The Price of Trust?

An Analysis of Emerging Digital Stewardship Models

31 March, 2020 Draft

David Fewer, Sarah Crothers, Brenda McPhail, Stephanie Perrin

Thanks goes to Kristen Kephalas, Sundeep Virdi, Sharli Chan, and Laura Garcia Vargas for research assistance.

Prepared with the generous support of the Contributions Program of the Office of the Privacy Commissioner of Canada.

TABLE OF CONTENTS

Introduction

Data is not the new oil, nor it is it the new plutonium.[1] However, like both of these resources, the failure to manage data properly may result in serious consequences. Economically, Canadian businesses and governments are being urged to tap into the power of data to improve competitiveness and realize efficiencies. Poor governance of data, however, carries with it more than mere economic risk. Data also implicates privacy interests, human rights, and intellectual property rights. Managed poorly, data collection and use can result in real harm to commercial, public and private interests.

Designing an effective data governance approach is a complicated undertaking. It is an over-constrained optimization problem requiring a solution that balances the tension between powerful competing forces. Perceived public benefits of data use can conflict with both commercial interests in exploiting data and personal privacy interests, not to mention the more nebulous concept of the public good. Accordingly, data governance must address numerous legal, financial and ethical ramifications, not all of which have emerged. A data governance system must therefore delicately balance competing priorities yet also flexibly adjust to new issues.

Appropriate data governance approaches are needed if Canada is to thrive in the data-driven economy.[2] But there is little consensus regarding what that should look like. Stealing much of the spotlight these days in the discourse is a mythical creature; the data trust. Born of the common law trust, most entities labeled as "data trusts" retain two key features: a critical element of fiduciary responsibility, and the principle that a trustee manages assets in the interests of a group of beneficiaries. Absent specific legislation, data trusts are not true legal trusts, partly because data is not recognized as an asset that can be held by a trust.

The label "data trust" has been applied to a wide variety of entities, many of which only share the trait that they possess data, creating confusion as to the very meaning of the

[1] Inspired by "Jim Balsillie: 'Data is not the new oil – it's the new plutonium'" (May 28, 2019), online: Financial Post < https://business.financialpost.com/technology/jim-balsillie-data-is-not-the-new-oil-its-the-new-plutonium>; also https://media.nesta.org.uk/documents/DECODE-2018_report-smart-cities.pdf at 8.

[2] Dan Ciuriak, "The Data-driven Economy: Implications for Canada's Economic Strategy", (June 2019), Policy Brief No 151, Centre for International Governance Innovation, online: cigi <https://www.cigionline.org/sites/default/files/documents/PB%20no.151web.pdf>.

term, “data trust”. The term itself carries with it the positive connotations of the term “trust”, and for that reason is vulnerable to misuse by interests keen on gaining access to data pursuant to governance tools that lack the safeguards of a true trust. For that reason, in this report we will focus on identifying responsible data stewardship practices rather than attempting to define an ideal trust model. The purpose of this report is to bring clarity to the question of what responsible data stewardship looks like. In doing so, we will examine desirable features of data trusts and compare a variety of emerging data governance systems. To that end, this report aims to define a framework that can be used to (i) identify a responsible data stewardship model and distinguish it from other data governance models, and (ii) classify different kinds of data governance models.

This report is structured as follows. Part I outlines the key features of a proposed data stewardship framework and provides a brief introduction to other data governance systems. Part II elaborates on each element of the proposed data stewardship framework. Part III evaluates several case studies (listed in Table 1) using the proposed framework. Part IV identifies different classifications or groups of data governance models and provides guidelines on the preliminary selection of a data governance system for a given set of data.

Table 1: Data Governance Systems Included in Comparative Analysis

Data Trust
Internet Corporation for Assigned Names and Numbers (ICANN) – Registration Directory Service (RDS)
Sidewalk Toronto Case Study 1: Personal Mobility Data, “Civic Data Trust”
Sidewalk Toronto Case Study 2: ICES Health Case Study, “Data Safe Haven”
TRUATA Platform
TRUATA Automotive Use Case
TRUATA Mastercard Use Case
First Nations Information Governance Centre (FNIGC) Regional Health Survey (RHS)
Barcelona DECODE

London-Open Data Institute (ODI) Collaboration
Silicon Valley Regional Data Trust (SVRDT)
National Health Information Network (NHIN)

Part I: Data Stewardship: Trusts and Other Data Governance Models

i. Proposed Data Stewardship Framework

Our discussion of data stewardship must start with data trusts, as this is the governance model that dominates the current public imagination. Data trusts first emerged as a governance tool in Hall and Presenti's influential 2017 report, *Growing the Artificial Intelligence Industry in the UK*.[3] The authors argued that the social and economic benefits of data and emerging technologies for managing data could only be achieved if data were shared more widely, and data trusts offered a tool for permitting data sharing in a more trustworthy model than the marketplace has to date offered citizens. In Canada, Sidewalk Labs seized on the idea of a data trust – what it termed a "civic data trust" – as a tool for overcoming the problems of obtaining meaningful consent for the collection, use and sharing of personal information collected in public spaces.[4] While the Sidewalk Toronto proposal has proven contentious, the federal Liberal majority government in the summer of 2019 was nonetheless sufficiently impressed with the features of the trust model as to propose amending PIPEDA, Canada's federal commercial private sector personal information protection law, to allow for the creation of data trusts. The government specifically envisioned using trusts to manage consents, on behalf of the public, to the collection, use and disclosure of personal information in circumstances in which consent would be difficult to obtain but the public interest nonetheless supports its exploitation.[5]

[3] Innovation, Science and Economic Development Canada, "Strengthening Privacy for the Digital Age", online: Government of Canada < https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html > [*ISED*].

[4] See the discussion of the Sidewalk Toronto case study in Part III, *infra*.

[5] *ISED*, *supra* note 3.

A data trust is not a true legal trust, but borrows the basic trust structure.[6] A traditional legal trust is an agreement involving three parties: (i) the trustor who contributes property to the trust, (ii) the trustee who manages the property, and (iii) the beneficiary for whom the trust is managed and receives a benefit. A core element of the legal trust is the fiduciary duty a trustee owes the beneficiary.[7] Fiduciary duties arise in relationships in which the fiduciary has scope for the unilateral exercise of power so as to affect the beneficiary's interests, and the beneficiary is peculiarly vulnerable to the fiduciary.[8] Breach of a fiduciary duty empowers beneficiaries to seek from courts powerful equitable remedies.

Trust law struggles to accommodate data trusts. For starters, data is not "property" in the sense that traditional trust law conceives of it. While there might be "interests" in data (such as personal information or expectations of confidence) and "rights" in data (such as copyright in an original compilation of data), in Canada data itself is not property that can be the subject of a trust. Similarly, trust law struggles with the kinds of "public interest" purposes that motivate data trusts. While discussion of these interesting issues lies beyond the scope of this Report, these issues help explain the multitude of governance structures held out as data trusts.

The Open Data Institute (ODI) defines a data trust as "a legal structure that provides independent stewardship of data."[9] The key similarity between a legal trust and a data trust lies in the purpose of stewardship: the idea that one party is managing an asset for the benefit of another. Along with stewardship, most data trusts also import a fiduciary-like

[6] BPE Solicitors, Pinsent Masons & Queen Mary University of London, "Data trusts: legal and governance considerations" (April 2019), online: the odi <https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf> at 12 [*ODI, Data trusts*].

[7] MaRS, "A Primer on Civic Digital Trusts", online: marsdd <https://marsdd.gitbook.io/datatrust/trusts/what-is-a-trust>.

[8] *Lac Minerals Ltd v International Corona Resources Ltd*, [1989] SCJ No 83, [1989] 2 SCR 574 at para 171 [*Lac Minerals*].

[9] Open Data Institute, "Data Trusts summary report" (2019), online: the odi <http://theodi.org/wp-content/uploads/2019/04/ODI-Data-Trusts-A4-Report-web-version.pdf> [*ODI*]; see also Max Pixel, "UK's first 'data trust' pilots to be led by the ODI in partnership with central and local government" (Nov 20, 2018), online: the odi <https://theodi.org/article/uks-first-data-trust-pilots-to-be-led-by-the-odi-in-partnership-with-central-and-local-government/>.

duty, implicitly, if not explicitly.[10] This obligation is generally imposed on the steward by a corporate structure or by contract, instead of through a legal trust structure.[11] Unsurprisingly, many data trusts carry forward the key terminology, defining a trustee, one or more trustors and one or more beneficiaries. The core elements of a data trust are those that maintain the spirit of a legal trust; the stewardship of an asset by a trustee and a fiduciary-like obligation. The most common situations in which data trusts are proposed have an important additional characteristic: there is a compelling public purpose motivating the effort to collect and share the data.[12] We take these to be core elements of our data stewardship framework.

The basic function of data stewardship is the management of data flows, which includes the collection, use, storage and sharing of data.[13] The features of data stewardship that are critical to its role are referred to as primary design features in this report and provide a framework for differentiating responsible data stewardship from other data governance structures. We summarize our primary data stewardship design features in Table 2.

Table 2: Data Stewardship Primary Design Features

Design Feature	Description
1	Independent stewardship of data flows, including collection, use storage, and distribution / disclosure of data
2	Imposition of fiduciary-like obligation
3	Objective is a public purpose

[10] Sean MacDonald, "Reclaiming Data Trusts" (March 5, 2019), online: cigi <https://www.cigionline.org/articles/reclaiming-data-trusts>.

[11] Rosario G Cartagena et al, "Building Ontario's Next-Generation Smart Cities Through Data Governance, Part 1: Health Data Safe Haven", online: orion <https://www.orion.on.ca/wp-content/uploads/2019/11/Smart-Cities_ICES_Health-Data-Safe-Haven.pdf> at 17 [*Cartagena*].

[12] Similar to: *ODI, Data trusts,, supra* note 6 at 13, although authors do not say public purpose is required.

[13] MaRS, "A Primer on Civic Digital Trusts", online: marsdd <https://marsdd.gitbook.io/datatrust/trusts/what-is-a-civic-digital-trust> [*MaRS Civic Trust*].

Secondary design features can be used to differentiate types of data governance models from one another. The most important of these features is the characteristics of the data itself, specifically the extent to which the data engages a privacy, proprietary, commercial or public interest. Additional secondary design features include: stakeholders, the self-sustenance model / structure, the technical architecture, the data access model, and the system for enforcement and offering remedies.[14] We summarize our secondary data stewardship design features in Table 3.

Table 3: Data Stewardship Secondary Design Features

Design Feature	Description
1	Data characteristics include privacy, proprietary and/or commercial interests that need to be protected
2	Stakeholders
3	Self-sustenance model and organizational / legal structure
4	Technical architecture
5	Data access model
6	System of Enforcement and Remedies

The proposed framework consists of primary and secondary design features that serve to both differentiate responsible data stewardship models from other data governance approaches and assist in the delineation of various types of data stewardships. Each feature will be further explored in the next section of this report.

ii. Brief Overview of Other Data Management Structures

The framework in this report defines three necessary primary design elements of a responsible data stewardship model as: (i) independent stewardship, (ii) a fiduciary-like obligation, and (iii) a public purpose. In the absence of any one of these features, the

[14] Alannah Dharamshi et al, "Building Ontario's Next-Generation Smart Cities through Data Governance, Part 2: Towards a Smart City Data Trust" online: computeontario <https://computeontario.ca/wp-content/uploads/2019/11/Smart-Cities_MaRS_Towards-a-Smart-City-Data-Trust-1-1.pdf> at 12-33 [*Dharamshi*].

system is not functioning as a responsible data steward. This section will discuss some of these alternate systems in an effort to clarify the boundaries of responsible data stewardship. The following discussion is not meant to be exhaustive, but will focus on approaches that are relevant to the present discussion.

a. Marketplace

In a data marketplace, access to data is granted in many cases in exchange for a fee. In other words, data is bought and sold. If access to the data is controlled only by the payment of a fee, and does not depend on oversight by an independent steward, it is better described as a data marketplace than as a data trust, much less as responsibly stewarded data.[15] Certainly, a data steward could charge fees in exchange for access to its database. However, a responsible data steward must offer independent oversight regarding selection and inclusion of data, who has access to it and for what purpose. In addition, our responsible data stewardship model requires a public purpose and the imposition of a fiduciary-like obligation. However, if a company is simply selling their own proprietary data for profit, then the system is unlikely to be functioning consistently with our stewardship framework.[16]

b. Contractual Agreements

Contractual agreements can create a data sharing arrangement that is similar to a marketplace. However, the data may or may not be exchanged for money, but instead traded or shared between entities for some kind of mutual benefit. For instance, these agreements may be formed between corporate entities who agree to share data only with each other for the purpose of developing a new product together.[17] Alternatively, these arrangements could be between multiple entities. Of course, contractual agreements may also be used in the context of our responsible stewardship framework to define how the

[15] MaRS, "A Primer on Civic Digital Trusts", online: marsdd <https://marsdd.gitbook.io/datatrust/trusts/technical-architecture-options> [*MaRS Technical Architecture*]; see also *Dharamshi, supra* note 14 at 33.

[16] We acknowledge that powerful contractors often act as coordinators for different actors using their systems/software to manage data. In other words, the contractor sees all the data, while the various contractees get the benefit of AI tools and statistical analysis, or benefit from specific tracking, but do not see each other's data. This is not often subject to independent oversight.

[17] *ODI, Data Trusts, supra note* 6 at 8, 14,

data provided by a particular donor or trustor can be used, or the conditions under which access can be granted to third parties.[18] However, in the absence of a public benefit to the data sharing, or the need for independent oversight, then contractual agreements are a more cost-effective solution compared to our responsible stewardship framework.[19]

c. Open Data

Open data is a concept that describes an approach that makes data publicly available with no restrictions. Weather data is often managed as open data, and made publicly available across most jurisdictions. Similarly, municipalities are managing more and more of their data under open data licenses. Open data may involve third party oversight since the data needs to be collected and published, but the data manager may or may not be independent. Also, a fiduciary-like obligation will likely not arise. Open data offers an intriguing data governance option, but is ill-suited to govern the kinds of multi-stakeholder demands that our responsible data governance framework is designed to address.

d. Data Co-operative

The data co-operative offers another option for managing data. A co-operative can closely resemble a data trust. The differentiator is that decisions regarding collection, use and distribution of data are made by the members of the co-operative rather than by an independent steward. In a co-operative, all members have an equal share in decision-making regarding what data can be shared, with whom, and for what purpose.[20] There is no independent oversight: the community is self-governing. Fiduciary duties are unlikely to arise. There may be a strong public purpose behind the formation of the co-operative. But in the absence of independent oversight, the data management system is not functioning as envisioned by our responsible data stewardship framework.

Distributed ledger technologies, or blockchains, provide one option that can be used in the context of a data co-operative to enable individual data donors to control access to their data. It is not a governance strategy in and of itself, but is rather a technology that enables certain governance strategies, such as a data co-operative. Users can give selective

[18] *Ibid* at 14.

[19] *Ibid* at 13.

[20] *Ibid* at 3, 16.

permission for their data to be shared with city officials, other government entities, third parties, and others.[21] Additionally, the technology allows users to give permission for use of their data for only specific purposes.[22] Blockchain is a trackable database that promises better security, making it a suitable technology for managing personal information.[23]

It is conceivable that a data steward could operate with a two-tiered system which incorporates a distributed ledger within a first level that allows users to grant permission for the use of their data by certain third parties. The second level could introduce an independent steward with a fiduciary-like obligation who acts as a second level of decision-making to ensure that data is used in the best interest of the beneficiaries and data donors. However, to date, the use of distributed ledgers has been primarily discussed in the context of a data commons or a data co-operative.

Part II: Design Features of a Trust-Like Data Stewardship Model

The key elements of trust vehicles are (1) the imposition of a fiduciary-like duty on (2) an independent steward, with (3) a public purpose to the stewardship endeavor. The characteristics of the data subject to the governance arrangement - which in turn may be subject to multiple and competing interests - also raises important considerations for the structure adopted. We consider these features in turn.

i. Imposition of Fiduciary Duty & Independent Stewardship

The first key element of a data trust is that the stewardship is independent. The second is the imposition of a fiduciary-like duty. These elements are closely related and so are discussed together.

A fiduciary duty is a legal obligation arising from the law of equity, and is characterized by relationships requiring trust; where one party is required to make

[21] Antonio Calleja-Lopez, Arnau Monterde & Xabier Barandiaran, "DECODE, Framework for democratic governance of distributed architectures" (Dec 2017), online: decodeproject <https://decodeproject.eu/sites/default/files/D%202.3.pdf> at 5 [*DECODE*].

[22] *Ibid* at 5.

[23] *Ibid* at 35. A description of how distributed ledger technologies work is beyond the scope of this report.

decisions in the best interests of another.[24] The fiduciary is required to act honestly, in good faith and only in the best interest of the beneficiary.[25] Some commonly recognized relationships in which such duty is imposed include trustee-beneficiary, parent-child, solicitor-client, and director-corporation.[26] The list is not closed. The Supreme Court has indicated that the critical feature of a fiduciary relationship is that "the beneficiary is peculiarly vulnerable to or at the mercy of the fiduciary holding the discretion or power."[27] There is a dependency to the relationship that prevents the beneficiary from protecting themselves from the decisions made by the fiduciary. In addition, fiduciary obligations are necessary where adequate remedies otherwise don't exist.

In contrast, situations in which the parties have equal opportunity to provide legal definition for their obligations, such as through a contract, do not give rise to fiduciary obligations.[28] Fiduciary obligation is to be reserved for situations in which "equity's blunt tool" is truly needed; when the available remedies under the laws of contracts or negligence are inadequate.[29]

The need for independent stewardship is related to the imposition of fiduciary obligations. The fiduciary must act only in the interest of the beneficiaries. So the fiduciary must be free of conflicts of interest.[30]

The importance of fiduciary obligations to a data trust become more apparent in the following sections when the data characteristics are more fully discussed. Fundamentally, the function of a data trust is to manage data. The trustee will decide what data can be shared, with whom and for what purpose. If the data set contains personal information, the data subjects are vulnerable to the decisions made by the data steward. Although the data donor may or may not also be a beneficiary, something similar to a fiduciary duty is engaged. To ensure that the steward's decisions are made in the best interests of the

[24] See *Lac Minerals*, *supra* note 8.

[25] M Litman, "Law of Fiduciary Obligation" (December 16, 2013), online: The Canadian Encyclopedia <https://www.thecanadianencyclopedia.ca/en/article/law-of-fiduciary-obligation> [*Litman*].

[26] *Lac Minerals, supra* note 8 at paras 30, 51, [1989] 2 SCR 574.

[27] *Ibid* at para 32.

[28] *Ibid* at paras 27, 31.

[29] *Ibid* at para 29.

[30] *Litman, supra* note 24.

beneficiaries and/or donors, the steward must also be independent. In some cases, data that engages a commercial or proprietary interest may also engage fiduciary-like obligations.

In the context of a data trust-like arrangement, fiduciary obligations may attach where the beneficiaries or the data donors are vulnerable to the decisions of the trustee. However, fiduciary obligations can also be imposed by statute. The reality is that the issue may need to be litigated before it is known to what extent fiduciary duty will attach to the management of data. For the purpose of this report, the term fiduciary-like will be used to indicate an obligation that is similar to fiduciary duty, regardless of whether it is imposed contractually, by statute, or if it is likely that the court would find that the relationship is of a fiduciary nature.

ii. Public Purpose

Data trusts are a response to the need to balance the risks associated with sharing certain types of data with the need to share it for the sake of advancing a public purpose. Where a fiduciary obligation is engaged, the implication is that the beneficiaries and/or data donors are placed into a position of vulnerability. Certainly, where beneficiaries and/or donors can be placed in harm's way by the decisions of the trustee, there should be a legitimate reason for doing so. Therefore, a trust-like data stewardship framework must require a public purpose.

On the other hand, if a fiduciary duty is not engaged, then alternative data governance options exist. For instance, two companies who wish to share data between themselves or sell the data to others for the sake of maximizing profit, can use a data marketplace or a contractual arrangement[31] to protect themselves from risk.

Trust-like data stewardship frameworks have been proposed in a variety of circumstances for a broad range of public purposes, including the management of municipal or regional traffic flows, advancing medical research, energy conservation, and

[31] *ODI, Data trusts, supra note* 6 at 9.

managing pollution and environmental protection.[32] Trust-like data stewardship frameworks could also be used to foster innovation by preventing data monopolies and ensuring a healthy competitive market in data-driven industries.

iii. Data Characteristics

a. Overview of Approach

The core function of a data stewardship framework is to manage the flow of data. The need for an oversight structure involving a steward with some degree of fiduciary-like responsibility stems from a desire to manage the tension that arises due to the various interests engaged by the data itself. For the purpose of this study, the characteristics that have been identified as most critical are: (i) privacy interests, (ii) commercial value, (iii) public interest and (iv) proprietary interests. In an attempt to quantify these various characteristics for a given set of data, each dimension can be assigned a value of low, medium or high. Colours or numerical values can be used to facilitate visualization—green (1) for low, yellow (5) for medium, and red (10) for high.

b. Privacy Sub-Index

The privacy interest sub-index takes into account a variety of factors, including:

- whether the data is personal information from either a legal or ethical perspective;[33]
- the degree of sensitivity of the information, e.g. a person's DNA profile is more sensitive than their chocolate bar preferences;
- whether the data has been de-identified and how easy it would be to re-identify it;
- the severity of the consequences in the event of a data breach; and

[32] Some examples mentioned in: Theo Bass, Emma Sutherland & Tom Symons, "decode, Reclaiming the Smart City: Personal data, trust and the new commons" (July 2018), online: media nesta <https://media.nesta.org.uk/documents/DECODE-2018_report-smart-cities.pdf> at 9 [*DECODE, Nesta*].

[33] By "ethical", we refer to information that a data subject would regard as personal despite the absence of a legal determination that the data is subject to lawful privacy obligations. Personal information in the hands of many provincial political parties, who outside of British Columbia are not subject to legislative oversight in respect of privacy practices, would qualify as such.

- the liberty and human rights implications of the loss of control over one's personal information.

Privacy interests in data can have a significant impact on the characteristics of the data set as a whole. Personal information in Canada is subject to significant and overlapping legal regulation. Accordingly, both data stewards and third parties interested in accessing or using data sets that include personal information expose themselves to significant potential legal liability. From the perspective of the data subject, protection of privacy interests in personal information in datasets may prove to be the linchpin of trust in a data stewardship arrangement. This is all the more so given the tremendous commercial interest in access to data, including personal information, that is part and parcel of the big data, data analytics, and smart city movements.

The privacy sub-index laid out here contemplates technical solutions to personal information liability issues and legal obligations, including de-identification strategies. Data de-identification strategies involve stripping data of elements that can potentially identify individuals. Data de-identification inevitably involves a trade-off between detail and privacy. Commercial interests in data inevitably want datasets that include more information, including more personal information. Privacy protection, on the other hand, requires sharing less detail, which may prove to limit the usefulness of the de-identified dataset. Accordingly, datasets subject to multi-stakeholder pressures inevitably face competing pressures.

The contentious nature of de-identification as a process, and as an alleged means of taking data out of the scope of many privacy laws, must be acknowledged. One prominent researcher, noted computer scientist Cynthia Dwork, has said that "de-identified data isn't".[34] There is emerging agreement that there can be no guarantee that de-identification strategies can prevent data from being re-identified by cross-referencing one dataset with

[34] Cynthia Dwork, "Differential Privacy: A Cryptographic Approach to Private Data Analysis," in Julia Lane, Victoria Stodden, Stefan Bender, Helen Nissenbaum, eds., Privacy, Big Data, and the Public Good, p. 297, Cambridge, 2014.

another that includes common fields and identifiers, an action facilitated by the large number of datasets for sale in the age of big data and data brokerage. Security and privacy advocates are critical of approaches to evaluating the risk of re-identification; poor security measures applied to data sets; and the underestimation of adversary knowledge, motivation, and resources. In particular, there is no evidence to support the claim that de-identification can effectively prevent location data from being re-identified. Nonetheless many data sharing initiatives still rely on de-identification strategies. While this report necessarily discusses de-identification practices and claims, such discussion is a means of critical engagement. Just as there is a range of positions regarding the possibility of effective de-identification in the privacy community, so too do the members of the research team bring a range of perspectives.[35]

Consent mechanisms built into Canada's existing laws offer a second layer of protection for data subjects' personal information. Canada's commercial sector privacy legislation requires consent to the collection, use and sharing of personal information. Provincial public sector privacy laws similarly require consent to the sharing of personal information, and consent to the use of personal information for a purpose other than that for which it was collected. While in theory consent could be an effective tool for individuals to control dealings with their personal information, in practice the utility of consent mechanisms has been undermined by the acceptance of complicated privacy policies and minimal notice as sufficient to obtain consent to the collection, use and disclosure of personal information. Accordingly, consent mechanisms seldom provide meaningful controls over the collection, use and sharing of personal information. From the perspective of the data processor, however, consent mechanisms can similarly pose problems. For example, consent can be difficult to gather from individuals where personal information is collected in a public space. Similarly, consent is required for the re-use of personal information for new purposes. However it can be extremely difficult to go back to the data subject to obtain the required content. This problem is particularly prevalent in the area of

[35] In particular, CCLA, is on record as objecting to data de-identification as an ineffective means of safeguarding privacy, and preventing privacy harms and violations.

big data analytics, since one of the primary values of big data approaches is to identify unexpected uses and applications of data. It is also problematic in the smart city context, where information collection is built into infrastructures, raising the possibility that individuals who live, work, or simply pass through parts of a city may not have a meaningful say in how, when, or where data about them is collected if complicated issues of consent remain unresolved.

c. Public Interest Sub-Index

The public interest sub-index indicates the strength of the public interest in the data set. Factors taken into account include the size of the population that stands to benefit, the importance of the objective, the consequence of collecting versus not collecting and sharing the data, and the consequence of a data breach. It is important to distinguish public interest considerations from commercial interest considerations. Frequently, commercial interests are offered as public interest justifications for exploitation of a data set. For example, opening up a data set to commercial research and development activity is frequently justified on the basis of the public benefit of new and useful commercial offerings. The "public interest" must be defined in each instance, and balanced against individual rights which may be abrogated.

d. Commercial Interest Sub-Index & Proprietary Sub-Index

The commercial interest sub-index accounts for the potential value of the data to private businesses. Factors considered include the number of large and small corporations active in the space and the potential for market growth. Proprietary interest is related, but specifically addresses whether the data may be subject to intellectual property protection, such as trade secret, copyright, or in some cases, database protection.[36] In these cases, there may be licensing agreements that need to be maintained, or respected.[37] Commercial or proprietary interest can apply to personal information, but it can also apply to technical data, such as that regarding a product's performance capability. In many contexts relevant

[36] *ODI, Data trusts, supra note* 6 at 22.

[37] *ODI, Data trusts, supra note* 6 at 10, 22, 30, 31.

to data governance, such as connected cars in smart cities, data can have both personal and commercial and/or proprietary aspects.

e. Summary of Data Characteristics Discussed in this Study

The data characteristics evaluated in this study are summarized in Table 4.

Table 4: Summary of Data Characteristics

Main Category	Factors Considered
Privacy	Sensitivity of data
	Level of de-identification
	Risk of re-identification
	Consequence of data breach
Public	Proportion of the population affected, balancing test
	Importance of the objective
	Consequence of not collecting and sharing the data
	Consequence of a data breach
Commercial	Number of large corporations active in the space
	Number of small corporations active in the space
	Market growth potential
Proprietary	Existence of trade secret protections on business-generated and business-collected data
	Existence of copyright protection on database compilation of business-generated and business-collected data
	Existence of licensing agreements

f. Examples

The following figures illustrate the approach used in this study. Figure 1 represents a data set consisting of the blood type of a collection of individuals. The data represented in Figure 2 is the pick-up and drop-off locations for a series of individuals who used a commercial ride-sharing service.

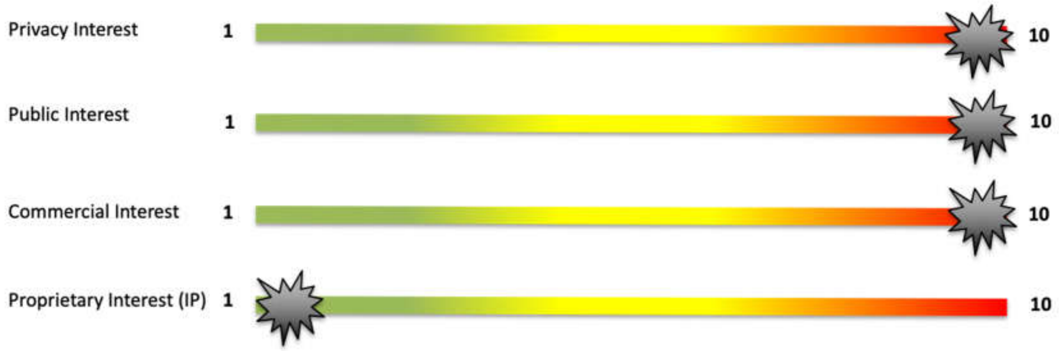


Figure 1 – Illustration of data characteristics for Example 1, blood type data. The resulting data index = (10,10,10,1).

The blood type data scores very high on privacy interests because a person's blood type is part of their biometric footprint. Certainly, blood type data is not as sensitive as a DNA profile, but is still highly sensitive. The public interest for this data is high because there is a legitimate public purpose in having a database of blood types. Indeed, the existence of databases of blood types for a region or country would be beneficial in the case of an emergency. In the event of an emergency, being able to quickly locate donors of the right blood type could mean the difference between life and death. However, blood type is highly personal biometric data, and so significant attention must be paid to protecting privacy interests. For instance, because blood can't be bought and sold in Canada, there is little commercial interest in blood.[38] Nevertheless, the data itself could be valuable if a private enterprise possessed the database. The commercial interest is therefore rated to be high.

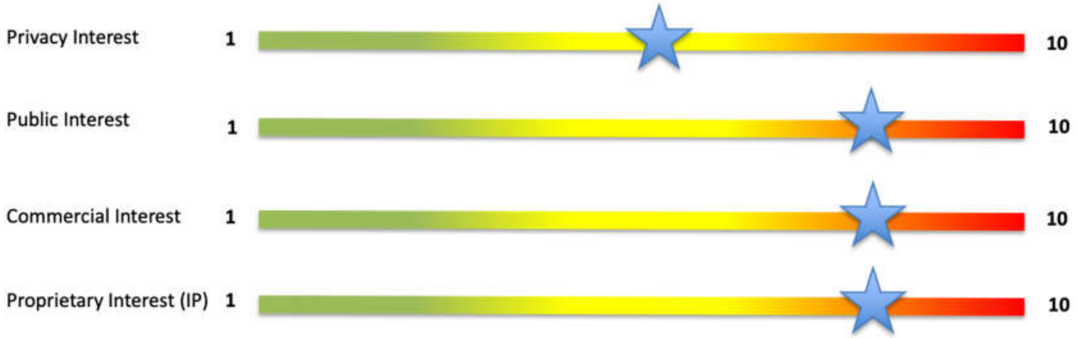


Figure 2 – Illustration of data characteristics for Example 2, ride share pick-up and drop-off data. The resulting data index = (5,7.5,7.5,7.5).

[38] *Trillium Gift of Life Network Act*, RSO 1990, c H20; also *Assisted Human Reproduction Act*, SC 2004, c2.

The pick-up and drop-off locations for ride share users is personal information, but arguably engages privacy interests to a lesser degree than biometric data.[39] A high degree of public interest is engaged because the data can be used to optimize traffic flows and energy conservation. Ride share user data is collected by commercial companies like Uber and Lyft, and so will have a proprietary aspect. In addition, the ride share data has a significant commercial value to other ride share companies that could use the information to optimize their own operations and business models. Figure 3 compares the data characteristics for blood type data (Figure 1) and ride share user data (Figure 2).

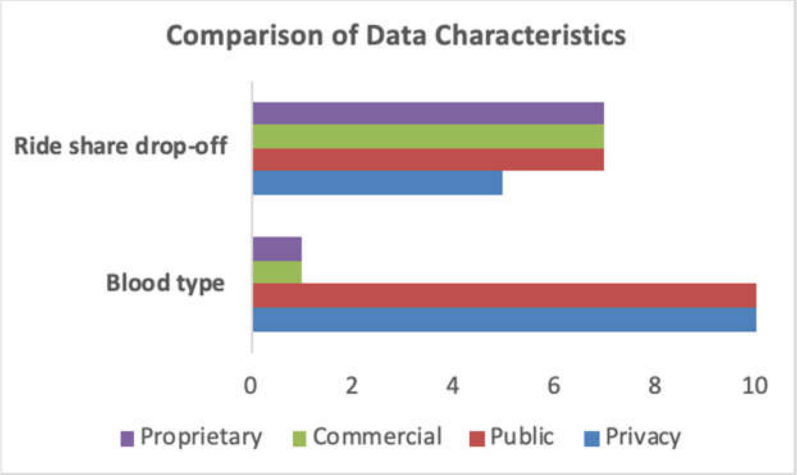


Figure 3 – Comparison of blood type data (Figure 1) and ride share data (Figure 2) examples.

g. Understanding Interactions

A key function of a data steward is to manage the tension inherent in sharing data that engages a privacy, public, commercial or proprietary interest. Therefore, once the data has been characterized across each dimension on its own, it can be helpful to bring the parameters together in order to observe interactions between variables. The relationships between the data characteristics provide additional insights that are relevant to the

[39] We acknowledge that this characterization of the relative privacy value of blood and ride share data is debatable, but offer it merely for the purposes of illustrating different approaches to thinking about data characteristics. It is not intended as dismissing the privacy interest in ride share data or location data in any way.

differentiation of trust-like data stewards from each other as well as from other data governance models.

For instance, a low privacy interest score combined with a high public interest score may not engage fiduciary duty, and so is likely not being managed by an independent third party. The data governance strategy is probably not trust-like. Weather data is a good example: there is a high public interest, combined with a low privacy interest. In most places, the data governance model employed for weather data is best described as open access. In this case, an open source approach is likely an efficient and effective solution.

Alternatively, a high value for privacy interest combined with a low value for public purpose may suggest that the risk of sharing the data outweighs the public good. In this case, a fiduciary-like duty may be engaged. However, in the absence of a public purpose, sharing the data may violate that fiduciary-like obligation. It is conceivable, but unlikely, that the data governance strategy used in this situation is trust-like.

A high privacy interest accompanied by a high commercial interest or a high proprietary interest can signal another kind of tension. When a dollar value is placed on information, it becomes an asset, to be bought and sold like any other property. Depending on the type of information, this can lead to the exploitation of individuals, and it is usually the most vulnerable people who will be most tempted to sell their privacy.[40] On the other hand, personal preference data of those with wealth is considered high value personal information, sought after by commercial actors. The lines could easily become blurred between commercial or proprietary interests and privacy interests. The high privacy interest likely engages a fiduciary-like obligation. Ideally, that means the data is being managed by an independent third party. The data governance strategy should therefore be trust-like. This situation illustrates the importance of independent stewardship to a properly functioning trust-like data governance structure. Understanding the tension between commercial interests and privacy interests is important to identify when a data governance strategy is actually functioning in a trust-like manner.

[40] Daniel Munro, “Should Tech Firms Pay People for Their Data?” (November 28, 2019), Centre for International Governance Innovation, Big Data, Platform Governance, online: cigionline < https://www.cigionline.org/articles/should-tech-firms-pay-people-their-data>.

Concurrently high scores in public interest and commercial or proprietary interests can also result in tension. If, for the same data, the privacy interest is low, the main concerns become the risk of monopolies and anti-competitive behavior where the data is not appropriately managed.[41] For example, usage rates of public transit between different locations has a high public interest, but can also have a commercial interest to ride share companies, taxis, and even potentially bicycle vendors.[42] If this data was collected only in aggregate, privacy interests would be relatively low. In this case, the general public would be the beneficiaries, and the general public would benefit from a data management strategy that prevents data monopolies. If the steward is independent, the governance model could likely be classified as trust-like. Indeed, a trust-like data steward would be an appropriate governance model.

Consider how the interactions of the data characteristics can be used to differentiate different data governance models for different types of data trusts. For the blood type example, the data scores high across all dimensions except proprietary interests, which is low. The combination of a high score in privacy interests and public interest, as mentioned above, can indicate a strong motivation to collect and share the data, as well as the need for caution to protect privacy. The concurrent high scores in commercial interests and privacy interests is cause for concern. Placing a value on information that should be private tempts the most vulnerable. Caution is needed to properly manage the commercial and privacy interests. A trust-like stewardship model would be a recommended approach in this situation.

In comparison, ride share user data has relatively lower privacy and public interest scores than blood type data, but higher commercial and proprietary interests. The lower privacy and public interest scores are still sufficient to justify a trust-like governance strategy. The proprietary and commercial interest scores also suggest that a trust-like model would be appropriate. However, the differences in their data characteristics suggest that different types of data trusts may be more or less appropriate in different circumstances.

[41] *ODI, Data trusts, supra note* 6 at 62.

[42] *Dharamshi, supra* note 14.

iv. Stakeholders

Stakeholder definition is an important design feature of a data trust. Indeed, identifying the beneficiaries, data donors (trustors), trustees and other interested parties is critical to differentiating data governance strategies and different types of data trusts. Assessing the characteristics of the data can provide valuable insight into who the stakeholders should be for a given data stewardship arrangement. Table 5 summarizes the likely stakeholders of interest based on the interests engaged by the data: privacy, public, commercial and proprietary, as discussed in the previous section. Indeed, Table 5 only provides a starting point in identifying categories of stakeholders. A thorough analysis of potential stakeholders is necessary to ensure that a data stewardship arrangement will achieve its intended purpose.

Table 5: Expected stakeholder categories based on data characteristics[43]

Data Characteristic	Individuals	Government	Businesses	Academia
Privacy	x	x		x
Public	x	x		x
Commercial			x	
Proprietary		x	x	

v. Business Models

A data trust-like governance model requires a business plan to define its operational functions, funding approaches and ownership model of the trust and of the data. There are six main organizational options for a data trust: a private not-for-profit structure, a private for-profit structure, a government agency,[44] a not-for-profit public-private partnership, a for-profit public-private partnership, and a charitable trust.[45] The right model option for a

[43] *Ibid*.

[44] *Ibid*.

[45] *Cartagena, supra* note 7 at 15.

given model will depend primarily on funding needs and opportunities, the mandate of the governance vehicle and the specific characteristics of the data to be managed.[46]

A thorough discussion of the positive and negative features of each business approach is out of scope for this report. Discussion is limited to the use of business models in differentiating trust-like structures from other types of data governance models and in classifying different types of data trusts.

vi. Technical Architecture

The defining feature of a data governance structure's technical architecture is location along the centralized versus decentralized spectrum.[47] In a centralized system, the data is collected, created, stored, and accessed in a single location. A decentralized system, in contrast, consists of a series of nodes on a network. A common set of standards may apply consistently and uniformly to each node, but the data is collected and stored using a series of individual repositories.[48]

The centralized approach facilitates the implementation of consistent and standardized collection, storing and sharing of data.[49] It is ideally suited for data sets with an interest that needs to be protected, such as privacy or commercial interests. A centralized architecture ensures universal implementation of any legal or ethical requirements.[50] On the other hand, a decentralized system is customizable for the needs of different sites, based on jurisdiction or variations in the data contained in the data set. A decentralized system may be necessary, for instance, when multiple government agencies with different policies and standards need to share data. From a security perspective, it is important not to move data until necessary. This favors a decentralized model. This also favors retention of individuals' constitutional rights.

Between the two extremes lies a hybrid approach, where there is a centralized platform or policy combined with multiple data storage nodes. Access is centralized.[51]

[46] *Ibid*.

[47] *Ibid* at 33; also *MaRS Technical Architecture, supra* note 11.

[48] *MaRS Technical Architecture, supra* note 11.

[49] *Ibid*.

[50] *Ibid*.

[51] *Ibid.*

vii. Data Access Model

The degree to which access to the data is restricted or controlled is one of the most important and defining features of a trust-like data stewardship model. There is a wide range of possibilities, based on the specific characteristics of the data and the mandate of the data steward. For instance, data consisting of highly personal and sensitive information can require strict access controls to meet legal and ethical obligations. Similarly, the existence of a fiduciary duty suggests the need for access restrictions. On the other hand, a strong public purpose that can be best satisfied by disclosure of the data may push in the other direction, favoring more open access.

Data access models used in trust-like data models lie on a spectrum. Open data exists at one extreme, and is characterized by no access restrictions or controls.[52] This model is unlikely to be utilized in a data trust (or at least not used exclusively). If there is no reason to limit access to the data, then there is little justification for adopting the structure and complexity of a data trust. At the other end of the spectrum, access to the data is severely restricted to a limited number of specific individuals or entities. Strict access controls may be required when the data is highly sensitive personal information, the data is proprietary or there is a significant commercial interest. Closed data access models can be based on data sharing agreements between parties or a marketplace, where data is bought and sold.[53]

viii. Enforcement & Remedies

The final design feature relevant to data trusts is the requirement of a system of enforcement and accompanying remedies. Enforcement is critical to accountability; to ensure that, at the system level, the data trust does what it is intended to do.[54] Remedies are required to ensure that any harms resulting from the operation of the data trust are compensated. Certainly, provincial and federal laws provide some protection. However,

[52] *Ibid*.

[53] *Ibid*.

[54] *ODI, Data trusts, supra note* 6 at 8.

federal and provincial privacy laws may leave gaps which the data trust must fill to be a realistic data management solution. A data trust may need a complementary system of enforcement and remedies.

Fiduciary or fiduciary-like obligations impose accountability on the trustee, but may not extend to others in the organization, such as individuals or businesses who are granted access to the data. For instance, users may be granted access to the data for one purpose, but may use it for another. In this case, enforcement measures may include revoking that user's access and financial penalties.[55] Additional enforcement measures should therefore be defined within the data trust framework.

Additionally, unauthorized breaches may not be due to a violation of fiduciary duty, but those harmed should still be compensated. Providing for remedies ensures that the data subjects who have entrusted their data to the trustee, either voluntarily or involuntarily, have options to correct or mitigate any harm done.

Part III: Case Studies

Having reviewed elements of common data governance models and identified those particularly essential to the functioning of data trusts, we turn to examine specific case studies in data governance that shed light on circumstances appropriate to the use of different trust models. We first undertake detailed examination of four cases before offering a briefer discussion of a number of additional data trust initiatives. We conclude with a summary of our observations.

i. In Depth Case Studies of Data Governance Initiatives

[55] *ODI, Data trusts, supra note* 6 at 38.

We begin our examination of data governance case studies with an in-depth examination of four very different initiatives: ICANN's Registration Directory Service, the Sidewalk Toronto Civic Data Trust proposal, Mastercard's Truata data governance scheme, and the First Nations Information Governance Centre's data governance initiatives.

a. Internet Corporation for Assigned Names and Numbers (ICANN) – Registration Directory Service (RDS)

This case study examines the potential for data trusts to solve problems at ICANN, the Internet Corporation for Assigned Names and Numbers, which manages the domain name system (DNS). ICANN is currently modifying its data management practices to conform to requirements of the General Data Protection Regulation (GDPR). ICANN originated under the authority of the United States Department of Commerce, but that oversight ended in 2016. ICANN is rather unique in that it is one of the first global experiments in a multistakeholder organization created to manage a finite public resource, names and numbers on the Internet. ICANN requires its contracted parties, those accredited to register names and put them into active status , to maintain a database of domain names and their owners, including contact information. ICANN is working towards a new data management approach that replaces the current public database with a two-tiered system. The first tier of information is publicly accessible. The second tier contains data that is only available to authorized requestors. Authorization is granted based on the purpose for the request and the identity of the requestor. The many stakeholders who want to continue having free access to personal and commercial data which is under ICANN control present various novel aspects of data interests for the current research project.

The key elements of this case study are as follows:

ii. Personal data has been available since ICANN was created, despite data protection law and the efforts of various data protection authorities to stop the publication of registration data. The potential of fines under the GDPR has forced a retrofit, and stakeholders who are used to free access must now justify access and figure out how to make access as simple and cost effective as possible.

- Third party access will now require accreditation and authentication of recipients. Questions arise as to who can be trusted to do this, and how far that trust goes (*i.e.*

does the accreditor guarantee only the identity of the requestor, or the validity of a request as well?)

- The registration data formerly available in the WHOIS registry is merely the tip of the iceberg, since most large registrars and registries who collect and provide access to the data have deeper relationships with registrants, often providing hosting, security, and other web services.
- This is a global ecosystem, raising interesting cross-border dataflow and enforcement issues.
- ICANN has as part of its mission a requirement to maintain the security, stability and resilience of the DNS. There are very legitimate concerns about the registration of domain names for the sole purpose of cyber-attacks and the release of malware, involving criminal organizations and sometimes complicit government actors. Managing these threats whilst protecting honest registrants is a balancing act.
- Oversight: ICANN has resisted the recommendations of data commissioners throughout its history, suggesting that reliance on complaints and investigation under data protection law might not be fruitful.
- ICANN maintains records, including recordings of public meetings and conference calls, of its activities, providing a rich resource of the deliberations and arguments surrounding this information over the past 20 years.

Background

ICANN is a non-profit corporation established in the state of California in 1998 by the US Commerce Department. The US government had been managing the assignment of domain names and IP addresses since the early days of the Internet when it evolved from the Arpanet, but as it became clear that the Internet was ready to be commercialized and made available to ordinary citizens and commercial entities and not just researchers and governments, it also became clear that it needed to be distanced from US control. The US Commerce Department issued two discussion papers in 1998, and a group of stakeholders came forward with a proposal that found favor.

When the corporation was founded, perhaps as a result of strong representations from the World Intellectual Property Organization (WIPO) and other business organizations, the registration of a domain name was made contingent on making certain details about the registration, notably name of registrant and contact information, freely available through an existing protocol known as WHOIS. It was considered extremely important that those who could establish a domain on the Internet and register a name be identifiable and traceable.   This WHOIS was made available for free and has been a source of contention since the birth of ICANN. As the Internet became a public space, the concept of public directories on the Internet became controversial, and the risks of such exposure became apparent.

Data Commissioners commented on the loss of privacy the directory represented. Spam and phishing attacks soon became more than a nuisance, and protecting email contacts and phone numbers a priority. Those exercising their fundamental rights of free speech might be targeted. Human rights defenders and educators were tracked down through WHOIS and harassed by those who wished to persecute them. While ICANN steadfastly refused to recognize that a right of privacy prevented this wholesale publication of personal data, the Registrars who served this new industry by registering domain names soon offered "privacy proxy services" whereby they protected the customer and put their own data in the WHOIS directory. Complaints about domains and websites were thus directed through the companies, and individual registrants protected.

In the meantime over the first 20 years of ICANN's existence, a vigorous debate went on inside this multi-stakeholder organization about the pros and cons of this public directory. ICANN decides on policy over the domain name system through a global multi-stakeholder community. A council comprised of elected representatives of different stakeholders, (contracted parties who profit from the registration and management of domain names, business, intellectual property, and internet service providers, and civil society as represented in the Noncommercial Stakeholders Group) have argued with several advisory groups representing end users, governments, and security experts over what ought to be done about the WHOIS.

With the coming into force of the General Data Protection Regulation in May 2018, it became clear to the Registrars and Registries who managed this personal data, that they as data controllers were at risk of being fined significant amounts if they continued to release personal data. They demanded that ICANN adjust their contracts and permit them to redact personal data from the WHOIS directory, and a new "temporary specification" was agreed in May 2018, just in time for the coming into force of the GDPR. ICANN struck a new committee, the Expedited Policy Development Process or EPDP, to find a permanent policy which respected data protection law and also provided access to third parties with legitimate interests in obtaining personal information.

This would not seem to be such an insurmountable problem, and it is true that progress has been made. However, several factors mitigate against a solution:

- **Legacy situation.** When the US Commerce Department insisted on a free and open directory, paid for by the contracted parties who were accredited to sell domain name registration and run registries for top level domains, it was favoring those actors who wanted access to name, address and contact data. Demands for more data and greater accuracy rested on this fundamental premise that third parties had a right to data, individuals and entities who operated on the internet did not have any expectation of privacy
- **Technical issues.** The WHOIS protocol soon became overloaded and new protocols were developed by the IETF, to operate on an increasingly large and complex internet. The RDAP (Registration data access) protocol permits much more complex searches and security, and was adopted by ICANN in 2019 but its implementation in compliance with policy is still in flux, while the policy is under development.
- **Growing cybercrime.** Domain names are put into operation by criminal actors for the sole purpose of distributing malware and harvesting data. Security actors need to know certain elements of data in order to stop such behaviour very quickly, and they tend to need access to all data that a contracted party has, because they are looking for patterns, not individuals.
- **Investigations into criminal behaviour.** Besides criminal activities using the resources of the Internet, regular criminal activity (theft, fraud, human trafficking, trademark abuse, etc.) occupy space on the Internet and need to be investigated by

law enforcement entities and other entities engaged in both criminal and civil investigations. Law enforcement agencies and governments claim that DNS registration data is essential for these investigations.

- **Market information.** The registration of a domain name is interesting information because domains and trademarks are important to business and innovation.
- **Facts of registration.** In order to put a domain into play on the DNS, certain information such as the IP address, the name itself, the registrar, need to be publicly available through other channels in the ecosystem, to ensure the entry is unique and the domain can be found. Those working at this level find it logical to have such elements in a public registry. The Chief Executive Officer of ICANN, while the EPDP was at work on a phase 1 report, established a group of technical experts to demonstrate that a public disclosure instrument could be built, using the RDAP protocol. This was not a multistakeholder process, it was a select group of dominant actors in the contracted parties, security, and business community. Their report was published April 20, 2019.[56] While the report is interesting from a technical perspective, it makes legal assumptions which are not necessarily sound, e.g. section 3.1.3: "ICANN org will be the sole party through which access to non-public domain name registration data is obtained in the gTLD space as part of a unified access model." By acting as the sole party as described above, ICANN org thereby reduces CPs' legal liability arising from disclosure of non-public gTLD domain name registration data." Privacy experts might disagree with these assertions.
- **Trust.** There are many reasons for the lack of trust that is apparent at ICANN among the various contenders for policy influence.

As of February 2020, ICANN has released a preliminary report of Phase 2 of the Expedited Policy Development Project to seek comments on its first imaginings of a new public access model. It is an opportune time to analyse the group's findings to see if a digital data trust could solve some of these issues, notably trust and oversight.

[56] https://www.icann.org/en/system/files/files/technical-model-access-non-public-registration-data-30apr19-en.pdf.

Deliberations on the Policy

To provide context, here is a summarized timeline of the debates about the WHOIS, with references to the documentation behind them.[57]

- First WHOIS committee in 2000
- First Task Force 2001-3
- Second Task Force2003-4
- Combined Task Force 2004-5
- WHOIS Review Team 2010-12[58]
- Experts Working Group 2013-14[59]
- Transition to Thick Registries 2011-13
- WHOIS conflicts with law implementation 2015-16
- Privacy Proxy Services Accreditation 2014-2015
- RDS PDP, 2015-18[60]
- RDS Review Team 2 2016-2019[61]
- Expedited Policy Development Process (EPDP) 2018-2020 (anticipated final report June 2020)[62]

There are many other reports and studies, mostly examining issues important to those seeking data access, such as on accuracy and the use of privacy proxy services. Even a brief scan of the material available will provide evidence of how intractable this issue has been over the years. Our central question is, how could a data trust help resolve data access and management issues, so in order to unpack the answers from this vast collection of arguments, we have isolated a few key themes and will explore the possibilities using that framework.

[57] See the WHOIS history for an overview, https://whois.icann.org/en/history-whois

[58] https://www.icann.org/resources/pages/whois-rt-final-report-2012-05-11-en

[59] https://community.icann.org/pages/viewpage.action?pageId=40175189

[60] https://community.icann.org/display/gTLDRDS/Next-eneration+gTLD+Registration+Directory+Services+to+Replace+Whois

[61] https://www.icann.org/resources/reviews/specific-reviews/whoisx

[62] https://community.icann.org/display/EOTSFGRD

Controllership and ownership of data

Over the many years of debate, ICANN had not acknowledged its controllership (using the term of the GDPR) of the data of registrants. While the contracted parties, the registrars and registries collect, use and disclose the data, ICANN sets the policy through its multistakeholder processes, policy which ultimately must be approved by the ICANN Board, in consultation with the Government Advisory Committee (GAC). ICANN negotiates the contracts with the contracted parties, and its staff in the Global Domains Division enforce those contracts. ICANN can remove the accreditation of these entities, thus putting them out of business instantly, and they are the party that has access to all the registrant data which the contracts require contracted parties to place in escrow.

In 2018 the EPDP sent ICANN a request for their input[63] which asks many good questions about the legal situation they are in. The Chief Executive Officer had been consulting the Belgian Data Protection Authority and the European Commission for advice on whether they can take on the liability risk for disclosure in a WHOIS replacement registry, but ICANN has not stated a final position as yet. A January 2019 legal memo[64] suggests that they have a role as an independent controller. Various members of the EPDP on the other hand believe that they are co-controllers, and therefore they need to negotiate a full co-controller arrangement with the contracted parties. The EPDP is expected to sunset in June 2020. Clarity about this issue of control is usually a first step in a privacy impact assessment, and it should have been established years ago.

There can be no doubt that the registrars, who maintain a customer relationship with the registrant, are controllers in their own right, because they gather data about their customers, including financial data. They also, in large part, make money in selling other

[63] This memo from a small team struck to examine this issue illustrates the confusion felt among the team regarding ICANN's role and controllership liabilities, as of December 2018; <unfortunately the questions have not yet been answered as of March 30 2020. https://community.icann.org/display/EOTSFGRD/2018-12-10+EPDP+Small+Team+--+to+further+discuss+controller+issues?preview=%2F100533806%2F100533808%2FJoint+Controllership+Questions+-+7Dec.pdf>.

[64] In response to these questions, ICANN's legal team prepared a draft paper outlining possible positions that they might take. The 20 page document, entitled ICANN org Response to Request from the EPDP Team for Additional Information re: Independent Controllers under the GDPR, Working draft for EPDP team review and discussion 14-January 2019 arrived shortly before the planned publication of the Phase 1 report. < https://mm.icann.org/pipermail/gnso-epdp-team/2019-January/001220.html >.

services, often acting as ISPS or web developers, and they also use networks of resellers (such as web designers). Maps describing these fuller data relationships have not been examined by the ICANN working groups, as this data would be considered out of scope. Big operators may share their data with, for instance, security service organizations, in order to outsource the significant security management issues they face. Some operators and registries operate the back ends for many other entities active in the DNS, including the country codes (e.g. .ca, .fr, .uk) and that activity also includes personal information. So it must be clear that in the current ICANN policy development process, all we are arguing about is the information ICANN has historically required contracted parties to collect, use, and disclose in a public registry.

This lack of transparency about the entire ecosystem and the control issues behind it does not inspire trust, but this is probably a fairly common situation in use cases for data trusts...one of the key reasons for contemplating the establishment of a data trust is to deal with power imbalances, and resultant lack of cooperation and transparency.

Rights of Registrants and the Legal Persons issue

A persistent problem that also has not yet been resolved, is the distinction between individuals, entitled to data protection under the GDPR, and legal persons (a definition that includes a variety of entities) who are not strictly speaking entitled to data protection under data protection law. Their employees, however, whose contact information might be released in a registration database, could be protected by data protection law, depending on local law. Informal organizations, small entrepreneurs, NGOS and religious organizations, might also be protected either under data protection law or other law, notably under Constitutional or Charter protection. Freedom of religion and political speech is usually protected, but organizations who register websites to engage in speech or association activities may be under threat, and their contact information needs to be protected. Commercial entities may have valid reasons, notably competitive or innovation reasons, to protect their identity in the context of names they have registered. In this respect, many intellectual property holders and commercial entities use lawyers as proxies to register their domain names, thus effectively insulating themselves from transparency. This is not something small entrepreneurs can usually afford to do.

Arguments have been made at ICANN that entities just need to be asked to self-identify as an individual (claiming privacy rights) or a legal person. In a global, multi-lingual environment with a wide variety of legal systems, and registrants unfamiliar with their potential data protection or confidentiality rights, this is a risk for contracted parties who rely on self-identification. It is also rather useless in dealing with the threat of registration by criminals, since they are unlikely to provide accurate information. However, the issue appears to be doomed to never-ending argument, as the sides have been clearly identified and the trenches dug for a long war; independent data trusts might be a useful forum to reach more independent, impartial decisions on such matters, by including more independent and neutral voices in the decision making.

Automation

There is a great deal of pressure during current discussions, to automate the request, evaluation and data release process. The Noncommercial Stakeholders Group is extremely skeptical about an end to end automation procedure. Requestors, however, refuse to stop pushing for it, despite the fact that contracted parties have been blunt about their refusing to agree to it as long as they hold the liability for disclosure. Tucows is a Canadian registrar that is now the second largest registrar of generic top level domains globally, and they have not only been among the first to implement the new RDAP protocol (in May 2018) but have been blogging about issues with requests for disclosure.[65] While they do not address automated disclosure specifically in the latest statistical report, it is clear that there is no way to evaluate trademark claims effectively in an automated fashion. It is also clear that at current volumes of only 2864 requests since May 2018, building an automated system that would be even remotely useful in performing a balancing test as required by section 6(1)(f) of the General Data Protection Regulation would be too costly to justify.

[65] https://opensrs.com/blog/2020/03/privacy-and-lawful-access-to-personal-data-at-tucows/.

Accreditation and Authentication

A key function in any disclosure of information to third parties under data protection regimes, is ascertaining effectively that the entity requesting the data is indeed the entity claimed, that the request is valid and contains all the required information and rationale, and that the request is for a legitimate purpose. Given the vast potential scope of requestors in a global domain name system, it is clear that different groups of requestors will have to be accredited in different ways. Accreditation also cannot be taken as synonymous with authorization; although there are many systems at the moment where groups of actors who trust one another share information informally, particularly threat information, this is a system that is ripe for more rigor under data protection law.

It has been suggested in these discussions at ICANN that certain cybersecurity data sharing organizations such as the Antiphishing Working Group (APWG) can accredit trusted actors according to certain standards, and that m embers of bar associations can accredit lawyers who act for trademark and copyright interests. Law enforcement agencies, working through the Public Safety Working Group of the Government Advisory Committee, have proposed a way for them to accredit members of law enforcement. While it is useful to be able to establish that a requestor is indeed a member in good standing of a professional association or recognized organization, and a token issued by an accrediting organization could suffice to present a request, the requests themselves would still have to be evaluated. The Technical Study Group, in their proposal for a unified access model, have proposed methods for how this could work technically, but the infrastructure for actually vetting and accrediting members, and the mechanisms for withdrawal of credentials have not been worked out. This problem is common to most use cases for data trusts as well, depending on the breadth and scope of the potential data requestors.

Jurisdiction and Transborder Dataflow

There has been very little discussion of the jurisdictional and transborder dataflow (TBDF) issues in the ICANN context, but these hardy perennials promise to raise difficult issues as we move towards implementation. To begin with, actors in the ICANN space tend to be scattered around the globe and active in many jurisdictions. Different processing activities may take place in different jurisdictions, depending on volume and cost. While a single disclosure request interface is envisaged, potentially under the control of ICANN, this processing activity might be outsourced, and there is no reason why it could not be regionally dispersed. This raises the questions of which data protection law applies, and what happens if a registrant resides in a jurisdiction without a data protection law, or a data controller (e.g. registrar) is established in such a jurisdiction. While we tend to think of this global registry as a single entity, a database, it is not and never has been. The RDAP protocol is an access protocol, the data resides on servers at the registrars or its agents, and therefore it travels, presumably back through the disclosure request mechanism and out to the requestor, whichever jurisdiction that entity resides in. Similarly, when data is escrowed, if it travels to the United States to the escrow agent recommended by ICANN (Iron Mountain) it would reside in the United States, and the registrants' constitutional protections in the matter of criminal procedures could not be guaranteed in the event that disclosure requests (subpoenas or other lawful instruments) came to the escrow agent. It should be noted that provisions for responding to such requests are anticipated in section 4.2 (Confidentiality) of the Registration Data Escrow Agreement (https://www.icann.org/en/system/files/files/iron-mountain-rde-template-18jul18-en.pdf).

Noncommercial stakeholders have pressed for a universal policy mapped to the GDPR, as ICANN ought to be administering the DNS in a way which respects human rights, and privacy is a human right which is recognized in most national Charters of rights, and is now guaranteed to be protected by law in the EU. However, national jurisdiction must be recognized, and it may be that the policy will be superseded by national laws in some instances, particularly those where free speech and freedom of association are not guaranteed. This should not mitigate against setting a high standard in a universal policy.

With respect to transborder dataflow, the current situation regarding where personal data might potentially travel is not well explained. Individuals in particular who are not aware of the extensive use of and access to the traditional WHOIS data may not be capable of assessing their risks from transborder dataflow .

Summary

The ICANN use case is instructive in our examination of data management challenges, because as a multistakeholder organization with extensive representation of customers or end users, it ought to have the following clear advantages in settling data rights issues:

- There is a great deal of transparency, records of official deliberations on policy are open and archived.
- There are opportunities for public comment.
- There is multistakeholder representation throughout ICANN's committees, the Generic Names Supporting Organizations' Council, and even the ICANN Board.
- The organization is global in scope
- The transborder dataflow issues permeate all aspects of the DNS ecosystem, and the system works despite potential jurisdictional issues.

However, from the perspective of civil society actors pushing for data protection and human rights in the NCSG, there are a number of failures in ICANN's management of this data rights issue:

iii. Intellectual property and business trademark owners have had a dominant power relationship with ICANN and its components since the original discussion papers were tabled in 1997.

iv. ICANN itself is steeped in a tradition of ignoring the protests of data commissioners regarding their treatment of personal data (see Perrin, S. The Struggle for WHOIS Privacy: Understanding the Standoff Between ICANN and the World's Data Protection Authorities) and it seems unlikely they will ever seriously embrace the concept that registrants have rights.

v. There is an alarming lack of transparency and fair play with respect to how the institution sees itself in terms of legal responsibility, controllership, accountability, and purpose of data processing.

vi. There is a lack of trust, particularly with respect to this policy issue, among the stakeholders at ICANN.

This prompts us to ask if a data trust could provide a more fair management structure and policy procedures for handling this data.

For the purpose of this report, the data management system of interest is the proposed replacement for the temporary specification 2018 as described in the EPDP Phase 1 Final Report, accepted by the ICANN Board in March 2019.[66] Brief descriptions of the design features of the ICANN data management system are included in Table 6.

Table 6: ICANN RDS Case Study Summary

Design Feature	Brief Description	
	Restricted Data	General Access Data
Mandate/Purpose	(i) provide contact information of registrants for the purpose of resolving issues & complaints, technical and otherwise, and also to facilitate business transactions; (ii) protect consumers; (iii) facilitate cybersecurity investigations; (iv) facilitate the protection of intellectual property; (v) academic research; (vi) inform general public[67]	
Independent stewardship	"multi-stakeholder" model; independent as of March, 2016; SSAD or Simplified Standardized Access model to be operated by a non-governmental third party as yet to be determined	
Fiduciary-like Obligation	Some, but limited to ICANN's narrow remit	
Data	Registrant name, registrant address, registrant phone number[68]	Domain name, abuse contact ID, registrant country,[69]
Privacy interests	Yellow (5)	Green (1)
Public interests	Yellow-red (7)	Yellow-red (7)
Commercial interests	Green-yellow (3)	Green-yellow (3)
Proprietary (IP) interests	Green (1)	Green (1)
Stakeholders	Domain name registrants & owners, general public, individual internet users, business-related internet users, law enforcement agents, intellectual property and trademark owners, security practitioners	
Trustor / Data Donors	Domain name registrants	
Trustee	ICANN	
Beneficiaries	All stakeholders are beneficiaries of a well-managed DNS, ; RDS data is essential for good management of the system	

[66] ICANN, supra note 49.

[67] *Ibid* at 8-9.

[68] *Ibid* at Annex E. List is not exhaustive, but is intended only to provide examples of restricted data available to authorized requestors.

[69] *Ibid*. List is not exhaustive, but is intended only to provide examples of basic data that is publicly available.

Business Model	Private not-for-profit	
Cost	Shared by Registrars and ICANN; costs will be passed down to registrants unless requestors are charged an access fee [70]	
Technical Architecture	Hybrid system with centralized rules and standards and common interface for handling data access queries. Access is centralized. Data collection, storage and sharing using distributed system involving individual Registrars in various jurisdictions around the world.[71]	
Data Access Model	Restricted data = requestors granted authorization based on permissible purpose	Basic (anonymized) data = open access
Enforcement & remedies	Penalties and other remedies such as loss of accreditation for inappropriate data use; contractual clauses used between ICANN and localized RDS providers to ensure compliance with policy	

The ICANN RDS has the primary design features of independence, and so should be categorized as a data trust. As of 2016, ICANN is independent of the authority of the government of the United States or any other country. Some level of fiduciary duty is engaged because the domain name registrants are at the mercy of ICANN in terms of whether or not their data will be shared, with whom, and for what purpose. The data itself is personal information, and in some cases, addresses and names could be considered sensitive information. However, this case study argues that the engagement of fiduciary-like obligations has failed. This prompts us to ask if a data trust could provide a more fair management structure and policy procedures for handling this data.

b. Sidewalk Toronto

A unique feature of the data trust proposed for the Sidewalk Toronto project is that the trust is not just for the management of data, but includes the entire "digital layer". This would include elements like the sensors, code and physical storage facilities, in addition to the data, standards, and the system interface.[72] However, for the purpose of this report, only the data elements will be considered.

Two use case studies were developed in order to prototype the proposed data trust. The purpose of the first use case was to define a data management system for mid- to long-range transportation planning for the Greater Toronto Area (GTA). The data of interest

[70] *Ibid* at 117-118.

[71] *Ibid* at 109.

[72] *MaRS, Civic Trust, supra* note 9.

included personal mobility, such as public transit user data, ride-sharing user data and navigational route mapping information.[73] (It is assumed that the latter set of data refers to the navigational data used by personal vehicles.) Uber, the Toronto Transit Commission (TTC), and Google would be primary data contributors in such a scenario.

The second use case was in the area of improving access to public health data. It is important to note that there is a significant upfront legislative challenge related to sharing health data, potentially requiring legislative amendments in order to permit broader sharing.[74] That said, the use case proposes the required changes and then evaluates a system assuming those changes have been made. The data set of interest is that currently held by the Institute for Clinical Evaluative Services (ICES), and includes the health records for all Ontarians qualifying for universal health coverage. The data is highly personal and highly sensitive, consisting of prescriptions, hospitals stays, emergency room visits, homecare, and narcotics use. The database also contains census, social services and immigration information. When the data is used for research, de-identification attempts are made by detaching the records from the identity of the person, and instead using a numerical identifier.[75] The stated purposes of increasing the access of researchers to the database are to further economic development and expand medical research. Commercial, privacy and public interests are all engaged to a high degree.

The proposed data management system for each use case is summarized in Table 7.

Table 7: Sidewalk Toronto Case Study Summaries

Design Feature	**Brief Description**	
	Case Study 1: Personal Mobility Data, "Civic Data Trust"[76]	**Case Study 2: ICES Health Case Study, "Data Safe Haven"**[77]
Mandate/Purpose	Ensure appropriate use of digital assets in the best interest of the public; mid- to long-range transportation planning across	Enhancing and improving the health of Ontarians;[79] "further economic development as it relates to access to health data";[80] innovation

[73] *Dharamshi, supra* note 10 at 11-12, 51.

[74] *Cartagena, supra* note 7 at 7.

[75] *Ibid* at 10-11.

[76] *Dharamshi, supra* note 10.

[77] *Cartagena, supra* note 7.

[79] *Cartagena, supra* note 7 at 15.

[80] *Ibid* at 22.

	GTA; a more collaborative city, increased innovation, reduced barrier for start-ups, improved mobility[78]	
Independent Stewardship	Yes, Board of Directors liable for breach of duty[81]	Yes
Fiduciary-like Obligation	Yes	yes
Data	Public transit user data, ride-sharing user data, driving route mapping[82]	Health records with directly identifying information removed (hospital stays, emergency room visits, prescriptions, etc), immigration records, census information [83]
Privacy interests	Yellow-red (7)	Red (10)
Public interests	Yellow-red (7)	Red (10)
Commercial interests	Red (10)	Red (10)
Proprietary (IP) interests	Red (10)	Green (1)
Stakeholders	Government (TTC, transportation policy dept), businesses (Uber, Google, TTC), individuals, shareholders	Government, individuals, researchers, businesses
Trustor / Data Donors	Google, Uber, TTC, users of Google, Uber, TTC	Ontario residents with universal health care coverage
Trustee	Governing Board of Directors; independent from government	ICES
Beneficiaries	Residents of the GTA, Google, Uber, other businesses with data access, government	Not clear, but includes researchers and patients, businesses
Business Model	Private not-for-profit corporation[84]	Charitable trust; not-for-profit[85]
Cost	Funding mechanism options:[86] (i) Endowment with data sharing transaction fee, tiered based on size of corporation or academic / government affiliation	Not discussed; assumed publicly funded and some costs can be recovered using license fees, etc.

[78] *Dharamshi, supra* note 10 at 11.

[81] *Dharamshi, supra* note 10 at 39.

[82] *Dharamshi, supra* note 10 at 11-12, 51.

[83] *Ibid* at 10-11.

[84] Ibid at 38-41

[85] *Cartagena, supra* note 7 at 16-17.

[86] *Dharamshi, supra* note 10 at 42

	(ii) Investment by smart city real estate developer	
Technical Architecture	Hybrid--decentralized model with centralized policies and standards and central access[87]	Hybrid; ICES would act as the Data Safe Haven, with links to other health data repositories, eg Toronto Public Health;[88] policies would be centralized
Data Access Model	Limited access based on transaction fees; data subjects have access to grant consent for use of their data for specific purposes	Access to directly identifiable data restricted to limited number of individuals; data released to researchers following Privacy Impact Assessment, if allowed under legislation
Enforcement & Remedies	Oversight by federal Privacy Commissioner,[89] but no independent system for enforcement and provision of remedies	Charitable trust overseen by the Attorney General and the Canada Revenue Agency; specific acts covering privacy of health data

The civic digital trust proposed by MaRS in the context of personal mobility qualifies as a data trust. Certainly there is a public purpose, and an independent Board of Directors making decisions regarding the collection, use and distribution of data. Access must be approved by the Board based on intended use. Fiduciary duty is explicitly included in the design of the data trust because the Board of Directors are liable for breach of duty.

The Data Safe Haven proposed by ICES also qualifies as a trust. ICES already acts as a steward of health and other data and shares what it is legally permitted to with researchers. Although it conducts research, it is independent from the other researchers who would request access to the data. The health data is highly sensitive and the data donors are also beneficiaries of medical research, so a fiduciary-like obligation will attach.

c. Truata

Trūata is a commercial third-party data privacy and analytics system that is structured as a legal trust. Trūata was launched in March 2018 as a partnership between Mastercard, a financial services firm, and IBM, an information technology firm, and works with California-based artificial intelligence company C3.ai as its software and technology

[87] Ibid at 50.

[88] *Cartagena, supra* note 7 at 37.

[89] *Dharamshi, supra* note 10 at 39.

developer.[90] Trūata serves as a central data repository for clients, who are other data collecting companies; Mastercard is both a creator and client of Trūata. They combine proprietary de-identification technology with legal and organisational safeguards in an attempt to provide privacy protections for data and allow for aggregation of ostensibly anonymized datasets across several companies within or across industries, in ways that purport to be either outside the scope of, or compliant with, data protection and privacy laws. It also segregates legal control of these datasets from other data holdings under participating companies' control, a formal separation that bolsters legal compliance with some data protection requirements.[91] The trust has been established under Irish law and is headquartered in Dublin;[92] as part of this trust structure, Trūata operates independently of Mastercard and IBM.

Use cases provided on the Trūata website span the automotive, financial, retail and telecommunications industries. The clients that Trūata attracts are those that have large amounts of data and wish to use it for secondary purposes. For instance, the automotive use case involves processing driver behavior data with a stated goal of helping manufacturers improve safety, customize user experience, provide traffic planning in the context of connected cars, or provide data for developing environmental policies. Although it is marketed to businesses, there is nothing precluding Trūata from taking on public sector clients.

The financial use case involves Mastercard's "merchant business intelligence solution" which merchants use to extract secondary insights from transaction data; Mastercard claims that this "competitive advantage" is worth approximately $30 million in

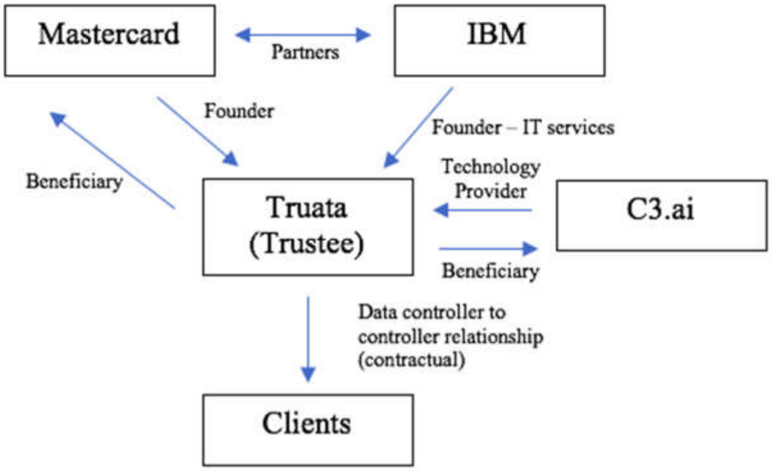
[90] C3.ai, "Trūata Selects C3.ai as Partner in GDPR Service Offering," press release, 23 May 2018, available at: https://c3.ai/Trūata-selects-c3iot-gdpr-service-offering/.

[91] See, for example, Case C-582/14, Breyer v Bundesrepublik Deutschalnd, October 19, 2016, (CJEU 2nd Chamber),http://curia.europa.eu/juris/document/document.jsf?text=&docid=184668&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=1130557

[92] Ireland is headquarters for many big technology firms; the data protection authorities there have been criticized for less rigorous investigations or enforcement actions under the GDPR in the popular press, such as for example, Nicholas Vinocur, "How one country blocks the world on data privacy," Politico April 24, 2019, available at https://www.politico.com/story/2019/04/24/ireland-data-privacy-1270123.

annual revenue, a revenue stream Trūata was created to protect.[93] The GDPR appropriately limits the ability of merchants to continue to use personal data in such a way and sets a high bar requiring rigorous attempts to render data subjects "not or no longer identifiable" for data to fall outside of the data protection regime.[94] Trūata attempts to achieve that bar; thus far, its compliance has not been challenged in court.

Stakeholders



As an entity held in trust, Trūata operates independently of Mastercard and IBM. Key actors on Trūata's core management team previously worked for Mastercard. Their Chief Privacy Officer is a privacy lawyer by training and she is responsible for carrying out tasks as described by GDPR articles 38 and 39.[95] Their in-house counsel, who also moved from Mastercard, oversees compliance with regard to Trūata's trust obligations. Trūata's Board Members include senior practitioners from financial services, academia, and private sector data privacy experts. The Board includes one former Mastercard executive and a current IBM General Manager.[96] Mastercard and IBM are beneficiaries of the Trūata trust

[93] Trūata Mastercard, online: <https://www.Trūata.com/2019/08/07/mastercard-business-intelligence-platform/> [*Trūata MC*].

[94] EU general data protection regulation 2016/679 (GDPR) Recital 26.

[95] Trūata, "Trūata appoints Chief Privacy Officer to its core team," press release, 18 June 2019, available at: https://www.Trūata.com/2018/06/18/Trūata-appoints-chief-privacy-officer-to-its-core-team/

[96] Trūata, "About Us – Board Members," available at: https://www.Trūata.com/about-us/board-members/.

deed.[97] This means that as shareholders, they are not allowed to have majority share in the company.

Trūata and the General Data Protection Regulation

Trūata is an interesting demonstration of the role that new regulation can play in spurring attempts at privacy innovation in the private sector where there might not otherwise be incentives to do so. Mastercard is explicit that it conceived of Trūata upon realizing it would need to change its privacy policies and practices in light of the GDPR, aiming to capitalize on the gap between existing company practices and the newly enforced rules. Trūata offers a service—as an independent data controller responsible for attempts to comply with strict standards of anonymization—that was previously unnecessary and is therefore unique in its specific applicability to the regulatory landscape of Europe.

Rather than regulating data directly, the GDPR instead relies on "enforced self-regulation" by placing emphasis on risk assessment and mitigation by the data controllers.[98] By acting as a data controller,[99] Trūata takes on the responsibility of managing and mitigating data privacy risks while still allowing its clients who wish to engage in particular kinds of data analytics using customer data to do so. Trūata, however, explicitly claims to be an example of privacy by design. As Trūata's CPO Aoife Sexton stated in an interview for this research, "privacy is essentially in the DNA of the company and has been there since day one [...] behind the rows of data are information about people and that should not be forgotten in how we are designing big data analytics."[100] Trūata has been looked to as a case study in a number of forums about business solutions to data privacy, in part because its trust structure is unique in the industry. For example, the United States

[97] Financial Times, "MasterCard and IBM to set up European 'data trust'," 15 March 2018, available at: https://www.ft.com/content/576171dc-27ab-11e8-b27e-cc62a39d57a0.

[98] Milda MaCenaite, "The "Riskification" of European Data Protection Law through a Two-Fold Shift," European Journal of Risk Regulation 8(506) (September 2017).

[99] Trūata, "Protecting Privacy, Powering Results," brochure, available at: https://www.Trūata.com/wordpress/wp-content/uploads/2019/03/Trūata-Anonymization-Solution-brochure.pdf

[100] Interview of Aoife Sexton by Kristen Kephalas, 24 March 2020.

Federal Trade Commission invited Aoife Sexton of Trūata to speak at a hearing about innovation in consumer privacy. [110]

Trūata's Legal Structure

Trūata embodies the characteristics of a data trust to varying extents: independent stewardship of data, a fiduciary-like obligation, and in some cases, a public purpose. Trūata is different from other models that might claim to be "data trusts" because it is a company that is held in trust. Connor Manning, a lawyer from Arthur Cox who helped create Trūata's trust structure described the legal relationship as "a corporate structure with a trust on top". [101] Trūata operates independently of its founders and is legally restricted by its trust deed. This legal separation gives Trūata the independent character that is integral to a data trust. According to Trūata's CPO Aoife Sexton, the trust deed constrains Trūata's actions. For example, she explains, "we [can] only take the data for the purposes of independent anonymization [...] the data is not ours, the data belongs to the customer. We must not attempt to re-identify the data." [102] The value added by the legal trust structure is that beyond Trūata's technical attempts to protect data privacy, Trūata's trust structure creates additional structural and legal safeguards of privacy.

As a trustee, Trūata has compounded legal obligations between trust laws and the GDPR to protect the data that it is entrusted with. It follows then that third parties are prevented from paying for access to Trūata's clients' data. Clients retain ownership of their data. Trūata describes its role as being a data controller, as opposed to a data processor. According to Article 24(1) of the GDPR, the controller "shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation".[103] A data processor

[101] Financial Times, "Data trusts raise questions on privacy and governance," 12 September 2019, available at:
https://www.ft.com/content/a683b8e4-a3ef-11e9-a282-2df48f366f7d

[102] Interview of Aoife Sexton by Kristen Kephalas, 24 March 2020.

[103] European Union (EU), General Data Protection Regulation (EU) 2016/679 (GDPR), article 24, available at: https://gdpr-info.eu/art-24-gdpr/

conversely, is subject to the direction of another entity.[104] As a controller, Trūata makes decisions about data anonymization processes and analytics independently of their clients. The client relationship is governed by contract, so any liability will arise from contract. However, the responsibilities of Trūata as a company held in trust that is restricted to its governing document and as a data controller under the GDPR create multiple layers of legal accountability for its practices. As the party in control of how and why data is processed, Trūata has the most liability when it comes to protecting customer data privacy. This is part of Trūata's appeal because it takes on the legal responsibility under the GDPR for its clients. Playing the role of data controller and bearing legal responsibility for data privacy leaves clients' data vulnerable to Trūata; there is the potential for misuse, and they must trust Trūata will control their data responsibly. This creates a fiduciary-like obligation. However, the obligation is not fiduciary in the strict common law sense as the client relationship is a contractual one.

The question of Trūata's public interest purpose is complicated because it is private and profit-driven and so cannot be truly publicly interested. This highlights a difficulty in categorizing a commercially-motivated data trust within a data stewardship framework; while public interest in personal transactional data, such as that processed by Mastercard, is high—what people buy, how much they spend, where they shop, these all are revealing details— it is the company who collected that data, not the data subjects, who are the formal beneficiaries of the trust relationship. While it is an innovative business model, there is a disjunction between the interests served, that of the companies collecting data, and those of the data subjects. This is particularly the case with regards to driver information, where the data is explicitly processed with an aim to monetizing it in the context of usage-based insurance,[105] attempting to avoid issues of user consent in a situation where drivers are perfectly aware that such data collection may contribute to changes in their insurance premiums. While the argument is that data subjects' privacy is

[104] European Union (EU), General Data Protection Regulation (EU) 2016/679 (GDPR), article 28, available at: https://gdpr-info.eu/art-24-gdpr/

[105] Trūata Automotive Use Case, Available at https://www.truata.com/industry-use-cases/automotive/#driver

enhanced because secondary data uses are only undertaken after data is passed through Trūata, individual data subjects would have little if any recourse to remedies directly from the trust, and are also unlikely to ever know whether their information was used in analytics that resulted in an effect on their lives. While Trūata has a "Customer Charter" that outlines the ethical use of data, such as not using data to undermine the democratic process and not to use it in a discriminatory manner, it does not form part of the legal contract.[106]

Table 8: Truata Data Trust Case Study Summary for Automotive and Mastercard Use Cases

Design Feature	**Brief Description**	
	Automotive Use Case	**Mastercard Use Case**
Mandate/Purpos e	<u>Commercial purpose</u>: maximize profit, in compliance with GDPR; <u>Public purpose</u>: traffic management of connected cars; on-demand car sharing; environmental policies[107]	<u>Commercial purpose</u>: Maximize profit, in compliance with GDPR[108]
Independent Stewardship	Yes, as the data controller, but holds no proprietary rights to the data.	Yes, as the data controller, but holds no proprietary rights to the data.
Fiduciary-like Obligation	Fiduciary-like obligation; data controller makes independent decisions about data	Fiduciary-like obligation; data controller makes independent decisions about data
Data	Driver and user behavior data, vehicle location, vehicle performance data that has been put through an anonymization process[109]	Mastercard transaction data that has been put through an anonymization process[110]
Privacy interests	Yellow	Yellow
Public interests	Yellow-red	Yellow
Commercial interests	Red	Red
Proprietary (IP) interests	Red	Red

[106] Interview with Aoife Sexton by Kristen Kephalas, 24 March 2020.

[107] *Truata Auto, supra* note 97.

[108] *Truata MC*, supra note 98.

[109] *Truata Auto, supra* note 97.

[110] *Truata MC*, supra note 98.

Stakeholders	Businesses, individuals, government	Businesses, individuals
Trustor	Businesses	Businesses
Trustee	Trūata	Trūata
Beneficiaries	Businesses, individuals	Businesses
Business Model	Both for-profit and not-for-profit	Private for-profit
Ownership Model	Client owns data (bilateral relationship); data could be shared among companies and government in smart city context if they contract for it	Merchant owns data and data trust, however as a company held in trust, merchant is precluded from majority shareholding and therefore cannot override decision making
Technical Architecture	Centralized or decentralized	Centralized or decentralized
Data Access Model	Not clear what flexibility exists to customize data access	Access is limited to merchants, Trūata and Mastercard
Enforcement & Remedies	Privacy laws apply to data collected by clients. Enforcement may vary by jurisdiction depending on the degree to which anonymization is recognized in statute and whether the statute applies subsequent to anonymization or to re-identification. No individual remedies directly from Trūata.	Privacy laws apply to data collected by clients. Enforcement may vary by jurisdiction depending on the degree to which anonymization is recognized in statute and whether the statute applies subsequent to anonymization or to re-identification. No individual remedies directly from Trūata.

In summary, Trūata as an entity is not a data trust; it is a for-profit business held in trust that offers a data management platform, although it could be used as an element of a data trust. Trūata maintains a data controller-to-controller relationship with its clients, meaning that it takes legal responsibility for all decisions related to data processes, making it the independent steward of the data for this purpose. Trūata maintains no proprietary rights over the data. This creates a fiduciary-like obligation in that clients must give up control over their data. As a company that is held in trust, Trūata's operations are restricted to the trust deed that governs it. This adds a layer of legal and structural protection over privacy interests. Depending on whose Trūata's client, there may be different levels of public interest engaged, but there is no direct accountability to data subjects. This commercially-motivated model effectively highlights the growing gap between issues of technical legal compliance, and issues of public trust and social license for data uses.

d. First Nations Information Governance Centre (FNIGC) – First Nations Regional Health Survey (FNRHS)

The First Nations Information Governance Centre is a non-profit organization that undertakes various data collection efforts within First Nations Communities.[111] One such initiative is the First Nations Regional Health Survey, which includes data sets collected during four surveys to date. The survey was the first implementation of the Ownership, Control, Access, Possession (OCAP) principles, which defines the governance standard for First Nations information..[112] The application of OCAP means that First Nations communities are in full control over the collection, use and distribution of their own data. Government access is controlled by a limited license.[113]

OCAP was developed as a response to negative research impacts drawn from data collected on First Nations without their consent or for their benefit. Research on First Nations is often undertaken by external individuals and institutions such as federal and provincial governments, universities, media outlets, and private companies. These agencies have produced and published research to the detriment of First Nations health, beliefs, and economic wellbeing. Research abuses include misuse of health information, collecting blood samples and genetic material and health data—without consent or exceeding consent, by using data approved for one purpose for another without permission. The results of these practices have been to exhibit, entrench, and exacerbate stereotypes of First Nations that further economic alienation and dependency on state services.

The RHS Cultural Framework published in 2005 establishes the criteria for First Nation objectives of health research. Defining Indigenous Intelligence in contrast to Euro-Western ideology, a critique of traditional approaches to health research lends way to a holistic and culturally meaningful framework for health researchers. Starting by defining a positive vision for a healthy Indigenous community, the RHS establishes physical, mental,

[111] First Nations Information Governance Centre, online: <https://fnigc.ca/about-fnigc/frequently-asked-questions.html> [*FNIGC FAQ*].

[112] First Nations Information Governance Centre, "Ownership, Control, Access and Possession (OCAP): The Path to First Nations Information Governance" (May 23, 2014), online: <https://fnigc.ca/sites/default/files/docs/ocap_path_to_fn_information_governance_en_final.pdf> at 4, 14 [*OCAP*].

[113] *Ibid* at 15.

emotional, way of living, harmony, relationship, sovereignty, culture, and environmental health goals. Visioning is the first of four steps to undertaking health research relevant for First Nations and is followed by relating, analyzing and building. Relating measures reality against the "ideal standard" of the vision and collects data in a culturally sensitive way with a whole picture of indigenous health. Analysis and interpretation interprets data to enlighten community understanding against its vision and standard of health. Building a healthy First Nations community provides direction towards improvements needed to actualize the vision for indigenous health.

The First Nations Regional Health Survey which originated with the Assembly of First Nations Chiefs Committee on Health mandated in 1994 that a First Nations health survey be implemented every four years across Canada.

The governance structure envisioned is complex:

- First Nations Information Governance Centre (FNIGC) is federally incorporated (Est 2010), mandated by the Assembly of First Nations Special Chiefs Assembly and governed by a Board of Directors, appointed by each First Nation Region.
- FNIGC governs data; Health Canada is a partner.
- RHS Cultural Framework lays out the indigenous worldview with which researchers should consider data collection, purpose and presentation "in a manner that is meaningful to First Nations peoples and communities."
- RHS Code of Research Ethics sets principles, policies and procedures for the RHS survey.
- Process for access requires approval by the respective level of First Nation government:
    - National level data must be approved by the national governing body;
    - Regional level data must be authorized by the regional First Nations organizations;
    - Community-level data requires direct consent of the First Nation community involved.

The data collected by the survey has expanded over time: The Regional Health Survey (RHS) pilot (1997) included regional-specific data that covered HIV/AIDS, suicide, mental health, residential schools, alcohol, drug use, sexual activity. The survey expanded in 2010 to include migration, food security, violence, care giving, depression, gambling, and health utilities index. The children's survey included questions on community wellness and immunization. Community participation is sought for both the design of data collection and analysis of results. Data is collected from approximately 30,000 individuals in 250 First Nation communities, involving 10 participating First Nations northern and on-reserve regions and the Yukon and Northwest Territories (in Phase 2). Inuit communities withdrew from participation.

Ownership of the data is central to the approach adopted by the First Nations Information Governance Centre. The survey is the first national survey fully owned, controlled and stewarded by First Nations. Control over the data is centralized and closed. Access to third parties is obtained via a limited license to use granted by First Nations to specific parties (e.g. Health Canada), and on a pay-per-use basis for academics, policymakers, and program planners. Aggregate data is published online for free. Commercial use of data is prohibited.

Privacy protections are extensive. Independent Privacy Impact Assessments are conducted on data collection and use. PIA risk assessments address not only the legal but also the moral and ethical issues around privacy. In RHS Phase 1 (2002/2003), Dr. David Flaherty was retained to conduct a Privacy Impact Assessment (PIA) of the data collection and storage plan for the RHS. The RHS received an Overall Grade of B to B+ on its 'Privacy Report Card' from Dr. Flaherty, with the RHS consent process being noted as particularly excellent. Dr. Flaherty's review was based on the ten privacy principles contained within the Canadian Standards Association (CSA) Model Code for the Protection of Personal Information, which is found in Schedule 1 of the Personal Information Protection and Electronic Documents Act, S.C. 2005, c. 5.

Since 2003, the RHS has further enhanced personal privacy protection with practical innovations, including development of: Privacy Breach Protocol; Personal Privacy Policy for use nationally and regionally; Criminal Records Check Policy for RHS fieldworkers; and a Protocol for the Return of RHS statistical reports to the participating First Nation. Personal privacy protection is continually considered and evaluated by the RHS National Team and the First Nations Information Governance Centre in an effort to maintain the highest possible level of personal privacy. This is necessary to honor the trust of the interviewees, as well as to protect the integrity and reputation of the RHS.

Access to data requires following the protocols for data use and dissemination:

(1) No data or statistical information will be released unless: “The RHS National Team has thoroughly complied with statistics release protocols (or data release protocols) approved by The FNIGC;” or “the RHS National Team has received direct approval of The FNIGC.”

(2) While collaborating with the RHS National Team, the Region staff and representatives may receive draft and internal copies of national tables, statistics, reports and other national RHS-related information. The Region will not release or disseminate the information except if: “the Region has thoroughly complied with statistics release protocols (or data release protocols) approved by The FNIGC;” or “the Region has received direct approval of The FNIGC.”

(3) The FNIGC and the Region will not release or disseminate any data or information from the RHS that identifies or could lead to the identification of a community without authorization from that community’s recognized leadership.

(4) The FNIGC will not release or disseminate any data or information from the RHS that identifies or could lead to the identification of a First Nations Region or group of communities (e.g. tribal council, treaty area) without authorization from the appropriately mandated First Nations authority.

(5) The FNIGC and the Region will not release or disseminate any data or information from the RHS that identifies or could lead to the identification of an individual except under circumstances described under the following section “Other First Nations Authorities and Third Parties.”

Consistent with the principles of OCAP, the Regions are mandated to return survey results to participating First Nation communities, provided that the respondents’ individual privacy rights are protected as required by this Agreement and the respondents’ consent forms. Statistical tables containing RHS community-level results will be returned to communities according to a protocol:

1. The First Nation will submit a request addressed to the applicable region for return of community-level statistical tables.

2. Upon receipt of the request, the RHS will begin preparing the statistical tables for return to the First Nation. The RHS will only return de-identified information.

3. Once the statistical tables for that First Nation are prepared, notice will be given and the First Nation will provide the applicable region with evidence of a duly executed First Nation Resolution (Band Council Resolution/BCR). The BCR will contain undertakings and conditions regarding confidentiality of the information, as required by the respondents' consent.

4. The statistical tables will be returned to the community, to the attention of the individual indicated in the BCR.

The data management strategy involved in the FNRHS aims to manage the health data of participating First Nations communities across Canada for the purpose of informing policy and identification of health issues that can be targeted to improve the health and wellness of community members.[114] The data engages privacy, public and commercial interests. Unlike the other case studies involving health data, the FNRHS data also engages a proprietary interest under the OCAP principles.

Fully post-processed, aggregated data from the FNRHS is available publicly on the FNIGC website. The raw data is presumably only accessible to the First Nations communities, the FNIGC, and licensed partners. Table 9 includes a brief description of the design features of the FNRHS data management system for the raw data.

Table 9: First Nations Information Governance Centre Regional Health Survey (FNRHS) Data Trust Case Study Summary

Design Feature	Brief Description
Mandate/Purpose	Advise health policy and decision-making, identify risks and develop programs to mitigate[115]
Independent Stewardship	Yes
Fiduciary-like Obligation	Yes
Data	Raw data: health data, including substance abuse and mental health, residential schools, violence, care giving, gambling, migration, food security[116]

[114] *Ibid* at 14.

[115] *Ibid* at 12.

[116] *Ibid* at 15.

Privacy interests	Red (10)
Public interests	Red (10)
Commercial interests	Red (10)
Proprietary (IP) interests	Green (10)
Stakeholders	Health Canada, 250+ First Nations communities[117]
Trustor / Data Donors	First Nations communities
Trustee	First Nations Information Governance Centre, which is governed by a Board of Directors & each First Nations governs their own data[118]
Beneficiaries	250+ First Nations communities
Business Model	not-for-profit
Ownership Model	Community owned and controlled
Technical Architecture	Centralized
Data Access Model	Closed
Enforcement & Remedies	Unclear

The FNRHS data management system is motivated by the public purpose of enhancing health and wellness policies and programs in First Nations communities. A fiduciary obligation is clearly intended by the OCAP principles. However, each First Nations community controls who has access to the data of its members, so there is not an independent data steward. There is a set of decision-makers within each community, but those decision-makers would likely also be data subjects and beneficiaries of the data management system. Thus, the FNRHS does not have independent stewardship, and so is not a data trust. It is better described as a data co-operative, although it is not clear whether individual members of each community can individually control who accesses their data or has an equal say in decisions related to sharing their data.

ii. Additional Data Governance Examples

Having reviewed specific case studies in depth, it is worthwhile examining a few notable data governance projects through the lens of our analytic framework. These include Barcelona's DECODE Project, the City of London's data governance partnership

[117] *FNIGC FAQ, supra* note 110.
[118] *OCAP, supra* note 111 at 17.

with the Open Data Institute, the Silicon Valley Regional Data Trust, and the U.S. National Health Information Network.

a. Barcelona Data Trust (Citizen Science Data Governance Pilot) Using DEcentralised Citizen-owned Data Ecosystems (DECODE)

The Barcelona Citizen Science Data Governance Pilot focuses on environmental data, such as noise levels and pollution, collected from sensors inside the homes of residents and around their neighbourhoods.[119] The data is encrypted, subjected to a de-identification process, and is shared with the community through a distributed ledger, at the discretion of each resident. The data donor gives permission for specific purposes for which their data can be used and whether third parties can be given access.[120] A sub-set of the shared data will be open access. The DECODE program describes the system as a data commons.[121] In the context of this report, it could also be classified as a data co-operative.

Because the data will be subject to a de-identification process, privacy interests may be lower. Certainly, it depends whether the noise measurements include audio recordings or just noise levels. (The standard approach is to measure the continuous equivalent sound level, L_{eq}, which only records the decibel level of sound on a running average basis.) The monitoring of environmental data over time could have significant public benefit. Air quality and noise levels are important to health and wellbeing, so paying attention to trends in air quality and noise levels could facilitate the identification of changes, specifically, degradation in conditions much earlier than would otherwise be possible.

Certainly, there are potential commercial and proprietary interests. However, by initiating the system as a data co-operative and granting open access to individuals and businesses alike, the proprietary interests in the data are reduced. It is not clear who owns the sensors and other infrastructure, though. Commercial interests persist, as there are opportunities to advance technology and develop software and hardware for environmental assessment using the data. In addition, the data can be used to ascertain health and wellness of the community.

[119] *DECODE, Nesta*, supra note 27 at 39, 50.

[120] *Ibid* at 7-8.

[121] *Ibid* at 50.

The proposed data management system the Barcelona DECODE case study is summarized in Table 10.

Table 10: Barcelona DECODE Case Study Summary

Design Feature	Brief Description
Mandate/Purpose	The purpose of the DECODE Pilot project to develop data management approaches and the associated technology. The environmental data (noise, pollution) collected through the Barcelona project will enable decisions made by city government, and will support the development of data management tools.[122]
Independent Stewardship	No; individuals manage their own data
Fiduciary-like Obligations	No; individuals manage their own data
Data	Environmental data: noise level and pollution from homes and neighbourhood
Privacy interests	Green (1)
Public interests	Red (10)
Commercial interests	Yellow-red (7)
Proprietary (IP) interests	Green (1)
Stakeholders	
Trustor / Data Donors	Residents; individuals
Trustee	Residents are given control of access to their data, and give permission for use that is specific to specific purposes
Beneficiaries	Residents, government, businesses
Business Model	Not clear
Technical Architecture	Decentralized; distributed ledger
Data Access Model	Permission granted by each user for specific purposes; data donors also determine who has access to their data, eg. third parties, the city, etc
Enforcement & Remedies	Not mentioned

The Barcelona data management system is not a data trust. Although it serves a public purpose, the data donors themselves control the contents of the database, with no additional oversight. There is not an independent trustee making decisions to safeguard the interests of the data donors or the beneficiaries. This system is more accurately described as a data co-operative or a data commons.

[122] *DECODE, supra* note 16 at 7-8, 39, 42-43.

b. City of London and Open Data Institute (ODI) Collaboration

London, England is a leader in the field of Artificial Intelligence. It is also very active in the creation of a data ecosystem, with several open data sources either in existence or under development.[123] The London Datastore is an open data resource containing a variety of anonymized data and statistics in areas such as health, transport, the economy, and the environment. There are additional open databases providing access to transportation data and air quality.[124] It is not surprising that data trusts are an active area of study in England.

The City of London has partnered with the Open Data Institute to further develop a functioning data trust. They have initiated three pilot studies, the first to collect and use city data, a second to track and minimize food waste, and a third to track illegal wildlife trade.[125] What is known to date about the proposed data trust is summarized in Table 11.

Table 11: London-ODI Case Study Summary

Design Feature	**Brief Description**
Mandate/Purpose	Collect and use data for public benefit while maintaining public trust[126]
Independent Stewardship	Yes
Fiduciary-like Obligation	Likely, due to nature of data being collected
Data	3 pilots: (1) city data, including energy use, electric vehicle parking space occupancy, weather (2) food waste, (3) illegal wildlife trade[127]
Privacy interests	Pilot 1, city data = Green-yellow (3) Pilot 2 & 3 = information not available
Public interests	Yellow-red (7)
Commercial interests	Yellow (5)
Proprietary (IP) interests	Green (1)
Stakeholders	Individuals, businesses, government

[123] Greater London Authority, "Smarter London Together" (June 2018), online: London < https://www.london.gov.uk/sites/default/files/smarter_london_together_v1.66_-_published.pdf>; See also, online: https://www.london.gov.uk/what-we-do/business-and-economy/supporting-londons-sectors/smart-london/smarter-london-together> [*London*].

[124] London Datastore, online: <https://data.london.gov.uk>.

[125] Open Data Institute, "Data trusts: lessons from three pilots" (April 2019), online: <https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit> at 11 [*ODI Pilots*].

[126] *Ibid* at 4.

[127] *ODI Pilots, supra* note 83 at 11.

Trustor / Data Donors	Individuals, businesses, government
Trustee	Independent 3rd party
Beneficiaries	Individuals, businesses, government
Business Model	Not clear
Technical Architecture	Pilot 1: centralized across city of London Pilot 2 & 3 = information not available
Data Access Model	Not clear
Enforcement & Remedies	Not mentioned

At this point, the London-ODI collaboration is not completely defined sufficiently to conclusively determine that this will be a data trust. It has a public purpose. It has not been specifically articulated that the data management system will include an independent steward, but ODI's definition of a data trust does include the element, so it is likely. Depending on the specifics of the data to be included, a fiduciary-like obligation will likely attach or be imposed.

c. Silicon Valley Regional Data Trust (SVRDT)

The SVRDT appears to be a company formed as a research center in association with the University of California, Santa Cruz. It also receives funding from the National Science Foundation (NSF).[128] The purpose of the data trust is to bring together data from various government agencies involved with children in some capacity, particularly children in poverty, in order to understand the success and failure of students. The agencies contributing data include the public school districts, Public Health, Child and Family Services (CFS), Mental Health, Juvenile Justice/Probation, and education technology companies.[129]

The inclusion of data from CFS, the public school system and Public Health suggest a very high engagement with privacy interests. However, trying to understand and improve the success rates of children in school serves an extremely high public interest. Proprietary interests are low, ideally. The commercial interest is not clear. There are "education technology companies" involved, but they are not named, so could be for-profit or not-for-profit.

[128] Silicon Valley Regional Data Trust, online: svrdt <https://www.svrdt.org> [*SVRDT*].

[129] Santa Clara County Office of Education, "Silicon Valley Regional Data Trust", online: http://go.stewardsofchange.com/rs/092-MYB-392/images/SOCI12-Day-Two-SVRDT-Final.pdf [*Santa Clara*].

Available information regarding the design of the data management system is summarized in Table 12.

Table 12: Silicon Valley Regional Data Trust (SVRDT) Case Study Summary

Design Feature	**Brief Description**
Mandate/Purpose	Overcome siloed functionality between government services by sharing data among agencies to solve complex problems related to children, especially children of poverty, by providing understanding of student failure and success[130]
Independent Stewardship	Yes; researchers who started the data trust are controllers[131]
Fiduciary-like Obligation	Yes, highly sensitive personal information, government agencies bound legally
Data	Students' education data, health information (including mental health), any home life issues reported to Child & Family Services, any criminal legal issues known through Juvenile Justice[132]
Privacy interests	Red (10)
Public interests	Red (10)
Commercial interests	Yellow (5)
Proprietary (IP) interests	Green (1)
Stakeholders	Children; government agencies in California, including public school districts, Public Health, Child and Family Services, Mental Health, Juvenile Justice/Probation; education technology companies
Trustor	Individuals via numerous public agencies in California, including public school districts, Public Health, Child and Family Services, Mental Health, Juvenile Justice/Probation, education technology companies
Trustee	SVRDT Company Directors
Beneficiaries	Students, children, families
Business Model	Public-private partnership; not clear if for-profit or not-for-profit[133]
Ownership Model	Not clear, assumed to be the company founders, who are also directors
Cost	Research center of the University of California; National Science Foundation grant
Technical Architecture	Hybrid, distributed data repositories at each participating government agency with centralized access[134]
Data Access Model	Access is two-way, so all organizations that contribute data have access to all data[135]
Enforcement & Remedies	Not mentioned, but California privacy laws should be sufficient

[130] *Ibid.*

[131] *SVRDT, supra* note 86.

[132] *Santa Clara, supra* note 87.

[133] *SVRDT, supra* note 86.

[134] *Santa Clara, supra* note 87.

[135] *Ibid.*

Based on the information available, the SVRDT is likely a data trust. It has a public purpose. Although not crystal clear, the SVRDT directors appear to fulfill the role of independent stewards. California laws will impose a suitable enforcement framework, which combined with the involvement of government agencies will result in fiduciary-like obligation if not legally enforceable fiduciary duties.

d. National Health Information Network (NHIN)

The NHIN is a healthcare data management system in the United States, under the authority of the Office of the National Coordinator for Health Information Technology. It is defined as a system of "technical, policy, data use and service level agreements and other requirements that enable [healthcare] data exchange, whether between two organizations across the street or across the country."[136] Essentially, various government agencies and private entities share healthcare data to improve the provision of services and the quality of care. Data sharing is governed by the Data Use and Reciprocal Support Agreement (DURSA), which defines the responsibilities of the participants, the purposes for which the data can be used, the hardware and software needed, the privacy rules that must be followed and that protection measures are required to ensure data security.[137]

Certainly, the healthcare data engages privacy, public and commercial interests. It is not clear to what degree proprietary interest are engaged.

A summary of the NHIN data management system is included in Table 13.

Table 13: National Health Information Network (NHIN) Data Trust Case Study Summary

Design Feature	**Brief Description**
Mandate/Purpose	Improve healthcare by enabling the exchange of healthcare information over the internet[138]
Independent Stewardship	Yes
Fiduciary-like Obligation	Yes

[136] Nationwide Health Information Network, online: <https://www.healthit.gov/sites/default/files/what-Is-the-nhin--2.pdf> [*NHIN*].

[137] NewSTEPS, "NHIN Data Use and Reciprocal Support Agreement (DURSA), online: <https://www.newsteps.org/resources/nhin-data-use-and-reciprocal-support-agreement-dursa>.

[138] *NHIN, supra* note 103.

Data	Healthcare data, social security information, social services, military services information[139]
Privacy interests	Red (10)
Public interests	Red (10)
Commercial interests	Red (10)
Proprietary (IP) interests	(?)
Stakeholders	Co-operative currently has 24 public and private entities; 7 are federal agencies.[140] Partial list: Centers for Disease Control (CDC), Centers for Medicare and Medicaid Services (CMS), Community Health Information Collaborative (CHIC), Department of Defense (DoD), Department of Veterans Affairs (VA), HealthBridge, Kaiser Permanente, Marshfield Clinic, MedVirginia, Regenstrief, Social Security Administration (SSA), Southern Pines Women's Health Center[141]
Trustor / Data Donors	Individuals through various government agencies and private entities, hospitals, pharmacies, doctors
Trustee	Office of National Coordinator for Health Information Technology (ONC)[142]
Beneficiaries	Individuals, government
Business Model	Public-private, not-for-profit
Technical Architecture	Hybrid; common software platform, with decentralized system of data repositories
Data Access Model	Closed to participating public and private entities only
Enforcement & Remedies	Data Use and Reciprocal Support Agreement (DURSA) defines the obligations of participants to protect the data

NHIN oversight is provided by the ONC, which is a government agency. The DURSA, in combination with federal and state laws imposes fiduciary-like obligation to a certain degree. Certainly, improving healthcare is a public purpose. Therefore, the NHIN possesses the key design features of a data trust.

iii. Summary of Case Studies

A summary of the data management system identified for each case study is summarized in Table 14, along with a summary of the presence or absence of the primary

[139] *Ibid.*

[140] *Ibid.*

[141] The Office of the National Coordinator for Health Information Technology, online: <https://www.healthit.gov/sites/default/files/factsheets/nationwide-health-information-network-exchange.pdf>.

[142] *NHIN, supra* note 103.

design features identified as necessary elements defined as "necessary" in the data trust framework.

Table 14: Summary of Data Management Systems for Case Studies

CASE STUDY	Independent Steward	Fiduciary Duty	Public Purpose	Data Trust?
ICANN RDS	X	X	X	YES
Sidewalk Toronto Personal Mobility Use Case	X	X	X	YES
Sidewalk Toronto Health Data Use Case	X	X	X	YES
Truata platform			X	NO
Truata Automotive Use Case	X	X	X	YES
Truata Mastercard Use Case				NO
FNRHS		X	X	NO
Barcelona DECODE			X	NO
London-ODI	X	X	X	YES
SVRDT	X	X	X	YES
NHIN	X	X	X	YES

Three data management systems studied in Part III were determined to not be data trusts because they lacked an independent steward. The Truata system also lacked a public purpose, but had high commercial and proprietary interests. In this case, contractual agreements were used.

The other two data management systems determined not to be data trusts were the Barcelona DECODE and the FNRHS. Both systems were classified as data co-operatives. In both cases, an important element of the data sharing effort was to ensure that data subjects were in control of decision-making surrounding the collection, use and distribution of their own data. This choice removed third party oversight in each system to different degrees.

The data management structure of the Barcelona DECODE project and the FNRHS raise an interesting question. In both cases, an independent steward could have been selected, converting the systems into data trusts. Instead, the choice was made to give data subjects and beneficiaries control over their data. It should be noted that this choice is largely a value judgement. Some would argue that individual autonomy should be the paramount consideration, and data subjects should always retain decision-making control over their data. Others would argue that the greater good or the protection of vulnerable groups should be given more priority. Resolving this issue is beyond the scope of this report. The key point of this discussion is that in many situations, the choice between a data trust and a co-operative will hinge on the need or desire for an independent steward, based on the specifics of the data, the data subjects, the purpose for which the data is being collected and shared, and the judgement of those designing the system. For the remainder of this report, the two options are considered to be largely interchangeable.

Part IV: Design Considerations for Data Trusts

i. Proposed Classification of Data Trusts and Other Data Management Systems

Data characteristics are central to the choice of data management systems, and can therefore dictate certain choices of technical architecture, business models or data access models. Therefore, one option is to categorize data management systems based on data characteristics. For instance, data with a high privacy or proprietary interest score may have strict access requirements to deal with due to the governing privacy law regime or a desire to protect trade secret information. Additionally, data with a high public interest score may encourage a not-for-profit business model to elevate the importance of public good over that of commercial gain. Even where there is a high commercial interest score, or possibly because of it, this may be true. Table 15 summarizes the type of data involved for each of the case studies in Part III, along with their business model and data access model.

Table 15: Summary of Data Characteristics, Data Access Model and Business Model for Case Studies

CASE STUDY	Data Governance System	Data Type & Characterization Index[a]	Data Access Model	Business Model
Sidewalk Toronto Health Data	Data trust	Health 10-10-10-1	Very restricted	Charitable trust (not-for-profit)
NHIN	Data trust	Health data 10-10-10-?	Very restricted	Not-for-profit
SVRDT	Data trust	Health, education, social services 10-10-5-1	Very restricted	Not-for-profit

Sidewalk Toronto Personal Mobility	Data trust	Personal mobility user data 7-7-10-10	Moderately restricted	Not-for-profit
Truata Automotive	Data trust	Personal mobility user data, vehicle performance 1-7-10-10	Moderately restricted (assumed)	For-profit or not-for-profit
London-ODI	Data trust	Energy usage, EV usage, weather 3-7-5-1	Not clear	Not clear
ICANN RDS	Data trust	Domain names, owner identification, contact info 5-7-3-1	Minimally restricted	Not-for-profit
Barcelona DECODE	Co-operative	Environmental noise and air quality 1-10-7-1	Some open access; some restricted access	Not clear
Truata M/C	Contractual agreement	Financial transactions 1-1-10-10	Very restricted	For-profit
FNRHS	Co-operative	Health, etc 10-10-10-10	Very restricted	Not-for-profit

a: Data characterization index = privacy-public-commercial-proprietary

Close inspection of Table 15 reveals some potential data trust classifications based on a combination of data characteristics, data access model and business model. Three main classifications emerge. The first category involves individual health data, which is highly sensitive personal information, but also engages a public interest. Therefore, it is not surprising that data management systems containing health data are highly restrictive in terms of their data access model, and use a charitable trust or not-for-profit business model. A similar model could be used for any highly sensitive personal data, or even data that engages a high degree of commercial or proprietary interest, if the motivation for the data trust is related to spurring innovation or preventing anti-competitive behavior, such as data monopolies.

A second main classification involves transportation data, which engages a moderately high degree of privacy interests and public interest. Transportation-related data can also engage a high level of proprietary and commercial interests if the data comes from a corporation like Google or Uber, or involves vehicle performance data. The reduced sensitivity of the data in comparison to the highly sensitive personal information contained in health data warrants a lower, but still moderate level of access restriction. For the case studies included in this analysis, the business model used or proposed is a not-for-profit structure, a decision likely driven by a strong public interest. However, a for-profit model could also be used. This second classification of data trusts would be suitable for any data that includes data characterized as having a moderate level of privacy, commercial or proprietary interests.

The remaining data trusts included in Table 15, that do not fit in either Category A or B, namely the London-ODI data trust and the ICANN RDS data trust, could form a third classification. These data trusts share similar data characteristics, in that they have a lower privacy interest, a moderate public interest, and a low proprietary interest. Because they have similar data characteristics, it is likely that their data access models will be similar. (The London-ODI data trust is not fully defined yet, so details of the data access model are not yet known.)

There are three systems that were found to not meet the criteria to be classified as data trusts: the Barcelona DECODE project, the Truata Mastercard use case and the FNRHS. The Truata Mastercard use case is the most unique in terms of its data characteristics. The motivation for the system is profit, and the data involved is proprietary financial data which has been de-identified using proprietary processes. The data also engages a high commercial interest for Mastercard. Access to the data was highly restricted and the business model was for-profit. This system was classified as a contractual model, and would be an appropriate option for other data sets with similar characteristics.[143]

The Barcelona DECODE and FNRHS systems were discussed above. Based on their data characteristics, they could have been designed as data trusts. However, the design choice was made to give decision-making control regarding data collection and use to the data subjects and beneficiaries.

[143] Leaving aside as out of scope the larger question regarding commercial uses of personal information for secondary purposes and the reasonable expectations of data subjects in that regard.

a. Comparison of Data Characteristics

Figure 5 compares the data characteristics for all case studies discussed in Part III, excluding the Truata platform. The case studies involving highly sensitive personal information are labelled as Category A, and the case studies involving moderately sensitive personal information are labelled as Category B. Category C includes case studies with a lower level of sensitivity.

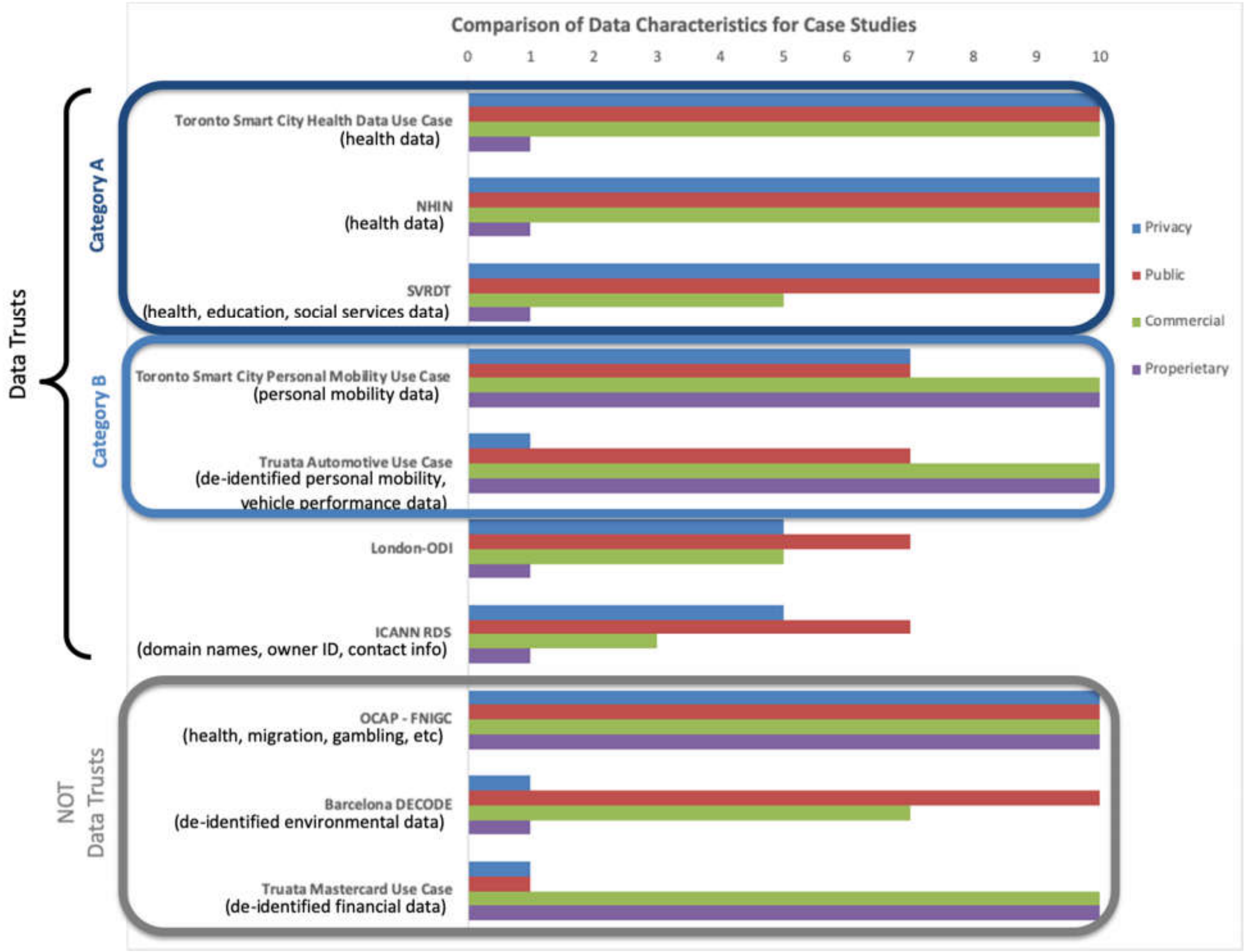


Figure 5 – Comparison of data characteristics for data trusts and other types of data governance models

b. Comparison of Data Access Model

Intuitively, the level of access restriction is expected to be related to the sensitivity of the data being accessed. As shown in Figure 6, data that has either a high privacy or a high proprietary interest is likely to require strict access controls.

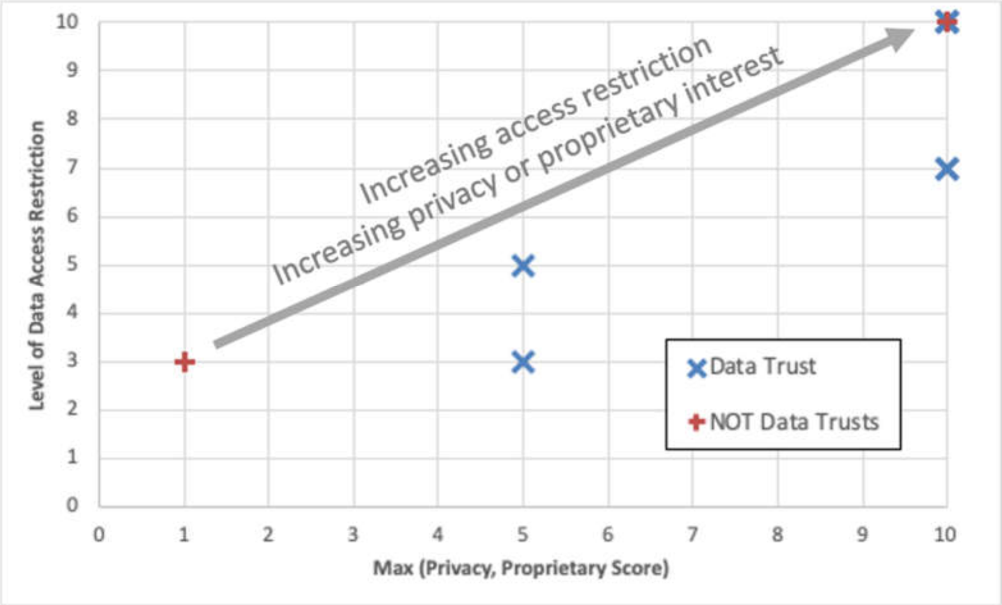


Figure 6 – Relationship between access restriction and privacy / proprietary interest

c. Comparison of Technical Architecture

As shown in Figure 7, there does not seem to be a correlation between technical architecture and type of data management system. The data co-operatives, Barcelona DECODE, and the FNRHS, are on opposite ends of the spectrum. Barcelona DECODE has a decentralized architecture because it uses distributed ledger technology. Such technology could be incorporated in almost any data management system, further supporting the observation that choice of architecture is not tied to choice of data management system. Additionally, the data trusts exist along a broad range of the spectrum, from centralized to hybrid. In the case of the health-related data management systems—FNRHS, Sidewalk Toronto Health Data Use Case, and NHIN and SVRDT—the choice of technical architecture was dictated by the fact that the data repositories already existed when the data management systems were networked or combined into a larger system.

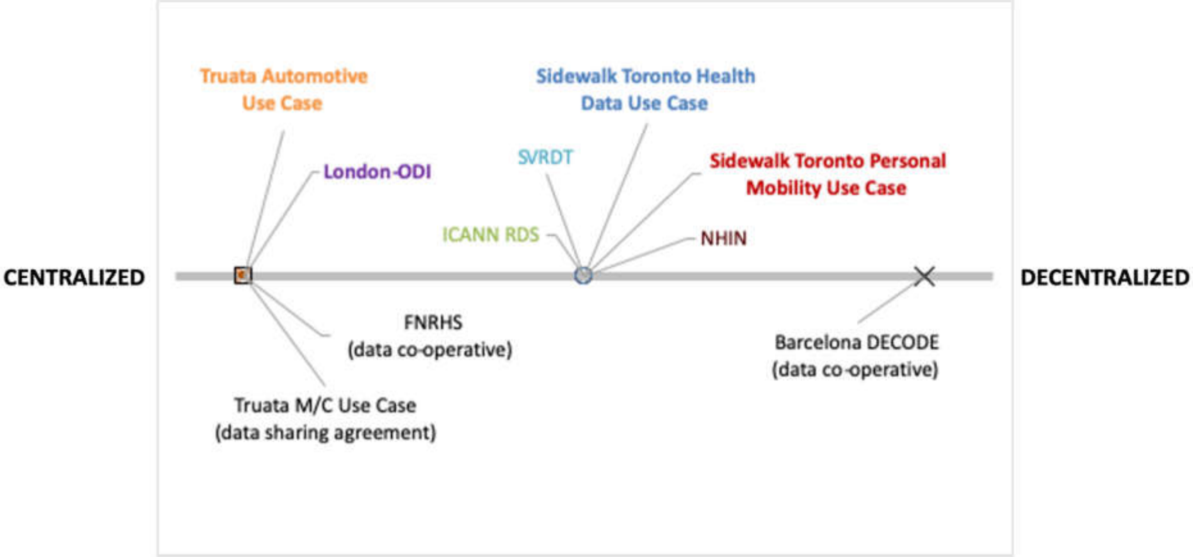

Truata Automotive Use Case
Sidewalk Toronto Health Data Use Case
London-ODI
SVRDT
Sidewalk Toronto Personal Mobility Use Case
ICANN RDS
NHIN
CENTRALIZED
DECENTRALIZED
FNRHS (data co-operative)
Barcelona DECODE (data co-operative)
Truata M/C Use Case (data sharing agreement)

Figure 7 – Technical Architecture vs Data Access Model[144]

ii. Summary of Proposed System of Data Management System Classification

Data characteristics are central to the choice of data management system and drive decisions related to secondary design features, such as the data access model. To a lesser degree, the business model may also derive from the data characteristics. However, in most cases, there are a range of business models which would be suitable.

A proposed data management system classification scheme is proposed in Table 16, based on a comparison of the design features of the case studies included in this report. The number of case studies investigated for each type of classification is limited, so ranges should be considered approximate.

Table 16: Summary of data Trust Classifications

DATA MANAGEMENT SYSTEM CLASSIFICATION	Data Characteristics	Data Access Model	Business Model
Data Trust (or co-operative) Category A – highly sensitive information	Privacy: 7-10 Public: 7-10	Very restricted	Charitable trust or not-for-profit

[144] Plot inspired by: Xabier E Barandiaran & Antonio Calleja-Lopez, "Decidim: political and technopolitical networks for participatory democracy" (2019), online: decidim docs < https://docs.decidim.org/whitepaper/en/decidim-a-brief-overview/.>

Data Trust (or co-operative) Category B – moderately sensitive information	Privacy: 5+ Proprietary/Commercial: 7+ Public: 5+	Moderately restricted	For-profit or not-for-profit
Data Trust (or co-operative) Category C – lower sensitivity information	Privacy: 3+ Proprietary/Commercial: 5+ Public: 5+	Minimal restricted	for-profit or not-for-profit
Contractual Agreement (could also be a marketplace)	Privacy: 1-3 Proprietary/Commercial: 7+ Public: 1-3	Very restricted	For-profit

iii. Preliminary Guidelines for Selecting Data Management Systems

Table 17 summarizes some general guidelines that can be used to select a type of data management system based on the combination of data characteristics. Due to the limited number of case studies investigated for each classification, these guidelines should be considered preliminary.

Data that engages a privacy, commercial or proprietary interest only minimally is suitable for an open access approach. This would be true for any level of public interest. However, if either the commercial or proprietary interests are elevated, a marketplace, contractual agreement, or data co-operative may be needed to ensure that commercial and proprietary interest are correctly managed. However, if public interest is also high, concerns over data monopolies and anti-competitive behavior should be considered before proceeding to share the data under a contractual agreement or in a marketplace. Instead, a data trust or a data co-operative may be better suited. In some instances, third party oversight may be warranted due to the nature of the data, making a data trust the optimal choice. However, if it is preferable, and possible, to have the subjects choose the fate of their data, then a co-operative is a better option.

A data trust or a data co-operative should be selected if the data engages at least a moderate level of privacy interest, regardless of the level of commercial, proprietary and public interests. However, if the public interest is low, then the risks associated with sharing the data may outweigh the benefits of sharing it. In that case, perhaps the data should not be shared.

Table 17: Guidelines for Selection of Data Management System Based on Data Characteristics

Privacy Interest	Public Interest	Commercial Interest	Proprietary Interest	Recommended Data Management System
Green	Green+	Green	Green	Open access
Green	Green	Any	Yellow+	Marketplace, data sharing agreement, co-operative
Green	Green	Yellow+	Any	Marketplace, data sharing agreement, co-operative
Green	Yellow+	Any	Yellow+	Data trust or co-operative (Category C)
Green	Yellow+	Yellow+	Any	Data trust or co-operative (Category C)
Yellow-Red+	Yellow+	Any	Any	Data trust or co-operative (Category A)
Yellow	Yellow+	Any	Any	Data trust or co-operative (Category B)
Yellow+	Green	Any	Any	Don't share data

Conclusion

This report proposed a framework for differentiating trust-like data stewardship mechanisms from other types of data governance systems and for classifying different types of data trusts. The framework consists of three primary design features and six secondary design features. Primary features are necessary components of a design trust, and include:

(i) independent stewardship,

(ii) imposition of a fiduciary-like obligation, and

(iii) a public purpose driving the decision to share the data.

Secondary design features provide additional detail, and in some cases, can be used to classify data trusts. A non-exhaustive list of secondary design features includes:

(i) data characteristics, including privacy, proprietary and/or commercial interests

(ii) stakeholders,

(iii) business model,

(iv) technical architecture,

(v) data access model, and

(vi) a system of enforcement and remedies.

This report conducted a comparative analysis of the design features for eleven data management systems. Seven of the systems were classified as data trusts, two as data collectives and one as a contractual agreement. The most useful parameter for classifying or grouping data trusts was data characteristics. The extent to which privacy, public, commercial and proprietary interests are engaged often dictates other features, such as the data access system, and to a lesser extent, the business model. Technical architecture, on the other hand, does not seem to trend with data management systems.

This study suggests three areas of consideration for ensuring data stewardship vehicles retain the benefit of trust-like data governance arrangements:

1. the character of the steward;
2. the nature of the steward's duty to beneficiaries,; and
3. technical, legal, and policy measures to protect privacy employed by the steward.

The Character of the Steward

The character of the steward is fundamental to its performance as a trust-like data governance arrangement. Simply, a faithless fiduciary cannot fulfill the role of a data trustee. In practice, this means the rigourous application of well-accepted governance principles. A trustee cannot function in a conflict of interest. Accordingly, architects of a data trust should vet potential trustees for actual and potential conflicts of interest. Similarly, principles of transparency and accountability should govern the functioning of the trustee.

The appointment of trustees - both initial and subsequent - raise similar concerns. A trustee imprinted with the duty to see to the best interests of beneficiaries ought to reflect the values and interests of those beneficiaries. Multi-stakeholder models raise particular concerns in this respect. Factors that help ensure a fair opportunity for everyone to give their best advice in a timely manner include:

- The trustee selection process must be fair, open, and accountable. Independent advisors could be sought to supervise the process.
- Compensation, if there is any, should be reasonable but not so lucrative that individuals are joining for the wrong reasons.
- Statements of interest should include why the individual wants to participate in the oversight group, and potential or actual conflicts of interest.
- It would be advisable to have a lengthy drying out period following a term on the committee to prevent immediate hiring by one of the stakeholder groups.
- Non-disclosure agreements must be as non-restrictive as possible. Individuals must be empowered to discuss disagreements with policy, and what their positions are.
- Meetings should be as public as possible, and records and archives well maintained and publicly accessible.
- The voting structure of the oversight group needs to be very carefully examined to ensure that one group cannot dominate.
- The limitations of civil society participation in multi-stakeholder, which occur for a number of reasons, e.g., inadequate resources, continuity of participation, research constraints, must be overcome to permit these stakeholders to participate.

The Steward's Duty to Beneficiaries

Trust-like data stewardship models contemplate a "fiduciary" duty to beneficiaries. This means that the trustee must understand and promote the fiduciaries best interests. Where data trusts function on behalf of the wider public this amounts to an obligation to promote the public interest. This raises a thorny question: how does a trustee divine the public interest? Common approaches to this challenge include appointing as trustees representatives from civil society or academia. Other approaches include robust public engagement and consultation.

At the core of a data steward's responsibility to beneficiaries lies an obligation to balance individual privacy rights with the broader public interest. The danger is that a trustee will redefine the public interest as symmetrical to the interests of other stakeholders, particularly commercial interests. While the public does have an interest in innovation and the fair supply of consumer products and services, these are far from the sum total of the public interest and in particular serve to undermine privacy interests. Governance structures that "trust-wash" commercial interests through the back door of a skewed definition of the public interest are to be opposed. Similarly, a data steward's duty to its beneficiaries includes a healthy skepticism towards data de-identification solutions. Techno-utopianism has no place in a data trust.

Technical elements of a data stewardship vehicle necessary to address privacy concerns

Finally, while this report has concentrated on governance structures, the technical elements of data stewardship architecture merit mention for their promotion of privacy interests. The fundamental job of a data steward is to control data flows pursuant to governance rules. If the governance vehicle lacks the technical controls necessary to enforce data flow rules, then even the best stewardship structures are vulnerable to failure. This means a data governance vehicle should feature strong access controls and comprehensive security protocols.

Further, a data stewardship vehicle should have the ability to impose rigorous data processing and use protocols, including but not limited to best effort de-identification standards. It should have the capacity and responsibility to assess re-identification risks as well as the risks to data subjects' privacy and other rights through the potential use of personal information even in de-identified form.

Concluding Thoughts

As Canada turns to reconsider its federal private sector privacy legislation, the topic of data trusts is top of mind. Structured rigourously, trust-like data stewardship vehicles could do much to promote innovative use of data while safe-guarding privacy values. If, instead, they are deployed to undermine such values in favour of other interests in personal data, data trusts may prove unworthy of the name.