

Data Trusts

A new tool for data governance



Acknowledgments

We would like to express our appreciation to the following list of experts for their active participation in the International Workshop on Data Trusts. Many also provided detailed comments on versions of this white paper, for which we are grateful.

Malik Ghallab

Research Director
CNRS

Rob van Kranenburg

Founder
Resonance Design & IoT
Council

Sylvie Delacroix

Professor
Uni. of Birmingham Law
School
Fellow
Alan Turing Institute

Jonnie Penn

PhD Student,
Cambridge, Affiliate,
Berkman Klein
Cambridge University
CFI, Berkman Klein

Christina Colclough

Director of Platform
and Agency Workers,
Digitalisation and Trade
UNI Global Union

Edward Santow

Human Rights
Commissioner
Australian Human
Rights Commission

John Howell

Director, Human Rights
Scrutiny
Australian Human
Rights Commission

Lillian Edwards

Incoming Chair of Law,
Innovation and Society
Uni. of Newcastle Law
School

Tabitha Goldstaub

Co-Chair (Council), Co-
founder, CogX
UK AI Council, Cog X

Sean McDonald

CEO
Frontline SMS

Amanda Clarke

Assistant Professor &
Public Affairs Research
Excellence Chair
Carleton University

Keith Porcaro

Fellow
Harvard Berkman Klein

Sean Mullin

CEO
Brookfield Institute

Salomé Viljoen

Fellow
Harvard Berkman Klein

Neil Lawrence

Professor
University of Sheffield

Amit Golder

*Assistant Vice
President, Strategy
New York City
Economic Development
Corporation*

Bertrand Paihlès

*National coordinator,
AI strategy and
deployment
Government of France*

Anne Carblanc

*Head of the Digital
Economy Policy Division
OECD*

Jack Hardinges

*Policy Advisor
Open Data Institute*

Bianca Wylie

*Founder
Open Data Institute,
Toronto*

Mylène Deschênes

*Ethics & Legal Advisor
Office of the Chief
Scientist of Quebec*

Harry Armstrong

*Head of Technology
Futures
Nesta*

The workshop benefited from a presentation on the Sidewalk Toronto project by Alyssa Harvey Dawson, General Counsel to Sidewalk Labs.

Special thanks to Dan Munro, Visiting Scholar and Director of Policy Projects at the Innovation Policy Lab in the Munk School of Global Affairs and Public Policy at the University of Toronto, who provided thoughtful feedback on the final draft.

Sponsors

Element AI and Nesta would like to express their gratitude to les Fonds de recherche du Québec, IVADO and the Ministry of Innovation, Science and Economic Development for their generous support and sponsorship.

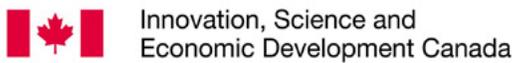


Table of contents

Introduction	6
Outline	8
Definitions	9
Part 1. Adapting to the pace of change	10
The traditional approach to regulation	10
Anticipatory regulation	12
Part 2. Theorizing data trusts	13
Disturbing the feudal approach to data governance	13
Legal trusts 101	14
The bottom-up approach	15
Part 3. Moving from theory to practice	17
Examples of data trust definitions – current case studies	17
Sidewalk Labs’ experimentation with data trusts	20
The need for data trust pilots	21
Part 4. Designing data trusts	22
Three use cases	22
Urban data	22
Health data	24
Data from online platforms	26
Part 5. Conclusions and recommendations	27

Introduction

Individuals have little control over their data—how it is collected, who collects it, and for what it is used. For many, the common experience with online platforms, mobile apps and other digital services is blindly accepting whatever demands they make of our data, which are often a necessary condition of use. Yet a new public awareness has grown amid news of scandals around the misuse of data and major data breaches, and it is clear that the private sector has failed to protect individual privacy rights through self-regulation.

The rise of these digital services, and the handful of companies that provide them, means that data is also becoming increasingly concentrated in the hands of a few private corporations. The impact of artificial intelligence threatens to exacerbate existing power imbalances between the tech giants, individuals and government regulators, as these companies not only have extreme wealth but now the power provided by AI insights from oceans of data.

The erosion of public trust and confidence in data-collecting organizations and in the technologies that rely upon this data (including AI) has provoked a backlash that threatens society's ability to access and use trusted data for the public good. There is a pressing need to explore new data governance models that give individuals a measure of control over their personal data, while industry and governments work to define, protect, and evolve concepts of digital rights.

In this context, on December 4 and 5, 2018, artificial intelligence products provider Element AI and U.K.-based global innovation foundation Nesta hosted an International Workshop on Data Trusts. While other possible solutions exist—such as revising consumer protection legislation to make it illegal for the public to consent to the collection and use of data in certain cases that would be contrary to public policy—the workshop was organized to better understand whether data trusts could be a way to enhance protection for individual privacy and autonomy, address

existing power asymmetries between technology companies, government and the public, and empower the latter to share in the value that data and artificial intelligence promise.

One of the goals of the workshop was to leverage the wide-ranging expertise from invited speakers and contributors to co-develop appropriate and practical data governance solutions that are fit for the future, have global relevance and can be tested today. In particular, participants were urged to consider whether data trusts can help move beyond mere compliance with existing privacy rules and promote public trust. To these ends, the workshop explored a number of questions, including: can data trusts be designed to empower the public? What entities should be

responsible for the oversight of such trusts? What are some key international best practices and experiences?

The purpose of this white paper is to advance the public policy conversation on data trusts by capturing the discussions of international experts and summarizing findings in the form of conclusions and recommendations. While the primary target audience for this piece is therefore policymakers, another purpose of this white paper is to sensitize the public and civil society organizations to current risks to privacy, data collection and use, and the need for improved data governance.

Outline

Part one of this paper outlines how the reactive approach to policy-making is inadequate to the challenge of regulating new technologies and their impacts on society. New anticipatory models of data governance are needed to co-evolve regulation and technology while protecting and advancing the public interest.

Part two examines theories of data trusts, the relevance of the data ownership model, as well as the bottom-up approach proposed by Sylvie Delacroix and Neil Lawrence.

Part three attempts to move from theory to practice, exploring implementable definitions of data trusts as well as the importance of testing models of data governance through pilot projects, including concrete examples presented by the Open Data Institute and Sidewalk Labs.

Part four explores the application of data trust models to three use cases—urban data, healthcare data, and data from online platforms—in light of the foregoing, while highlighting outstanding questions or issues for further reflection.

Part five summarizes the workshop's conclusions and includes recommendations for policymakers, industry and civil society regarding the relevance of data trusts to the elaboration of national strategies for data governance.

Definitions

As data trusts, data commons and other forms of data sharing models are not well-defined and often conflated, the workshop used the Open Data Institute's definition of data trusts as a starting point, namely, that a data trust must have a clear purpose; a legal structure, constitution and trustees; (some) rights and duties over stewarded data; defined decision-making processes; a description of how benefits are shared; and sustainable funding.

Consent-based models of data governance place the onus on the consumer to determine whether a company's extensive policies regarding the collection and use of their personal information are fair and reasonably respect their preferences and comfort with respect to particular uses. While consent-based models aim to address legitimate concerns about liberty, privacy and fair arrangements, there are several flaws with these models.

First, companies are often unaware of or do not communicate all the potential uses of the data that they collect, rendering consent either meaningless or unreliable in many cases. Second, consumers have no ability to negotiate the terms

and conditions of a company's collection and use of their personal information, which take the form of a contract of adhesion: if a consumer disagrees with the terms, he or she has no recourse other than to decline the service and select another provider. Often, there is simply no alternative.

In the world of online platforms, an absence of competition has discouraged innovation in data governance beyond the consent-based model of data privacy, which remains consumers' only choice.

Furthermore, studies have estimated that it would take the average person 244 hours to read all the privacy policies of websites he or she visits in a year.¹ Finally, privacy policies are notoriously complex: approximately 30 per cent of Fortune 500 companies' privacy policies require a postgraduate degree to understand, and only one per cent of them are understandable to audiences with a high school education or less.²

¹ McDonald, A. M. & Lorrie, F. C. (2008). The Cost of Reading Privacy Policies. *Journal of Law and Policy for the Information Society*, volume 4,3.

² Ibid.



Part 1. Adapting to the pace of change

The traditional approach to regulation

Governments using traditional methods of developing regulation are struggling to cope with the increasing pace of technological change: by the time new policies and methods of oversight are implemented, advances in technology, shifts in markets or socioeconomic change will have introduced new variables. This regulatory challenge is particularly acute in the realm of data governance. Over the last two decades the rise of online platforms and the Internet of Things have given technology companies unprecedented power over data and individual privacy.

Governments have often taken a reactive approach to regulation and governance; first letting the market develop freely and then instituting incremental change as threats and problems emerge. While this may work in slower-moving fields, such approaches are insufficient to deal with the challenge of data governance and AI, where new business models can scale very rapidly, reaching millions of consumers worldwide in only a short time. This means problems can manifest and scale just as quickly. Considering that these models affect some of our most basic rights, including the right to privacy but also our freedom of expression and others (think of the

“searchability” of an online post, its ranking in search engines, etc.), such reactive regulatory approaches seem unfit for purpose.

Daniel Munro, a visiting scholar in the Munk School of Global Affairs and Public Policy at the University of Toronto, observes that we are dealing with a particular case of the Collingridge dilemma: “while it is easier to regulate technologies when they are new—because they are not yet reflective of sunk costs and vested interests—uncertainty about their effects makes it hard to know exactly what to do. By contrast, when technologies are more developed and diffused throughout society, their consequences may be clearer but efforts to regulate them will be harder. In the early days of emerging technologies, we have power but insufficient clarity to act. In later days, we have more clarity, but declining power.”^{3,4}

³ Munro, D. Risk, Uncertainty and the Governance Dilemma for Artificial Intelligence. *Dan Muro*. Retrieved February 23, 2019 from <https://www.danmunro.ca/blog/2019/1/16/risk-uncertainty-and-the-governance-dilemma-for-artificial-intelligence>

⁴ Munro, D. Governing AI: Navigating Risks, Rewards and Uncertainty. *Public Policy Forum*. Retrieved February 23, 2019 from <https://ppforum.ca/wp-content/uploads/2019/01/Governing-AI-PPF-Jan2019-EN.pdf>

Anticipatory regulation

To confront the social changes inherent in digital transformation, regulators and government agencies must adopt governance frameworks that help them look forward and manage risks. Nesta has shown leadership in describing this type of proactive regulatory approach, known as “anticipatory regulation”.⁵ Anticipatory regulation includes at least three features that are important to keep in mind when considering how new governance models such as data trusts might be useful as practical regulatory tools.

First, anticipatory regulation is inclusive and collaborative; it is designed to engage the greatest number of stakeholders in the conversation on governance. In this way, anticipatory regulation reflects the principle that those affected by the rules should be able to participate in modifying them. By including more and different people, gains in quantity and quality of information may be made—a consideration that is especially significant for the collection and use of data in the context of artificial intelligence. Anticipatory regulation requires policymakers to engage with the people whose interests are affected by questions of governance. In the context of data trusts, then, we must ask: who are these people? How should they be engaged? Are data trusts the correct way to address their needs?

Second, anticipatory regulation must be future-facing. In contemplating how data trusts might be implemented, anticipatory regulation requires us to focus not only on situations as they currently stand, but how they might evolve in the future. This feature of anticipatory regulation recognizes that certain factors, including data usage, the relationships between actors in the system, or even legislation, may change over time, introducing new dynamics into the governance model being tested.

The nature of a data trustee’s fiduciary duty ensures a forward-looking commitment to managing risks associated with data and the assets they are used to create.

Third, anticipatory regulation involves a proactive, experimental approach that can foster innovation. In what ways can data trusts be used a regulatory tool that enables innovation? How easily can data trusts adapt to accommodate future change?

⁵Nesta. *Anticipatory regulation*. Retrieved January 27, 2019 from <https://www.nesta.org.uk/feature/innovation-methods/anticipatory-regulation/>



Part 2. Theorizing data trusts

Disturbing the feudal approach to data governance

Many modern businesses are driven by the accumulation of data and the power it generates. The emerging dominance of a handful of technology companies has led to the monopolization of data in certain domains, especially with regards to personal information. Some fear that the dominant technology companies have come to represent a new form of feudal oligarchy—and that it is time for an online Magna Carta.

The current power imbalance in data mirrors medieval feudalism: either data is openly available, such as in the information commons, or it is managed on our behalf by a digital oligarchy in which data subjects play the role of vassals.⁶ But while the concept of ownership applies easily to land, it does not to data—in respect of which we can neither have full knowledge regarding the nature of its

scope and use, nor readily assert control. Delacroix and Lawrence emphasize the limits inherent in an ownership approach to data: at most, data ownership confers the kind of access rights that are similar to water rights.

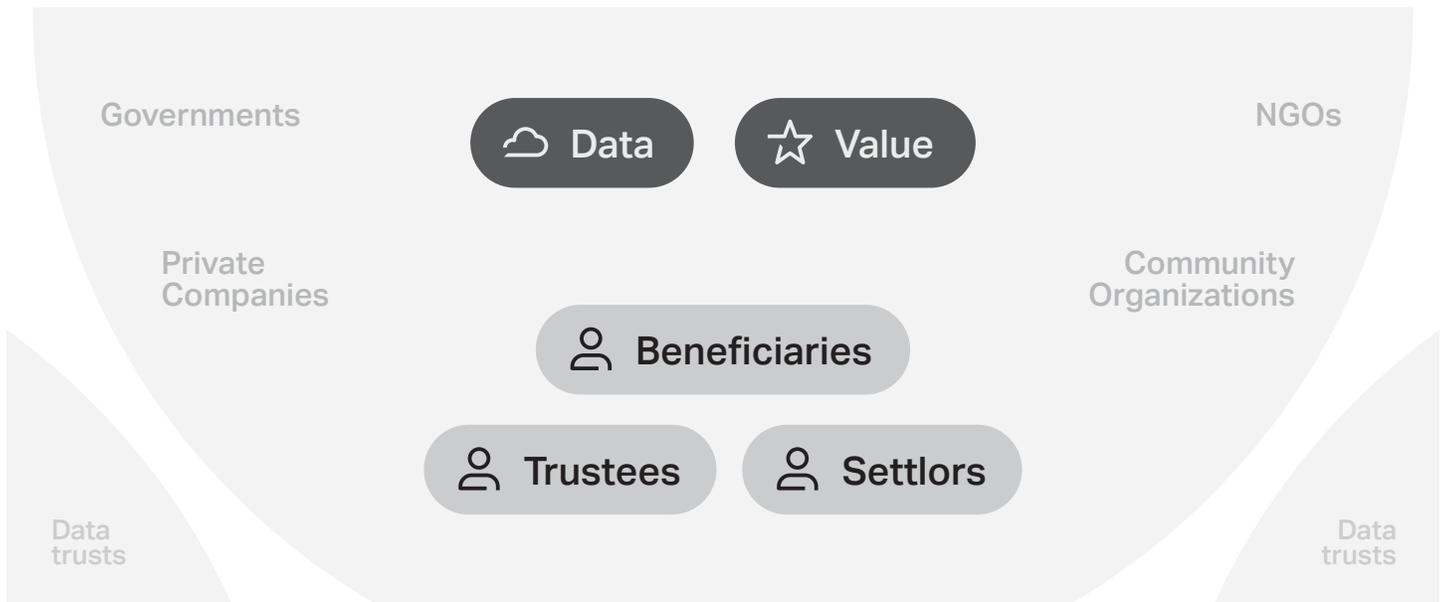
The goal of re-claiming data ownership, for instance, from online platforms, fails to address the more important questions of the management of data access and use.

Specifically, the ownership model of data can impede our ability to develop policies that ensure data is accessed and used in a manner that respects human rights.⁷ Numerous exchanges during the workshop supported this idea. Accordingly, just as feudalism gave way to representative democracy, a participatory model of data stewardship is now needed.

⁶ Delacroix, S. & Lawrence, N. (2018). Disturbing the 'One Size Fits All' Approach to Data Governance: Bottom-Up Data Trusts. *SSRN*. Retrieved from <https://ssrn.com/abstract=3265315>

For a comprehensive discussion of the impact of technology on human rights, please see the following:

⁷ Australian Human Rights Commission. (2008). *Human Rights and Technology Issues Paper*. Retrieved February 20, 2019 from <https://tech.humanrights.gov.au/sites/default/files/2018-07/Human%20Rights%20and%20Technology%20Issues%20Paper%20FINAL.pdf>



Legal trusts 101

In basic terms, a trust creates a legal way to manage rights in an object for the benefit of another person. The trust begins with the object—which in our context could be data, code, or technology, etc.—that a “settlor” places into a trust. A trust document or charter stipulates the purpose and terms of the trust, which exists to benefit a group of people, known as the “beneficiary”. The beneficiary may be a person, a group of people, or a person who may be identified in the future—but it must be identifiable. The settlor appoints a trustee organization to manage the object in the beneficiaries’ best interests, according to the terms of the trust.

In this regard, the trustee bears strict legal duties—a fiduciary duty, which includes duties of loyalty,

prudence and diligence—as well as the obligation to take legal action on the beneficiaries’ behalf. In the event of a claim for some alleged harm, moreover, the trustee bears the burden of demonstrating that he or she has acted in the best interests of the data-subjects, as outlined in the trust’s charter. It was noted that the duty of loyalty could remove the need to create economic incentives for trustees, which would risk introducing a conflict of interest with the latter’s fiduciary duty.

The trust model includes a great measure of flexibility, as the terms of the trust may be tailored to a particular data set, problem or purpose that the trust aims to address. During the workshop, participants discussed how different models of data trusts could be designed to address three specific contexts: the collection and processing of urban data, health data and data collected from online platforms.

The bottom-up approach

Data trusts can be structured in a number of different ways depending on their specific purposes, the terms established by the trustors, or what is required to protect the best interests of the beneficiaries in question.

University of Birmingham professor Sylvie Delacroix and University of Sheffield professor Neil Lawrence propose a “bottom-up” approach to data trusts, in which data-subjects pool their own data into a legal structure (the “trust”) for a social or economic benefit of their choosing. In this approach, data-subjects tend to be both the settlors as well as the beneficiaries for whom the trust and its terms have been established.

Delacroix and Lawrence argue that a bottom-up approach to data trusts could “reverse the power imbalance that currently exists between individuals and the corporations that use their data for their benefit.”⁸

An ecosystem of data trusts, each with different constitutional terms, could enable data-subjects to select an approach to data governance that mirrors their privacy preferences and values.

This became a point of consensus during the workshop, as participants explored how models of data trusts could be designed to provide the public with greater access to their personal data as well as control over the purposes for which it is being used.

Delacroix and Lawrence list four essential conditions to ensure the viability of their proposal. First, the creation of a data trust must be straightforward. Second, the data trust must provide for the safety and security of the data. Third, trustees must be able to compel the erasure of data pertaining to their beneficiaries, notably, in situations where the data is being stored or used contrary to the trust’s terms.⁹ Fourth, the trustees must be empowered to exercise beneficiaries’ portability rights on their behalf, which would include the ability to share the trust’s data with other data trusts, as well as other public and private sector entities that conform with the policies of the trust.

Discussions during the workshop explored challenges identified by Delacroix and Lawrence with respect to the implementation of data trusts. The first relates to the need to raise the public’s awareness of the risks inherent in current data sharing arrangements, and the potential benefits of alternative models. Improving data literacy

⁸ Ibid.

⁹ At present, the right to erasure (or “to be forgotten”) appears to be only partially backed by Article 17 of the European Union General Data Protection Regulation (GDPR).

will be critical to encouraging demand for an ecosystem of data trusts that adequately reflects consumer privacy preferences.

The second challenge relates to the trustees' ability to cover costs related to damages. One possibility would be to require trustees to hold liability insurance, financed from data licensing fees or covered by the state. Questions related to the identity of the trustees and how they should be selected were addressed by participants during the discussions on the three use cases, described below.

A third challenge was identified regarding data whose provenance is shared, and the difficulty of distributing such data when only one of the settlors exits the trust. Portability rights will also need to be clearly defined to enable settlors to move from one trust to another. While the workshop focused on the collection and use of personal data, participants observed that portability rights would be affected by legal requirements for the free movement of personal data in the General Data Protection Regulation and non-personal data in free trade agreements, such as in the Comprehensive Economic and Trade Agreement between Canada, the European Union and its member states, and in the new European Regulation on a framework for the free flow of non-personal data in the European Union.

A fourth challenge related to ensuring a viable market for the trusts, since the proposed model essentially relies upon the negotiation power of the aggregated data of large numbers of individuals to improve the terms and conditions under which the settlor-beneficiary can access certain services that rely upon their personal data. Without a sufficient number of participants, the data trust model may not take off. It was noted that governments could help spur the market by requiring that personal data only be accessed through data trusts, in particular in situations where the state already has a licensing process in place (ride-sharing/taxi licences, for instance, as per below).

A fifth challenge related to the level of data management required of individuals, as they assess their participation in different trusts suited to different purposes. Retirement savings plans provided a useful analogy here: certain individuals may wish to manage each trust to which they adhere (similar to those who manage their retirement through day trading), while others may leave it to "trusts of trusts" to manage their data under meta-terms, such as "maximizing my privacy" and "ensuring positive social outcome," somewhat like mutual funds.

Part 3. Moving from theory to practice

The workshop benefited from the experience of the Open Data Institute, which has worked to synthesize diverse conceptions of data trusts into a common approach that may be tested in practice. Sidewalk Labs' presentation of its own experience developing a proposal for data trusts in Toronto led to important discussions regarding different approaches to structuring data trusts, particularly in the context of public sector projects. Participants discussed the need for clarity around definitions and the importance of pilot projects in evaluating assumptions about the effects of certain data trust models.

Examples of data trust definitions – current case studies

In 2017, a U.K. Government-commissioned independent review into artificial intelligence fueled a great deal of interest in data trusts in the U.K. and around the world.¹⁰ The independent

review recommended data trusts as a way to “share data in a fair, safe and equitable way” and indicated that they would play an important role in growing the AI sector in the U.K.¹¹ The review did not include a common definition for data trusts or a consensus on what form they should take, however, creating a risk of confusion.

After reviewing existing literature, the Open Data Institute produced a report in July 2018 summarizing the range of different uses and interpretations of the term.¹² This included data trusts as “a repeatable framework of terms and mechanisms”; “a mutual organization”; “a legal structure”; “a store of data”; and “public oversight of data access”. The ODI later synthesized and adopted a definition of a data trust as “a legal structure that provides independent third-party stewardship of data”.¹³ This interpretation of a data trust builds on existing work and definitions used by others to apply what has been learned from

¹⁰ Hall, D. W. & Presenti, J. Independent Report: Growing the Artificial Intelligence Industry in the UK. *Government of the United Kingdom*. Retrieved January 28, 2019 from <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk>

¹¹ Ibid.

¹² Hardinges, J. (2018). What is a data trust?. *Open Data Institute*. Retrieved January 27, 2019 from <https://theodi.org/article/what-is-a-data-trust/>

¹³ Hardinges, J. & Wells, P. (2018). Defining a 'data trust'. *Open Data Institute*. Retrieved January 27, 2019 from <https://theodi.org/article/defining-a-data-trust/>

trust law to the governance of data, such as Lilian Edwards' research on the potential role of data trusts¹⁴ and Sean McDonald and Keith Porcaro's descriptions of civic data trusts.¹⁵ It is also aligned with the work of organizations such as the Centre for International Governance Innovation.¹⁶

The ODI is working with the UK Government's Office for AI and other partners on three data trust pilots. The pilots will address data collected, shared and used in a number of different contexts to establish whether data trusts represent a useful approach in managing and safeguarding data that could enable and stimulate more data sharing between organizations. The pilots explore the data trust model and its usefulness for both personal and non-personal data in different contexts at the local, national and global level.

The ODI is taking a practical approach to address a series of research questions across these pilots. In each, it is working with an organization, or group of organizations, seeking to increase access to the data they hold along with third party experts. The activities it is undertaking for each of these pilots include:

-  user research and engagement to understand the data holders', potential data users' and other stakeholders' objectives, requirements and desired outcomes for a data trust;
-  legal analysis to explore the requisite legal personality, and subsequent process for and implications of incorporating a data trust;
-  designing a decision-making process for a data trust based on different deliberative and engagement techniques;

¹⁴ Edwards, L. (2004). The Problem with Privacy. *International Review of Law Computers & Technology*, volume 18,3, 263-294. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1857536

¹⁵ McDonald, S. & Porcaro, K. (2015, August 4). The Civic Trust. Retrieved January 27, 2019 from <https://medium.com/@McDapper/the-civic-trust-e674f9aeab43>

¹⁶ Wylie, B. & McDonald, S. (2018, October 9). What is a Data Trust?. *Centre for International Governance Research*. Retrieved January 27, 2019 from <https://www.cigionline.org/articles/what-data-trust>



designing a data reuse process that potential data users would use to discover, seek to gain access and gain access (or not) to the data via a data trust;



assessing the technical architecture that could be used to underpin and enable access to data via a data trust;



research to explore how the benefits of data access could be distributed equitably to the different stakeholders of a data trust;



assessing the viability of implementing a data trust in that particular context.

These activities will produce a design of a data trust for each pilot, along with a recommendation on whether to proceed to an implementation phase (and if not, what other data access models or approaches may be more relevant). Alongside the outputs for each pilot, the ODI will also publish a final synthesis report and an independently-commissioned assessment of the programme and set of recommendations for data trusts.

Sidewalk Labs' experimentation with data trusts

For many participants, Sidewalk Labs' experience developing a data trust as part of its proposal for an urban revitalization project in downtown Toronto represented a cautionary tale about public consultation, the independence of trustees and the importance of the structure of data trusts in promoting public confidence.

Questions emerged regarding Sidewalk Lab's policies on the collection and use of urban data. While the company had adopted a data privacy statement and responsible use guidelines, certain questions lingered: did Sidewalk Labs intend to monetize the public's data? Did the project intend to serve as a source of data for Sidewalk's parent company, Alphabet Inc.? How would consent be obtained for data collected in the physical environment? How would such data be protected and governed? Who would own the data? Would Sidewalks' partners be governed by this framework, too?

After exploring a variety of collaborative governance models, including data cooperatives and data commons, Sidewalk Labs ultimately proposed the creation of a "civic data trust"—an

entity which in its view would be responsible for protecting the public's interest in Sidewalk Toronto's data governance. The choice reflected a recognition of the unique character of the "urban data" it would collect, which it considered to be a public asset.

Sidewalk Labs proposed that the trust follow the privacy policy and guidelines for responsible data use it had developed. The trust would contain a charter ensuring that any data collected would be used to benefit the community, spur innovation and investment. In Sidewalk Labs' view, the trust would also play an important role in curtailing the private ownership of data that might be considered a public asset, while ensuring compliance with Canadian data privacy legislation. The trust would enable other organizations to share certain data with entities across different sectors for their mutual benefit, thereby driving new opportunities for innovation.

Sidewalk Labs introduced additional governance mechanisms to account for sensitive nature of collecting data in public spaces. All corporate partners would be required to submit an application as well as a responsible data impact assessment to the board of trustees in order to obtain authorization for the use of potentially invasive equipment, such as pedestrian counters or adaptive traffic lights. As part of the application process, the trust could also require additional privacy protections, including de-identification at source in real-time or the deletion of personal information.

The need for data trust pilots

The discussion above exposes some of the challenges of moving from theory to practice in the implementation of data governance models. First, Sidewalk Labs' experience illustrates the importance of a public dialogue in establishing a productive framework for public data governance experimentation. In the context of public data governance, there is a clear need for an iterative process between communities, governments and corporations, committed to co-evolving regulation and technology through a governance model that recognizes a plurality of interests.

Second, an enabling environment should not sacrifice the public's right to the technology or data it participates in creating.

Data trusts should be used to ensure that the public's interest in the intellectual property created (e.g. the code, data and/or the technology) for some public good is protected, while allowing time for our understanding of new technologies and their interactions with humans to mature.

Third, Sidewalk Labs' proposal highlighted the importance of data trust design to building public trust, in particular, regarding the identity of the settlor

and trustees of the trust and the responsibility of trustees to ensure that any data entrusted to a trust is in fact the settlor's to give. In contrast to the bottom-up approach proposed by Delacroix and Lawrence, the Sidewalk experience demonstrates that, even with the best of intentions—responsible data use guidelines and trust charters included—a private corporation that serves as a settlor of a data trust risks provoking public apprehension. Indeed, workshop participants felt very strongly that this model failed to address the types of power imbalances at the core of the issues being discussed, and further exemplified the disenfranchisement of citizens in the decision-making process as to how their personal data is to be used, as the terms of the trust were chosen by Sidewalk Labs in the first place.

Fourth, Sidewalk Labs' experience illustrates the importance of pilot projects in shaping our assumptions about models of anticipatory governance.¹⁷ Data trust pilots can assist in highlighting existing as well as new issues—for instance, how personal information can be exploited by third parties. Currently, public policy discussions on the collection and use of personal information by third parties have focused mainly on privacy issues. Piloting data trusts may help to bring greater clarity to other impacts of third-party data use, such as the potential for discrimination on the bases of age, race or sexual orientation.

¹⁷ Leurs, B. & Duggan, K. Proof of concept, prototype, pilot, MVP – what's in a name. Nesta. Retrieved January 27, 2019 from <https://www.nesta.org.uk/blog/proof-of-concept-prototype-pilot-mvp-whats-in-a-name/>

Part 4. Designing data trusts

Three use cases

Participants were invited to anticipate potential data trust pilots by exploring the implementation of data trust models in three different contexts: urban data, health data, and data collected from online platforms.

Bearing in mind anticipatory regulation's emphasis on inclusivity, participants began by discussing the different personas that may or may not wish to engage with data trust models in each of these contexts. Next, participants were asked to define the problem statement regarding data sharing in the context of personal data, and to explore the asymmetries of power that exist between data-subjects and data-controllers. Discussions focused on identifying the type of data being collected, how it might be used, by whom, the value it represents, and for whose benefit. Fundamental to each of these questions were issues related to data trust implementation, including the identity of the settlors, trustees, beneficiaries as well as financing.

Urban data



A person's decision to engage with an urban data trust will depend on a range of factors, including their privacy preferences, perception of the trust's value, who the trustees are and how they've been selected.

Urban residents, for instance, may have an interest in participating in an urban data trust whose purpose is to build a better urban environment, reduce pollution or improve public transit services. Government, civil society organizations and universities may be equally supportive for similar

reasons. However, participants noted that some individuals might have higher thresholds for data privacy because of individual preference, social or cultural values, or experiences that justify suspicion of centralized data collection such as historical or current discrimination. Others indicated that the very same people might share a greater desire to pool their data, in order to debunk myths and highlight their concerns.

This part of the discussion underscored the reality that technology is not neutral and may even reinforce social inequality. To begin to address these power imbalances, urban data trusts may adopt a socially inclusive mandate and provide broad representation on their board of trustees, including civil society organizations, urban mobility experts and community representatives. Additional questions related to the management of how public value would be extracted from urban data at the city or regional level. To increase public oversight, participants also suggested that the board of trustees of an urban data trust could be subjected to access to information legislation, to promote transparency of decision-making.

Questions arose regarding the effect of municipal politics on the stability of data trusts. If the trustee of an urban data trust is the local government, for instance, what happens to the trust if a new government with a vastly different agenda is elected? Can the new government change the terms of the trust? Trust structures and their terms

are normally designed to withstand such changes, as the settlors have dictated the terms of the trust at the point of settlement, and the trustee is forever bound by these terms. In other words, data trusts that accommodate ongoing changes to the terms of the trust would not be “trusts” in the current and traditional legal sense, but perhaps closer to a corporate governance model with its articles of incorporation, by-laws, and processes by which shareholders must approve changes called for by the directors.

Questions were raised about geographically defined data collection arrangements, such as in the case of Sidewalk Labs, where citizens consent to the collection of their data simply by entering the physical space. Can members of the public who enter these zones unaware of the nature of the arrangements request to have their data removed from the trust? Would removal provide meaningful reparation if their data has already been used to train a model?

Participants indicated that given the invasive nature of data collection in physical spaces, residents may quickly decide to discontinue their participation in the urban data trust if it is unable to demonstrate that it has made progress on achieving its goals. It was noted, however, that even if the trust is unable to deliver on its terms, the trust model provides a way to manage succession of the assets—i.e. the data, the code and the technology—in a manner that protects the public interest.

Health data



The mission of the health data trust plays a significant role in determining the types of people that may be interested in participation. For instance, individuals suffering from a medical condition may be more willing to pool sensitive information into a trust that supports research of that condition than a healthy person who has had little interaction with the medical system. Hospitals, medical professionals and researchers may also be more inclined to engage with a health data trust, recognizing the potential to optimize service delivery, promote early diagnosis, improve quality of care or advance scientific research. Other individuals may be reluctant to engage with the trust, fearing that sensitive health data could be leaked to their insurers.

A health data trust could serve as a coordinating body that would manage requests for licences for data to conduct research, for instance, or to improve medical care and service delivery.

In some cases, there could be several beneficiaries: members of the public could benefit from better healthcare; better data access could enable research labs to develop more sophisticated models and obtain funding on this basis; and medical professionals could be empowered to provide better care to patients. Some indicated that having the public as a beneficiary could incentivize people to regularly participate as subjects of medical research, leading to ethical issues and potential conflicts of interest that would have to be addressed in the trust's terms.

In light of the diverse interests at stake, it was noted that there is a strong incentive in the medical field to establish multiple trusts, each being representative of a particular group of stakeholders. Concerns were raised as to whether or not the fragmenting of the datasets across multiple trusts could weaken the utility of the data or limit the possibilities for data sharing in the event of the incompatibility of trust terms. One possible solution could be the creation of "meta-trusts", as noted above—in this context a health-focused trust that would negotiate with a

variety of smaller, generalist trusts to pull together a larger dataset while complying with the terms of each individual trust.

The discussions raised challenges related to data whose provenance is shared, such as genetic data, in which patients, hospitals and the medical professionals who administered the imaging might have an interest. Would each of these groups have to consent for the medical image to be shared into a trust? Shared provenance data also creates problems for the revocability of consent, including the rights to erasure and portability mentioned above. It is also important to note that revocability would have to reside with the trustee to constitute a traditional legal “trust”, as noted above. If only one of the entities wishes to withdraw the shared provenance data, how will the trustee decide?

The choice of an appropriate trustee centered around the multidisciplinary knowledge and sensitivity such a complex task would require, as well as the power the trustee would wield. While some participants suggested that the government could fill this role in the health context, particularly in countries that have public healthcare, others feared that it lacked the necessary expertise and advocated for the establishment of a new order of

data-governance professionals, who could sub-specialize. To include representation from the public on the board of trustees, it was theorized that a hybrid trust-corporate governance structure, in which settlors are treated as shareholders in the trust, benefitting from all associated rights and remedies, would be worth exploring.

Participants noted the difficulty for settlors to ascertain the value a trust could create in the case of cancer research where goals are either uncertain and at best long-term. In such cases, a top-down regulation could impose a general obligation to share medical data under certain conditions and safeguards, regardless of the trust that individuals have joined. Participants observed more generally that the top-down regulation of data trusts by government could help ensure that trust terms and conditions respect human rights and promote the public good.

Regarding financing, it was proposed that a trust could generate revenue by charging licensing fees for continued access to the trust’s data. Others warned that ethical issues would emerge if incentives were created for people to profit from health data, and that state funding might be more appropriate.

Data from online platforms



Workshop participants focused their discussions on ride-sharing companies as an example of an online platform. Ride-sharing companies collect data from drivers and riders. City officials and residents may be interested in engaging with a trust whose purpose is to improve urban planning, for instance by using aggregated data to identify the need for new bus routes. Others may be interested in joining a ride-sharing data trust if its mission was to generate revenue for the citizen-consumers. Several participants registered strong objections to this idea, which they feared would risk creating a new data privacy class system in which those who can afford not to share their data may benefit from greater privacy rights than those who cannot. In either case, the trustees would negotiate on the data-subjects' behalf to ensure that the data accessed by the data-controller was being used in accordance with the purpose of the trust.

If ride-sharing platforms were required as part of their certification process to negotiate access to data with an independent trust organization, the latter could serve as an important counterbalance to the power asymmetries that currently exist between the platforms and their drivers or riders.

For instance, a ride-sharing data trust could protect drivers' privacy by preventing platforms from leeching data from other apps and services installed on the users' mobile devices, such as social media platforms, or from having their data used to train algorithms for self-driving cars, unless specific compensation is negotiated for this use.

Certain data trusts could also be used to promote transparency regarding responsible business practices. Drivers' unions or employee associations could access the trust data to monitor discrimination and advocate for better working conditions. Riders may be more inclined to participate in data trusts that support socially responsible ride-sharing companies. If the ride-sharing company were required to establish an independent trust organization—as Sidewalk Labs did—rules could be created to regulate the terms of the trust and to require appropriate representation as a condition of certification.



Part 5. Conclusions and recommendations

The discussions from the workshop elicited a series of conclusions and recommendations that should be taken into account by policymakers in the elaboration of national strategies for data governance.

Conclusions

- Consent-based models of data governance fail to protect the public against privacy violations and the unethical collection and use of personal data.
- In some sectors, a lack of market competition has given consumers no alternative choice for the protection of privacy, and none is likely to appear.
- The monopolization of data by a limited number of corporations has limited the extraction of public value from data.
- The concept of ownership is not easily applicable to data and fails to address the more important questions of management of data access and use and the impact of the latter on human rights.
- There has been an erosion of public trust and confidence in data-collecting organizations and the technologies they employ.
- Maintaining the status quo risks provoking a public backlash that would imperil the ability to use data to equitably distribute resources.
- There is a pressing need to explore new data governance models that provide individuals with some control over data and the technologies that use them and advance the public good. It also risks dragging technologies that rely upon that data – including AI – down with it.

-
- There is a need for government and industry to develop inclusive and forward-looking models of data governance that encourage innovation through an iterative process.
 - Clear definitions of new data governance models enable governments, industry and society to pilot their applications and learn through experimentation.
 - Data trusts offer a flexible and inclusive model that enables government and industry to co-evolve regulation and technology, allowing time for concepts of digital rights to mature while immediately strengthening the rights of citizen-consumers.
 - Data trusts can protect the public's intellectual property rights in data against monopolization by private interests, enabling the sharing of public value.
 - Data trusts leverage existing legal governance structures, such as trustees' fiduciary duty, to provide the public with stronger protection against privacy violations and the unethical collection and use of their personal data.
 - Data trusts are not a panacea, and in some circumstances other legal structures or models may be more appropriate (corporate governance models came up several times).
 - Conflating all solutions with trusts is not useful.
 - Bottom-up data trusts can be a useful tool to partially address power imbalances that exist between corporations, government and citizen-consumers, promoting a more equitable distribution of resources and the further protection of rights.
 - Top-down regulation of data trusts by government could help ensure that trust terms and conditions respect human rights and promote the public good.
 - An ecosystem of data trusts would enable the public to choose a data governance regime that reflects their privacy preferences and supports their values.

Recommendations

- Governments, industry, trade unions and civil society should collaborate to pilot data trusts in order to improve upon the consent-based model of privacy, especially in sectors where an absence of competition has left consumers with no viable alternative.
- In collaboration with stakeholders, governments should initiate a public education campaign to improve data literacy and promote innovation and experimentation with data trusts.
- Governments should introduce legislation in certain circumstances requiring companies to negotiate with data trusts for the collection and use of the public's data.
- Data trust pilots should explore the application of a variety of data trust models to particular use cases, notably, to identify solutions with respect to their structure and financing.
- Governments should implement data trusts as tools to increase access to data and promote a more equitable distribution of its economic value.
- Governments should explore ways to make the right to privacy meaningful in the digital context, for instance, by adjusting legislative frameworks to empower trustees with the right to exercise revocability, portability and erasure on behalf of a trust's beneficiaries.