherence to agreed standards). *See* MOBILE DATA FOR SOCIAL GOOD REPORT, *supra* note 3, at 10, tbl. 2.

106. Niño et al., *supra* note 105, at 1097.

107. *Id*.

sharing; the interest may arise because the data holder collected the data or otherwise acquired intellectual property or contractual rights in the data. While one might assume that this group includes the data subjects,[108] i.e., those generating the relevant data to be analyzed, from a legal perspective, such an assumption is mostly incorrect. U.S. law treats notice and consent as the legal threshold for lawful collection of information, but does not recognize any proprietary interest of data subjects.[109]

Skill Holders are equipped with the technological expertise to extract actionable insights from the data. These data experts make up for the data illiteracy and lesser technological expertise of other stakeholders.[110] They possess data analytics skills and translate the results to the problem holders so the latter can then devise relevant policy solutions or practical strategies to address the social issue. Examples of skill holders include the UN Global Pulse,[111] DataKind,[112] and academic researchers.

These categories are not firmly distinct; in fact, they often merge. For example, when a public organization or an NGO has been gathering information on a social issue it is attempting to address, it would act as both a problem holder and a data holder. Similarly, data holders may, at times, operate as skill holders too. For example, in 2016, Facebook analyzed data pulled from Brazilian users' posts about Zika.[113] The social media giant then shared the insights with UNICEF, which incorporated the findings into a successful ad campaign, resulting in more awareness and preventive measure-taking in the Brazilian population.[114]

Differentiating between the various stakeholders, however, is essential to guarantee a balanced data philanthropy partnership. Each of the stakeholders hold interests that are indispensable for a data-based collaboration for the greater good: without the problem holders, there is no one to safeguard the interests of affected communities; without the data holders, there is no data to analyze; and without the skill holders, data is useless.

4. Sharing Incentives

Incentives to collaborate vary among the groups of stakeholders. As representatives of affected communities, problem holders work with other

108. In fact, Niño and others make specific reference to the data subjects as part of the data holders' group, "Sometimes the people or collective in need are indeed a data source in themselves. For instance, their interactions with different systems or services could generate relevant data to be processed and analyzed in the project." *Id*.

109. *See, e.g*., Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013) [hereinafter *Privacy Self-Management*].

110. Niño, *supra* note 105, at 1097.

111. Global Pulse is a flagship innovation initiative of the United Nations Secretary-General on big data. *See* UNITED NATIONS GLOBAL PULSE, https://www.unglobalpulse.org/ (last visited July 27, 2019).

112. DATAKIND, *supra* note 49.

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

114. *Id*.

stakeholders to solve the social issues that interfere with or risk the lives of people from those communities. Their motivation is mostly altruistic and is reinforced systematically through their form of governance—commonly, public agencies or NGOs. Depending on their organizational affiliation, skill holders may share a similar set of altruistic interests, offering their analytics skills as a donation of services to promote a socially beneficial end.[115] Skill holders can also collaborate to promote an individual goal with or without an altruistic aspect. For example, while academic researchers often pursue certain studies from public-spirited perspectives, they are also motivated by professional aspirations.

These motivations are broad data-for-good motivations. As the term "philanthropy" suggests, data philanthropy mostly points to the sharing motivations of data holders, especially to the extent they operate in the private sector. The information economy flourishes because data is a valuable financial asset. Access to privately-held data and data-driven insights have a price tag in the market, and when data holders choose to give them away for free, they lose potential income. Sharing also exposes data holders to a variety of risks, as explained below,[116] making the motivation question even more puzzling: why would private businesses share their data for socially beneficial purposes at no cost?

Corporations are inclined to allow access to, or sharing of, collected data for a variety of reasons. Thilo Klein and Stefaan Verhulst map out different business incentives for data-based collaborations.[117] According to Klein and Verhulst, sharing motivations depend on the context of the sharing, the questions posed, and the corporate and legal culture of the firm.[118] Even though their taxonomy is aimed at data collaboratives and not data philanthropy, the classification works well to describe most of the interests' dynamics in data philanthropy. Klein and Verhulst suggest six categories of sharing incentives: reciprocity; research, recruitment and insights; reputation and public relations; increasing revenue; regulatory compliance; and responsibility and corporate philanthropy.[119]

Reciprocity: Klein and Verhulst describe two types of reciprocity scenarios. The first involves pure business interest—the sharing could produce mutual benefits for the data holder and the other stakeholders, especially if the combination of the data holder's data and data from other sources provides some advantage to the former.[120] This setup is common when the collaboration involves only private sector for-profit actors, such as the Accelerating Medicines Partnership, where pharmaceutical companies share genetic and molecular data

115. DATAKIND, *supra* note 49 (epitomizing donation of data-analytics services for socially beneficial purposes).

116. *See infra* Subparts I.C.1, I.C.2 and accompanying text.

117. Klein & Verhulst, *supra* note 61, at 9–13.

118. *Id.* at 9.

119. *Id*. at 9–13.

120. *Id*. at 9.

in a data pool.[121] Although such collaborations foster innovation and promote social interests, it is questionable to classify them as data philanthropy. It is a business partnership that happens to benefit the public, with no problem holder to communicate the problems of the affected population and safeguard its interests. If a problem holder is present in this kind of collaboration and the social issue, as opposed to the business interest, was seen as instrumental in the process, this type of reciprocity motivation would apply.

The second type of reciprocity, which is more applicable to data philanthropy, is one of compensation. Companies that collect personal information give back to counterbalance what they have taken from individuals and society at large.[122] This kind of data philanthropy acts as a non-mandatory surveillance tax, similar to the carbon tax: a company that pollutes society with surveillance "pays" by donating some of its data for socially beneficial causes.

Research, Recruitment, and Insights: Opening up data may help companies generate new insights that they cannot, as a practical matter, or should not, as a financial matter, generate in-house.[123] Sharing data allows companies to enjoy data analytics expertise that their employees may not possess, often free of charge.[124] By exploring the data in new ways, outsiders could expose potentially promising business models and identify themselves as talented hires for the sharing business.[125] The Orange D4D challenge exemplified this incentive,[126] as does the Spanish Bank BBVA Innova Challenge, where participants gain access to some of the bank's data.[127] Past winners of the challenge have developed socially beneficial applications, such as one that predicts overcrowding in city buildings, as well as commercial solutions for the bank—for example, better ways to support customers.[128]

Reputational Advantage: Sharing data for socially beneficial purposes could do wonders for a business's reputation. When sharing is extensively publicized, private sector actors gain media attention, which could increase their exposure to potential users, investors, and business partners.[129] For example, Orbital Insights, a data analytics company, shared satellite data and geo-

121. *Accelerating Medicines Partnership*, NAT'L INST. OF HEALTH, https://www.nih.gov/research-training/accelerating-medicines-partnership-amp (last visited July 27, 2019).

122. Klein & Verhulst, *supra* note 61, at 9.

123. *Id*. at 10.

124. *Id*.; *see also* Jordana George et al., *Data Philanthropy: An Explorative Study*, *in* PROC. OF THE 52ND HAW. INT'L CONF. ON SYS. SCI. 5858, 5864 (2019) ("UPS gained additional logistics data, algorithms, and experiences from working in these high-risk regions. Such work also increased employee satisfaction & retention, particularly for talented data scientists such as Soldner Freeman.").

125. Klein & Verhulst, *supra* note 61, at 10.

126. *See supra* notes 73–74 and accompanying text.

127. *Innova Challenge Big Data Highlight*, BBVA, https://bbvaopen4u.com/en/actualidad/innova-challenge-big-data-highlights (last visited July 27, 2019); *see also* Thomas Hale, *The BBVA Bank Vaults That Hoard Data Instead of Bullion*, FIN. TIMES (Feb. 25, 2015), https://www.ft.com/content/bbbfebc4-b79f-11e4-981d-00144feab7de.

128. Klein & Verhulst, *supra* note 61, at 10.

129. Klein & Verhulst, *supra* note 61, at 11.

analytics in collaboration with the World Bank to track poverty around the world.[130] This partnership spawned significant investor interest for Orbital, which, as of recently, has raised a total of $78.7 million.[131]

The positive publicity is also a helpful asset for businesses in their ongoing relationship with policymakers. When lobbying or otherwise attempting to convince policymakers to pursue a certain course of action, companies can point to their data philanthropy history to demonstrate an alignment between the public interest and their business interest.

Increasing Revenue: Klein and Verhulst explain that, in the course of data collaborations, corporate data is sometimes offered for sale, not shared for free.[132] This direct form of increasing revenue is characteristic of data collaboratives but not of data philanthropy, which requires unpaid sharing of, or access to, privately-held data. Nevertheless, even though data philanthropy does not generate direct increase in revenue, it does, as the other incentives discussed in this Subpart demonstrate, indirectly cut costs and increase revenues for private sector actors in the short and long term.

Regulatory Compliance: Some companies generate data to secure regulatory compliance, and sometimes data sharing is required under sectorial regulations.[133] Repurposing data that is already collected or that must be shared for socially beneficial uses increases the value of the compliance-motivated investment.[134] With this logic in mind, Apple, Cisco, Dell, and Google (among others) release the data they submit in their Employer Information Report (EEO-1), a compliance survey mandated by the Equal Employment Opportunity Commission.[135]

Responsibility and Corporate Philanthropy: Data philanthropy overlaps with traditional corporate philanthropy and corporate social responsibility principles. Corporate donation is a company's way of giving back to the community, but is also valuable for the business, as the act of giving indirectly improves the competitive business environment.[136] Sharing data for socially beneficial purposes could similarly better the operation ecosystem for the company.[137]

---

130. *Id.* at 11–12.; *see also* Orbital Insight, *Leveraging Commercial Applications to Help the World Bank Map Poverty*, MEDIUM (Jan. 4, 2017), https://medium.com/from-the-macroscope/leveraging-commercial-applications-to-help-the-world-bank-map-poverty-79bca51814ee.

131. *Orbital Insight Raises $50 Million to Track Economies from Space*, BLOOMBERG (May 2, 2017), https://www.bloomberg.com/news/articles/2017-05-02/orbital-insight-raises-50-million-to-track-economies-from-space (last visited July 27, 2019).

132. Klein and Verhulst, *supra* note 61, at 12.

133. *Id*.

134. *Id*.

135. *Id*.

136. *Id*.

137. Klein and Verhulst, *supra* note 61, at 12 ("A classic example of such an ecosystem-supporting responsibility is a company that contributes data to help improve education, which could eventually improve the labour pool from which they hire staff."); *see also* Matt Stempeck, *Sharing Data is a Form of Corporate*

5. What Is "Good?"

Data philanthropy is praised for advancing the greater good through the reuse of private-sector data for socially beneficial purposes. But what is good? The literature is currently lacking an established comprehensive definition of both the descriptive and normative aspects of the public good. Instead, commentators have pointed to examples of "good" collaborations and used them as case studies. For example, using data to predict dengue fever outbreaks more quickly in Pakistan and tracking human migration in Nepal following an earthquake are socially beneficial uses of mobile data.[138]

Commentators have also proposed categories of social action where data has been used for the common good. For example, in one study the authors refer to studies of forced migration, disease, poverty and economic stagnation, ethnic divisions, and ecological and environmental crises as instances of socially beneficial data collaborations.[139] Another article identified additional application areas, discussing humanitarian crises, global health care and health disparities, ecology and global-scale environmental issues, rural development, human rights, crime prevention, and child welfare.[140] However, these lists are illustrative rather than exhaustive and do not provide solid guidelines for identifying the "social good" threshold.

The data philanthropy practice calls for a comprehensive definition of social good and a set of concrete guidelines to instruct collaborations. As in other areas of the law, devising a definition for what should be considered a social good is not only challenging, but also highly context-dependent and entirely normative. While such definition is beyond the scope of this Article, it is important to note that until a definition is formally constructed, it is up to the various stakeholders to characterize a certain undertaking as promoting the social good. There are clear social good cases, such as those of established humanitarian efforts, and there are borderline cases where the contribution is unclear or attenuated. As the decision to donate data involves various motivations, among which are commercial profit-maximizing interests, controversial characterizations of social good should be scrutinized rigorously.

6. Reuse Outside the Business Model

This Article adds an additional illuminating layer to existing definitions of data philanthropy: in data philanthropy collaborations, the socially beneficial data reuse resides outside the scope of the sharing entity's business model. This addition highlights the difference between socially beneficial data use that is integral to a business's design, and socially beneficial data reuse that is made outside the scope of a sharing business's activity.

Philanthropy, HARV. BUS. REV. (July 24, 2014), https://hbr.org/2014/07/sharing-data-is-a-form-of-corporate-philanthropy.

138. MOBILE DATA FOR SOCIAL GOOD REPORT, *supra* note 3, at 7.

139. Chandy et al., *supra* note 47, at 704.

140. Niño et al., *supra* note 105, at 1095.

Business use of data-driven insights may be beneficial for consumers[141]: effective ad targeting could mean exposure to more applicable commercial content; a useful search engine would rank results based on their relevance to the individual consumer; and the use of patient data by hospitals has shown to improve monitoring and the accuracy of patient medical histories.[142] Business use of data-driven insights could also promote public ends. For example, Finland-based Enevo optimizes waste collection and recycling by collecting and analyzing data from refuse containers around the world.[143] While these data uses are socially beneficial, none of them qualify as data philanthropy. Data philanthropy operates on an incentive system that, while likely to generate positive spillover effects on a business reputation or otherwise promote commercial interests, is external to the core business model. Offering a socially beneficial product or service to a business's consumers, as well as promoting the greater good through for-profit social ventures, produces social value that is directed by the design of the commercial enterprise. The laws and social norms governing these beneficial uses of data account for their incentives structure. Data philanthropy, which looks to extract additional social value outside the territory of the sharing entity's business model, is scrutinized differently based on costs, benefits, and the different incentive structure on which it operates.

This additional requirement not only identifies different incentives in terms of a business's goals and allocation of resources, but also implies the involvement of an external party. Data reuses that are not formally contemplated could broaden engagement to include external stakeholders. In this sense, the additional requirement aligns perfectly with the designated roles of stakeholders in data philanthropy: a socially beneficial collaboration between a problem holder, a skill holder, and a data holder implicates external parties, which add an additional layer of oversight and accountability.

C. THE CHALLENGES OF DATA PHILANTHROPY

Data philanthropy is a useful instrument for socially beneficial actions. Yet, it raises a variety of challenges, including business risks and individual or group injuries that may trigger broader social harm. Four types of risk are commonly identified with data philanthropy: costs and competitive disadvantage; privacy, security, and ethics; legal constraints; and error and bias in private sector data.

---

141. *See, e.g.*, Amy J. Schmitz, *Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots,"* 2014 MICH. ST. L. REV. 1411, 1452 (2014); Jessica A. Wood, *The Darknet: A Digital Copyright Revolution*, 16 RICH. J.L. & TECH. 14, 54 (2010); Wullianallur Raghupathi & Viju Raghupathi, *Big Data Analytics in Healthcare: Promise and Potential*, 2 HEALTH INFO. SCI. & SYS. (2014), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/.

142. Avi Goldfarb & Catherine Tucker, *Privacy and Innovation*, 12 INNOVATION POL'Y & ECON. 65, 70 (2014).

143. ENEVO, https://enevo.com/ (last visited July 27, 2019).

1. Costs and Competitive Disadvantage

For many profit-driven, private sector actors, the notion of voluntarily sharing privately held data free of charge is largely implausible. Sharing the data, or otherwise allowing access to it, exposes the information to security risks, human error, and abuse. If the data falls into the hands of a business' competitor, it could quickly result in destructive financial loss. Risking a competitive edge without strong evidence of significant gains from the sharing activity makes no business sense. Clearly, not all privately-held data sharing generates a similar level of risk—the degree of the risk depends on the type of data and how close it is to the core confidential segments of a company's operation.[144] But sometimes it is the data from which information about a business's customers or strategy can be extracted that is essential for promoting a certain social good. Businesses engaging in data philanthropy are also likely to incur costs, like labor and setup, none of which can be justified as a core business-related need.[145]

2. Privacy, Security, and Ethics

Many commentators have identified privacy as one of the greatest challenges to data philanthropy.[146] Data collected by private sector actors often contains personal and sensitive details about individuals' lives, from their physical whereabouts and social interactions to their shopping preferences and financial standing.[147] In some of the sharing models of data philanthropy, individuals or entities that were rarely authorized by the data subjects to review or analyze their data get access to it.[148] A commonly cited concern in this context

144. Robin et al., *supra* note 101, at 8. For a discussion of similar concerns in the context of cancer research, see Michael Mattioli, The Data-Pooling Problem, 32 BERKELEY TECH. L.J. 179, 209–214 (2017).

145. *Mobile Data for Social Good Report*, *supra* note 3, at 13.

146. *See, e.g.*, Silja M. Eckartz et al., *A Decision Model for Data Sharing*, in ELECTRONIC GOVERNMENT. 253, 255 (Marjin Janssen et al. eds., 2014) ("In settings where data is shared with or between private organizations, most barriers to data sharing are related to privacy or to competition regarding economically sensitive data."); FPF REPORT, *supra* note 17, at 11 ("Privacy and security were cited as the top concern for companies that hold personal data because of the serious risk of re-identification."); Sean Martin McDonald, *Ebola: A Big Data Disaster: Privacy, Property, and the Law of Disaster Experimentation*, THE CTR FOR INTERNET & SOC'Y (2016) (discussing the "marked tension in the debate around experimentation with humanitarian technologies and the impact on privacy"); MOBILE DATA FOR SOCIAL GOOD REPORT, *supra* note 3, at 14 (citing "[l]ack of common approach to data privacy and risk mitigation associated with data use" as one of the main concerns around uses of mobile data for social good); Niño, *supra* note 105, at 1078 ("Despite the growing concern about access to and sharing of personal information, this field lacks a clear and effective framework to address legal, ethical and privacy issues related to the use of personal and sensitive data [sic]."); Klein & Verhulst, *supra* note 61, at 6 ("[I]n reality, gaining access to the data is often a formidable challenge due to privacy, confidentiality, and security concerns, as well as cross-jurisdictional regulatory incompatibilities in how data may be owned and transferred.").

147. *See generally* Nizan Geslevich Packin & Yafit Lev-Aretz, *Big Data and Social Netbanks: Are You Ready to Replace Your Bank?*, 53 HOUS. L. REV. 1211 (2016) (discussing the expansion of big data companies and social network in the financial services market and considering the ramifications of bringing in massive troves of consumer data collected in a variety of contexts into the financial context).

148. When collection purposes are determined, the scope of the authorization from the user is set. Because of the serendipitous nature of data at that point, it's hard to know which potentially beneficial uses can later arise. *See* discussion in Part IV.

is identification of specific individuals and the sensitive information associated with them in the shared data.[149] Two decades ago, the ultimate technological cure for these kinds of privacy concerns was anonymization. Because personal data could be stripped of any reference to its subjects, sharing anonymized data was considered safe and harmless.[150] Anonymization, however, turned out to be far from identification-proof. Researchers have repeatedly demonstrated that data subjects may be identified from anonymized datasets.[151] Stories like that of Thelma Arnold, the sixty-two-year-old AOL customer identified by the New York Times in anonymized AOL datasets,[152] stand as iconic reminders of the anonymization failure. Once anonymized data is shared and aggregated, users can often be reidentified, either because the patterns of behavior recorded in the data are unique to a particular individual or because the anonymized data, when combined with external data sources, unveil the identity of its subjects.[153]

An alternative technological safeguard for personal information is differential privacy, a mathematically-driven solution to reidentification risks.[154] Under this framework, query results are altered by adding noise to the dataset and making it difficult to identify individuals with high certainty.[155] Even though many advocates of differential privacy claim that a well-designed model can provide robust anonymization while allowing for rich statistical analysis, some argue that this model, too, is not identification-proof.[156]

Furthermore, privacy harms materialize not only for individuals, but also for groups. Even if the dataset is less granular, identification of demographic groups could be risky for these groups, as they can become the target of discriminatory or otherwise harmful policies.[157]

149. *See, e.g.*, Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1744–48 (2010); Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?* 10 AM. J. BIOETHICS 3, 5–7 (2010); Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 WASH. L. REV. 703, 720 (2016) (discussing the risk of reidentification in the course of genetic research); Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J. L. & TECH. 1, 3–4 (2011).

150. Ohm, *supra* note 149, at 1716.

151. Arvind Narayanan and Vitaly Shmatikov were able to re-identify individuals from an anonymous dataset provided by Netflix as part of a contest to improve the company's movie recommendation engine. *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, *in* 2008 PROC. OF IEEE SYMP. ON SEC. & PRIVACY 111 (2008).

152. Ohm, *supra* note 149, at 1717; *see also* Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), http://www.nytimes.com/2006/08/09/technology/09aol.html.

153. Ohm, *supra* note 149, at 1723–25.

154. Cynthia Dwork, *Differential Privacy*, 33 INT'L COLLOQUIUM ON AUTOMATA, LANGUAGES & PROGRAMMING, 1 (2006).

155. Rubinstein & Hartzog, *supra* note 149 at 718.

156. MOBILE DATA FOR SOCIAL GOOD REPORT, *supra* note 3, at 16.

157. *Id*.

3. Legal Constraints

In data philanthropy, privacy harms are also tied to the data subject's consent or lack thereof. In some cases, private sector actors cannot share their data because their users simply did not agree to the secondary use of the data. The repurposing of data, regardless of the socially beneficial motivation behind it, does not adhere to the "respect for context" principle, under which data should only be used for the purpose for which it was collected.[158] In the absence of informed consent, all stakeholders also risk violating fundamental autonomy principles.[159] However, in a world of standard form contracts and online terms of service, these cases are somewhat rare, because corporate players often stipulate to the use of collected data for very broadly defined purposes.[160] As many privacy advocates rightly note, this kind of formal consent rarely resembles true, clear, and informed consent.[161] Most people do not read privacy policies and terms of service, with reasons ranging from lack of interest and difficulty understanding the legal language, to the time-consuming nature of reading those contracts and consumers' nonexistent bargaining power.[162] Users are more likely to avoid reading contracts when a great number of consumers are bound by the same terms, because they assume that the terms must be reasonable.[163]

When personal information about the business's consumers is exposed, those individuals or groups suffer an invasion of their privacy that may impose emotional, physical and/or economic harm. The sharing business may also be subject to harms like: criminal investigations or civil legal liabilities; regulatory fines; loss of regulatory licenses or certifications; crucial reputational harm; drops in share prices or increases in cost of capital; massive departures of existing customers; lower employee recruitment, productivity, and retention; and an overall increase in operating expenses.[164] Privacy harms may also bring down social action actors. Take the case of inBloom, a non-profit organization aimed at making student data available for approved third-party applications and software for educators.[165] The collection and use of personally identifiable

158. Jane R. Bambauer, *All Life Is an Experiment. (Sometimes It Is a Controlled Experiment.)*, 47 LOY. U. CHI. L.J. 487, 490 (2015).

159. *Id*.

160. Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1432–35 (2000); Mayer-Schönberger & Padova, *supra* note 16, at 322; Jeff Sovern, *Opting In, Opting Out, or No Options at All: The Fight for Control of Personal Information*, 74 WASH. L. REV. 1033, 1072–74 (1999); Strandburg, *supra* note 92, at 142; Richard Warner, *Undermined Norms: The Corrosive Effect of Information Processing Technology on Informational Privacy*, 55 ST. LOUIS U. L.J. 1047, 1084–86 (2011); Tal Zarsky, "*Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE. J. L. & TECH. 1, 33–34 (2003).

161. *Privacy Self-Management*, *supra* note 109.

162. Packin & Lev-Aretz, *supra* note 147, at 1279.

163. *Id*.

164. Klein & Verhulst, *supra* note 61, at 15.

165. *Id*.

information about students resulted in a major backlash, and ultimately inBloom's demise.[166]

In addition to privacy risks, shared private sector data is subject to information security risks. Deficient, outdated or inflexible security protocols give rise to data vulnerabilities that could make hacking effortless and may lead to negligent leakages.[167] In the course of sharing data, as information sometimes leaves a business's servers or as more parties are involved in processing or analyzing the data, these risks are exacerbated.

4. Error and Bias in Private Sector Data

Another set of informational concerns with respect to data philanthropy is the perpetuation of existing inequalities and the creation of new ones due to data bias and error. Inaccurate data affects its quality and generates erroneous data output. Unofficial data sources can sometimes be decentralized, unstandardized, unstructured, and unrepresentative.[168] The risk of finding irrelevant or bogus correlations with statistical significance is endemic to big datasets, and the potential for errors increases greatly when multiple data sets are combined.[169] Furthermore, due to the disparity in worldwide technology proliferation, many data sources suffer temporal and spatial restraints, which, if not acknowledged appropriately, could result in significant errors.[170]

Data can also be incomplete or otherwise non-representative, overlooking "data invisibles."[171] The risk of partial representation is mostly common in social networking data because, although there are still many individuals that do not socially network online, social networking data is widely used for research and analyses and assumed to provide a fairly representative sample of the general population.[172] Errors can also result from alterations to the context and semantics of the data in the course of collection or analysis.[173]

166. *Id.*; *see also* Natasha Singer, *InBloom Student Data Repository to Close*, N.Y. TIMES (Apr. 21, 2014), https://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/.167.Klein & Verhulst, *supra* note 61, at 15.

167. Klein & Verhulst, *supra* note 61, at 15.

168. Robin et al., *supra* note 101, at 8.

169. *See* Danah Boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO. COMM. & SOC'Y 662, 668 (2012).

170. Anwaar Ali et al., *Big Data for Development: Applications and Techniques*, BIG DATA ANALYTICS, July 1, 2016, at 11.

171. Klein and Verhulst, *supra* note 61, at 15.

172. Boyd & Crawford, *supra* note 169, at 669 ("Twitter does not represent 'all people', and it is an error to assume 'people' and 'Twitter users' are synonymous: they are a very particular sub-set. Neither is the population using Twitter representative of the global population. Nor can we assume that accounts and users are equivalent.").

173. A famous example of context-less errors is the multiple contexts of the word "smoking:" "without further rules to refine that term, the keyword will retrieve plenty of content about "smoking marijuana," "smoking ribs," and "smoking hot girls." *See* Mark Myslín et al., *Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products*, 15 J. MED. INTERNET RES. (2013); Ashley Sanders-Jackson et al., *Applying Linguistic Methods to Understanding Smoking-Related Conversations on Twitter*, 24 TOBACCO CONTROL 136 (2015).

In addition to error, all phases of data collection, processing, and use are susceptible to human bias.[174] Input bias, which stems from biased or lacking source data, is a concern for data philanthropy models that facilitate access to raw data.[175] Some data also goes through a cleaning process, filtering data that is deemed "dirty." Cleaning the data sometimes requires subjective, non-technical decision-making, which may inject further bias into the data.[176] For example, a Catholic health system provided a dataset of patient records to the U.S. government for cancer research.[177] As it turned out, transgendered and transsexual patients labeled themselves as being of "UNKNOWN" sex and gender, but in-house informaticists imputed or inferred their sex based on other available data, such as their height and weight.[178]

The risk of bias extends to training and programming as well. Training bias results from poor definition of baseline data or inadequate research strategy, while programming bias presents itself in the algorithmic design.[179] Both can lead to misinterpretation of the data and flawed decisional inferences that could lead to ineffective, discriminatory, or otherwise harmful actions.

Importantly, human bias is often unintentional, unconscious, and unavoidable: "[e]ven in situations where data miners are extremely careful, they can still effect discriminatory results with models that, quite unintentionally, pick out proxy variables for protected classes."[180] Flickr's auto-tagging of black men as "animal" or "ape" in users' photos,[181] as well as Google's targeting search results for black-sounding names with ads about criminal activities,[182]

---

174. Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward A Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 99 (2014) (discussing potential instances of discrimination in big data predictions); Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677 (2016); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1262 (2008) ("The biases of individual programmers can have a larger, accumulating effect, because, in a complex software system composed of smaller subsystems, the actual bias of the system 'may well be a composite of rules specified by different programmers.'"); Citron & Pasquale, *supra* note 22, at 4 ("Because human beings program predictive algorithms, their biases and values are embedded into the software's instructions, known as the source code and predictive algorithms."); Helen Nissenbaum, *How Computer Systems Embody Values*, COMPUTER, Mar. 2001, at 119 (explaining that seemingly objective systems can generate unfair discrimination).

175. *See generally* Karen R. Chinander and Maurice E. Schweitzer, *The Input Bias: The Misuse of Input Information in Judgments of Outcomes*, 91 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 243 (2003) (describing input bias and its consequences on data).

176. Nizan Packin and Yafit Lev-Aretz, *Learning Algorithms and Discrimination*, *in* RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 88, 91 (Woodrow Barfield & Ugo Pagallo eds., 2018).

177. Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 561 (2014).

178. *Id*.

179. Barocas & Selbst, *supra* note 174, at 683–84 ("Because data mining relies on training data as ground truth, when those inputs are themselves skewed by bias or inattention, the resulting system will produce results that are at best unreliable and at worst discriminatory.").

180. *Id.* at 675.

181. Alex Hern, *Flickr Faces Complaints Over 'Offensive' Auto-Tagging for Photos*, GUARDIAN (May 20, 2015), https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos.

182. Lauren Kirchner, *When Discrimination Is Baked into Algorithms*, ATLANTIC (Sept. 6, 2015), https://www.theatlantic.com/business/archive/2015/09/discrimination-algorithms-disparate-impact/403969/.

perfectly demonstrate both the difficulty of warding off biased uses of data and how such uses can dangerously reinforce existing stereotypes.

II. MAKING ROOM FOR DATA PHILANTHROPY

Responsible data philanthropy can contribute greatly to social causes, to the progress of science, to human advancements, and even to saving lives. Responsible data philanthropy can also contribute to a more just allocation of resources among different groups in society. But the key word here is "responsible." What makes data philanthropy responsible and what role should the law play in setting the limits?

A. DATA PHILANTHROPY: A LEGAL PERSPECTIVE

Data philanthropy currently operates in a legal vacuum. Many questions are left open and decided on the go by private sector actors and other participants. What guidance should the law provide to distinguish between data sharing that promotes the greater good and donations that may end up generating more evil than good? How can the law encourage data sharing and, at the same time, circumscribe it to prevent risks from materializing ex-ante and to mitigate them ex-post?

Many complex legal and ethical aspects of data philanthropy are ripe for exploration. For example, to what extent should the law treat donations of data as a charitable contribution for tax breaks? How should the non-rivalry quality of data affect the quantification of the donation? Should intellectual property protections play a role in reinforcing competitive advantage in data philanthropy initiatives? Is data philanthropy a digital age articulation of traditional philanthropy? Should the law group data philanthropy with other corporate social responsibility initiatives? Are there instances of data sharing for socially beneficial purposes that should not be instigated voluntarily but legally mandated? Can proprietary interests of data subjects, which have been rejected as a basis for intellectual property rights, justify such a mandate for sharing? Would data philanthropy be better facilitated through a centralized model in which the government plays a democratic oversight role? Should we distinguish between in-house corporate research, corporate-funded research, and traditional independent academic research? What would such substantive differentiation look like? Should a company's motivation for sharing impact the classification of the sharing as data philanthropy? Are acts of data philanthropy that are primarily intended to bolster relationships with policymakers any different than lobbying? Who can be a legitimate problem holder, skill holder, or data holder? How should legal tools prevent data philanthropy from turning into a trump card to support unlimited data retention? How can the law minimize bias in data-driven social insights and mitigate harmful effects ex-post?

These are just some of the questions that data philanthropy raises, and surely many more will arise over time.[183] While future projects might provide further guidance on these issues, this Article centers on what many have listed as the most pressing challenge to data philanthropy—privacy risks. However, before exploring the legal toolkit to find the appropriate legal device for reconciling data philanthropy with privacy values, this Article offers a skeptical view of the claim that privacy is a considerable legal hindrance to data philanthropy.

B. PRIVACY: A PROBLEM OR A SYMPTOM?

Most current discussions around data-for-good and data philanthropy pit the social benefits from the reuse against various legal risks—primarily privacy.[184] However, are privacy risks truly interfering with data philanthropy in practice? What is it about privacy that stands to stop or limit data philanthropy initiatives?

Before answering this question, it is important to highlight an often-neglected fact: many socially beneficial data reuses, especially in the context of environmental and agricultural initiatives, do not involve human subject data.[185] For example, in 2014, Intel shared data from sensors—located in crops and other strategic areas to monitor soil and air moisture levels—with researchers from the Earth Research Institute at the University of California.[186] Nothing in this data set involved human information or risks to individual privacy.

Nevertheless, most data philanthropy initiatives have specific interest in human subject data, an interest which, according to the current sentiment in the field, leads to serious privacy risks. In this context, privacy risks could mean either the real-life materialization of privacy risks or the legal liability that might be attached to their unfolding. In some cases, privacy risks are addressed by regulations that ensure adequate protections, while in other cases, privacy harms do not give rise to legal liability. For example, the sale of information about one's shopping preferences to third parties is, to many, a privacy-intrusive practice, yet it is a common, legally valid transaction facilitated through contractual consent. In other words, data philanthropy and privacy might conflict in two possible ways: when data philanthropy interferes with businesses' compliance with privacy regulations, or when data philanthropy

183. For a list of existing guiding legal sources on the data reuse, as well as their limitations in the humanitarian context, see U.N. OFFICE FOR THE COORDINATION OF HUMANITARIAN AFFAIRS, DATA RESPONSIBILITY GUIDELINES WORKING DRAFT (2019), https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf.

184. *See supra* Subpart I.C.2.

185. For example, Global Forest Watch offers a variety of data and tools to monitor forests. *See* GLOBAL FOREST WATCH, https://www.globalforestwatch.org/.

186. Lyndsey Gilpin, *How Intel is Using IoT and Big Data to Improve Food and Water Security*, TECHREPUBLIC (June 13, 2014), https://www.techrepublic.com/article/how-intel-is-using-iot-and-big-data-to-improve-food-and-water-security/.

threatens individuals' privacy interests, notwithstanding the existence or application of a privacy regulatory framework to a given reuse of information.

Compliance concerns are overstated. A recent report by the Future of Privacy Forum has found that, in regulated domains, some companies have pointed to privacy regulations as a barrier to sharing information with researchers.[187] Other companies, whose shared data was covered by privacy regulations, reported building regulatory compliance into the sharing process.[188] This discrepancy may imply that privacy compliance costs are sufficiently high to disincentivize voluntary sharing for socially beneficial purposes but could also indicate that the reference to privacy compliance costs is used to mask other, less publicly laudable reasons for non-sharing.[189] Furthermore, many data markets are not subject to privacy regulations that limit data sharing. In those markets, most businesses rely on broadly defined terms of service to allow various uses and reuses of the data, and the same could and has been done to cover instances of data philanthropy.[190]

The difficulties around privacy compliance, therefore, do not seem to explain the emphasis on privacy risks constantly voiced in discussions around data philanthropy. An alternative possibility could be concerns around privacy harms that do not result in legal liability, or where legal risks are secondary to the reputational effects of potential privacy harms. Here, too, the voiced concerns seem overstated. In virtually all data philanthropy collaborations, personally identifiable data is successfully shared with privacy safeguards.[191] Scholars from various academic disciplines have also been engaging in proposals for new or improved privacy protections in data-for-good exchanges.[192] Institutional and structural qualities of stakeholders in data

187. FPF REPORT, *supra* note 17, at 11.

188. *Id*. Similarly, BBVA's Data and Analytics team has shared financial data with the UN Global Pulse to measure communities' resilience after a natural disaster. BBVA shared its customer data in an anonymized and aggregated form to comply with national laws and regulations. Press release, BBVA, UN Global Pulse, BBVA Announce Partnership and New Project Measuring Economic Resilience to Disasters with Financial Data (Sept. 13, 2016), https://www.bbva.com/en/un-global-pulse-bbva-announce-partnership-new-project-measuring-economic-resilience-disasters-financial-data/.

189. A similar observation was made in the context of sharing cancer data, "Interestingly, several experts suggested that HIPAA provides a plausible excuse for institutions that do not wish to share data for reasons unrelated to privacy, such as reputational concerns. This argument is 'particularly hard to argue with,' one subject stated." Mattioli, *supra* note 144, at 209.

190. *See Data Policy*, *supra* note 16.

191. While imperfect, these privacy safeguards include aggregation and anonymization, as well as limited access. See for example, the Yale School of Medicine's Open Data Access (YODA) Project, which facilitates the access of researchers and physicians to medical device and anonymized clinical trial data from Johnson & Johnson, and Harvard School of Public Health's malaria tracking research, where phone company Safaricom shared de-identified data that was then used to model the travel patterns of cell phone users. Press release, Harv. Sch. of Pub. Health, Using Cell Phone Data to Curb the Spread of Malaria (Oct. 11, 2012), https://www.hsph.harvard.edu/news/press-releases/cell-phone-data-malaria/.

192. *See, e.g.*, Yves-Alexandre de Montjoye et al., *Enabling Humanitarian Use of Mobile Phone Data*, ISSUES IN TECH. INNOVATION, Nov. 2014, at 7–8; Linnet Taylor, *The Ethics of Big Data as a Public Good: Which Public? Whose Good?*, 374 PHIL. TRANSACTIONS ROYAL SOC'Y, Dec. 28, 2016, at 10 (2016); Katherine J. Strandburg, *Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context*, *in*

philanthropy collaborations also often decrease the prospects of unethical exploitation of personal information by direct participants. Profit-maximizing businesses have their reputation on the line, researchers in academic institutions are subject to Institutional Review Board approval and strict ethical research standards, and public institutions are committed to promoting the public interest in a transparent and accountable manner.

Voiced concerns around privacy could also be the result of bad rhetoric that fails to take into account the complexity around corporate data sharing and points to an obvious, relatable concern. In an oft-cited example, the media implied that the sharing of mobile data for Ebola tracking purposes was thwarted due to privacy hysteria notwithstanding the immense potential social benefit.[193] However, as shown above, data philanthropy involves a number of sharing incentives and disincentives, as well as interests of various stakeholders.[194] Pointing to privacy as the only or most pressing impediment to data philanthropy is patently wrong.

As the above discussion shows, data philanthropy involves privacy risks, but they can be avoided, minimized, and mitigated. Most importantly, privacy risks—in terms of compliance and potential reidentification—are wrongly cited as major impediments to data philanthropy. Instead, I argue business and not-for-profit actors have highlighted privacy concerns for two main reasons: the expanding definition of privacy, and the need for legal acknowledgment.

While many commentators refer to privacy risks, in the context of data philanthropy, as risks related to reidentification of data subjects, threats to privacy materialize through a variety of information-related risks. Because data philanthropy has been explored in various disciplines, the meaning of the term privacy has not been uniformly applied. Privacy has acquired a broad meaning in the legal, philosophical, and social context: whereas in the past, privacy was perceived as the risk of human observation and subsequent judgment,[195] over the years the term has broadened to include a variety of information-related concerns such as error, bias, manipulation, and discrimination.[196] While they are

---

PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 1, 25–29 (Julia Lane et al. eds., 2014).

193. Taylor, *supra* note 192, at 6, 8.

194. *See supra* Subpart II.B.

195. Recall Warren and Brandeis' characterization of the right to privacy as a right to be "let alone" and unbothered by other humans, to decide to what extent a person's "thoughts, sentiments, and emotions shall be communicated to others." Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195, 198 (1890). The emphasis on the "others" is crucial. *Id.* at 199. William Prosser's formulation of the four privacy torts protected only against privacy injuries that incorporate human observation and judgment: (1) public disclosure of private facts, (2) intrusion on seclusion, (3) depiction of another in a false light, and (4) appropriation of another's image for commercial gain. William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383, 389 (1960).

196. Yafit Lev-Aretz, *Privacy and the Human Element* (unpublished article) (on file with the author). For broader interpretations of the right to privacy *see, e.g.*, Crawford & Schultz, *supra* note 174; David Gray & Danielle Citron, *The Right to Quantitative Privacy*, 98 MINN. L. REV. 62, 83–87 (2013) (describing how Fourth Amendment search and seizure has been narrowed to respond to a broader concept of personal privacy); Mary Madden et al., *Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans*, 95 WASH. U. L.

more nuanced and complex than individual identification, these risks are often labeled privacy risks. And these risks may be exacerbated in the complex ecosystem of data philanthropy.

The call to provide a framework for privacy protections in data philanthropy not only misidentifies the risks but also the required response. What is often called for is not necessarily a framework for balancing information-driven risks with social good, but a formal legal acknowledgement to signal that data philanthropy follows market standards and social norms and is endorsed by governing institutions. In other words, the demand is not for legal intervention for the sake of guiding stakeholders on privacy matters; it is a demand for legal intervention for the sake of legal recognition of data philanthropy. Privacy, in this context, provides an easily identified tool for legal recognition.

The legal guidance offered in the next Parts will address some of the broader informational problems that the term privacy represents as well as the lack of legal acknowledgement. After analyzing the current legal landscape, this Article hones in on the FIPs as the best-suited legal lever for providing both an acknowledgement of and guidance on informational risks in data philanthropy.

C. CURRENT LEGAL LANDSCAPE

Article 12 of the Universal Declaration of Human Rights protects individuals from arbitrary interference with their privacy and calls for establishing legal protections against such interferences.[197] Around the world, legal systems have translated this abstract commitment into real-life legal interventions through a variety of privacy protections at different conceptual and practical levels. In the United States, these laws, regulations, and policies are derived from separate, overlapping authorities—state and federal statutory laws, agency regulations, industry best practices, and private contractual agreements.[198]

Federal privacy law lacks comprehensive legislation that addresses informational privacy across all industries. Instead, privacy protection in the commercial sphere is applied and enforced through a sectoral approach.[199] Commonly, these laws are narrowly customized to specific categories of data in specific industries or practices. Examples include the Telecommunications

---

REV. 53, 64–67 (2017) (discussing the disparate treatment poorer communities experience in relation to privacy risks and concerns); Theodore Rostow, *What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers*, 34 YALE J. ON REG. 667, 676–69 (2017) (describing statutory privacy protections in the commercial sphere in response to consumers' broader definition of privacy).

197. G.A. Res 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948).

198. Rostow, *supra* note 196, at 676.

199. *See, e.g.*, DANIEL J. SOLOVE & PAUL SCHWARTZ, INFORMATION PRIVACY LAW 792–94 (5th ed. 2015); Michael C. James, *A Comparative Analysis of the Right to Privacy in the United States, Canada, and Europe*, 29 CONN. J. INT'L L. 257, 260 (2014); Omer Tene, *Privacy Law's Midlife Crisis: A Critical Assessment of the Second Wave of Global Privacy Laws*, 74 OHIO ST. L.J. 1217, 1217 (2013) [hereinafter *Privacy Law's Midlife Crisis*].

Act,[200] the Fair Credit Reporting Act (FCRA),[201] The Health Information Portability and Accountability Act (HIPAA),[202] The Gramm-Leach-Bliley Act (GBLA),[203] the Family Educational Rights and Privacy Act (FERPA),[204] and the Electronic Communications Privacy Act (ECPA).[205]

Federal agencies have also been involved in privacy rulemaking. For example, in 2016, the Federal Communications Commission passed a set of landmark privacy protections for internet users—now repealed[206]—requiring ISPs to disclose information collection practices and obtain consumers' consent for the selling of that information.[207] The Federal Trade Commission, which is authorized to target "unfair or deceptive acts or practices,"[208] has no rulemaking authority. It nevertheless has been active on the privacy front through issuing numerous privacy complaints against private sector actors and entering consent decrees with corporate players like Facebook and Snapchat over unfair privacy practices.[209]

There are also theoretical, general, and non-binding policies to guide the collection and use of personal information. One such source is the FIPs, a set of widely accepted principles listing protections for personal information. The FIPs were first introduced in a report by the advisory committee to the Secretary of Health, Education, and Welfare in 1973, in response to the growing use of data banks and other recordkeeping systems storing and processing personal information.[210] The report prescribed a list of fair information principles, including transparency, use limitation, access and correction, data quality, and

200. Under the Telecommunications Act, Internet Service Providers (ISPs) may not use, disclose, or permit access to identifiable customer network information for purposes outside the provision of the services from which the information is derived. 47 U.S.C.A. § 222(c)(1) (West 2017).

201. The FCRA establishes certain duties for consumer reporting agencies and affords protections for personal credit information. Fair Credit Reporting Act, Pub. L. No. 91-508, § 601, 84 Stat. 1114, 1128 (1970) (codified as amended in scattered sections of 15 U.S.C.).

202. HIPAA imposes obligations on doctors and medical services when handling their patients' data. Health Information Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified as amended in scattered sections of 26, 29 & 42 U.S.C.).

203. The GBLA regulates data practices in financial services. Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999) (codified as amended in scattered sections of 12 & 15 U.S.C.).

204. FERPA institutes fair information practices in the education sector. Family Educational Rights and Privacy Act, 20 U.S.C. § 1232g (2012).

205. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (1986) (codified as amended in scattered sections of 18 U.S.C.).

206. Brian Fung, *The House Just Voted to Wipe Away the FCC's Landmark Internet Privacy Protections*, WASH. POST (Mar. 28, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/03/28/the-house-just-voted-to-wipe-out-the-fccs-landmark-internet-privacy-protections/?utm_term=.83b93df4d869.

207. Press Release, Federal Communications Commission, FCC Adopts Privacy Rules to Give Broadband Consumers Increased Choice, Transparency and Security for Their Personal Data (Oct. 27, 2016), https://www.fcc.gov/document/fcc-adopts-broadband-consumer-privacy-rules.

208. 15 U.S.C.A. §§ 45, 52 (West 2017).

209. Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 600, 610 (2014).

210. U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS AND THE RIGHTS OF CITIZENS: REPORT OF THE SECRETARY'S ADVISORY COMM. ON AUTOMATED PERSONAL DATA SYSTEMS (1973).

security. Over time, different constructions of the FIPs have been articulated and incorporated into a number of data protection regimes around the world.[211]

The most important and oft-cited restatement of the FIPs is the *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, published by the OECD in 1980.[212] The OECD Guidelines specify eight fair information principles:[213]

The Collection Limitation Principle limits the collection of personal data, calling for it to be obtained by lawful and fair means with the knowledge or consent of the data subject.

The Data Quality Principle requires that collected data be relevant to the purposes for which it will be used and is accurate, complete, and kept up-to-date.

The Purpose Specification Principle entails ex-ante specification of the collection purposes.

The Use Limitation Principle prohibits the disclosure or use of data for purposes other than those specified at the time of collection unless the consent of the data subject has been obtained, or the disclosure or use is required by the authority of law.

The Security Safeguards Principle ensures that reasonable security safeguards to personal data are in place.

The Openness Principle demands transparency about data collection practices and policies including the main purposes of the data use and the identity and location of the data controller.

The Individual Participation Principle gives individuals the right to know which data is collected about them; to access the data within a reasonable time, in a reasonable manner, and in a readily intelligible form; and to challenge the data and have inaccurate data erased, rectified, completed, or amended.

The Accountability Principle requires that data collectors be accountable for complying with the principles stated in the guidelines.

In addition to being governed by federal and state statutes and the soft-law provided by the FIPs, the collection and use of personal information is also governed by contracts. This is largely because both legal and FIPs protection of information privacy place a crucial emphasis on individuals' consent as legitimizing "nearly any form of collection, use, or disclosure of personal

211. Gellman, *supra* note 27, at 1.

212. ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, OECD GUIDELINES ON THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA (1980) [hereinafter OECD GUIDELINES], http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflows ofpersonaldata.htm. Subsequent U.S. versions include the FTC and Homeland Security privacy guidelines. *See* DEPT. OF HOMELAND SECURITY, PRIVACY POLICY GUIDANCE MEMORANDUM (2008), http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf; FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICY MAKERS (2012), https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers.

213. OECD GUIDELINES, *supra* note 212.

data."[214] As a result, private sector actors can and often do, include broad data collection purposes in their terms of service.[215] Users, who rarely read the terms of service that they agree to, formally accept these stipulations, which in practice act to legally legitimize repurposing of collected data. As argued above, with consent-based privacy protection, such stipulations often allow data sharing relatively immune to legal risk.

III. DATA PHILANTHROPY AND THE FIPS

Current legal protections of privacy have allowed for many instances of data philanthropy. Indeed, without statutory or regulatory attention, data philanthropy is mostly governed by contractual agreements that may authorize data sharing for socially beneficial purposes. Why not leave the specifics of data philanthropy to data collectors and users? After all, some users willingly agree to responsible and safe sharing of personal information about themselves for socially beneficial causes, and it is safe to assume that most people, when presented with the opportunity to share data in a responsible manner, would agree as well.[216] Why should the law interfere further if it currently allows data philanthropy to exist?

The answer is that the law allows data philanthropy to exist, but nothing more. Despite a growing interest in the utilization of data-for-good and many examples of data philanthropy that have benefited societies around the world, the legal community lags behind. There are no works on how data giving can be done responsibly, no discussions of the practical outcomes of engaging in data philanthropy, no explorations of various sharing structures and players, and no guidance that could both incentivize safe sharing and make it safer.

The contractually manageable state of data philanthropy has contributed to the current legal vacuum. But another plausible reason for this relative silence on the legal front could be the struggle that many privacy scholars encounter when trying to think seriously about data philanthropy. After all, in a culture of constant surveillance, endless data collection, monetization of personal information, and little respect for privacy, it feels normatively uncomfortable to discuss the legal facilitation of beneficial uses of this "dirty" data. To borrow from another legal discipline, this data has the appearance of the fruit of a very poisonous tree.[217] The term philanthropy, which is very appealing to the private

214. *Privacy Self-Management*, *supra* note 109, at 1880; *see also* Lisa M. Austin, *Enough About Me: Why Privacy Is About Power, Not Consent (or Harm)*, *in* A WORLD WITHOUT PRIVACY: WHAT LAW CAN AND SHOULD DO? 131, 132–33 (Austin Sarat ed., 2015); Woodrow Hartzog, *The Inadequate, Invaluable Fair Information Practices*, 76 MD. L. REV. 952, 964–66 (2017); *Privacy Law's Midlife Crisis*, *supra* note 199, at 1218–19.

215. *Data Policy*, *supra* note 16, and accompanying text.

216. *See* Subpart I.A (listing examples of individual data philanthropy).

217. Under the exclusionary rule, if primary evidence in a criminal case was illegally obtained and is thus tainted, then all evidence derived from it may be subject to the same flaw as they are all fruits of the same poisonous tree. *See* Nardone v. United States, 308 U.S. 338, 340 (1939); United States v. Hernandez, 670 F.3d 616, 620 (5th Cir. 2012); United States v. Galaviz, 645 F.3d 347, 354 (6th Cir. 2011).

sector, is condemned and rejected by many privacy advocates. How, they ask, can data that was lawfully but often immorally obtained and used be re-channeled as a charitable form of giving?

These concerns are understandable, and the intuitions behind them are valid. They cannot, however, stand in the face of a scholarly investigation of and normative work on data philanthropy. By leaving data philanthropy unexplored we fail to increase social value: fighting to limit immoral collection practices is socially desirable, but socially beneficial reuses of private sector data are equally so. We should use an alternative effective term to replace "philanthropy," we should disapprove of immoral information practices, and we must fight for a better privacy-protecting future. At the same time, we must acknowledge that information that has already been collected can be effectively reused for socially beneficial causes. This data languishes in servers while it could be repurposed to advance research and promote humanitarian causes. Ignoring this data because we disagree with the way it was obtained or used is the worst of all worlds—it does not prevent future damage or mitigate existing harms, and it leaves independent academic views outside the ongoing discussion about a working definition of responsible data philanthropy.

Data philanthropy raises important legal questions that are being answered offhand, on a case-by-case basis without the proper involvement of legal scholars and those who are affected by these decisions. The latter group is bound by contracts of adhesion that are rarely read, never negotiated, and which make these unpremeditated decisions around data philanthropy into an approved legal standard. A legal academic discussion of data philanthropy must take place and must, for the most part, be done independently of concurrent advocacy and policy efforts to stop harmful information practices. Extreme cases of highly illegitimate information practice combined with a pressing social need would require balancing work, but as a general rule, we should attempt to keep questions about the legitimacy of the collection and the legitimacy of the reuse utterly separate. The fact that data is used for a socially beneficial purpose does not immunize its collection and use from being challenged on moral, ethical, and legal grounds. At the same time, data that has already been collected should not be off-limits for socially beneficial uses merely because it was obtained illegitimately.

This Article advocates for a data philanthropy exception to the FIPs. In the next Subsections, it explains why the FIPs are currently the best-suited channel for the legal governance of data philanthropy, and why broad interpretation cannot achieve the same purpose as a designated exception. This Article concludes by proposing a general framework for legal checks on privacy protection in data philanthropy through the FIPs.

A. WHY THE FIPS?

Data philanthropy could be legally tackled through legislative reform. However, three objections loom large. First, designing statutory tools require a

deep and comprehensive understanding of the data philanthropy universe and the various challenges that the law has to address. Data philanthropy, however, is a relatively young phenomenon. While there are many examples of socially beneficial data sharing, they are diffused across different data holders, skill holders, and problem holders, various levels of exigency, and diverse social benefits. Put differently, because data philanthropy is nascent and administered impromptu, further study and experimentation is required before prescribing rigid guidelines through legislation.

Second, privacy could be potentially undermined in the legislative process because of pressure from interest groups. As the public choice theory proposes and as history confirms, the lawmaking process implicates organized interest groups who strive to promote their agenda.[218] The resulting legislative product is determined by relative group strength—the group with most political capital commonly exercises superior influence on the lawmaking process.[219] In the context of privacy, public choice concerns are exacerbated. Devoting special legislation to address data philanthropy would bundle privacy together with other data-giving issues in a manner that could prevent appropriate balancing. Privacy has failed as a policy goal because of similar bundling in the past, which has framed privacy as an individual interest that has to bend in the face of societal objectives.[220]

Third, federal legislation is costly, could take years to complete, and cannot be updated promptly. To turn a bill into law, several stages of internal consideration must come to pass: legislators in congressional committees must approve the proposal, which must then be discussed in hearings, examined in debates, approved by majorities in both houses of Congress, and then either approved by the President or supported by a veto-override in both houses.[221] State legislation would likely entail shorter and less complex processes, but would still involve significant costs and time to finalize. Furthermore, state legislation invites an additional ground for challenge—that of different legal treatments and lack of harmonization across jurisdictions.

This is not to say that data philanthropy should be exempted from statutory governance. In the future, comprehensive uniform or sectoral legislation is likely to offer better facilitation of data philanthropy. But in the current political climate around privacy and data-for-good uses, and with data philanthropy in its

218. Yafit Lev-Aretz, *Copyright Lawmaking and Public Choice: From Legislative Battles to Private Ordering*, 27 HARV. J.L. & TECH. 203, 213-16 (2013) [hereinafter Lev-Aretz, *Copyright Lawmaking and Public Choice*].

219. *Id*.

220. PRISCILLA M. REGAN, LEGISLATING PRIVACY 22-23 (1995) ("The policy process began with an emphasis on the value of privacy, and much of the policy debate was framed in terms of an individual interest — privacy — in conflict with a societal interest — government efficiency, law enforcement, and an honest work force. In policy debates, the individual interest was on weaker footing than the societal interest. Privacy was on the defensive because those alleging a privacy invasion bore the burden of proving that a certain activity did indeed invade privacy and that the individual privacy interest was more important than the societal interest.").

221. Lev-Aretz, *Copyright Lawmaking and Public Choice*, *supra* note 218, at 243-44.

early stages of development, public legislation is unlikely to quickly and effectively materialize. While less ideal, devising initial intervention in data philanthropy through the combination of industry self-regulation and soft-law levers would be more realistically achievable at this point. Specifically, a data philanthropy exception to the FIPs would act as a subtle form of intervention to both limit instances of risky data sharing and incentivize responsible data philanthropy.

The FIPs are not without problems. Privacy scholars have had a "love-hate" relationship with the FIPs since their inception because, as Professor Woodrow Hartzog stated, "[w]hile the FIPs have been remarkably useful, they have painted us into a corner."[222] Since their early days, the FIPs have been widely criticized for their substance and for the way in which they have been implemented.[223] The FIPs were regarded as efficiency principles that sought better functioning and fairer (but not necessarily fair) information systems without providing true privacy protection in the age of ubiquitous surveillance.[224] Not only have the FIPs not contributed to a more privacy-respecting culture, but they have, in fact, worsened the state of privacy because they give the illusion of protection while laundering surveillance through formalistic compliance.[225] The FIPs terms, preoccupied as they are with individuals' consent, can legitimize destructive information collection practices.[226]

In addition to these commonly voiced critiques, Hartzog points to other, less discussed complaints about the FIPs, such as their limited scope that overlooks the effect of design signals and transaction costs on trust, obscurity and autonomy.[227] The FIPs also neglect an important set of relationships in the age of networks and double-sided markets—that of information intermediaries[228]—and seem to be detached from humans' susceptibility to manipulation and the extent to which information, in the hands of those who wish to manipulate individual choice, can be a dangerous tool.[229] Similarly, automated decision-making represents a blind spot for the FIPs, which do not attend to the structural problems of automated systems, which include bias, discrimination, and the mistaken perception that humanly created automated systems generate objective facts.[230]

---

222. Hartzog, *supra* note 214, at 953.

223. *Id*. at 964.

224. *Id*. (referencing JAMES RULE ET AL., THE POLITICS OF PRIVACY 93 (1980)).

225. *Id*.; *see also* Fred H. Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF THE 'INFORMATION ECONOMY' 341, 342 (Jane K. Winn ed., 2006).

226. Hartzog, *supra* note 214, at 973; Cate, *supra* note 225, at 342; *Privacy Self-Management*, *supra* note 109, at 1881–82.

227. Hartzog, *supra* note 214, at 967.

228. *Id*. at 968.

229. *Id.* at 969–70. For more on this risk. *See* Yafit Lev-Aretz, *Free Choice Must Be Free*, TECHCRUNCH (June 29, 2017), https://techcrunch.com/2017/06/29/free-choice-must-be-free/.

230. Hartzog, *supra* note 214, at 972.

The FIPs also place a crucial emphasis on users' control. This emphasis, however, fails to empower users to make informed decisions about the flow of their personal information, and instead leaves them "bewildered, hopeless, and agreeable to anything."[231] Control does not scale and individuals cannot reasonably exercise the control they get through notice and consent models—reading all the terms of service and privacy policies that people are bound by in today's information economy is virtually impossible.[232] Even when people do read and have a good sense of their rights and obligations, they feel helpless in the face of corporate power and their inability to opt-out.[233] And as the FIPs continue to promote the control fixation in privacy practice, other important principles are forsaken and disused.[234]

Indeed, the FIPs have many shortcomings. And many of their flaws are heightened because of the FIPs' established standing and the fact that privacy policy, at this point, cannot do without them.[235] But this is exactly why they are the best place to start addressing issues around data philanthropy. Here, too, we must keep exploring to find the optimal regime to promote more transparency, accountability, and an appropriate mixture of privacy by design and user control. Improvements to the FIPs do not nullify the criticism they have rightly received. But against the backdrop of poor alternatives and the particular virtues of the FIPs that could adequately address some of the unique qualities of data philanthropy, they currently represent the best legal home for data philanthropy.

A legal intervention in data philanthropy should promote everything that the FIPs stand for: respect for privacy, pragmatism, a global focus, and sufficient open-endedness to allow jurisdictions to study data philanthropy and implement the legal administration of it as they see fit. As Hartzog rightly points out, the FIPs are "the closest thing the world has to a universal privacy touchstone."[236] Even though overreliance on the FIPs has led to many problems, the FIPs remain prevalent.[237] Their pragmatic nature, and their influence over information practices around the world, has turned the FIPs into an essential tool in the globalization of privacy policy.[238] The FIPs have similarly served to harmonize states' legislation in the United States and inspire across-the-board privacy standards in the industry.[239] Currently, anyone who speaks about privacy uses

231. *Id*. at 977.

232. *Id*. at 972–77.

233. *See generally*, Yoan Hermstrüwer & Stephan Dickert, *Sharing Is Daring: An Experiment on Consent, Chilling Effects and A Salient Privacy Nudge*, 51 INT'L REV. L. & ECON. 38 (2017) ("Salient ex ante consent options may lure people into giving up their privacy and increase their compliance with social norms."); Lior Strahilevitz & Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers?*, 45 J. LEGAL STUD. S69 (2016).

234. Hartzog, *supra* note 214, at 974.

235. *Id*. at 965.

236. *Id*. at 959.

237. Kirsten Martin & Helen Nissenbaum, *Measuring Privacy: An Empirical Test Using Context to Expose Confounding Variables*, 18 COLUM. SCI. & TECH. L. REV. 176, 187 (2016).

238. Hartzog, *supra* note 214, at 960–61.

239. *Id.*

some variation of the FIPs language. For data philanthropy to be guided in a privacy-respecting direction and concurrently incentivized to spread, a data philanthropy exception would be best introduced through this common and established model of fair information practices.

Before becoming the bedrock of information privacy protection, the FIPs had to make a choice between several competing notions of privacy. Among these competing notions, privacy as control prevailed in the FIPs, a choice that many have lamented and criticized—rightly, for the most part—for its unraveling effect on the state of privacy.[240] Still, for informational privacy to be practically administrated, a choice had to be made and maintained over time. Thanks to this choice, the FIPs could propose measures of privacy protection that, while lacking in many senses, still offer more than mere intuition and bolster other privacy values such as autonomy and fairness.[241]

One of the commonly cited reasons for the FIPs' longevity is their flexibility. The FIPs were first introduced at a time when big data, smartphones, artificial intelligence, and mass surveillance were the exclusive domain of science fiction.[242] Since then, all of these technologies and phenomena have become a reality, and the FIPs still bear relevance in the face of these dramatic changes. The FIPs' technology neutrality and timelessness are strongly linked to their reliance on open-ended principles. Adhering to some basic principles of information privacy in the abstract, the FIPs model is one of standards rather than rules.[243] They thus inspire new regulatory proposals and animate a piecemeal development of privacy protection.[244] New information practices like data philanthropy require both sufficient breathing room for development and high-level guidance for spread and growth. The FIPs' elasticity is thus crucial for guaranteeing privacy protection that could respond and be better tailored to future technological, legal, and social changes. The high-level articulation of the FIPs not only allows, but also entails, further deliberation over the details in specific contexts.

A data philanthropy exception to the FIPs would provide much needed guidance and instigate a scholarly conversation about the appropriate governing rules for responsible data sharing. Such exception would also act to reinforce the FIPs by acknowledging a domain in which some of these principles must be waived or partially enforced. The inclusion of a data philanthropy exception in a set of established, widely-accepted, and cross-jurisdictional privacy standards would increase the visibility of data philanthropy. The data philanthropy exception would introduce the possibility of engaging in responsible data

240. Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 254–55 (2011); Neil Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180, 1182–83 (2017); Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431, 436–37 (2016).

241. Hartzog, *supra* note 214, at 963.

242. *Id*. at 982.

243. Paul M. Schwartz, *Information Privacy in the Cloud*, 161 U. PA. L. REV. 1623, 1658–59 (2013).

244. Hartzog, *supra* note 214, at 961.

sharing with smaller private sector actors who may not have been aware of data philanthropy, or who have been aware but were deterred by its lack of legal acknowledgment.

B. BROADER INTERPRETATION VERSUS AN EXCEPTION

Of the different values of the FIPs, the purpose specification principle—requiring ex-ante particularization of collection purposes—as well as the use limitation principle—allowing disclosure or use of collected information only for those specified purposes—are mostly at odds with data philanthropy.[245] Both purposes necessitate ex-ante stipulation of specific uses to guide collection and use activities, while data philanthropy relies on repurposing already-collected data. In light of this conflict, guiding data philanthropy through the FIPs means that data philanthropy either must be incorporated into the FIPs as an exception to these principles or that the FIPs must be interpreted broadly to legitimize data philanthropy.

Because the purpose specification and use limitation principles aim to provide a framework for treatment of users' data that meets individuals' expectations, it makes sense to argue that a broader interpretation of the FIPs should acknowledge a spectrum of socially beneficial repurposes to which users would normally agree. Where the social benefit is significant and the privacy risk is low or non-existent, most individuals, if asked, would agree to repurpose collected information about themselves for the greater good.[246] Such constructive agreement is entirely within the realm of reasonable interpretations of the FIPs, as they constantly engage in balancing privacy with other values and aligning privacy practices with privacy expectations.

However, the interpretative approach suffers a number of weaknesses that cloud its appealing merits. First, data philanthropy is compatible with the themes of the FIPs but is greatly antithetical to the purpose specification and use limitation principles. Reconciliation of those principles with data philanthropy requires a broad interpretation to the point of overriding the language prohibiting unauthorized reuse. Applying such broad interpretation could set a precarious precedent leading to similarly broad readings of the FIPs in other contexts and effectively eroding the protections they grant. Second, a designated interpretation-based exception for data philanthropy would exacerbate existing inconsistencies between different versions of the FIPs. The informal application of the FIPs to data philanthropy is likely to be mapped out differently around the globe. Opening the door to such broad interpretation could also result in similar moves in other contexts, curtailing the common ground of different versions of the FIPs and making the FIPs less effective as a common set of guiding principles. Third, opting for an interpretative approach would make the already decentralized FIPs even more decentralized. An interpretative accommodation

245. OECD GUIDELINES, *supra* note 212.

246. *See supra* Subpart I.A (listing examples of individual data philanthropy).

of data philanthropy, unlike formal amendments, delegates the rulemaking power to private sector actors, who strategize their policy internally according to their business interests. Different preferences of different private actors are certain to produce significantly different interpretative exceptions and applications. However, such broad interpretation of the FIPs is unlikely to take place to begin with, because a significant departure from market standards as prescribed by the established language of the FIPs exposes businesses to heightened legal liability. Only dominant market players can afford a risk of legal liability of this magnitude, and only adoption by a sufficient number of market actors would make the interpretation-based exception an entrenched standard. Furthermore, policy strategies around data philanthropy may not be appropriately communicated at scale to guide smaller market players in their data collaborations, even if an interpretation-based exception gains sufficient traction to be considered a market standard from a legal perspective. And, lastly, integrating a constructive agreement into the FIPs is especially dangerous in the privacy context. Traditionally viewed as an individual right, privacy has commonly yielded to other societal interests such as safety and efficiency.[247] An interpretative approach that construes individual consent in the name of the greater good without guaranteeing a broad and informed perspective, as well as offering a contextual strategy for the application of the exception, ignores the social value of the privacy right.[248] The social value of data philanthropy cannot be overstated, but neither can the social value of privacy, which risks being degraded in the absence of a clear formalization of data philanthropy.

A formal data philanthropy exception might also suffer a number of flaws. First, the FIPs are commonly praised for their flexibility and global nature, but these benefits lessen the possibility of amending them. At the moment, several institutions have offered their non-statutory versions of the FIPs, including the OECD and the Canadian Standards Association outside the U.S.; and, the Federal Trade Commission, the Department of Homeland Security, and the Department of Health and Human Services in the U.S.[249] The most influential and oft-cited account is the OECD version of the FIPs, which since its introduction in 1980, has been revised once in 2013 to reflect twenty-three years of changes in international privacy activities, privacy laws, and privacy policy.[250] Although data philanthropy theoretically could be introduced into future revisions of the OECD frameworks, given that it has been only a few years since the sole revision, it is unclear when such revision might happen again. Second, assuming that the data philanthropy exception would not be added to all versions of the FIPs concurrently, the risk of shrinking the common ground of the different FIPs versions remains. Third, like the interpretative exception, the

247. REGAN, *supra* note 220, at 22–23.
248. *Id*.
249. Gellman, *supra* note 27, at 9, 10, 19, 20.
250. *Id*. at 9.

formal data philanthropy exception could end up inviting additional exceptions that may interfere with the FIPs generality and weaken their protections.

Nevertheless, a formal data philanthropy exception could successfully reconcile the purpose specification and use limitation principles within the democratic oversight of a public institution. A formal amendment would involve a broad, informed perspective, attention to diverse interests including those of underrepresented groups, and a comprehensive consideration of the societal risks and benefits. Unlike the interpretation-based exception that is conceived internally by private sector actors, the process of formally accommodating data philanthropy in the FIPs includes institutional assurances of fairness, transparency, and accountability. Governmental public institutions are in a unique position to adopt a data philanthropy exception because many of them can engage the public as well as a wide array of stakeholders in the amendment process. Specifically in the United States, a number of public institutions have already engaged in discussions about the FIPs and their application to changing technological and social realities, including the Federal Trade Commission,[251] the U.S. Department of Health and Human Services,[252] the Department of Homeland Security,[253] and the Department of Commerce.[254] Importantly, some of the U.S. versions do not impose any purpose specification and/or use limitation requirements. For these versions, the data philanthropy exception should be added, not as an exception, but as an additional clause or clarification to instruct private sector players in sharing corporate data for the social good. A federal agency or governmental department's formal acknowledgment of data philanthropy under the FIPs could stir a global and local conversation and eventually contribute to tailored regulation of data philanthropy.

C. A DATA PHILANTHROPY EXCEPTION TO THE FIPS

Scholars have both lauded and criticized the FIPs—often in the same breath.[255] This ambivalence is easily understood: the FIPs have been an indispensable part of privacy policy, but they have also exhibited many weaknesses within their scope, which, in the face of new technologies, seems too narrow. Consequently, scholars have advocated rethinking the FIPs internally and externally, recommending improvements to the language of the FIPs and suggesting complements outside of them.[256] The proposal submitted

251. Gellman, *supra* note 27, at 20, 28.

252. *Id*. at 19, 30.

253. *Id*. at 10, 21.

254. *Id*. at 25.

255. Hartzog, *supra* note 214, at 983 ("Recognizing the FIPs as a vital part, but not the whole, of privacy regimes is the only path to a sustainable future for privacy."); *see also* Andrew Proia et al., *Consumer Cloud Robotics and the Fair Information Practice Principles: Recognizing the Challenges and Opportunities Ahead*, 16 MINN. J. L. SCI. & TECH. 145, 158–63 (2015).

256. Hartzog, *supra* note 214 (exemplifying Hartzog's work on the FIPs, which represents such ambivalent approach); *see also* Woodrow Hartzog, *Social Data*, 74 OHIO ST. L.J. 995 (2013). For additional analyses of the FIPs protective and toothless power, see Bamberger & Mulligan, *supra* note 240, at 255; Justin

here is similar in the sense that it commences by advancing an internal modification to the FIPs that will mirror the FIPs' idiosyncratic qualities—a standardized, general, and flexible exception. At the same time, this Article calls for further studies of data philanthropy to reevaluate strategies internal to the FIPs and sow seeds for future legal intervention through mechanisms external to the FIPs.

Because the purpose specification and use limitation principles stand to prohibit instances of data reuse without authorization, data philanthropy should be introduced as an exception to these principles. The purpose specification and use limitation principles instruct data handlers to have a clear and articulated vision of their collection motivation and to limit the use of collected data to those stated purposes only. Any use outside the scope of the initial purposes necessitates additional or updated consent. As the above discussion has repeatedly emphasized, consent provides a legal way to engage in data philanthropy.[257] However, contractual consent should not be designated as the sole path to legitimate data reuse. This strategy further incentivizes overbroad contractual stipulations that make the purpose specification ineffective because broad terms can cover a huge universe of uses. As some socially beneficial uses cannot be anticipated, consent also increases transaction costs that can frustrate data philanthropy. Ex-ante consent can also have chilling effects and compromise the authenticity of human behavior, thus spoiling data-driven research.[258] Consent is widely criticized for being ineffective in driving true participation of the data subject in the decision-making process.[259] In other words, consent, as mandated through the purpose specification and use limitation principles, imposes barriers to data philanthropy in the name of an unachievable ideal.

An effective data philanthropy exception should promote three ends. First, it should provide meaningful guidance to industry players who have already engaged in data philanthropy initiatives, helping them differentiate between different uses, different purposes, different levels of privacy risks, and different social-good justifications for reuse. Second, it should signal to players that have yet to join the collaborative efforts that they, too, could donate their data responsibly for socially beneficial purposes. Third, a data philanthropy exception hosted in one of the formal versions of the FIPs would stir a global

---

Brookman, *Protecting Privacy in an Era of Weakening Regulation*, 9 HARV. L. & POL'Y REV. 355 (2015); Nikhil S. Palekar, *Privacy Protection: When Is "Adequate" Actually Adequate?,* 18 DUKE J. COMP. & INT'L L. 549, 567 (2008); *Privacy Self-Management*, *supra* note 109, at 1884–85.

257. *See supra* Subpart III.B.

258. Graham Crow et al., *Research Ethics and Data Quality: The Implications of Informed Consent*, 9 INT'L J. SOC. RES. METHODOLOGY, 83 (2006) (examining approaches of research governance and their effects on the quality of the data collected); Lloyd Lueptow et al., *The Impact of Informed Consent Regulations on Response Rate and Response Bias*, 6 SOC. METHODS & RES., 183 (1977); Eleanor Singer, *Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys*, 43 AM. SOC. REV. 144 (1978).

259. *See, e.g.*, James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137, 1181–82 (2009); Julie E. Cohen, *Privacy, Ideology, and Technology: A Response to Jeffrey Rosen*, 89 GEO. L.J. 2029 (2001) (showing how current consent models are largely meaningless).

discussion around data philanthropy—a conversation that would contribute to a more detailed operationalization in the future.

A possible construction of the exception could resemble the EU General Data Protection Regulation (GDPR) treatment of socially beneficial data reuse.[260] Having taken effect in May 2018, the GDPR mandates its own purpose specification and use limitation rules: Collection of personal data must be conducted for "specified, explicit, and legitimate purposes" and any further processing of collected information must be compatible with those purposes.[261] Unlike the FIPs, the GDPR offers an exception for the further processing of personal data for the performance of tasks carried out in the public interest, including archiving purposes, scientific and historical research purposes, or statistical purposes.[262] These purposes are not considered incompatible with the original processing purposes.[263] According to Article 89, further processing for one of those purposes is subject to appropriate safeguards for the rights and freedoms of the data subject to technical and organizational measures.[264] Article 89 also allows Member States to come up with appropriate statutory derogations from some rights set by the GDPR, when those rights "seriously impair the achievement of" those purposes and the derogations "are necessary" for their fulfillment.[265]

A data philanthropy exception could be similarly comprised of purposes in the public interest (for example, research, journalism, and healthcare) and outline distinct rules for each category. The GDPR categories seem reasonable at first. Even the fiercest privacy advocates would agree that in some cases, privacy must step back and be balanced against societal values like public health, law enforcement, national security, economic efficiency, and environmental protection.[266] Nevertheless, the GDPR approach fails to provide sufficient context for the balancing mission. The GDPR instructs Member States to interpret "processing for scientific research purposes" broadly to include, for example, "technological development and demonstration, fundamental research,

260. Council Regulation 2016/679, of the European Parliament and of the Council of 27 Apr. 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Council Directive 95/46/EC, 2016 O.J. (L 119) 1 [hereinafter GDPR].

261. *Id.* at art. 5(1)(b).

262. *Id*.; *see also* recital 50.

263. *Id*.

264. *Id*. at art. 89(1). This clause emphasizes data minimization as one of the end goals of the safeguards. By so doing, it directly links the original collection with the secondary possessing of the information, a logic that is criticized here. It also highlights de-identification of the data subjects as an appropriate safeguard: "Those measures may include pseudonymization provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner." By so doing, it ignores broader risks associated with generating collective knowledge, as outlined in *infra* Subpart III.C.1.

265. *See* GDPR, *supra* note 260, at art. 89(2) (explaining scientific or historical research purposes or statistical purposes); *see also id.* at art. 89(3) (explaining) archiving purposes in the public interest).

266. *See generally* Regan, *supra* note 220 (explaining that, while privacy has to step back in the face of other interests, we rarely see other interests stepping back to protect privacy).

applied research and privately funded research."[267] The definition of statistical purposes is similarly broad and covers "any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results."[268] By so doing, the GDPR ignores institutional differences and groups together uses with varying degrees of social-good justifications for the processing of all personal data, regardless of the different privacy risks they pose.

As an alternative, this Article proposes a graduated exemption model that better captures competing interests and considers elements like time and costs associated with allowing and restricting reuses of personal information. Specifically, instances of data philanthropy should be assessed through three use privileges: Exigencies, Responses, and Collective Knowledge. After fitting the requested reuse into the relevant category of use privileges, stakeholders should conduct a risk assessment and consider the potential privacy violations as well as the potential harm resulting from subjecting the reuse to privacy safeguards. The risk assessment should be contextualized through the prism of social norms and individuals' privacy expectations in the particular instance. Helen Nissenbaum's theory of contextual integrity provides a useful decision-making heuristic for informing data philanthropy risk assessment in the three categories of use privileges.[269] The data philanthropy exception concludes by mandating that no data be retained beyond the time required to complete the socially beneficial reuse within the relevant use privilege category.

1. Use Privileges Categories

The data philanthropy exception does not provide a blanket authorization to eradicate privacy rights. Like the FIPs, this exception is designed to engage in the balancing of competing interests around personal information. Thus, a preliminary condition for the exception to apply is the existence of a conflict between an individual's or a group's privacy interest and the requested reuse. To qualify for one of these categories, privacy safeguards or privacy-related costs must hinder the beneficial use or otherwise make it less beneficial. Socially beneficial uses that can be completely fulfilled alongside privacy protections or privacy-related costs are beyond the scope of the data philanthropy exception: when privacy protections or costs have minimal or no impact on the use, the exception would not apply.

Where socially beneficial uses face burdening costs or cannot fulfill their purpose when privacy safeguards are integrated, the potential reuse would enter the exception's domain. Under the exception, a justification for reuse may be classified as one of three purposes: Exigencies, Responses, or Collective Knowledge.

---

267. *See* GDPR, *supra* note 260, at recital 33, 159.
268. *Id*. at recital 162.
269. Nissenbaum, *supra* note 23.

Exigencies: The law often facilitates the balancing of competing interests. In this task, exigencies occupy a singular place. When exigency materializes, the law acknowledges the urgency and allows a certain activity, which is usually subject to legal limits or requirements, to be completed in violation of these rules. For example, when an ambulance is on its way to an emergency scene, we do not expect it to obey speed limit rules. In fact, we want the ambulance to get there as fast as it can while accepting some increase in risk to other drivers, such as, when the ambulance runs a red light. These intuitions take concrete constitutional and statutory forms. The first is the Fourth Amendment, which offers two exceptions to its warrant requirements, both of which conceptualize some form of exigency.[270] The second is the Disaster Relief and Emergency Assistance Act of 1974 (Stafford Act),[271] under which—when emergencies are declared—provides a congressional grant of power, which is customarily reserved to Congress,[272] to the president,[273] and sets a legislative mechanism by which the president could suspend or override other laws to cope with a crisis.[274] Many legal scholars have embraced emergency exceptionalism over the years.[275] While the temporary degrading of rules has its limitations,[276] it is an indispensable part of any legal regime and human intuition.

Following the same logic, it is normatively justified to remove some privacy protections that may hinder or retard the use of personal information in states of emergency. The exigency use privilege highlights a pressing need that is commonly linked to a shortage of time. Often, emergencies pose a risk that has either materialized or is about to happen and that can be prevented or mitigated, but only within a short time frame. For example, if, following a natural disaster, mobile data can help governments and emergency aids identify and quickly get to affected areas, demanding additional specific consent for the repurposing of the data could be impossible, immensely costly, or impractically time consuming. Even if anonymizing the data is practically feasible, but the personal information of mobile users could help in finding them and responding

270. For the exigent circumstances exception, see New York v. Quarles, 467 U.S. 649, 653 n.3 (1984) ("We have long recognized an exigent circumstances exception to the warrant requirement in the Fourth Amendment context."). For the emergency-aid exception, see Mincey v. Arizona, 437 U.S. 385, 392 (1978) ("Numerous state and federal cases have recognized that the Fourth Amendment does not bar police officers from making warrantless entries and searches when they reasonably believe that a person within is in need of immediate aid.").

271. 42 U.S.C. §§ 5121-5206 (2006).

272. U.S. Const. art. I, § 9, cl. 7 ("No Money shall be drawn from the Treasury, but in Consequence of Appropriations made by Law . . . .").

273. 42 U.S.C. § 5122.

274. *See id.* § 5170.

275. Sanford Levinson, *Constitutional Norms in a State of Permanent Emergency*, 40 GA. L. REV. 699, 713–15, 726–27, 747 (2006). For philosophical analysis of emergencies and the law, see Cass R. Sunstein, *National Security, Liberty, and the D.C. Circuit*, 73 GEO. WASH. L. REV. 693, 693–94 (2005); Adrian Vermeule, *Holmes on Emergencies*, 61 STAN. L. REV. 163 (2008).

276. Justice Holmes' account on emergencies acknowledges checks on governmental rights. *See* Vermeule, *supra* note 275, at 164–65.

better to their needs in emergencies, prioritizing privacy interests that could delay or prevent the use would be wrong.

The exigency use privilege of data philanthropy represents the weakest form of privacy protection—the level of immediate risk justifies a higher degree of privacy invasion. Even though the exigency use privilege occupies one extreme on the spectrum of privileged uses, it does not grant a blanket authorization for data reuse in emergencies. Like the FIPs, the exigency use privilege requires balancing competing values and interests. At times, the privacy harms caused by data reuse in emergencies could generate a subsequent set of emergencies for individuals or groups at risk, such as by identifying individuals living in domestic violence shelters or participating in witness protection programs.

Responses: The responses category of data philanthropy use privileges covers data reuse intended to generate insights in tackling a social problem or addressing a social need. In the response category, the need for the data is less urgent than it is under the exigencies category and usually includes responses to a social problem that are not limited by a critically short time frame. The risk assessment would thus require a greater social benefit to trump a privacy harm that could have been excused under the exigency use privilege. The universe of responses to social problems hosts a range of pressing issues. But the pressure to address a pressing issue does not necessarily match the critical pressure associated with exigencies. For example, using mobile data immediately after an earthquake to learn about population displacement for a more targeted humanitarian aid would qualify as an exigency, whereas using email data to identify suicidal teenagers would qualify as a response. In the former example, the use is required at a certain point of time, for a targeted short-lived effort, while in the latter, the use addresses an ongoing social problem. Responses may turn to exigencies if, for example, the analysis produces knowledge of an immediate risk. Similarly, exigencies may become responses, such as in the aftermath of a terror event when the immediate danger has subsided.

Reuses under the response category would enjoy more flexible privacy-protecting rules, without strict adherence to the purpose specification and use limitation principles. As the use moves away from the exigencies use privilege to the responses category, time constraints would be less relevant to justifying reuse without privacy safeguards. Instead, discussion would home in on other costs associated with maintaining or adding privacy protections as well as specific properties of the response that may require accessing and/or processing personally identifiable information.

Collective Knowledge: The greatest promise of data philanthropy lies with breaking silos, and the most far-reaching effects of silo breaking are generated through research. Using datasets from private sector actors, researchers work to create new knowledge. Studies have used shared data in various disciplines and interest areas. In the wake of the Cambridge-Analytica scandal, in which an academic researcher took advantage of privileged access to Facebook user data,

researchers and private sector players have attempted to come up with ethical guidelines for data sharing for research. While these efforts could end up providing useful guidelines for researchers, most of them are currently offered on a high level of generality and are still at their infancy.[277]

Unlike exigencies and responses, collective knowledge has a steadier and more continuous effect over time. It addresses a research question through a learning process that requires significant time to complete. The resulting work is usually published or otherwise shared with others in the relevant research community and turns into another building block of collective knowledge. While the production of collective knowledge does not lead to immediate real-life actions as with the other use privileges, the consequences of knowledge production are potentially more substantial and long-lasting. Under the exigencies use privilege, an action must be taken immediately, and the consequences of that choice will be analyzed ex-post when time is not of the essence. The response category allows for more deliberation prior to a responsive undertaking, dividing the learning process between ex-ante prediction and ex-post examination over time. The collective knowledge category allows for the most ex-ante deliberation prior to the implementation of any real-life pursuits. Consequently, the collective knowledge use privilege holds the greatest long-term promise, but also the greatest risk of being cemented as fact notwithstanding possible bias or resulting privacy harms. Also, while it is difficult to pinpoint the exact influence of a certain study on a certain strand of scholarship, research and human knowledge have always exhibited a cumulative nature.

Because collective knowledge is rarely generated under pressure and enjoys the longest ex-ante learning process, this use privilege requires greater justifications for relaxing privacy safeguards. Unique research needs may represent a relevant justification under the collective knowledge use privilege of data philanthropy.[278]

Data philanthropy privileged use categories are not clear-cut. To begin with, data-driven socially beneficial uses are rarely unambiguous—they can be motivated by an exigency, turned into a response, and over time, developed into collective knowledge. A reverse development may also materialize, as an existing research field has to respond to social problems that become

277. *See, e.g.*, *Guidelines for the Responsible Use of Social Media Data in Research*, LANCASTER UNIV., http://wp.lancs.ac.uk/social-media-research-ethics/guidelines-for-the-responsible-use-of-social-media-data-in-research/ (last visited July 27, 2019). Facebook's initiative for election-related research is also an example. *See, e.g.*, Gary King and Nathaniel Persily, A New Model for Industry-Academic Partnerships (Feb. 2, 2019) (unpublished manuscript), https://gking.harvard.edu/partnerships; Ian Lundberg et al., Privacy, Ethics, and Data Access: A Case Study of the Fragile Families Challenge (Sept. 1, 2018) (unpublished manuscript), https://arxiv.org/pdf/1809.00103.pdf; Molly Jackman & Lauri Kanerva, *Evolving the IRB: Building Robust Review for Industry Research*, 72 WASH. & LEE L. REV. ONLINE, 442 (2016); Elliot Schrage & David Ginsberg, *Facebook Launches New Initiative to Help Scholars Assess Social Media's Impact on Elections*, FACEBOOK NEWSROOM (Apr. 9, 2018), https://newsroom.fb.com/news/2018/04/new-elections-initiative/.

278. *See, e.g.*, de Montjoye, *supra* note 192, at 6 ("We also considered cases where specific individuals could be contacted based on criteria applied to the data.").

increasingly urgent, even up to the point of facilitating crucial immediate relief. Following the flexible, general quality of the FIPs, these use categories would sometimes overlap. The data philanthropy exception balances privacy safeguards with data philanthropy needs but leaves room for stakeholders to exercise significant discretion. Over time, as data philanthropy becomes more widespread, this discretionary power will be subject to additional evolving standards and lessons learned from previous instances of data philanthropy.

2. Risk Assessment

After matching a certain use with one of the use privileges, the data philanthropy exception moves to mandate a risk assessment. The risk assessment essentially looks at harms and benefits in the context of the specific use. The assessment does not entail quantitative comparison and requires nuanced consideration of different harms and benefits. In broad strokes, the risk assessment makes allowances for harms expected from data reuse including identification, perpetuation of bias, and the introduction of illegitimate discrimination or access barriers. The risk assessment also recognizes harms expected from barring data reuse in the specific context and the benefits expected from allowing such reuse.

The risk assessment process highlights the privacy expectations of the data subjects by applying Nissenbaum's contextual integrity decision heuristic.[279] The contextual integrity theory offers a framework for modeling intuitive judgments when information flows undergo radical changes.[280] A practice would be deemed to violate contextual integrity when it transgresses context-relative informational norms. Those norms are understood through four identifiers of the information flow: the relevant contexts;[281] the actors, including the sender and receiver of the information and the information subject;[282] the attributes, which refer to "the kind and degree of knowledge;"[283] and the transmission principles that set the conditions under which information should transfer.[284] When one of the identifiers of the information flow changes, the change is flagged as a prima facie breach of contextual integrity.[285] Next, moral and political factors implicated by the changes in flow are considered, followed by an evaluation of these factors in the specific context, and concluding with a final judgment as to the compatibility of the information practice with contextual integrity principles.[286]

Recognizing the four contextual integrity identifiers in the context of the original collection and the requested reuse would help stakeholders to better

279. NISSENBAUM, *supra* note 23, at 180–81.
280. *Id*. at 180.
281. *Id*. at 141.
282. *Id*. at 141–43.
283. *Id*. at 143–45.
284. *Id*. at 145–47.
285. NISSENBAUM, *supra* note 23, at 148–50.
286. *Id*. at 162–69.

understand the information flow and scrutinize the changes generated by data philanthropy. Because data philanthropy always involves repurposing collected data, it would always be considered a prima facie breach of contextual integrity. However, looking at broader moral principles that may underscore a need to address social issues, prevent looming harm, or manage a crisis, could end up justifying reuse from a data subject's perspective and/or from a social perspective. Considering the specific norms within the relevant context would provide a clearer picture as to the expectations of data subjects and the conditions for a socially acceptable information flow.

Admittedly, contextual integrity does not offer practical guidance on the assessment of risks in a given case. Identifying the benefits and risks of data reuse prior to the reuse is a hard task and there is no public consensus as to the relative weight of particular benefits and values.[287] Furthermore, the line between different types of benefits and risks is an elusive one. For example, a single reuse can yield public and private benefits that are not mutually exclusive, and some reuses might be highly beneficial on the individual level but only moderately helpful to the public (and vice versa).[288] It is also hard to identify all relevant stakeholders: should, for example, the interests of individuals who were not subjects of the data sets but who are nonetheless negatively affected by knowledge accrued from the use of consenting subjects' data, be taken into consideration in the risk assessment?[289]

These difficulties are yet another support to the path advocated for in this Article. Because the line drawing for data reuse for social good has not been successfully undertaken by policymakers, notwithstanding the increasing presence of data philanthropy, initial guidance must be offered to market participants. Contextual integrity is an excellent tool in this context, as it helps to think through an isolated case of data reuse from a broader policy perspective, map out the social expectations, and balance conflicting interests.

In the course of assessing the risk through the prism of contextual integrity, assessors should also consider the type of information used and its level of sensitivity: from information that, if disclosed or accessed without proper authorization is unlikely to negatively impact data subjects or other affected parties, to information that, if disclosed or accessed without proper authorization, is likely to damage data subjects or affected parties, or frustrate efforts to address the relevant exigency or social problem.[290]

When contextual integrity places a specific privacy risk within the realm of social expectations, a risk assessor is instructed to give that privacy risk less weight in the overall analysis. This is how, for example, stakeholders can

287. Fluitt, *supra* note 28, at 7–9.

288. *Id*.

289. *Id*.

290. This spectrum of data sensitivity is inspired by the United Nations Office for the Coordination of Humanitarian Affairs Data Responsibility Guidelines, which states that "a set of principles, processes and tools that support the safe, ethical and effective management of data in humanitarian response." DATA RESPONSIBILITY GUIDELINES WORKING DRAFT, *supra* note 183.

account for the difference between data subjects who suffer privacy harm but are also the target of the social benefits of the data reuse, and individuals whose privacy is sacrificed for the sake of saving or helping others. In the first instance, individuals would be far more willing and expected to waive their privacy rights for the benefits they will accrue; whereas, in the second instance individuals may be more reluctant, especially if the privacy harm is significant. By evaluating values and morals in context, contextual integrity would also account for institutional trust mechanisms, such as the mandated review of an Institutional Review Board in the academic context, and flag them as risk-lowering factors in the assessment.

3. Post-Reuse Retention

After matching the scrutinized reuse with the appropriate use privilege and conducting a risk assessment, stakeholders would move to decide whether a socially beneficial reuse outside the scope of the initial collection purpose is allowed. However, even when a reuse is validated as qualifying for the data philanthropy exception, post-reuse retention would not be allowed beyond the time required to fulfill the purpose of the reuse. The retention time allowance usually correlates with the degree of privacy protection in the use privileges—the higher the protection, the longer the retention. The data philanthropy exception would typically allow the shortest retention time for uses in the exigencies category: once the emergency subsides, the data is no longer useful in the exigencies domain and, unless the situation moved to the responses category, there is no justification for retaining the data outside the original scope of the reuse. If the use has moved to the responses category, more retention time is required until the social issue has been addressed or until attempts to address it through data philanthropy are halted. The longest retention time is commonly needed in the collective knowledge category, as knowledge production entails the longest time to complete. Additional reuses, such as those required for replication and validation studies following the first reuse, would have to be reviewed as independent reuses under the collective knowledge use privilege and would thus be granted their own retention time allowance. Even though collective knowledge enjoys the longest time retention, further attention should be given to retention and dissemination standards. In many cases, data used for research or scholarly articles are left unsecured and stored on open servers that could be accessed by anyone.[291] This data can be easily reproduced, leaked outside the academic circles and potentially sold to third parties.[292] Retention must adhere to acceptable cybersecurity and access standard within the researching institution.

291. Sheera Frenkel, *Scholars Have Data on Millions of Facebook Users. Who's Guarding It?*, N.Y. TIMES (May 6, 2018), https://www.nytimes.com/2018/05/06/technology/facebook-information-data-sets-academics.html?smprod=nytcore-ipad&smid=nytcore-ipad-share.

292. *Id*.

CONCLUSION

Socially beneficial uses of private sector data hold great promise. Since 2011, when the data philanthropy discourse was informally launched, the practice has evolved, with new examples of private sector data sharing emerging every day. This Article traces the development of data philanthropy and introduces the growing conversation around data philanthropy in other research disciplines. Following a detailed description of data philanthropy, including existing sharing models, sharing incentives, and relevant stakeholders, this Article centers on what has been acknowledged as the most pressing legal challenge for data philanthropy: privacy risks. This Article recognizes that it is not privacy compliance and reidentification concerns that drive current demands for a privacy framework in data philanthropy discussions. Instead, privacy represents a broader set of informational concerns, as well as the need for general legal acknowledgement.

As statutory legislation to regulate data philanthropy would be ineffective at this early stage, facing very low prospects of completion within a reasonable time and an inability to adapt quickly to technological and social changes, this Article proposes a data philanthropy exception to the Fair Information Practice Principles. Such exception would provide guidance to incumbent stakeholders and expose smaller players to the possibility of responsible data philanthropy. The proposed exception is structured as a graduated model of use privileges, ranging from exigencies, where privacy safeguards are the weakest, to responses to social problems, where protection increases, to generating collective knowledge, where the privileged use must comply with the highest level of privacy safeguards within the exception. The use categories reinforce a contextualized analysis of socially beneficial uses of data and are followed by a risk assessment that incorporates various interests and considers individual and social expectations. Correlating the time allowance for data retention with the time frame of the use privileges further guarantees an appropriate balance between the privacy interests of the data subjects and broader social benefits in data philanthropy.

***