Data & Policy (2021), 3: e14
doi:10.1017/dap.2021.13

COMMENTARY

How data governance technologies can democratize data sharing for community well-being

Dan Wu[1], Stefaan G. Verhulst[2], Alex Pentland[3], Thiago Avila[4], Kelsey Finch[5] and Abhishek Gupta[6]

[1]Immuta Inc., College Park, Maryland, USA
[2]The GovLab, New York University, New York, New York, USA
[3]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[4]Faculdade Estácio de Alagoas, Maceió, Brazil
[5]Future of Privacy Forum, Washington, District of Columbia, USA
[6]Montreal AI Ethics Institute, Montreal, Quebec, Canada
*Corresponding author. E-mail: wu12345@gmail.com

Received: 17 June 2020; **Revised:** 29 May 2021; **Accepted:** 02 June 2021

Keywords: Data Strategy; Data Governance Technologies

Abstract

Data sharing efforts to allow underserved groups and organizations to overcome the concentration of power in our data landscape. A few special organizations, due to their data monopolies and resources, are able to decide which problems to solve and how to solve them. But even though data sharing creates a counterbalancing democratizing force, it must nevertheless be approached cautiously. Underserved organizations and groups must navigate difficult barriers related to technological complexity and legal risk. To examine what those common barriers are, one type of data sharing effort—data trusts—are examined, specifically the reports commenting on that effort. To address these practical issues, data governance technologies have a large role to play in democratizing data trusts safely and in a trustworthy manner. Yet technology is far from a silver bullet. It is dangerous to rely upon it. But technology that is no-code, flexible, and secure can help more responsibly operate data trusts. This type of technology helps innovators put relationships at the center of their efforts.

Policy Significance Statement

Public interest data produced by private companies represent an opportunity to promote community well-being, especially by improving key economic indicators. With new data sharing arrangements, less-resourced innovators—including individual researchers, citizen developers, local communities, and small-and-medium-sized enterprises—can access sufficient data to fuel artificial intelligence and data analytics, reframe problems, and solve them in new ways. This will also have the effect of empowering people to solve problems locally, potentially creating more responsible and trustworthy solutions that better meet the needs of their communities. This piece identifies key challenges policymakers are likely to face in data sharing arrangements, providing a map of technical solutions that they must consider to create trustworthy data sharing by default.

Google's Sidewalk Toronto has propelled new data sharing efforts to the headlines of popular publications like Vice[1] and The Atlantic.[2] These efforts are a promising strategy for civically minded companies and governments. Yet experts and innovators—who are leading data sharing efforts—have raised concerns about the ability of sharing initiatives to protect[3] user data, comply with increasingly complex regulations such as California's Consumer Privacy Act (CCPA) and the European Union's General Data Protection Regulation (GDPR), and further solidify the market power of a few.

In light of this, new data governance technologies may help. They can help more, especially those with fewer resources, develop safer and more effective new data sharing arrangements. Sharing data responds to a concern levied against powerful technology companies—one that goes beyond standard privacy and security risks they pose. Critics argue that a few special organizations, due to their data monopolies and technical resources, are able to decide which problems are solved and how.[4]

With these new data sharing arrangements, less-resourced innovators—including individual researchers, citizen developers, local communities, and small-and-medium-sized enterprises (SMEs)—can access sufficient data to fuel artificial intelligence and data analytics, reframe problems, and solve them in new ways. This will also have the effect of empowering people to solve problems locally, potentially creating more responsible and trustworthy solutions that better meet the needs of their communities.

In the following, this paper identifies (a) the potential benefits of data sharing, especially for community well-being and then launches into (b) key challenges and solutions from these data sharing efforts, using data trusts as a key case study. Section 1 identifies a major benefit of these data sharing initiatives, particularly the potential to develop better metrics. Section 2 dives into the experiences of one important community-based data sharing approach—the data trust—and explores technology interventions to safeguard well-being.

1. The Promise of Data Sharing

Public interest data produced by private companies represent an opportunity to promote community well-being, especially by improving key economic indicators. In the last decade, public interest data have multiplied. Previously, government institutions positioned themselves as the largest producers of public data around the world, with an emphasis on collection for the production of census data.

Now, private companies are a major source of similar data. This is driven by the growth of the digital economy, pervading basic determinants of community well-being, such as healthcare, housing, transportation, education, and food. Furthermore, these data are being produced at a rapid rate—from a census that occurs every decade to one that is updated in real time.

Take one data sharing example that has important well-being implications: the Gross Domestic Product (GDP) used by countries worldwide. Agricultural, industrial, and service data inform it, providing a window into the health of various economies, like states and cities. Yet the GDP misses many other vital economic and social sector data,[5] including the distribution of resources, childcare work, tourism-based service work, and happiness.

Tourism-based service work, for instance, is now readily captured by numerous digital services, such as Airbnb,[6] allowing governments to monitor the use of accommodation services. Take Airbnb's

[1] Why Does Google Want to Hand Its Smart City Data to a …—Vice (2018, October 16). Available at https://www.vice.com/en_us/article/vbknkj/google-wants-a-civic-data-trust-for-toronto-smart-city-sidewalk-labs (accessed 22 January 2020).

[2] Privacy Advocates Are Criticizing Google Sidewalk Labs (2018, November 21). Available at https://www.theatlantic.com/technology/archive/2018/11/google-sidewalk-labs/575551/ (accessed 22 January 2020).

[3] Toronto, Civic Data, and Trust—Sean McDonald—Medium (2018, October 17). Available at https://medium.com/@McDapper/toronto-civic-data-and-trust-ee7ab928fb68 (accessed 22 January 2020).

[4] The Seductive Diversion of 'Solving' Bias in Artificial Intelligence (2018, December 7). Available at https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53 (accessed 22 January 2020).

[5] What's Missing From GDP? | Demos (2013, January 29). Available at https://www.demos.org/research/whats-missing-gdp (accessed 22 January 2020).

[6] Airbnb. Available at https://www.airbnb.com/ (accessed 22 January 2020).

recent decision to share data with the City of Portland. This decision not only allows the city to evaluate vital economic data missing from GDP better, but also mitigate the impact of these new innovations on key aspects of community well-being, such as housing affordability or violations of local housing laws.[7]

Finally, opportunities to make data sharing easier advance innovation for the public good. Before data exchanges, organizations like hospitals or health technology startups would face significant legal barriers to obtaining personal health information.[8] Now, these organizations can purchase these directly from patients, who may have voluntarily shared and pooled their data into an exchange, like a cooperative data trust, a concept explored further in the following section. Now that patients have more bargaining power to benefit and control the use of their data, they may be incentivized to share more data, reversing prior trends where patients would rarely benefit from transfers of their data. More data supplies may also mean more organizations, not just the most resourced, can access that data. As more organizations have better insight into critical problems, like healthcare, societies may end up with more dynamic, innovative ecosystems.

2. Problems and Solutions in Data Sharing: Case Study of Data Trusts

While there are multiple definitions about their scope and form,[9] data sharing generally has two forms. The first involves efforts to pool data into a central organization, allowing more people to access or analyze that data directly. The second type is not pooled into a central repository. Instead, limited amounts of data are shared with select third parties—such as a local government and nonprofit—from one source, like Airbnb.

To examine what those common barriers are, one type of data sharing effort—data trusts—are examined. Data governance technologies have a large role to play in democratizing safe data sharing efforts like these.

Neil Lawrence, a professor of machine learning at the University of Sheffield, argues that data trusts are a vehicle to promote "data democracies,[10]" where more can participate in a data-driven society. In a world of more data sharing, organizations can empower citizens and "bring them along for the technology ride."[11]

But with great power comes great responsibility. Managing such data sharing efforts comes with a variety of operational barriers, which span technology burdens and legal risks. Due to these issues, those with the most resources are set up to drive the design and operations of data trusts, much like they are able to in other highly regulated technical spaces.[12]

While it is dangerous to rely on technology entirely, governance technologies like these have an important role to play in helping innovators spend less time on administrative tasks and center their work on building relationships and identifying the most critical use cases.

2.1. Problem 1: The lack of "data literacy and skills"

Data trusts are technically complicated to operate. One design architecture to consider is the use of a no-code data policy platform. This problem, at its heart, is about the sophistication of data skills needed to

[7] Airbnb agrees to share data for over 17000 NYC …—Engadget (2019, May 26). Available at https://www.engadget.com/2019/05/26/airbnb-nyc-listing-data-sharing-agreement/ (accessed 22 January 2020).

[8] "What Managers Need to Know About Data Exchanges." MIT Sloan Management Review. José Parra-Moyano, Karl Schmedders, and Alex "Sandy" Pentland (2020, Summer—Forthcoming).

[9] UK: The first data trust pilot projects | DataGuidance. Available at https://platform.dataguidance.com/opinion/uk-first-data-trust-pilot-projects (accessed 22 January 2020).

[10] Data trusts could allay our privacy fears | Media Network | The …. (2016, June 3). Available at https://www.theguardian.com/media-network/2016/jun/03/data-trusts-privacy-fears-feudalism-democracy (accessed 22 January 2020).

[11] Smart Cities: Securing Privacy and Meeting Responsibilities …. (2017, April 6). Available at https://pid.samsungdisplay.com/en/learning-center/blog/smart-cities-securing-privacy-meeting-responsibilities (accessed 22 January 2020).

[12] How to save the third wave of technology from itself …. (2019, March 21). Available at https://techcrunch.com/2019/03/21/how-to-save-the-third-wave-of-technology-from-itself-regtech/ (accessed 22 January 2020).

provision access to data, create workflows for multiparty data sharing, and connecting data from multiple sources.

University of London's legal data whitepaper[13] calls data literacy "one of the most significant challenges." Open Data Institute's (ODI) report on its data trust pilot, GovLab's reflections on data collaboratives,[14] and Center for International Governance Innovation's analysis[15] of data trusts also highlight this concern.

Tools like policy engines that can control data without code can help mitigate the barrier of this skills gap. Freed of the need to code, nontechnical data trust operators can create policies and authorizations themselves. This means companies can respond to data regulations and access and generate insights from data from months to days.

Consider one data trust's data sharing agreement, which would hide sensitive data from everyone except those in the compliance role. Instead of writing custom code, a data governance official can choose drop-down options to create this policy. The policy statement can state: *mask all data tagged "sensitive data" for everyone except those in "data compliance."* Once executed, this policy—if combined with tools like data unification that connects databases—will apply to all relevant datasets.

Turnkey legal frameworks further address the literacy gap. Inspired the legal frameworks credit unions use to manage millions of their members' data, Alex Pentland's group at MIT has developed software systems for implementing them more easily[16]. Governance issues ranging from express member directives to terms of use are automatically executed to mitigate the barriers to using these frameworks. Importantly, the incentives of the cooperative entity are aligned because of its information fiduciary responsibilities to its individual members[17], facilitating trustworthy data sharing.

2.2. Problem 2: Navigating privacy laws

The problems do not end with the lack of technical skills to execute the protection of data. Experts also find it challenging[18] to design legally justified *methods* to protect data adequately, especially in relation to increasingly complex privacy laws, such as CCPA and GDPR.

One solution is to invest more in applied methods[19] to obscure, generalize, or aggregate data, especially in methods that better navigate the privacy-utility trade-off.

University of London's whitepaper identifies two[20] critical legal constraints to data sharing. The first concerns whether trusts can share and use the data at all. Privacy regulations, like the GDPR, require organizations to limit their processing based on the purpose of the data collection. In other words, using data for new purposes is prohibited unless there is a legal justification for processing, such as fresh consent or legitimate business interests.

[13] Data trusts: legal and governance considerations—Open Data …. Available at https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf (accessed 22 January 2020).

[14] How to Scale Data Collaboratives? Data Stewards Network …. (2019, March 26). Available at https://medium.com/data-stewards-network/how-to-scale-data-collaboratives-323087cf8fe1 (accessed 22 January 2020).

[15] Reclaiming Data Trusts | Centre for International Governance …. (2019, March 5). Available at https://www.cigionline.org/articles/reclaiming-data-trusts (accessed 22 January 2020).

[16] Building the New Economy. Edited by Alex Pentland, Alexander Lipton, and Thomas Hardjono (2020, Forthcoming). Available at https://wip.mitpress.mit.edu/new-economy (accessed 22 January 2020).

[17] Trusted Data: A New Framework for Identity and Data Sharing. Edited by Thomas Hardjono, David L. Shrier, Alex Pentland (2019, October 18).

[18] Building Trust Through Data Foundations (2019, December 2). Available at https://cdn.southampton.ac.uk/assets/imported/transforms/content-block/UsefulDownloads_Download/69C60B6AAC8C4404BB179EAFB71942C0/White%20Paper%202.pdf (accessed 22 January 2020).

[19] nistir 8053—NIST Page. Available at https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf (accessed 22 January 2020).

[20] Data trusts: legal and governance considerations—Open Data …. Available at https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf (accessed 22 January 2020).

Trusts help overcome these legal constraints because beneficial owners have likely consented to specific uses in advance. In BrightHive's data trust, for instance, data subjects will agree to a data trust agreement that governs the usage of that data, including, for instance, who can access the data, how much data, for what purposes, how long, how the data will be protected by default, and the procedures the trust must take to make changes to these rules.

Regardless of this consent, most data sharing approaches will take a risk-based approach, often defaulting to aggregate statistics for most use cases and allowing the use of fine-grained statistics when the benefits outweigh the risks, which are mitigated through additional controls and applied methods to reduce re-identification risks, as further discussed below.

This issue leads to the second concern: whether trusts will sufficiently protect the confidentiality of sensitive data. Companies must make sure that data are being viewed only by those who must. Furthermore, those users only see the minimum level of data they must, not more than necessary. A complementary approach is the use of use-based privacy (Bagdasaryan et al., 2019), concretely implemented in frameworks like Ancile.[21]

Aside from complying with legal requirements, confidentiality will help third parties contribute data. The ODI, for instance, identified "concerns about reputation impact" as a critical[22] barrier to providing access to data. Organizations and individuals were worried that data users would use their data for the wrong purposes or see too precise data that allow them to re-identify data subjects and paint them in a "negative light."

Applied methods, such as masking, sampling, k-anonymization, and differential privacy, help address these reputational concerns. In contrast, companies that only rely on security—as Uber likely did in 2014 —might be tempted to violate privacy by providing cleartext access to all employees who simply meet security requirements. Uber employees were found to be spying on politicians, celebrities, or even ex-girlfriends without the right privacy controls in place[23].

Instead, de-identification tools can wisely parcel access based on user attributes and purpose. Such tools enable organizations to avoid black-and-white security controls, like Uber's prior system, where all users have access to all data or none at all. Role-based access control (Ferraiolo et al., 1995), a staple technique in the cybersecurity world, allows for a graduated approach to assigning permissions and levels of access to the data that are stored in a data trust. The utilization of this methodology will improve trust of the users who provide their data to the data trust.

Consider a situation where only those in the group "Human Resources" might see related Human Resources datasets. Or only those with the purpose of "Improving Vehicle Operations" might see datasets tagged with that purpose. Or those in the group "data scientist" with the purpose "Healthcare Treatment Analytics" can only see personal health information that has been k-anonymized (Sweeney, 2002),[24] where the distinguishability of records has been minimized, making it difficult for attackers to identify individuals that are part of the dataset. By using these techniques, organizations can better meet key privacy requirements of regulations like the GDPR, while maintaining enough utility to meet their end users' needs.

Beyond de-identification tools, responsible data sharing entities will have policies that protect data by default. For instance, a data trust should strongly consider nulling or making differentially private all sensitive data by default, except for those with special attributes or purposes. If users do not meet these conditions, their view of the data are limited to anonymized data or aggregate statistics. As a result, even if a successful hacker

[21] Ancile—Use-Based Privacy for applications, Github. Available at https://github.com/ancile-project/ancile (accessed 10 June 2020).

[22] Data trusts: lessons from three pilots (report)—The ODI (2019, April 15). Available at https://theodi.org/article/odi-data-trusts-report/ (accessed 22 January 2020).

[23] Uber settles federal probe over 'God View' spy software (2017, August 15). Available at https://nypost.com/2017/08/15/uber-settles-federal-probe-over-god-view-spy-software/ (accessed 22 January 2020).

[24] "A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release."

breaches the account of a typical user or discovers the encryption key, hackers will only see a hash or blank. With less valuable data to expose, de-identification mitigates the harm of data incidents, like a data breach. In other words, the "surface area" of attack has decreased. The additional benefit of using a data trust is that it safeguards data disclosure even in the face of new complementary datasets that become available for a hacker to mount an attack based on the mosaic effect and other associated data correlation techniques (Gupta, 2018). This also helps mitigate the problem of storing data in a centralized repository which becomes a more attractive target for hackers, constituting a single point of failure (Simmonds et al., 2004).

Applied methods are far from perfect. Yet these tools help data trusts align with the expectations of a data sharing entity's members, ultimately producing more trustworthy data systems.

2.3. Problem 3: Auditing data access

Even if data trusts have implemented the best processes, failure is always a possibility and, thus, auditing is a necessity. To address this issue, data sharing arrangements can use secure sandboxes, where data usage is continuously monitored and auditable, to access and manipulate data. As identified in the risk management model[25] "three lines of defense," auditing processes are the third line to ensure harm to the trust and its data users are limited. Similarly, the principle of accountability—of which auditing is a critical application—is a critical feature of GDPR.

This focus on auditing is echoed in various reflections about existing data sharing projects. GovLab, for instance, identified[26] how many felt that "regulatory oversight and audit processes should be established… [otherwise] there is potential for irresponsible data processes to perpetuate and create harms." Similarly, Sean McDonald[27] critiqued Google's data trust for lacking few "resources, investigatory powers, and punitive authority" for auditing its activity. For leaders in the data trust space, auditing is clearly top-of-mind.

Yet auditing is operationally difficult. Databases likely have different policies about who can access these databases and why. As a result, users and third parties may be violating the purpose limitations of highly sensitive data. Manual processes, furthermore, make it difficult to consistently document and audit data user behavior. With the costs and delays of auditing, companies may be tempted to shortchange this behavior, increasing the risk of violating the trust of data sharers.

In order to provide consumers with this information, organizations can consider secure sandboxes to make auditing easier to manage. A sandbox refers to a security mechanism to execute an untested process, such as a data sharing arrangement. Instead of releasing sensitive data to third parties without further involvement in is use (i.e., what some call the "release-and-forget[28]" model), a secure sandbox maintains adherence to policies and regulations by design.

A critical component of a sandbox is an identity protocol. Such a protocol attempts to solve[29] the lack of a unified system to authenticate personal or digital identities online. In the physical world, tools such as driver's licenses and passports, help solve this problem, albeit imperfectly, once one considers the rise of identity theft.

While there is no perfect or universally-accepted digital solution now, digital data sharing systems can use "identity and access management" tools to define, manage, and authenticate the roles and access

[25] GRC technology: Enabling the three lines of defense: PwC. Available at https://www.pwc.com/us/en/services/risk-assurance/library/grc-technology-three-lines-defense.html (accessed 22 January 2020).

[26] How to Scale Data Collaboratives? Data Stewards Network …. (2019, March 26). Available at https://medium.com/data-stewards-network/how-to-scale-data-collaboratives-323087cf8fe1 (accessed 22 January 2020).

[27] Toronto, Civic Data, and Trust—Sean McDonald—Medium (2018, October 17). Available at https://medium.com/@McDapper/toronto-civic-data-and-trust-ee7ab928fb68 (accessed 22 January 2020).

[28] Building Trust Through Data Foundations A Call for a Data …. (2019, December 2). Available at https://cdn.southampton.ac.uk/assets/imported/transforms/content-block/UsefulDownloads_Download/69C60B6AAC8C4404BB179EAFB71942C0/White%20Paper%202.pdf (accessed 22 January 2020).

[29] Blockchain for Identity Management—The Department of …. (2016, December 11). Available at https://www.cs.bgu.ac.il/~frankel/TechnicalReports/2016/16-02.pdf (accessed 22 January 2020).

privileges of online identities accessing secure sandboxes. While these tools are not global or unified, they nonetheless provide a reliable method of authenticating identities to highly sensitive data.

Building on a meaningful identity protocol, data sharing initiatives using secure sandboxes can also use a single access and control layer for all of its data. Because all data passes through a unified data layer —often assisted by data virtualization technologies[30]—all data activities can be monitored and reported in a more efficient manner, without having to move data manually or make data copies. An example of access and control is the nulling of data columns tagged "personal data," except for groups with special permissions or purposes.

To track and report on the usage of these data, immutable audit logs are a critical tool. These logs can be reported to key stakeholders, who can verify compliance with internal policies and regulations. Immutability, furthermore, "enhances confidence that the audit trail has not been tampered with, even by the database administrator."[31]

Again, technology cannot resolve what some call the "ugly"[32] questions with regard to auditing data trusts. For one, to require audits of such magnitude, data trusts would have to garner "a relatively rare amount of enforcement power for someone other than a government regulator."[33] While technologies like the above do not guarantee compliance, by at least reducing the costs of auditing, enforcement may be more likely.

2.4. Problem 4: Developing evidence of need

While data governance technologies cannot directly address this barrier, it can reduce administrative burdens so that innovators can focus on building relationships and identifying evidence of need for data sharing. The ODI identifies that the first barrier to a trust is a "lack of evidence" about the utility of sharing data. It found that the barriers to sharing data often outweighed their perceived benefits. The technology stacks above may help reduce the perceived costs of sharing data, such as confidentiality concerns and the lack of data skills, bolstering more data sharing projects.

Yet, those who start data trusts require deeply human techniques—not technology—to develop evidence that they are solving a real, painful problem. Despite its near-cliche status, framework[34] like "minimum viable product" may help guide such efforts to develop that evidence cheaply and quickly.

Active listening is required to identify pain points of stakeholders. Creativity is required to design "minimally viable" prototypes to test whether the data trust will solve identified problems. Perseverance is required to iterate on deficiencies. And judgment—aided by frameworks like the ODI's data ethics canvas[35] and Future of Privacy Forum's open data risk assessment[36]—is required to mitigate problems before they arise. Again, data governance technologies enable such prototypes. They reduce the costs of otherwise prohibitive investments into information technology, legal, and data officers.

But relationships must be at the center of any data sharing arrangement. Building trust via transparency in these relationships between the data fiduciaries and the data subjects will be critical for the success of a data trust.

[30] Data Virtualization Defined: How it Helps Organizations …. (2018, November 15). Available at https://www.dataversity.net/data-virtualization-defined-organizations-succeed/ (accessed 22 January 2020).

[31] Nation At Risk—Markle Foundation (2009, March 4). Available at https://www.markle.org/sites/default/files/20090304_mtf_report.pdf (accessed 22 January 2020).

[32] Toronto, Civic Data, and Trust—Sean McDonald—Medium (2018, October 17). Available at https://medium.com/@McDapper/toronto-civic-data-and-trust-ee7ab928fb68 (accessed 22 January 2020).

[33] See note 32.

[34] Methodology—The Lean Startup. Available at http://theleanstartup.com/principles (accessed 22 January 2020).

[35] SHARED ODI Data Ethics Canvas User … Google Docs. Available at https://docs.google.com/document/d/1MkvoAP86CwimbBD0dxySVCO0zeVOput_bu1A6kHV73M/edit (accessed 22 January 2020).

[36] FPF Publishes Model Open Data Benefit–Risk Analysis (2018, January 30). Available at https://fpf.org/2018/01/30/fpf-publishes-model-open-data-benefit-risk-analysis/ (accessed 22 January 2020).

2.5. Recommendations

In sum, four data governance technologies help address key challenges in sharing data. From our survey of challenges and solutions, future data sharing initiatives should consider these problems and solutions:

1. Problem 1: Significant data engineering skills to control data. *Solution*: A no-code data policy platform to control data can reduce the engineering burden to manage and control data.
2. Problem 2: Complex data protection laws. *Solution*: Automated usage of applied methods to obscure, generalize, and/or aggregate data.
3. Problem 3: Difficulty of auditing data usage. *Solution*: Secure sandboxes to share and manipulate data, including systems to authenticate identities and provision access and control.
4. Problem 4: Identifying evidence of need for data sharing. *Solution*: Data governance technologies reduce administrative costs of sharing data safely, so that innovators can build meaningful relationships and identify use cases.

To be sure, technology is not a silver bullet. As Problem 4 also identifies, data sharing at its heart is about building new people-centered relationships.

In other words, data governance technologies cannot automate away key data governance questions. These include (a) *why* organizations and communities want to share and analyze data (e.g., the use cases for community well-being), (b) *how* to share data (e.g., policies and risk-based assessments), and (c) *how* decisions will be made (e.g., the creation of a charter, the designation of a data protection officer, and inclusive decision-making bodies). These other aspects of data governance are discussed at length in other work.[37] Nonetheless, data governance tools can streamline execution—actually controlling data and mitigating risk based on these rules.

Ultimately, data governance technologies have an important opportunity to improve community well-being. Social problems such as housing or economic vitality require cross-sectoral solutions. Data sharing across sectors, especially the private and public, can inform better policies and uplift people safely.

Competing Interests. The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or nonfinancial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript.

Dan Wu works at a data governance technology company. While he refers to *categories* of designs and architectures related to his work at Immuta, specific technologies and companies are never mentioned in this piece.

Author Contributions. Conceptualization, D.W., S.V., A.S.P., and K.F.; Writing-review & editing, D.W., S.V., A.S.P., T.A., K.F., and A.G.

Data Availability Statement. Data availability is not applicable to this article as no new data were created or analyzed in this study.

References

Bagdasaryan E, **Berlstein G**, **Waterman J**, **Birrell E**, **Foster N**, **Schneider FB and Estrin D** (2019) Ancile: Enhancing Privacy for Ubiquitous Computing with Use-Based Privacy. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society*, pp. 111–124.

Ferraiolo D, **Cugini J and Kuhn DR** (1995) Role-based access control (RBAC): Features and motivations. In *Proceedings of 11th Annual Computer Security Application Conference*, pp. 241–248.

Gupta A (2018) The evolution of fraud: Ethical implications in the age of large-scale data breaches and widespread artificial intelligence solutions deployment. *International Telecommunication Union Journal*, *1*, 0–7.

[37] Building Trust Through Data Foundations A Call for a Data …. (2019, December 2). Available at https://cdn.southampton.ac.uk/assets/imported/transforms/content-block/UsefulDownloads_Download/69C60B6AAC8C4404BB179EAFB71942C0/White%20Paper%202.pdf (accessed 22 January 2020).

Simmonds A, **Sandilands P and Van Ekert L** (2004). An ontology for network security attacks. In *Asian Applied Computing Conference.* Berlin, Heidelberg: Springer, pp. 317–323.

Sweeney L (2002). k-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10*(5), 557–570.

Cite this article: Wu D, Verhulst S. G, Pentland A, Avila T, Finch K and Gupta A (2021). How data governance technologies can democratize data sharing for community well-being. *Data & Policy*, 3: e14. doi:10.1017/dap.2021.13