DE-RISKING DATA

EQUITABLE PRACTICES IN DATA ETHICS AND ACCESS

AMY O'HARA AND STEPHANIE STRAUS

INTRODUCTION

Data on individuals are collected on almost every facet of our lives: our location, well-being, purchases, and interests. Despite this fact, many of us do not understand the ways in which our data are being used (see art installation in Designboom 2018[1]). Whether working for pay, using social media, or scrolling our phones, we often mindlessly agree to terms of service and data use, not rifling through the reams of legalese or considering the benefits of the data use at hand. Similarly, data are highly valued for many secondary uses, including research and evaluation in the government and nonprofit space. Yet these governments, nonprofits, and philanthropy organizations that enable this secondary data analysis do not always communicate with their data subjects and the greater public *why* they are using individuals' data. Individuals may not have been given an informed choice about their data being used or considered how their data will be repurposed for program evaluation, trend analysis, strategic planning, or predictive modeling. This is an issue because this lack of transparency undermines the public's trust that data will be used for greater good, which hampers future data efforts and precludes proper community engagement.

In addition to the lack of public transparency, there are no adequate, widely accepted guardrails for responsible data use in a big data world focused on evidence building. Ethical review from the biomedical world may be a poor fit for assessing responsible data use by government agencies, requiring principles more pertinent for program participants and communities represented in surveys. For example, the Menlo Report (2012) affirmed that the Belmont principles of beneficence, justice, and respect for persons, from the medical ethics world, were a sound fit for information and communication technology research, and added a fourth principle (respect for law and public interest), and encouraged development and implementation of ethical impact assessments. From these principles sprung the Institutional Review Boards (IRB) that govern all ethical data use on human subjects. However, for secondary data uses, IRBs do not always apply, and data users are left to their own devices to ensure their data subjects are being properly protected and their communities properly informed.

There are many points of discussion relating to private sector, administrative, and research uses of data. This document focuses on administrative and research uses of data not motivated by concerns about monetization. *This chapter summarizes the landscape of ethical, trusted data use as it currently exists in the research and evaluation ecosystem in the United States to discuss the current blind spots and what they mean for equitable data practices. We suggest that fair and equitable practices around data ethics and access are essential to the sustainability of administrative research uses of governmental, private and public data—and the risk of not using data for these purposes far outweighs the risk of using them. We recommend ways to improve the usage, access to, and provisioning of these datasets, highlighting real-world examples that, although promising, represent isolated instances and so must be properly scaled to produce true high-level impact.*

ROOM TO IMPROVE ON PUBLIC INTEREST, TRUST, AND TRANSPARENCY

The Administrative Data Research Network in the United Kingdom (ADRN-UK), the primary government data intermediary for their Office of National Statistics, found that the public is broadly supportive of their data being used as long as: 1) the work is in the public interest; 2) data

privacy and security needs are being met; and 3) there is trust and transparency (Waind 2020). The United States often is quite strong in data privacy and security, but has lagged in establishing what the public interest is, and is equally weak in creating trust and transparency for its data subjects and stakeholders. The National Institute of Standards and Technology (NIST) routinely establishes the baseline that federal information technology systems must meet to prevent unauthorized access. NIST continually monitors the needs of the government, updating its standards (NIST 2020a), and develops new frameworks, such as the Research Data Framework (NIST 2020b). The National Institutes of Health launched All of Us[2] in 2018, aiming to build a massive database containing the electronic health records, biomarkers, and survey responses from 1 million participants to improve precision medicine. All of Us has privacy and trust principles, as well as data security policy principles. These principles, along with their data security framework and certificates of confidentiality, aim to protect privacy for the people in this longitudinal study.

Across these initiatives, however, there has been less a focus on *why* this data must be used. This gap in explaining why using administrative data is in the public interest, and a parallel lack of transparency about current and planned uses, is significant. These initiatives are needed to create an environment of trust between data owners and data subjects, which feeds directly into the concept of social license. *Social license* exists when the public trusts that data will be used responsibly and for societal benefit. Social license pertains to the reuse of government records, as well as data held by other organizations, such as healthcare systems, post-secondary institutions, and private sector companies. It requires an understanding of what safe use would be, belief that data security terms and conditions will be met, and trust that enough value will be created through data uses (Data Futures Partnership 2017). Data users must earn *and* maintain trust. This requires continuous communication and engagement to align user intentions and data subject preferences (O'Hara 2019). However, there often are power imbalances between data subjects, controllers, and users. Additionally, there are not always opportunities for direct communication with or consent from research subjects about secondary data uses. Some key questions in building social license include:

Public Interest	Trust	Transparency
• How do researchers and evaluators make clear what they are going to *do* with the data?	• How do we shift from disempowered users (who feel their data are out of their control) to empowered data users?	• How do we increase knowledge about data use across a wide range of subjects of various ages and cultures, and who communicate in a multitude of languages?
• Can they explain their findings, why they matter, and what they plan to do next?	• How can data use be seen as contributing to bettering health, communities, and society?	
• Does the audience understand how their data were used? How are the learnings applied to real-life issues?	• How do we make sure data collection and retention are not exploitative?	

INVESTING IN SOCIAL LICENSE

To address public interest, trust, and transparency, we need a balance of norms that apply to all evidence-building data uses, with sensitivities specific to each type of data and how it is used. This could involve a combination of government regulations and standards, as well as norms about data use and public involvement. Like the American Humane organization, with its "No Animals Were Harmed"® certification in film productions that meet a rigorous standard of care for animal actors, we need standard-setting followed by compliance monitoring. The film industry knows that allegations of noncompliance will be investigated and that productions failing to meet standards will be sanctioned. Similarly, in government and philanthropy, oversight bodies can ensure regulations and standards are met, and gather input on evolving concerns of the communities contributing data and affected by the data uses. Supporting groups include the federal government, state and local governments, and philanthropy.

Federal Government

Government can aid in the development and introduction of standards and policies that boost transparency, and can help define the public interest

through deeper use of data for evidence building. Under federal regulations, data users already must abide by standards from the National Institutes of Health and inform Institutional Review Boards (IRB), which often act as the only ethical checkpoint before big data analyses. IRBs are helpful for research on human subjects in outlining proper informed consent to reflect respect for persons and in applying the ethical principles of justice and beneficence.

However, many uses of secondary data are exempt from IRB review, and there is no federal standard for assessing data ethics.[3] We can do better. Looking to our international peers, New Zealand has a digital government strategy that produces standards and guidance for online engagement, and the Information Commissioner's Office (ICO) in the UK has a data sharing code of practice that informs researchers on what they need to tell data subjects (New Zealand government 2020; Information Commissioner's Office 2019).

State and Local Governments

State and local governments can pass laws and incorporate transparency and trust building into policies. For example, in 2008, Seattle, Washington, passed an executive order on inclusive outreach and public engagement (Nickels 2008). An outreach and public engagement liaison from each city department now helps community members with the translation and interpretation of policies using data and with understanding study specifics and broader public health issues. They have tools to support public engagement, including an evaluation template to gauge the effectiveness of their engagement efforts. The city of Fort Saskatchewan developed a public engagement framework centered on the representation of diverse voices and encouraging dialogue with its citizens to develop solutions for issues affecting their lives (The Praxis Group 2012). The Actionable Intelligence for Social Policy (AISP) at the University of Pennsylvania serves networks of state and local governments using data to improve service delivery. Their learning cohorts benefit from their Toolkit for Centering Racial Equity Throughout Data Integration,[4] which highlights best practices in advancing racial equity through data sharing and integration. With activity templates, it also guides users in identifying which stakeholders they should engage from within their community.

Efforts to work across governments, with academic support, are growing, as well. For example, the Societal Experts Action Network,[5] a collaboration

between the National Science Foundation and the National Academies of Science, Engineering, and Medicine, connects expert researchers with mayors and city officials to develop evidence-based recommendations to support local, state, and national responses, particularly in the wake of the COVID-19 pandemic. However, there are over 19,000 cities in the United States, and there are 1.5 million nonprofits. What networks must form to scale such activities? A challenge is that community engagement, sharing and opening data, exploring consent issues, and—most importantly—clearly communicating how and why data are used are activities spread across teams, seldom falling to one role to do this important trust-building work. Additionally, many of these efforts are jurisdiction-specific. While these state and local data ethics/public engagement efforts are excellent examples, they represent isolated instances across the U.S. data use landscape and, coupled with the lack of standards and incentives to build trust and transparency, amount to limited progress overall.

Philanthropy

What can philanthropists do to encourage better practices? They can require attention to trust-building and transparency, just as they require accountability for expenditures, evaluation after convenings, and data archiving. Philanthropists also can encourage norms and systems that hold researchers accountable for appropriate data use and clear communications with stakeholders, learning from less successful endeavors (Carter and others 2015; Dahl and Saetnan 2009). Philanthropic organizations also can influence the public perception of data use through targeted messaging, such as the Data Saves Lives[6] campaign led by the European Patients' Forum (EPF) and the European Institute for Innovation through Health Data (i~HD). Data Saves Lives is an initiative that shares relevant information and best practices on the use of health data to help both health patients and the general public understand the importance of health data use and what safeguards the health community has in place. In addition to targeted messaging, nonprofit organizations can work in collaboration with government organizations to create more application-based recommendations to facilitate data users in changing their practices. The UK Anonymisation Network (UKAN), a nonprofit organization that works with the UK's ICO, designed an operational method of planning called the Anonymisation Decision-Making Framework, which data users can reference

to anonymize data and remain compliant to UK legislation. In the United States, the nonprofit organization Thrive! uses data to facilitate equity audits and help local governments identify programs to invest in that will reduce disparities, thus disrupting the generational cycle of poverty. Thrive! currently is launching local government pilot programs in Massachusetts, Vermont, and New York (Gardizy 2021).

Philanthropy can insist on translational work, communicating what was learned and its relevance. Grantees can be required to communicate their methods to safeguard data during and after use, and stress the benefits gained relative to the minimal risks of using the data. The community of funders, data controllers, and data users must articulate how data use is less risky than non-use; that is, that the risk of not knowing whether a treatment is effective outweighs the managed risks of using data. That message must be heard by policymakers, regulators, stakeholders, the media, and the general public. Beyond grant reports and scholarly outputs, this message can be delivered through Hill briefings, development of draft legislation, and op-eds.

EQUITABLE DATA PRACTICES

Marginalized and vulnerable communities face data equity challenges. Organizations are acknowledging and addressing current inequities, including institutional, financial, and technical barriers that prevent these communities from accessing data or conducting analyses of interest, as well as the actual and potential harms that stem from misuse of data, even in efforts for evidence-based policymaking.

Through the Urban Institute's Elevate Data for Equity project,[7] briefs and resources are available that encourage researchers and communities to manage data through its life cycle. These briefs and reports contain actionable items for researchers to incorporate, such as seeking communities' interests in research design elements, accounting for the potential social risk of research publications in reinforcing inequities, and returning research results to community members in open-access journals. As described above, more tools are available from AISP's Toolkit for Centering Racial Equity throughout Data Integration[8] projects. Pew Charitable Trusts is engaged in a Civil Legal System Modernization project,[9] focusing on open, efficient, and equitable courts—with the individuals involved in the court system at the center. Equitable courts encourage transparency and access to justice,

regardless of representation status, race, ethnicity, economic status, disability, and language spoken.

The Civil Justice Data Commons,[10] part of Pew's Civil Legal System Modernization project (see text box below), has shaped its product to protect the marginalized individuals present in civil court data. By collaborating with nonprofit and community advocacy organizations, social service providers, and courts, who lend voices to those involved in the civil legal system, we have built a technical infrastructure and data governance model with an equity lens, with elements such as a systematic research proposal approval process for desired users of the Commons, and thorough de-identification and disclosure avoidance protocols to guard against re-identification of data subjects.

Philanthropic organizations also are pursuing equitable data practices. The Robert Wood Johnson Foundation (RWJF) established the National Commission to Transform Public Health Data Systems,[11] which soon will publish recommendations to improve health equity. The commission used a framework of truth, racial healing, and transformation, analyzing how current and historic institutional racism and discrimination (for example, against people of color, of those with different abilities, or based on sexual orientation and gender identity) impact laws and policies. Another way in which philanthropy can help lead equitable data practices is through community-based participatory research, which treats community members as research partners, not just data subjects, involving them in the entire process from research question development to data analysis (Lief 2020). American Indian tribes have successfully used this participatory research, partnering with local universities and research institutions in Texas and South Dakota, for example, to take ownership of the economic development data collected on their communities, correct inaccuracies in existing federal government data, and produce actionable solutions tailored to their on-the-ground needs.

CIVIL JUSTICE DATA COMMONS

As part of the Pew Legal System Modernization initiative, we have founded a Civil Justice Data Commons[12] that applies the best practices of data governance to civil court data. We aim to create a

(continued)

secure, robust repository for civil legal data, gathered from courts, legal service providers, and other civil law institutions, that will enable stakeholders, researchers, and the public to better understand the civil legal system in the United States. By working with stakeholders in legal aid, social services, and advocacy organizations, we are building fair and equitable access to court data. We also are working with the courts to address their knowledge gaps, particularly surrounding fairness, equity, and access to justice. Our project relies on philanthropic support from the Alfred P. Sloan Foundation and Pew Charitable Trusts, as well as the National Science Foundation, to develop capacity to build the evidence—in a system where resources are lacking in individual courts or within state court systems. This project will have implications beyond courts alone, as researchers, nonprofits, and government organizations alike can apply for access to the CJDC to examine the connections of civil court involvement to economic, labor, health, and other social outcomes.

RETURNING TO "NO SUBJECTS WERE HARMED"

Clear, ubiquitous messaging is needed to explain that we can build evidence without harming people, creating social license. Evidence is a public good, and building it comes with broad societal benefit. Individuals, groups, and communities should discuss the harms, actual and perceived, that could come from data use. Discussing these harms should be a dialogue, not a conversation ender. Could a No Harm certification work with personal data? Only if the public recognizes and believes in it. We must work together to incentivize data controllers and users to adopt practices the public can recognize. We must strive for social license, producing evaluations showing that data use can be additive to our knowledge, not just extractive from the data subjects and their communities.

NOTES

1. See "Artist Visualized the Lengthy Terms of Services of Large Corporations like Facebook and Instagram," Designboom, May 7, 2018, www

.designboom.com/readers/dima-yarovinsky-visualizes-facebook-instagram
-snapchat-terms-of-service-05-07-2018/.

2. See NIH, https://allofus.nih.gov/.

3. GSA released a framework to support federal leaders and data users in 2020. See https://resources.data.gov/assets/documents/fds-data-ethics-frame work.pdf.

4. See https://aisp.upenn.edu/wp-content/uploads/2022/07/AISP-Toolkit _5.27.20.pdf.

5. See Societal Experts Action Network website, www.nationalacademies .org/our-work/societal-experts-action-network.

6. See Data Saves Lives website, https://datasaveslives.eu/.

7. See Marcus Gaddy and Kassie Scott, "Principles for Advancing Equita-ble Data Practice," Urban Institute, June 2020, www.urban.org/sites/default /files/publication/102346/principles-for-advancing-equitable-data-practice_0 .pdf.

8. See https://aisp.upenn.edu/wp-content/uploads/2022/07/AISP-Toolkit _5.27.20.pdf.

9. See Project Civil Legal System Modernization, Pew Trusts, www .pewtrusts.org/en/projects/civil-legal-system-modernization.

10. See Georgetown Law website, www.law.georgetown.edu/tech-institute/.

11. Robert Wood Johnson Foundation, Better Data for Better Health, www .rwjf.org/en/library/collections/better-data-for-better-health.html.

12. See Georgetown Law website, www.law.georgetown.edu/tech-institute/.

REFERENCES

AISP. A Toolkit for Centering Racial Equity Throughout Data Integration, https://aisp.upenn.edu/wp-content/uploads/2022/07/AISP-Toolkit_5 .27.20.pdf.

Bailey, M., Dittrich, D., Kenneally, E., and Maughan, D. "The Menlo Report." *IEEE Security & Privacy Magazine* 10, no. 2 (2012): 71–75. https:// ieeexplore.ieee.org/document/6173001.

Carter, P., Laurie, G. T., and Dixon-Woods, M. "The Social License for Re-search: Why Care Data Ran into Trouble." *Journal of Medical Ethics* 41 (2015): 404–09. https://jme.bmj.com/content/41/5/404.citation-tools.

Dahl, J. Y., and Sætnan, A. R. "'It All Happened So Slowly': On Controlling Function Creep in Forensic DNA Databases." *International Journal of Law, Crime and Justice* 37, no. 3 (2009): 83–103.

Data Futures Partnership. A Path to Social License: Guidelines for Trusted Data Use. 2017. www.aisp.upenn.edu/wp-content/uploads/2019/08 /Trusted-Data-Use_2017.pdf.

Designboom. "Artist Visualizes the Lengthy Terms of Services of Large Corporations like Facebook and Instagram." May 7, 2018. www.designboom.com/readers/dima-yarovinsky-visualizes-facebook-instagram-snapchat-terms-of-service-05-07-2018/

Gardizy, A. "Building an Escape Route from Poverty." *Boston Globe*. January 17, 2021. www.bostonglobe.com/2021/01/17/business/building-an-escape-route-poverty/.

Information Commissioner's Office. Data Sharing Code of Practice: Draft Code for Consultation. 2019. https://ico.org.uk/media/2615361/data-sharing-code-for-public-consultation.pdf.

Lief, L. How Philanthropy Can Help Lead on Data Justice. *Stanford Social Innovation Review*. February 6, 2020. https://ssir.org/articles/entry/how_philanthropy_can_help_lead_on_data_justice#.

Mackey, E., Elliot, M., and O'Hara, K. The Anonymisation Decision-Making Framework. UKAN Publications. 2016. https://fpf.org/wp-content/uploads/2016/11/Mackey-Elliot-and-OHara-Anonymisation-Decision-making-Framework-v1-Oct-2016.pdf.

National Institute of Standards and Technology. SP 800-53 Rev. 5: Security and Privacy Controls for Information Systems and Organizations. 2020a. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf.

National Institute of Standards and Technology. Research Data Framework (RDaF): Motivation, Development, and a Preliminary Framework Core. 2020b. January 1, 2020. www.nist.gov/system/files/documents/2020/12/01/Preliminary%20RaF%20final_12-01-2020.pdf.

New Zealand Government. Standards & Guidance: Engagement. October 20, 2020. www.digital.govt.nz/standards-and-guidance/engagement/.

Nickels, G. J. *Executive Order 05–08: Inclusive Outreach and Public Engagement*. Office of the Mayor, Seattle. April 4, 2008. http://clerk.seattle.gov/~CFS/CF_309282.pdf.

O'Hara, K. *Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship*. WSI White Paper #1. 2019. https://eprints.soton.ac.uk/428276/1/WSI_White_Paper_1.pdf.

The Praxis Group. Collaborative Governance Initiative: City of Fort Saskatchewan Public Engagement Framework. 2012. www.fortsask.ca/en/your-city-hall/resources/Documents/Public-Engagement-Framework.pdf.

Waind, E. Trust, Security and Public Interest: Striking the Balance. ADR-UK Strategic Hub & Economic & Social Research Council. May 2020. www.adruk.org/fileadmin/uploads/adruk/Trust_Security_and_Public_Interest-_Striking_the_Balance-_ADR_UK_2020.pdf.