Understanding data stewardship: taxonomy and use cases

Siddharth Manohar
Astha Kapoor
Aditi Ramesh

aapti institute

OMIDYAR NETWORK INDIA

Contents

Glossary

Data

Any representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means. *(IT Act, PDP Bill)*

Consent

Given by the individual to an entity for processing of their data; must be free, informed, specific, clear, and capable of being withdrawn. *(PDP Bill)*

Data Fiduciary

An entity, company, or any individual who alone or in conjunction with others determines the purpose and means of processing of personal data. *(PDP Bill)*

Data principal /user

Persons, both natural and legal, to whom any data relates. *(PDP Bill)*

Data Requestor

Individuals/entities seeking to access data from data principals and fiduciaries.

Entity /Entities

Any legal person(s), including corporate persons, who uses a data steward to hold data.

Third Party/Parties

Any legal person(s) who accesses data through a data steward.

Restricted/Unrestricted

Mode of access to data: whether access is restricted to certain companies or organizations, or accessible generally by third parties.

Public

Data that is on public record, online or offline; e.g. court records, election data, etc.

Open

Data that is freely accessible, and can be used, shared, and built-on by anyone, anywhere, for any purpose.

Personally Identifiable Information (PII)

Any information relating to a uniquely identifiable individual. *(IT Rules)*

Non-Personally Identifiable Information (Non- PII)

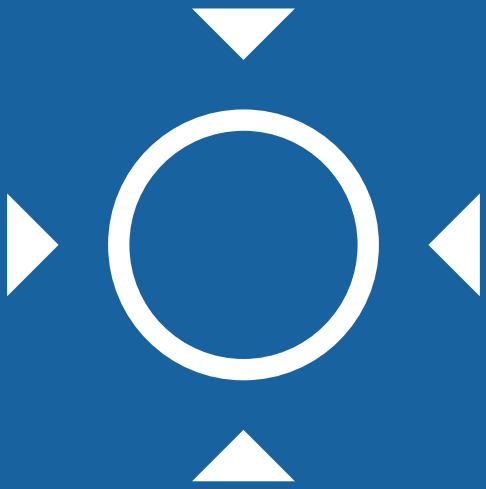
All data that is not personally identifiable. *(PDP Bill)*

Sensitive Personal Information (SPI)

A subset of personal information consisting of important details such as financial details, sexual orientation, medical history, etc. *(IT Rules)*

Section

1



Context for Data Stewardship

As more and more data is generated, it has become critical for creating value for businesses, providing insights for policy-makers, and empowering individuals with improved choice and, more targeted goods and services. With its growing centrality to decision-making, the ubiquity of data also raises concerns about unfair competitive advantages, invasion of privacy, and poor utilization of data. The current scenario of data sharing and governance lacks structure in a majority of jurisdictions. This is required in order to protect the interests of stakeholder groups in the context of data sharing and usage patterns.

In the last decade, there has been tremendous collection and aggregation of data by governments and private actors. As of June 2019, there were 4.4 billion internet users across the world, generating data on social media, e-commerce, search engines, ride-sharing apps, email services, internet of things etc. This data generated can help businesses better understand users, build applications for societal impact, and provide more tailored goods and services.[1] Insights into consumer behaviour is a significant competitive advantage and has been the central use case for large tracts of data.

Companies use data as capital, a critical raw material that can be used to generate more wealth. For example, Netflix analyses user data to understand preferences and viewing for specific genres, and updates its algorithms to recommend the best programs in a targeted manner.[2] This understanding of consumer preferences also allows Netflix to produce

[1] "How Much Data is Created on the Internet Each Day?", Jeff Schultz, Micro Focus blog, (August 6, 2019) retrieved December 9, 2019 from https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/.
[2] "Big Data for Social Innovation", Kevin C. Desouza & Kendra L. Smith, Stanford Social Innovation Review, Summer 2014, retrieved November 30, 2019 from https://ssir.org/articles/entry/big_data_for_social_innovation.

content that is well received and gives the company returns on investment – in the 2019 Oscars just six years after it had started producing content, Netflix-produced films had 15 nominations.[3]

Unsurprisingly, 84% of S&P 500 companies in 2018 drew their value from intangible assets i.e. data,[4] which is evidence of this advantage. The increasing centrality of data to businesses which use it to create efficiencies in their business, build new products, services, and explore new markets, is a cause of some concern for the free-market and competition in digital services.[5] Data accumulation can create monopolies that prevent new competitors from entering the market, or succeed in ones where they are already present. This hurts the economy, disempowers consumers, who may face lack of choice and over time, high prices.[6]

A counter trend to the growth of big technology companies is the demand for greater individual control over personal data which may improve safeguards to privacy, misuse of data, and reduce discrimination and exploitation. This movement also believes that the benefits of data (as capital) should be reaped by individuals as well. Data governance to enhance individual agency and check against the abuse of centralised power can create avenues for innovation and value propositions in terms of greater access and impact.

---

[3] "If anyone won the Oscars this year it was Netflix – the prize for its industry disruption", Louis Brennan & Paul Lyons, The Conversation, (February 26, 2019) retrieved November 30, 2019 from https://theconversation.com/if-anyone-won-the-oscars-this-year-it-was-netflix-the-prize-for-its-industry-disruption-112448.

[4] "$21 Trillion in U.S. intangible assets is 84% of S&P 500 value – IP rights and reputation included", Bruce Berman, IP Close Up (June 4, 2019) retrieved November 30, 2019 from https://ipcloseup.com/2019/06/04/21-trillion-in-u-s-intangible-asset-value-is-84-of-sp-500-value-ip-rights-and-reputation-included/.

[5] "Facebook, Exploitative business terms pursuant to Section 19(1) GWB for inadequate data processing", Bundeskartellamt (2019), retrieved October 25, 2019 from https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=4.

[6] "Sources of Tech Platform Power", Lina M Khan, Georgetown Law Technology Review, retrieved October 25, 2019 from https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Khan-pp-225-34.pdf.

Further, there is a growing realization of the value of data for addressing society's "wicked problems" – complex, intricate questions that involve a number of stakeholders, and have the potential to create significant impact for stakeholders. For example, the global water crisis could benefit from data driven insights on use, wastage, and enable better utilization of water infrastructure. However, governments, non-profits and businesses working on solving this issue have limited data access and are unable to utilize large data sets to build solutions that are tailored to individuals and communities. Beyond availability, quality of data remains a challenge that is compounded by the ability of smaller firms to process large sets of data.

It is undeniable that there is a need to unlock the value of data by sharing, such that it is released from the monopolies of big technology companies, and used to empower individuals and address societal problems. It is, therefore, time to build systems and processes that allow for easy and safe data sharing in ways that enable innovation without compromising individual rights and security and to derive public good. This balance of societal good, market innovation, and privacy is at the core of the questions on data governance, which needs to create data sharing apparatuses that are equitable, accountable, and just.

In this context, we believe that data stewardship, where an intermediary facilitates or holds consent and decision-making on behalf of users, is a viable solution that can balance individuals data rights and the use of data for societal good. Stewards are also responsible for ensuring data is unlocked to generate societal value, and maintain the security standards and quality of the datasets. Data stewardship can potentially enhance accountability of platforms, user control over their own data, and consequently, trust in processes of data use and analysis.[7]

[7] "Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship", Keiron O'Hara, Web Science Institute, retrieved October 25, 2019 from https://eprints.soton.ac.uk/428276/1/WSI_White_Paper_1.pdf.

That being said, the current vocabulary on data stewardship is complex, and often conflicting. Terms and concepts are used loosely and inter-changeably, making adoption of a specific taxonomy complicated. Creating a common taxonomy is a necessary step to kick-off pilots, for large-scale implementation, and for regulation. This report builds on the existing work regarding the practice of data stewardship, its definitions, models and use cases, and aims to unify the language of stewardship and present a framework of analysis that encompasses all existing models of stewardship. It also explores the legal and technical features of these models, and examines a few specific takeaways for India.

This early report is one of overarching thinking that connects siloed discourse on data stewardship throughout the world. Its value lies in bringing together divergent perspectives and examples on stewardship into a singular document, and building a collective taxonomy. It sets the up the conversation on data stewardship for broader and more in-depth exploration.

This report is structured as follows: first, we introduce the idea of data stewardship, explore the role of data stewards, and the value of the steward for various parties. Thereafter, we attempt to build a uniform vocabulary and create a framework for analysis. In the next section, we deep-dive into the models of stewardship, bringing in various examples currently in practice. Finally, we explore the applicability in India, and how stewardship models fit in with existing laws and practices. We conclude the report with questions that can help frame short- and medium-term conversations on stewardship.

We expect that through the reading of this report, data stewardship and its possibilities will be better understood. We imagine that in the coming 2-3 years data stewardship pilots will gather pace, and this document can be a handy go-to as innovators/companies begin to think about the issue.

Introduction to Data Stewardship

Data sharing is the concept of transferring data from one entity to another. The transfer or sharing of data between two private entities is generally governed by a data sharing agreement. At present, these agreements are often exclusive contracts between entities–such as, when an online service provider shares data regarding preferences of a set of users with advertisers. Advertisers then use this data to target users better. The principal vehicles of data sharing today are such data sharing agreements; this has led to a troubling pattern - a handful of companies have become concentrated nodes of aggregated user data, due to centralised market patterns in online services.[8]

From this emerge several distinct issues: First, large internet and technology companies accumulate large data sets on individual preferences and are able to build detailed profiles which might be used for targeted content, which in turn dictates individual decision-making. Second, smaller companies that do not have large sets of data, are at a significant disadvantage as they are unable to access these large datasets to enter or compete. Accumulation of data stifles innovation.[9]

[8] "Should we be Worried about Data-opolies?", Maurice Stucke, Georgetown Law Technology Review, retrieved October 25, 2019 from https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Stucke-pp-275-324.pdf.

[9] "The Antitrust Case Against Facebook", Dina Srinivasan, Berkeley Business Law Journal Vol. 16, Issue 1, retrieved November 30, 2019 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3247362

Even governments, and non-profits working on people-centric issues do not have sufficient access to user data. Finally, given that the business model of internet and technology companies is data extractive, users are forced to constantly part with large parts of their data on a regular basis in return for conveniences, and have no way to negotiate with technology companies once they enter into an agreement at the point of sign-up. Users also lack control on how their data is further used and shared once they consent to the terms of a service, often at the point of sign up, and have no readily accessible recourse against violation of the agreement or of applicable laws. Revocation of consent is made difficult.

In this context, a data steward, who manages the sharing of data, without an interest in the data, becomes important. Simply put, a data steward is an intermediary who works on behalf of the users/entities to manage data and its sharing. Stewardship as a concept is designed to structure data flows and data sharing according to certain defined incentive structures – thereby creating models that promote societal good.

Stewards hold varying degrees of fiduciary responsibility towards the users/entities. This is related to how fiduciaries are being defined in law. For example, in the Indian context, the soon to be introduced Personal Data Protection Bill (2018) defined fiduciaries as "any person, including the State, a company, any juristic entity or any individual who alone or in conjunction with others determines the purpose and means of processing of personal data."[10] The focus of the Bill, however, is primarily on personal data and the rights, duties, and governance structures around it. Discussions have also been initiated by the central government to lay out a policy to regarding the collection and usage of non-personal data. To this end, the Indian Ministry of Electronics and Information Technology (MEITY) has constituted a committee to look into the issue in depth and submit a report

[10] "The Personal Data Protection Bill, 2018", Indian Ministry of Electronics and Information Technology (2018), retrieved October 25, 2019 from https://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf.

on its recommendations.[11] This was followed by comments from the Chair of the committee exploring the possibility of usage of such data for public benefit.[12]

This increasing conversation on unlocking the value of data through sharing is not confined to India. The EU has also passed a regulation to support the free exchange of non-personal data, and the conversation continues to pick up pace in the global context. The discussions focus on the harms of localization of data within countries, and the limits on competition arising – which arise from the lack of mobility of data across vendors who may maximize its value.[13]

Countries around the world are beginning to explore and discuss possibilities of using data for societal and policy purposes by state and non-state actors. Regulatory and policy frameworks to achieve this are being actively considered to this end, such that data can be protected from misuse and capture but also be opened for use. Data stewardship is a possible solution to this complex problem.

History of data sharing

Data sharing has been undertaken for a long time to enable mutual benefit, as countries and industries benefit from cooperating and learning about one another. Sharing of meteorological data for example, has helped the study of weather and climate patterns generating benefits across borders.

[11] "Indian govt forms committee to recommend governance norms for non-personal data, Infosys' Gopalakrishnan to head it", Aditi Agrawal, Medianama, (September 16, 2019) retrieved October 25, 2019 from https://www.medianama.com/2019/09/223-meity-non-personal-data-committee/.

[12] "Benefits of data rules must reach citizens, not just companies: S Gopalakrishnan", Bharani Vaitheesvaran, The Economic Times (September 16, 2019), retrieved October 25, 2019 from https://economictimes.indiatimes.com/tech/internet/benefits-of-data-rules-must-reach-citizens-not-just-companies-s-gopalakrishnan/printarticle/71142724.cms.

[13] EU Regulation on a framework for the free flow of non-personal data, 2018, retrieved 30 November 2019 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN.

Similarly, the practice of governments sharing data with academics for analysis has been instrumental to the study of economics. Sharing data is instrumental in the development of knowledge, research, innovation, and cooperation that help us better understand our society, economy, and polity. Some data, therefore, is rightly regarded as a shared resource that benefits society at large.

Health data has seen the earliest growth in formalized sharing. In some ways, the marginal benefit from sharing data is easily communicated as it leads to collaborations and outcomes that have impact the lives of people significantly.[14] Health data for research is used through anonymized sharing between hospitals and researchers, who navigate these relationships through data sharing agreements. Consent from users or patients is acquired at the point of care. While patients may be informed of purpose and safeguards to their privacy, they are unlikely able to negotiate the terms of sharing. However, given the contribution to research, and possibly in areas that impact the users/their families personally, the incentive to share is high.

With growing awareness on safeguards for privacy, hospitals are becoming increasingly conscious of setting up mechanisms that do not compromise the data. One of these mechanisms are data rooms, or high security data centres, that allow for patient health information sharing through trustworthy channels. That said, there is a need for a more organized approach to collection and sharing of data in the context of health. More and more countries are setting up clear policies for data sharing in health, such as Norway, which has put together a strategy for access and sharing of research data. The strategy is anchored in the need to share and reuse data more widely, but combines this openness with essential safety and security guidelines that afford the highest protection to users.[15]

[14] "A beginner's guide to data stewardship and data sharing", Marcel P. Dijkers, Spinal Cord, retrieved October 25, 2019 from https://www.nature.com/articles/s41393-018-0232-6.
[15] "National strategy on access to and sharing of research data", Norwegian Ministry of Education and Research (2018), retrieved October 25, 2019 from https://www.regjeringen.no/contentassets/3a0ceeaa1c9b4611a1b86fc5616abde7/en-gb/pdfs/national-strategy-on-access_summary.pdf.

Benefits of Unlocking the value of data through sharing

Once unlocked, public, personal and non-personal data can benefit society while safeguarding individual rights in the following ways:

Data can be used to innovate for societal good and generate value for a large number of people. Sharing mechanisms can build a balance between societal good, and individual privacy such that incentives to share for companies and individuals are clear. Eg: the Chicago Data Collaborative[26] provides a central access point for all public information on multiple levels of the justice system.

Solutions to problems such as urban planning, environmental data collection, public health, etc. can be made publicly accessible as opposed to being kept restricted in proprietary forms. E.g.: The Humanitarian Data Exchange provides a platform for data sharing across organisations and countries in relation to crises and disaster management.

Data can be held in the interest of a specified set of users/entities, who can exercise greater agency over how it is used. Consent can be aggregated so that individuals have more negotiation power with technology companies, through a collective representation of their interests.

Functions of a Data Steward

Data Stewards are designed to be trusted intermediaries that lie between users, fiduciaries and requestors and can ease the process of sharing. A Data Steward has four main functions:

Collaboration: the data steward provides a platform to individuals and entities to pool and share their data and bring different data sets together, for the benefit the public. Disaggregated silos of data are collated and organized by the data steward to make organized and structured systems to store, share, and use the data that is entrusted to it.

[16] The Chicago Data Collective, retrieved October 25, 2019 from https://chicagodatacollaborative.org/.

Management: The steward makes data available in a usable format. To this end, the data steward is tasked with meeting certain standards in maintenance of data quality. This finds application in terms of standards for input/accepting data, its storage, as well as in sharing. Depending of the structural role of the steward, it may be in a position to prescribe data formats and technical specifications that apply to a certain type of data or purposes of usage.[17]

Accountability: The data steward, is entrusted with data by users and other companies with separate liability. It is required to follow disclosure norms and offer accountability mechanisms to end users. It communicates the fiduciary responsibility to users through disclosures on data use, safety and security standards and practices, and transparency regarding the structure and activities of the steward. The specifics of the accountability measures applicable vary according the model of data stewardship as well as the sectoral regulations applicable in the relevant jurisdiction.

Intermediation: The data steward, based on the terms on which data is collected, manages data, including consent, on behalf of users and participating companies. This includes management of sharing and providing third-party access to this data as well. This function positions the steward as an intermediary of data sharing and access: between participating users and companies on one hand and third-party entities on the other (See Figure 1). Being in this position of a trusted intermediary, the steward builds mechanisms of enforce the rights of users over their data. Broadly, the steward protects the interests of the users here and incentive structures around the steward are designed to account for this. The aggregation of data with the steward, acting as representative of its constituent users, presents a unique scenario of the collective interests of the users being represented by one entity. Stewards can negotiate with data requestors and fiduciaries on behalf of users, a task which is difficult for users to take up individually.

[17] "The Three Goals and Five Functions of Data Stewards", Stefaan Verhulst, The GovLab (19 September, 2018), retrieved October 25, 2019 from http://thegovlab.org/the-three-goals-and-five-functions-of-data-stewards-2/.

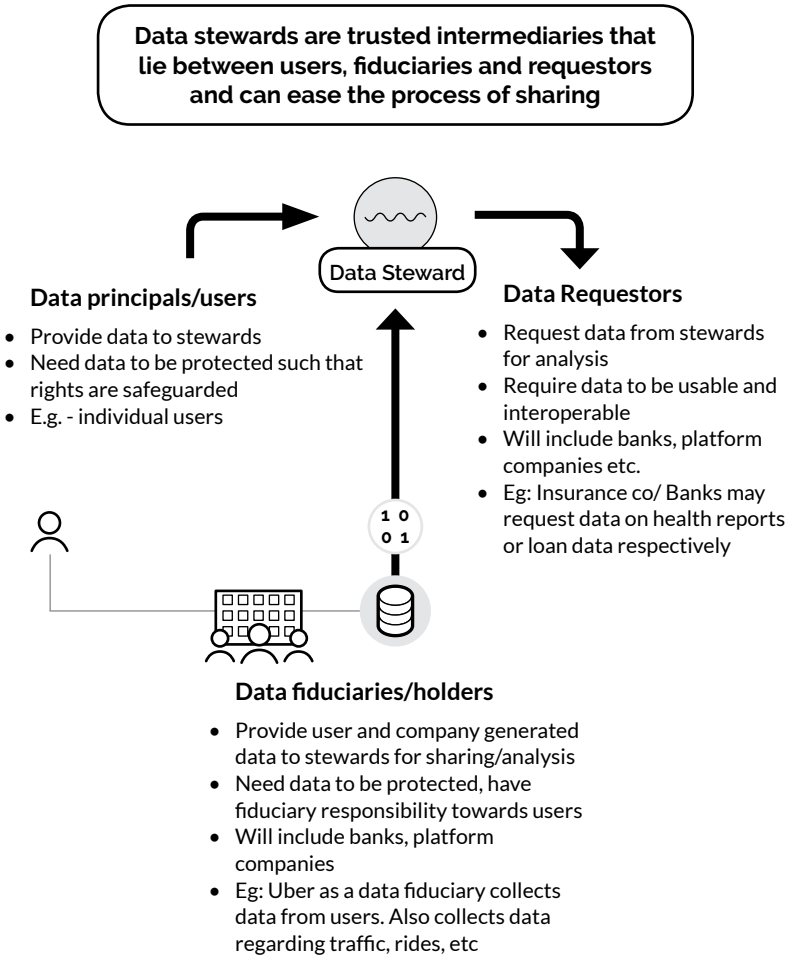


Figure 1. Role of a data steward

Value add of a data steward

Value for data principals and fiduciaries

Trust in data sharing processes: The role of a steward is envisaged as a trusted intermediary designed to protect user interests and manage data better to minimise the possibility of harm to the user. This is done by giving the user more control over the sharing and usage of their data. The incentive structure does not lead to uncontrolled usage of data, but gives users control over limiting the use of their data, in terms of purpose, time, or parties. The stewards may advertise the value of sharing the data to users, including for creation of societal value over and above individual benefit. The priority given to privacy and security of data is also expected to be a draw for users to approach stewards to manage data for their respective objectives.

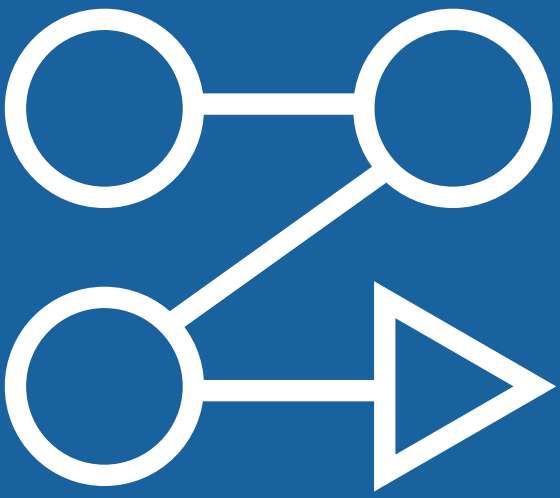
Value for data requesting entities

Availability of reliable, quality datasets in usable formats/quality: Stewards curate collaborations and maintain quality of data, making sharing between multiple parties easier. A standardized and repeatable set of terms for access and usage makes data available for purposes that may not yet be economically feasible. The data is made accessible for a wider variety of uses, which may not have as yet been carried out. This increases the value of the data as a whole as well as to certain stakeholders specifically. The availability of large sets of data otherwise in non-shareable domains, can be utilized for generating public value.

Reduced cost of accessing data: Data requestors can work directly with stewards as opposed to engaging in multiple data-sharing agreements with different fiduciaries and users. This reduces transactions costs across the value chain.

Section

2



Methodology

Objectives

Through this report, we seek to create a comprehensive understanding of the relevant questions in the subject. This would include: a consolidated view of global precedents of data stewardship - for public good - in both public and private sectors; and a comprehensive study of the applicability of different models of data stewardship in India, and the opportunities and challenges therein.

Our initial approach was shaped by the fact that the subject remains as yet relatively unexplored. Data stewardship, as defined in this report, is practiced in many different forms for a variety of purposes and is referred to with varying nomenclature in different contexts, both in commercial terms as well as in the emerging body of research on the subject.

While the lack of existing research in the area means there is no single authoritative source on the subject, studies carried out by the Open Data Institute (ODI), GovLabs, Nesta, and Centre for International Governance Innovation have been instrumental in building our understanding of the breadth of issues in data stewardship and in building a framework to understand how the various forms of stewardship interact with and relate to one another.

On the application and practice side of data stewardship, there are challenges in observing existing models as a large number of projects remain in the pilot phase or other forms of limited operation. This makes it difficult to draw major conclusions regarding impact and possible challenges from a study of their practices.

The varied application of these ideas has led to an inconsistent taxonomy across existing literature, with seemingly parallel models exhibiting varying degrees of overlap when broken down to their constituent elements. The inverse has also been true, where two significantly different systems and

contexts of data stewardship share a similar nomenclature. Some of these projects include Truata, MiData, Solid, and Apple Health Records which are all called personal data stores.

Compounding this issue is a more common problem with research on proprietary services – the methods, standards and practices remain privileged information in many cases, making examples of commercial application of data stewardship models less conducive for a thorough study. In cases where these services make the relevant resources publicly available, there may be challenges in determining the exact technical and security protocols employed to index, secure, share, and use data. Therefore, we started our research with a few anchor issues to understand the major concerns around data stewardship – data collection, consent, sharing of data, data security and privacy – and proceeded to look at existing use cases of data stewardship to build an understanding of how the concept is operationalised. This approach helped us overcome challenges in disparate and inconsistent language currently used in discussing data stewardship.

In our research, we analysed over 100 use cases of stewardship – all of varying size, purpose, geography, and business models – to carry out this study (See Figure 2). We used this sample set to build our understanding of how these use cases were similar and different to each other. A study of commonalities and differences allowed us to determine crucial points of difference and identify defining characteristics for distinct models of data stewardship. Using these defining characteristics, we created an initial framework of six models of data stewardship.

This initial analysis yielded six models of data stewardship. However, significant concerns and overlaps remained – and, we conducted in-depth interviews with experts (Annexure 1) in order to clarify and unpack these issues, which brought us to the framework now discussed in Section 3.

A few forms of data sharing are excluded from our analysis: chiefly Open Data and direct bilateral data sharing agreements. Whilst these are relevant, and in fact the dominant modes of data sharing prevalent today, they are not, based on our analysis, stewarded data. Which is to say, there is no one entity that remains responsible for maintenance of the data, and no intermediating entity that can represent the interests of users. Our analysis focuses on models wherein a steward is appointed to maintain and manage the data and has certain responsibilities attached to it.

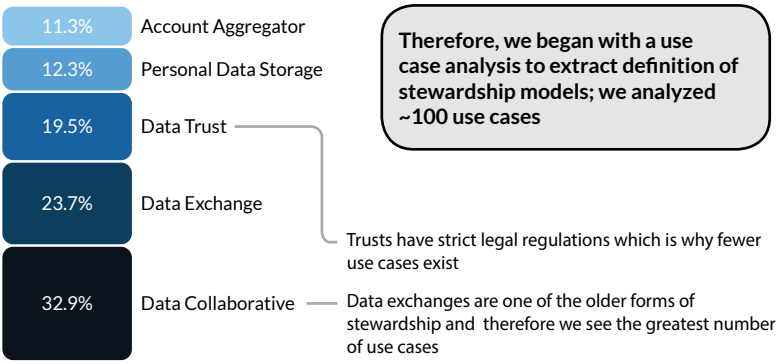


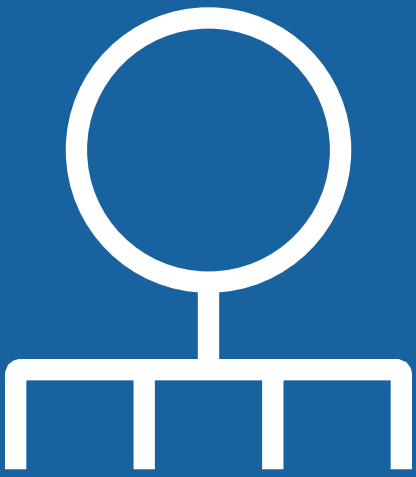
Figure 2. Use case analysis

Open Data: Open Data is data that can be freely used, shared, and built-on by anyone, anywhere, for any purpose. Being a dataset that is released into the public domain, Open Data is not stewarded by an entity. Rather, it is accessible from its existing infrastructure. There are two key dimensions to the "openness" of data which are distinctly differently from stewarded data:

	Open data	Stewarded Data
Availability	Data must be available as a whole and in a convenient form	Steward manages and regulates the final form of the data
Re-use and redistribution	Data must be provided under terms that permit re-use	Requires trustee, individual, or nominee consent for data use
Universal participation and access	Everyone must be able to access data (no discrimination or restrictions)	Access is restricted by trustee, individual, or nominee

Section

3



Creating the Taxonomy: Defining Stewardship Models

Our definition chart is built from the ground-up, based on our analysis of use-cases. Our study of the existing examples of data stewardship drew out salient features that distinguished the models from each other to create a table of definitions (Figure 3), with six models varying from one another on six different metrics. While we found overlap on a few parameters, the models also have distinctive features that help providing a unique definition for each.

Our table of definitions is useful for several things: it is a representative landscape of the prevalent models of data stewardship operating across the globe; it pulls out the primary constituent traits of the models, which help in analyse overlap and divergence from one another; and it helps get an idea of which models would be useful to the various stakeholder interests at play.

However, what it did not do is provide a framework for those using this research to map out a suitable option for stewarding data. In order to address this feedback, and build a framework that allows individuals and companies to analyze models more conveniently in relation to one another, we undertook expert interviews with industry and academic (see Annexe 1). These interviews led to a reworking of the framework, building a more coherent structure in which the overlap between the models is addressed.

We demonstrate this evolution of definitions to highlight the methodology of categorizing the stewardship models. We believe that this analysis lends to our objective - arrive at a common understanding of the definitions. These definitions, derived from analysis, can be used by all participants in the ecosystem, and can serve to simplify some of the complexity in the vocabulary.

	Where is the data collected from?	What is the access type?	What is the purpose of the data?	Who are the key stakeholders?	What types of data are used?	What is the legal framework?
Data Trust	Participating entities, Individual Users	Restricted	Defined by policies; shared with third parties as per user consent	Participating users and entities, third party clients	SPI, PII, Non-PII, Public	Trust-specific policy; Laws to enforce user rights
Data Collaborative	Participating entities	Unrestricted	Public good; Specific uses defined by policies	Entities, third party clients	PII, Non-PII, Public	Laws for large projects; supplementary rules and policy
Data Exchange	Participating entities	Restricted	Defined by agreement of participating entities	Participating entities, data principals	PII, Non-PII, Public	Existing laws; Data-sharing agreements
Account aggregator	Individual Users	Restricted	Transferred to requesting entity	Users, source entities, requesting entities, regulators	SPI, PII	Specific regime(RBI regn, PDPB, Technical stds, consent framework
Personal data store	Individual Users	Restricted	Controlled by individual users	Participating users, third party clients	SPI, PII, Non-PII	Store-specific policy; laws to enforce user rights
Data marketplace	Individual Users	Restricted	Transferred to third party clients	Participating users, third party clients	SPI, PII, Non-PII	Marketplace-specific policy; laws to enforce user rights

Figure 3. Definition chart of six stewardship models

Unbundling the framework

Foundational question 1: What is the intent of the steward?

Our study of stewardship models includes an analysis of the various uses and applications of stewardship as well as the details of the model employed. Through our study, we organised the various applications of stewardship under three large brackets recognizing that it is possible that stewardship projects may change, modify, or take on additional roles. The initial intent however remains a defining feature for their formation.

- **Creation of Societal Value:** Data is shared, exchanged, and stewarded to create value that contributes to society, and drives towards a "common good". Data is used for the benefit of society as a whole, rather than a limited section thereof.
- **Generation of Commercial Value:** Generating value for private entities in the form of commercial or monetary gains through sale of anonymised or derivative data, leveraging business innovations, business analytics, etc.
- **Empowerment of Individuals:** Individuals are given effective and consistent control over their data, and decision-making power over its access and usage. Individuals are then able to generate value from their data as they deem fit.

Foundational question 2: What is the type of data that is being stewarded?

A crucial question for any data steward is what kind of data it will steward. It may choose to steward – 1) Non-PII data that is anonymous and cannot be traced back on an individual identity 2) PII, data that can be used to identify an individual; and 3) SPI, which includes certain sensitive elements of data relating to an individual, including financial details, sexual orientation, medical history, etc. The standards and policies suitable for a steward vary significantly according to the type of data it stewards. Further, the governance structures are dependent on the law of the law and

therefore vary – a steward of SPI and PII for example is required to provide user control and access for revocation of consent and purpose limitation as per data protection laws across GDPR jurisdictions.

Data stewardship analysis framework

The analysis carried out through our foundational questions led to the creation of a more precise taxonomy of data stewardship. We whittled down our earlier six categories of data stewards to forming four distinct models. The following chart (Figure 4) provides a concise representation of our framework as described above. It captures the unique, defining characteristics of the four models of data stewardship as explained in our analysis. The axes are explained below, and represent the most critical features of any stewardship model.

On the x-axis, the role of the steward is represented. There are three possible roles for the steward in this framework. It may remain data-blind and simply redirect data flow as per user instruction and policies; it may alternately provide a data stock, where it retains data collected from users, with further policies on specifics; or else it may also provide a function of stock and flow, where it stores data and also provides the function of sharing this data with third parties based on its policies and user consent.

The role of the steward significantly impacts the measures necessary for it to securely collect, handle, use, and share data. For instance, the framework of an Account Aggregator, which remains a data- blind pass-through mechanism, will need less specifics on data handling than a Data Trust which manages and potentially makes use of the data in its care. Similarly, Personal Data Stores offering comprehensive user controls for sharing data will differ significantly in architecture and security measures from Data Collaboratives that may deal in de-identified data or metadata.

On the y-axis, who decides access and use of data is represented. The mode of data access adopted by the model varies in terms of the point of decision-making. The permission may come from the users themselves

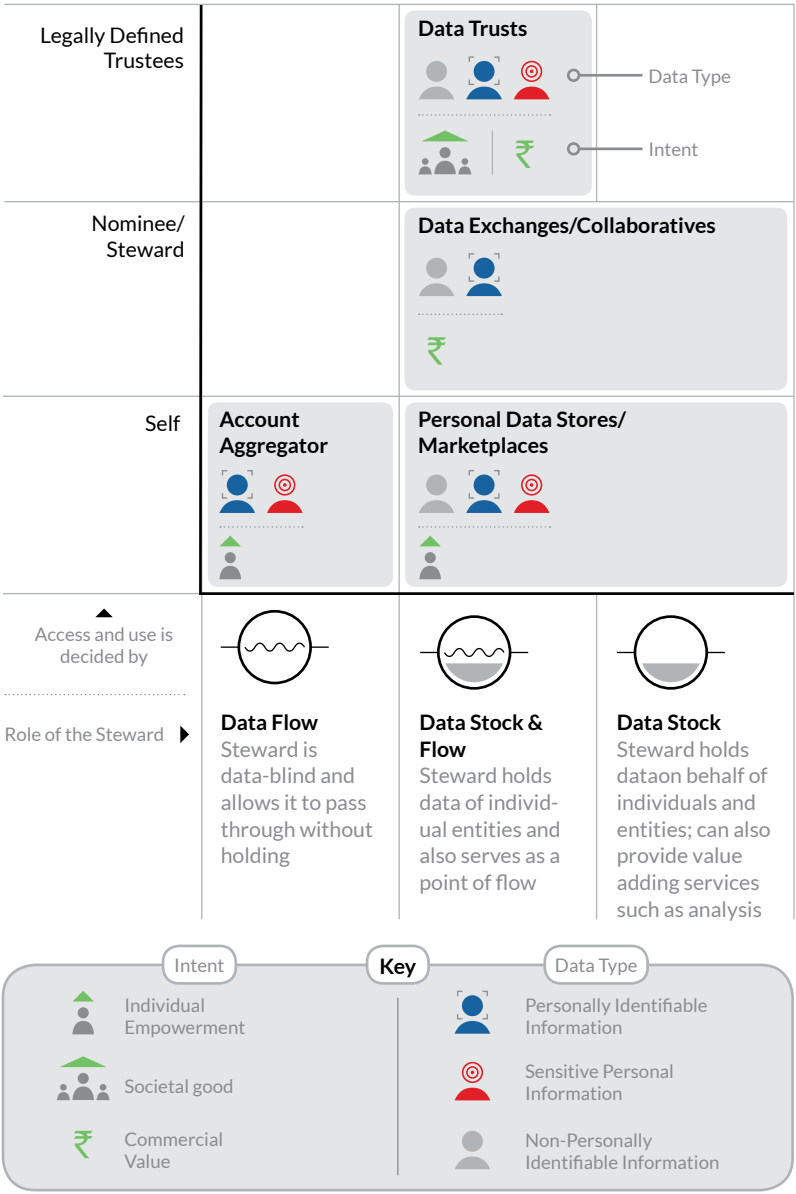


Figure 4. Modified framework & substantive definitions

where the steward plays the role of consent management; it may come from an entity or individual nominated on behalf of the users; or it may come from an authority defined by legally applicable standards. This metric is critical as it indicates the terms of the consent obtained by the steward with respect to usage and sharing of data. The framework of access and use also indicates the broader governance structure around the collected data, and will impact the kind of recourse available to users in order to pursue any rectifications or remedies against misuse of data.

This framework has led to a set of four models, each decidedly distinct from each other on the two sets parameters. The defining traits of these models are formed from how the models interact with the axis parameters.

Design Choices

While the above framework identifies the major choices that define the nature of a steward, there are various issues on which a steward must make decisions based on which its nature, purpose, and and applicable policies would vary. These are the practical concerns that are likely to come into focus when a steward wants to start operations. Represented in Figure 5 is a non-exhaustive illustration of a list of design choices that stewards may need to make for successful implementation of their mandate. Questions of business models, type of data, and interactions with third parties play a role in the practices of the steward. These choices impact one another, as well as other outcomes such as the minimum security standards applicable, sectoral regulations, and other relevant compliances.

It is important to note that there may be further technical and structural measures that a steward may adopt beyond those specified. For instance, the technical possibilities for security continue to evolve, as do possible business models. The specifics of these decisions may be modified while continuing to adhere to minimum standards of security and user control.

Taxonomy of stewardship models

1. **Data Exchanges/Collaboratives:**
Data Exchange is a steward that collects data from either individuals or companies, performs functions of holding as well as sharing the data, has access models restricted by purpose or participation, and is regulated through its defined role of a nominee/representative steward. Eg: Chicago Data Collaborative, Amsterdam City Portal, Financial Data Exchange

2. **Data Trust:**
Data Trusts are stewards that collects data from either individuals or companies, performs the function of holding data with a view to share it further according to predefined purposes and policies, and is governed by a legal layer of the framework of a trust. Eg: Transport for London, Truata, MiData.

3. **Personal Data Stores:**
Personal Data Stores are a type of steward that collects data from individuals, performs functions of sharing this data based on user permissions as per certain predefined purposes, and is controlled directly by the user who avails of its services. Eg: Solid, Digi.me, Meeco.

4. **Account Aggregator:**
Account Aggregator, the working model for which is currently being developed in India, is a steward that does not hold data. It operates as a centralized consent engine, and communicates instructions initiated by the user, to transfer their data from one fiduciary to another. Eg: CAMS Finserv, Perfios, Estonia Data Exchange Layer.

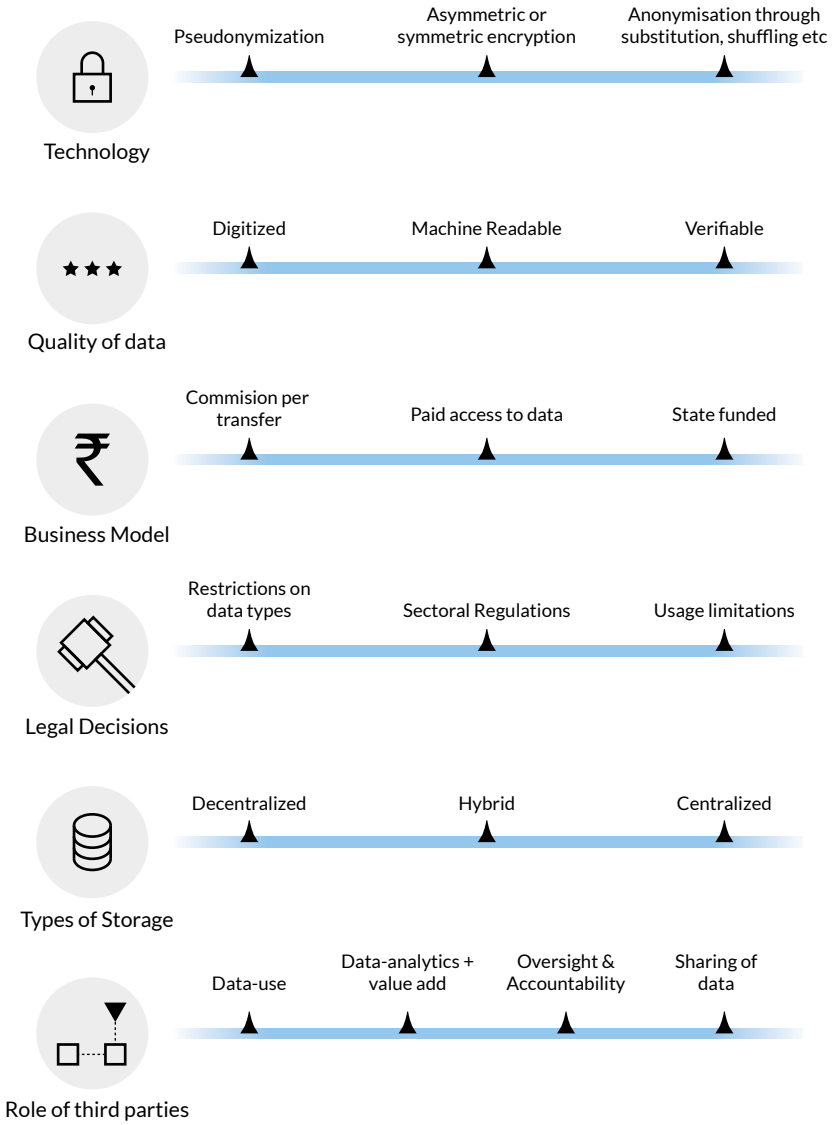


Figure 5. Design choices for data stewards

Section

4

Deep Dives

In this section we will further explore the specifics of data stewardship models, the participants, and the nodes in each model along with their respective roles. The use cases for each model are also explained.

1

Account Aggregators

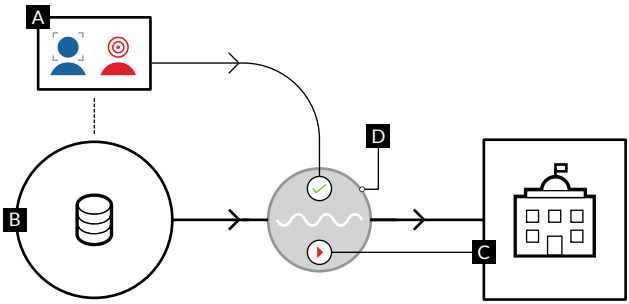
Account Aggregators ("AAs"), in reference to our framework are a flow mechanism where access and use is decided by the individual. Account Aggregators represent a homegrown model of stewarding data in the Indian context. The Reserve Bank of India in 2016 notified the Account Aggregator Directions[18] to create a legal framework for entities to take up the service in a predictable policy environment. The Directions define Account Aggregators as non-banking financial institutions that offer services of such as retrieving/collecting the user's financial information; and consolidating, organizing and presenting such information to the customer, as may be specified by the bank.

The primary role of the steward is to aggregate user consent, redirect user data and act as a conduit between data fiduciaries that hold and share user data. It is a central point of contact for users who entrust the Account Aggregators to manage their data while maintaining control oversharing permissions. The data is encrypted and redirected and the steward does not store data internally (see Figure 6). Eight service providers are listed as official licensees by Reserve Bank of India to provide Account Aggregation services.[19]

Potential Use Cases of Account Aggregators:

The uses of Account Aggregators are restricted, for now, to the financial services sector in the Indian context, due to the intention behind its regulatory design, and its predefined scope of commercial use.

[18] "Account Aggregator Directions, 2016", Reserve Bank of India (2016), retrieved October 25, 2019 from https://rbidocs.rbi.org.in/rdocs/notification/PDFs/MD46859213614C3046C1BF9B7C F563FF1346.PDF.
[19] "Account Aggregators in India, DigiSahamati Foundation, retrieved November 30, 2019 from https://sahamati.org.in/account-aggregators-in-india/.



A. User provides consent for sharing of data from one fiduciary to another
B. Participating entity (Data provider) provides data requested
C. Entity requesting data submits request to AA for specific data
D. AA transfers requested data to the participating entity (data requestor)–AA manages consent engine.

Figure 6. Graphical representation of the Account Aggregator model of stewardship

In addition, Account Aggregators can prove useful to services such as insurance, where providers may use the steward to request relevant financial data or medical records. The model can also be used in e-governance, where tax filings and citizen documentation etc. can be shared through the architecture. Data is collected with customers' explicit consent. Transactions are backed by agreements covering the Account Aggregators and data requesting entities and data providing entities and are routed through users.

The chief requirement in the Account Aggregators rules is to have in place a consent architecture to process customer permissions for sharing data across fiduciaries.[20] The consent architecture must account for a number of details regarding the customer, including their identity, the information

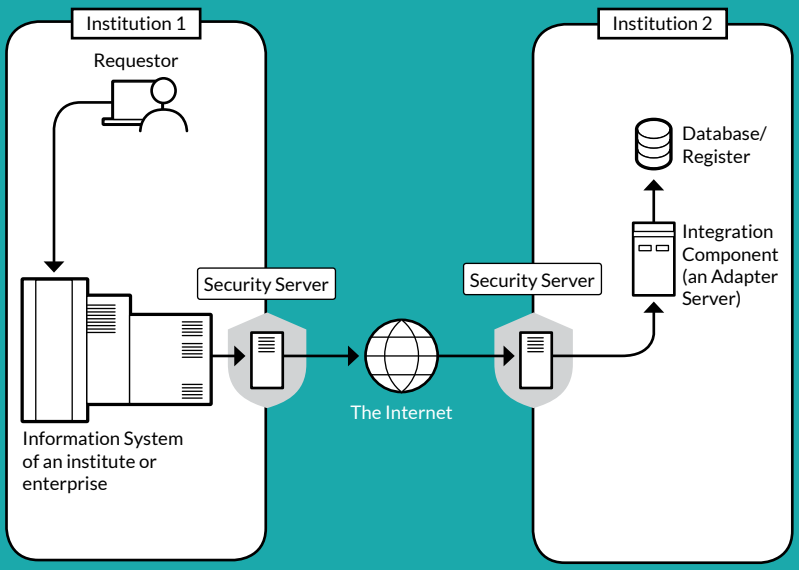
[20] "Electronic Consent Framework", Ministry of Electronics and Information Technology (2016), retrieved October 25, 2019 from http://dla.gov.in/sites/default/files/pdf/MeitY-Consent-Tech-Framework v1.1.pdf.

requested by the data requestor, the purpose of such collection, the identity of this data requestor, all notifications related to the transaction, and the timestamp and authentication details of the transaction.

This model, as it stands, has several pros. First, the steward is data blind and does not have access to user data. This means, that the steward can, at no point, exploit data to create products and provide services. Second, the steward plays a transient and temporary role, and its usage can be terminated easily by the users. Finally, the role of the steward is limited to transferring data, it makes no decisions about purpose, and leaves those decisions entirely to user, who is empowered to give and revoke consent as required. However, there are several cons as well. It is unclear whether the user will be able to view the data once consented. There are also challenges to large scale adoption as the consent burden is squarely on users. Consent management is the only service offered by Account Aggregators, which may make sustainability difficult. With these pros and cons in mind, Account Aggregators can be explored as possible models of data stewardship for specific use cases.

The Estonia Data Exchange Layer

Estonia has set up a data exchange for governance services and citizen information that users may access through multiple service points.[21] Datasets contained in the system are not centralized. Entities using the system and users logging on at the nodes will retain data locally.

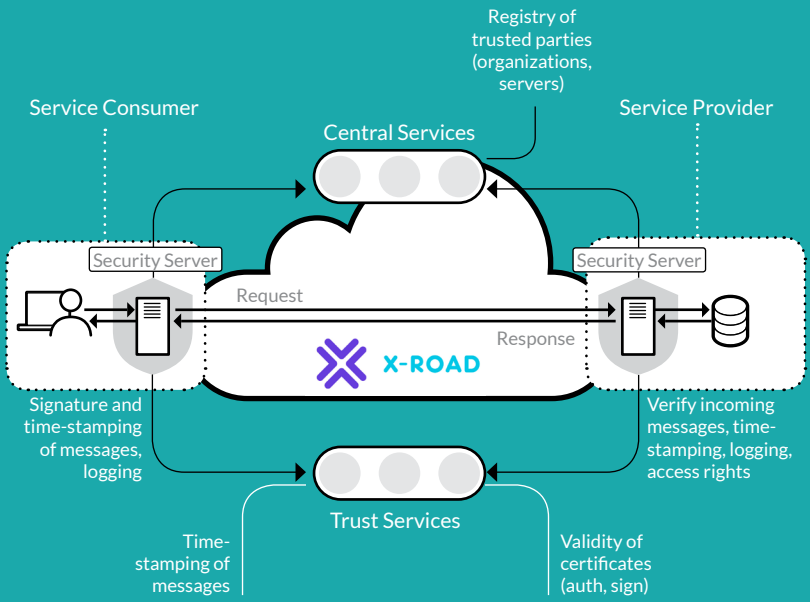


Source: Republic of Estonia Information System Authority[22]

[21] "e-Estonia X-Road", Enterprise Estonia, retrieved October 25, 2019 from https://e-estonia.com/solutions/interoperability-services/x-road/.

[22] "Introduction of X-tee", Estonian Information System Authority (2016), retrieved October 25, 2019 from https://www.ria.ee/en/state-information-system/x-tee/introduction-x-tee.html.

However, the system is centrally managed. The central log maintains a set of identities in order to carry out verification for all transactions and transfers. All outgoing data is digitally signed and encrypted, and all incoming data is authenticated and logged. Similar systems of data management have been implemented in Finland, Kyrgyzstan, Faroe Islands, Iceland and Japan.



Source: X-Road as a Platform to Exchange MyData, MyData Journal [23]

[23] "X-Road as a Platform to Exchange MyData", Petteri Kivimäki, MyData Journal (31August, 2018), retrieved October 25, 2019 from https://medium.com/mydata/x-road-as-a-platform-to-exchange-mydata-d1e9f250a89a.

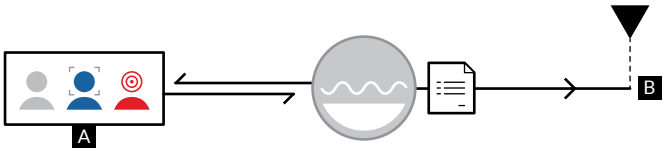
2

Data Trusts

Data Trusts are a stock and flow mechanism, where use and access is decided by a trusted intermediary. They are a model of stewardship that involve storage and collection of data under the direction of an underlying policy or legal framework (see Figure). The model attempts to create an architecture where data is under the control of an entity tasked with certain obligations, a trustee, who does not have a vested interest in access to the data. The obligations of the Data Trust are codified in a policy governing its formation and functions. The obligations defined by this policy have the backing of the legal concept of a trust where the beneficiaries (in this case end users) can enforce their rights against the trustee, i.e. the data steward.

The Data Trust is the only model of stewardship with a defined framework of legal oversight included as part of its structure. Here, data is stored "in trust" on behalf of individuals and entities; this data is used and shared according to defined purpose of the trust. Data Trusts are still being piloted and explored, but there are examples of data stewards offering storage services while being available to third parties for use, such as MiData. The Open Data Institute has carried out pilot projects in using mobility data in civic projects to determine issues and outcomes. They recommend purpose-specific policies on legal structures and possible business models.[24] Their reports cover issues of the possible economic effects of data trusts,[25] and possible legal structures to support their development.[26]

[24] "Greater London Authority and Royal Borough of Greenwich Data Trust Pilot", Ed Parkes, Sonia Duarte & Rachel Wilson, Open Data Institute (2019), retrieved October 25 from https://docs.google.com/document/d/1QT1sA9LFDJZBChfEPdZE4ZENManoMaOsJx1A-gq30d0/.
[25] "Independent assessment of the Open Data Institute's work on data trusts and on the concept of data trusts", London Economics, 2019, retrieved November 30, 2019 from https://theodi.org/wp-content/uploads/2019/04/Datatrusts-economicfunction.pdf.
[26] "Extended ODI Data Trust Report 5: Further use cases to consider", Open Data Institute, 2019, retrieved November 30, 2019 from https://theodi.org/wp-content/uploads/2019/04/BPE_PITCH_EXTENDED_ODI-FINAL.pdf.



A. Data is shared and accessed by participating entities and individuals
B. Data accessed by third parties based on policy and governing forms of the trust

Figure 9. Graphical representation of the Data Trusts

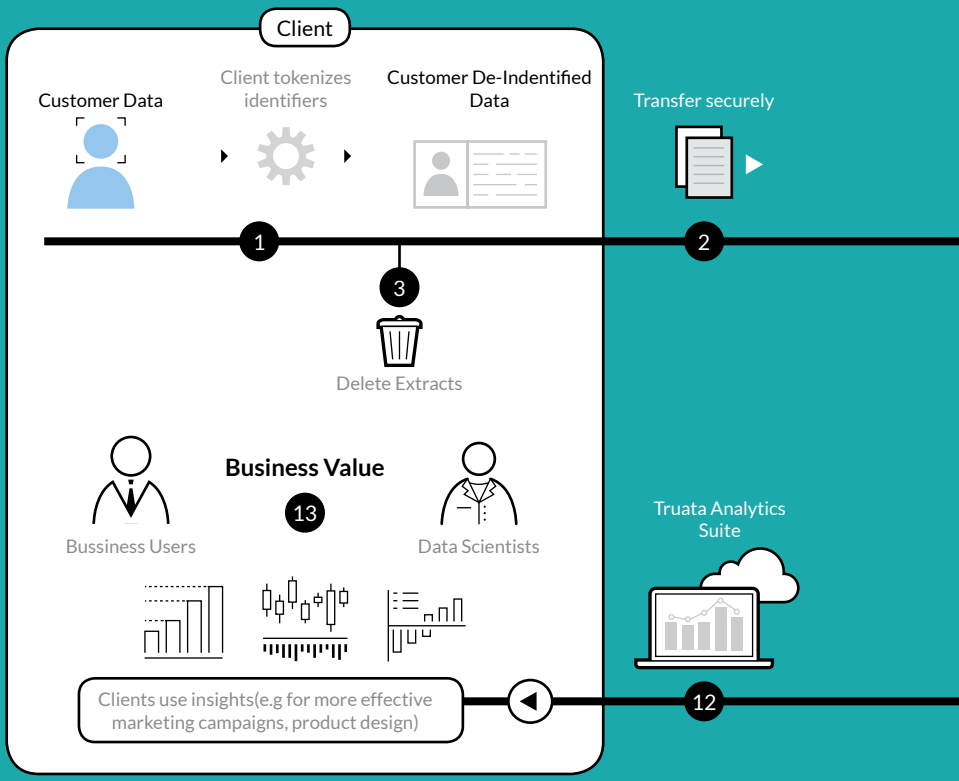
Use cases for data trusts can vary across public and private purposes. They can be established for specific public purposes, such as civic projects or e-governance. Examples of this are ODI's pilot on urban mobility and waste management which offer data-based solutions for both these civic services to better manage processes and reduce costs. Data Trusts can also be setup for specified private interests, as in Truata's model of anonymised data-holding and analytics services. If there are existing datasets available, Data Trusts can leverage anonymised data to carry out research in the relevant sector. They can be used by companies for analytics and research, as per the data usage policies of the steward.

The legal structure of a Data Trust allows large datasets to be shared in order to benefit the person sharing their data. The idea of a data trust draws from the traditional concept of a trust – property (in this case data) held and managed by a designated trustee (the data trust) in the interest of certain beneficiaries (users). The trust allows the use of existing legal structures to hold data trusts accountable by invoking their fiduciary responsibility to settlors, or in this case, the user-beneficiaries.

The specific features of the legal structure are determined by multiple factors: the nature of the data being shared; the purpose of the data trust itself; beneficiaries of the data trust; the level of access of data envisaged.

The terms establishing the data trust define its function and powers and this makes it possible to develop different models of data trusts for different purposes. Trusts can vary in terms of specific rights and responsibilities regarding access to data, usage, and user control.

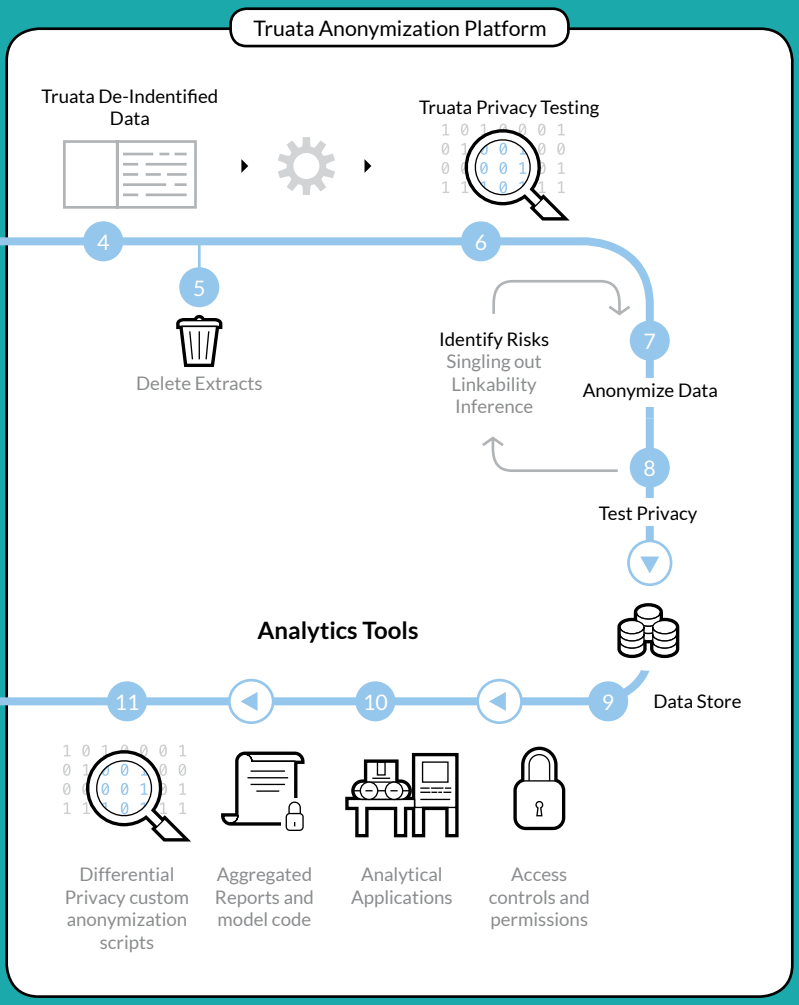
Data Trusts have several pros. First, the appointment of Trustees means that specific entities have an enforceable fiduciary responsibility to users. Second, Trusts are purpose specific, and users can make informed decisions on which Trusts to use for what purposes. Third, Trusts can also enable collective bargaining, as they stock data of multiple users, and make decisions on behalf of users on purpose and use of this data. That said, the Trusts are not a perfect model and some open questions on their efficacy remain. There are questions on large scale adoption of Data Trusts, given that Trust laws differ in various countries. Trusts are currently an intellectual exercise which exist in controlled pilot stage implementations – it is difficult to ascertain their functionality outside these environments.



Trūata Data Anonymisation Solution

Truata offers a data management and analytics service to help clients leverage data while protecting customer privacy. This is done by companies in possession of data removing identifiers from their dataset and transferring this data to Truata. The service carries out its independent anonymisation procedure in order to remove any possibility of using the data to trace back to an individual. Truata then carries out analytics on this anonymised dataset and provides outcomes of the analysis back to the client company. The value offered by Truata is in its privacy-safe service of data analytics through an anonymisation solution.[27]

[27] "Truata Anonymisation Solution brochure" (2018), Truata Limited, retrieved October 25, 2019 from https://www.truata.com/wordpress/wp-content/uploads/2019/03/Truata-Anonymization-Solution-brochure.pdf.



Source: Truata Anonymisation Solution brochure[28]

[28] Truata Anonymisation Solution, Truata Limited, retrieved October 25, 2019 from https://www.truata.com/solution/how-it-works/.

3

Personal Data Stores

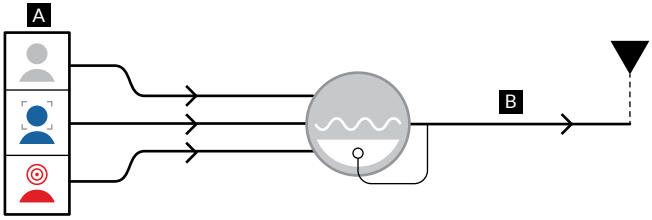
Personal Data Stores ("PDSs") are a stewardship model where access to data is managed by the individual whose data it is, and are applicable for stocks and flows of data (see Figure). PDS' allow individuals to have greater control over their data, and leverage it for empowerment and convenience. While most PDS are relatively early stage and not mature at present, there are a variety of potential uses cases for them in relation to SPI, financial, health data, etc. In PDS, data are held with the individual the storage holds and aggregates the data for the user.

Personal Data Stores offer a single storage point for an aggregated set of social media information from various platforms. Applications which sit on individuals' devices are able to interface with the PDS and other applications. A PDS can also be used to aggregate financial data such as bank statements, credit scores etc., or health data that involves anything from health records to daily fitness activity. It can be used to aggregate personally relevant documents such as passport details, flight details, credit card details, etc.

Use cases for Personal Data Stores are similar to private Data Trusts, because they accord users with a high degree of specific control over usage and sharing of data. They find use as a personal stock of data as well as a sharing mechanism. The comprehensive controls over data sharing ensure that users know the exact mode of sharing data with third parties. This empowers users to benefit from their data whilst businesses rely on this data.

Personal Data Stores are significantly empowering for individuals as it simplifies self-control of data and decisions. The Personal Data Store is mainly a managing consent on behalf of the users, and stores user data safely. Its role, like the Account Aggregators is transient. It also empower

users to make decisions of their data, including drawing monetary and other benefits from it. However, in most countries Personal Data Stores are not regulated, and their role and potential is not entirely understood. The idea of the data marketplace many Personal Data Stores are structured to facilitate might wrongly incentivize users to sell their personal data. While several Personal Data Stores now exist, their use as a data stewards needs to be explored further. As things stand, they are largely instruments of individual control over personal data.

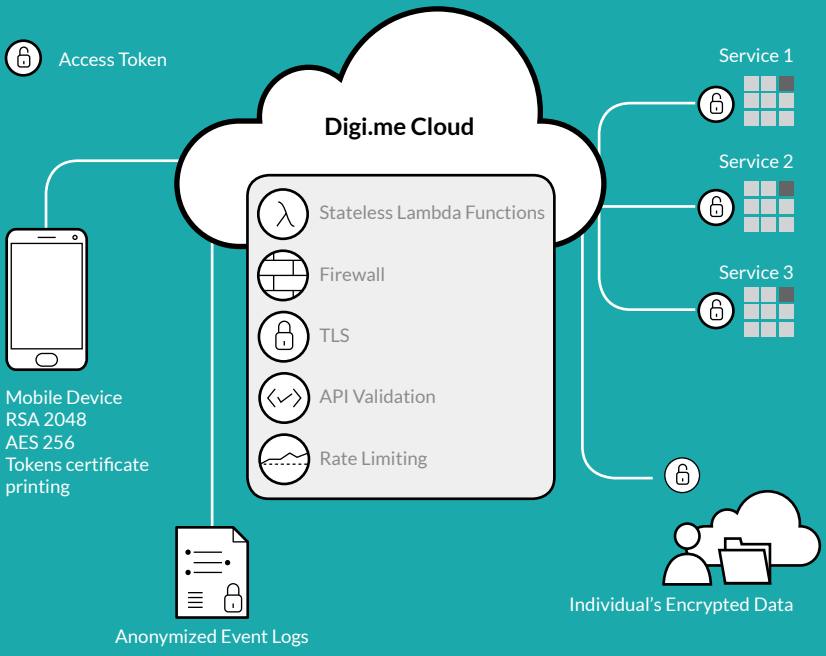


A. Individual shares his/her data with personal data storage
B. Data can be sold/shared with third parties or simply stored in the PDS

Figure 11. Graphical representation of the Personal Data Store

Digi.me Private Sharing Application

The Digi.me "private sharing app" allows users to upload and store their data securely, and exercise control over the exact type of data shared and the entities it is shared with.[29] The service provides a central point for users to manage the sharing and usage of their data by third parties. The platform also connects with third party companies that are able access user data through the platform, thus providing value to all stakeholders that form part of the data sharing chain. The solution provided is a sharing mechanism for the data market that maximizes users' privacy and their control over personally relevant data



Source: Digi.me Private Sharing Technical Overview[30]

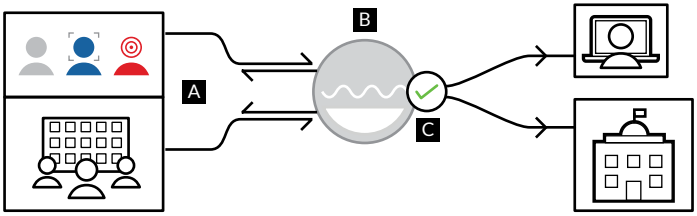
[29] "What is Digi.me", digi.me Limited, retrieved October 25, 2019 from https://digi.me/what-is-digime/.

[30] "Digi.me Private Sharing Technical Overview", digi.me Limited, retrieved October 25, 2019 from https://digi.me/user-centric-architecture/.

4

Data Exchanges/Collaboratives

Data exchanges/collaboratives and are a mechanism for stock, as well as stock and flow, and a nominee or steward makes decisions on use and access. They aim to share data amongst participating entities for a shared goal or purpose and create value, allowing for the creation of public value by looking at issues from different data perspectives (see Figure). Large existing data sets are pooled, either for public benefit or commercial purposes. Private companies hold data that can be unlocked for greater value. GovLabs and New York University have highlighted the value of Data Collaboratives for public value. The World Bank and the UN Office for the Coordination of Humanitarian Affairs have set up data collaboratives and demonstrated their value for development.[31]



A. Data is shared and accessed via participating entities
B. Collaborative/Exchange stewards data
C. Data can be accessed by others as well (consent granted by collaborative/exchange

Figure 13. Graphical representation of the Data Collaborative

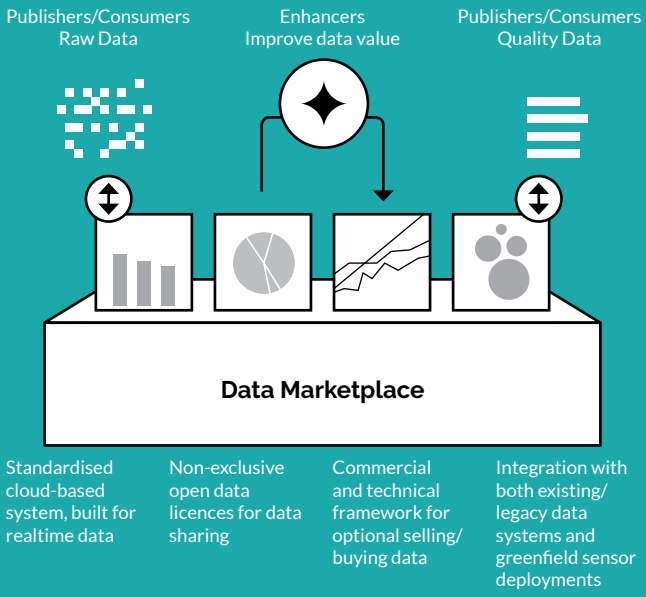
[31] "Humanitarian Data Exchange", United Nations Office for the Coordination of Humanitarian Affairs, retrieved October 25, 2019 from https://data.humdata.org/.

At present, data collaboratives range from pilot stages to advanced. Among stewardship models, collaboratives are the most well-used and evolved framework. Data exchanges have existed for longer than other models and are now in early to advanced stages (eg., the Financial Data Exchange, Chicago Data Collaborative).

Data collaboratives/exchanges are the most well-tested model of stewardship in the world. They are deployed both for private gains and public good, and have demonstrated their value to businesses. Several prominent use cases can be studied across the world, and therefore, adoption of exchanges/collaboratives as a model for stewardship is easier to institutionalize. At the same time, exchanges/collaboratives are also open to misinterpretation and misuse, especially since there are no legal and regulatory guidelines around them. Exchanges/collaboratives need to be explored as potential models of stewardship that can safeguard the interests of users while unlocking data.

The oneTRANSPORT Data Marketplace

The service, based in the UK, offers a platform for companies and organisations to upload datasets to be discovered and leveraged by others.[32] The data provided by these companies can then be monetised and also be used by transport authorities. The transport authorities themselves manage their data through usage of the platform as well. The service provides analytics toolkits for data providers on the platform to derive greater value from the data already in their control. This incentivises more companies and transport agencies to adopt the platform and make their datasets accessible to other users. The aggregated dataset on the platform allows authorities to make use of it for better municipal solutions.



Source: oneTRANSPORT Guide to Data Marketplaces[33]

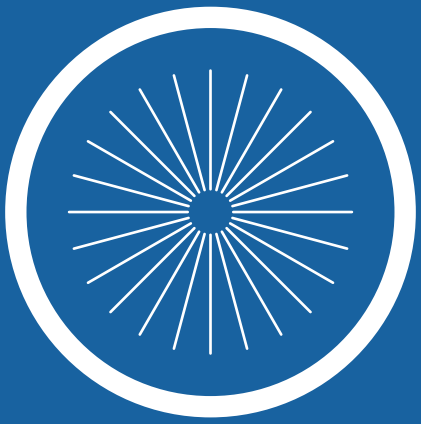
---

[32] Beginners Guide to the oneTRANSPORT Data Marketplace Service Portal, accessible at http://onetransport.io/wp-content/uploads/2019/01/Beginners-Guide-to-the-oneTRANSPORT-commercial-v0.5-VS-final1.pdf

[33] oneTRANSPORT Guide to Data Marketplaces, accessible at http://onetransport.io/wp-content/uploads/2019/01/Quick-Start-Guide-double-1-24_V2.pdf.

Section

5



Stewarding Data in India

Examples and Pilots in India

Data Stewardship, while still at an early stage in India as in much of the world, is starting to be applied. A major such example is the Indian Urban Data Exchange (IUDX).[34] The IUDX provides is a platform that connects providers and users of data. The data made available on the IUDX is described by the platform as non-PII. The intent of the project is to make urban datasets accessible to a wider set of entities in order to maximize value from data that currently lies in silos. As per the stewardship framework, the IUDX project performs functions of a data flow.

In addition, India has seen regulatory activity on Account Aggregators, led by the Reserve Bank of India guidelines on their establishment and operation.[35] Eight Account Aggregators that have obtained license to operate from the Reserve Bank of India, but the impact of the service on user privacy and convenience is yet to be measured.[36]

Legal and Policy Landscape in India

The legal landscape in India has been taking steps in recent times to adapt to changing modes of data regulation. Chief amongst these efforts is the Personal Data Protection Bill, 2018, in which is likely to be law by early 2020.[37] The Bill was released alongside an in-depth report (the "Srikrishna

[34] "Indian Urban Data Exchange", Ministry of Housing and Urban Affairs, retrieved October 25, 2019 from https://www.iudx.org.in.
[35] "Account Aggregator Directions", Reserve Bank of India (2016), retrieved October 25, 2019 from https://rbidocs.rbi.org.in/rdocs/notification/PDFs/MD46859213614C3046C1BF9B7CF563FF1346.PDF.
[36] "Account Aggregators in India", DigiSahamati Foundation, retrieved November 30, 2019 from https://sahamati.org.in/account-aggregators-in-india/.
[37] "The Personal Data Protection Bill, 2018", Indian Ministry of Electronics and Information Technology (2018), retrieved October 25, 2019 from https://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf.

Committee Report") charting out the vision for an operational data economy.[38] Significantly, both the Bill and the report adopt the framework of data fiduciaries and a consent-centric regulatory approach.

As another point of interest, the report attempts to identify "community data" as a natural resource and defines it as "a body of data sourced from multiple individuals, over which a juristic entity may exercise rights". Non-personal data and community data are increasingly becoming topics of relevance in the Indian discourse on data regulation. The earliest significant regulatory effort in this regard was the National Data Sharing and Accessibility Policy, 2012.[39] The policy set out guidelines for governmental agencies with respect to data handling and sharing. The policy intended to make public data more openly accessible. It prescribed processes for access to data collected and controlled by government agencies with different levels of restrictions.

The concept of "Community Data" also finds mention in the 2019 draft for the National E-Commerce Policy released by the Department for Promotion of Industry and Internal Trade.[40] The policy attempts to make commercial datasets generated in India available to Indian companies and SMEs. It attempts to do this through measures such as provisions on data localization and mandatory data sharing.
Most recently, MEITY has constituted a Committee of Experts to Deliberate on a Data Governance Framework.[41] The Srikrishna Committee Report had signaled that specific regulations on non-personal data would be required,

[38] "A free and fair digital economy: Protecting privacy, empowering Indians", Committee of Experts under the Chairmanship of Justice B.N. Srikrishna (2018), accessible at https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf.
[39] "National Data Sharing and Accessibility Policy", Indian Department of Science and Technology (2012), retrieved October 25, 2019 from https://nsdiindia.gov.in/nsdi/nsdiportal/meetings/NDSAP-30Jan2012.pdf.
[40] "Draft National E-Commerce Policy", Department for Promotion of Industry and Internal Trade (2019), retrieved October 25, 2019 from https://dipp.gov.in/sites/default/files/DraftNational_e-commercePolicy_23February2019.pdf.
[41] "Office memorandum: Constitution of a committee of experts to deliberate on data governance", Ministry of Electronics and Information Technology (2019), retrieved October 25, 2019 from https://www.medianama.com/wp-content/uploads/data-governance-framework.pdf.

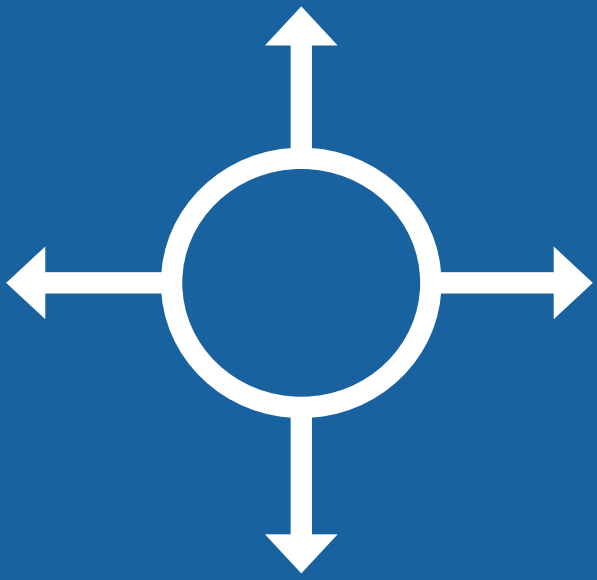
and the formation of the MEITY committee is an initial is a step in this direction. Further clarity on this issue is expected once the further plans and activities of the committee become clearer.

Applicability of Indian Trust Law

Beyond data related regulation, it is also pertinent to examine Indian Trust Law. Indian laws do not recognise trusts as a separate legal entities, but rather as an obligation of the trustee to hold and own property, and manage the property for the benefit of specified beneficiaries. The trust framework is seen as a set of obligations arising out of ownership of property which are set by the owner (known as the settlor). This means that a data trust must be set up on the basis of the purpose which it is designed to serve. It cannot be defined as a trust merely in terms of an entity, but must be attached to certain obligations and a set of third-party interests that it serves (the users, in this case). While the application of trust law in the Indian jurisdiction is an application of common law principles through judicial interpretation rather than based solely on the legislation, public trusts may nonetheless be explored as an area of interest as the object of public trusts can be varied: promotion of education, art, culture, spreading spiritual awareness and objects of public good for the benefit of a large section of the public. This opens up the possibility of designing a data trust in a similar mould with a defined purpose for an aspect of public good from a variety of sectors.

Section

6



Way forward

This report explores the landscape of stewardship and aims to arrive at a common taxonomy for stewardship models. In the process of doing so, it deep-dives into models of stewardship, their use cases and potential future use. Our framework of analyses stewards through access and use decisions and roles of the steward, and a number of potential data stewardship models.

Through this research, one thing has been made certain – the need for data stewardship is pressing, and there is a significant opportunity to unlock societal value of data – while safeguarding the privacy and control of individuals. At the same time, there are significant unanswered questions with regard to stewardship. Most critically, there is a lack of clarity on the governance structures on stewards as well as the business models, which are both important to ensure that the steward is responsible, fair, and ethical. Related to this, there is a need to further probe the principles or qualities of a "good steward".

The next phase of inquiry on data stewardship will aim to define these characteristics of a "good steward", given that stewardship is likely to exist in the private and public sector, there is a need to understand responsible data handling practices, regulations and values of a good steward. This phase will also focus on codifying these principles, and socializing them into a widely accepted guidelines/checklist for "good data stewardship". Further, appropriate use cases for data stewardship will be identified in India/South East Asia.

To do this, we aim to set up the Data Economy Lab, a hub for data stewardship. The Lab will map and build principles for good stewardship, contemplate responsibility and regulations for data stewards, as well as governance and business models. It will also focus on data stewardship governance on the lines of user-centric function and account for data minimization, storage limitations, and purpose limitations. It will also socialize the idea of stewardship within the ecosystem, and build a

community of concern to think about data stewardship in different contexts. The Lab will also discover opportunities and use cases, and look to incubate pilots on stewardship with key stakeholders over time.

Thereafter, the Lab will delve into governance and technical processes across various data types (PII, non-PII, inferred data, machine data, etc.) and suitable models of stewardship for different use cases will be explored. Related to this, the Lab will also define appropriate governance and technical solutions for different stewardship models such that lawfulness, fairness, and transparency are maximized.

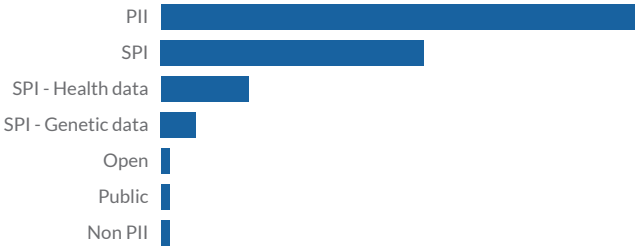
The Data Economy Lab will pull together different strands of concerns, opportunity, and experiments on stewardship into one place.

Annexes

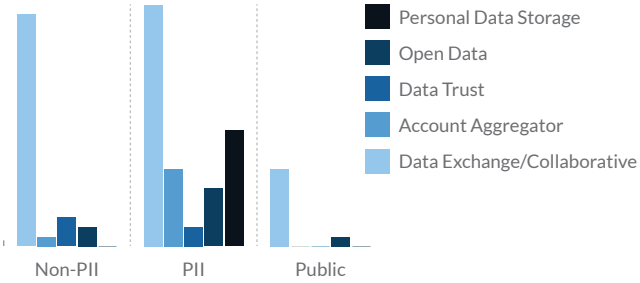
Annexe 1: List of Interviewees

Expert	Organisation
Pramod Verma	Ekstep/ISPIRT
Saurabh Panjwani	ISPIRT
Prof. Bharadwaj Amrutur	IISc
Prof Kapil Vaswani	IISc/MSR
Inder S Gopal	IISc
Arun Babu	IISc
Parminder Singh	IT for Change
Jeni Tennyson	Open Data Institute (ODI)
Saikat Guha	Microsoft Research
Sean McDonald	Digital Public
Stefaan Varhulst	GovLabs

Annexe 2: Graphical Representation

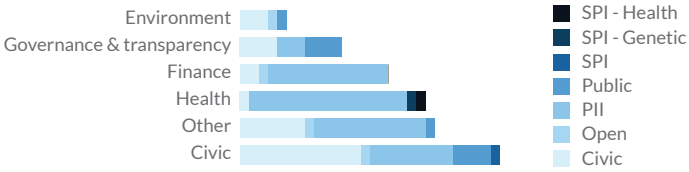


Type of Data Stewarded

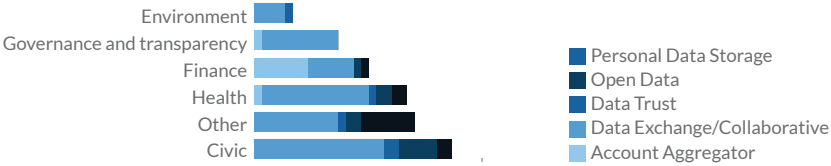


Type of Data by Model Type

Based on database of 100+ data stewardship use cases



Sector of Use Case by Data Type



Sector of Use Case By Model Type

Type of Model by Where Use Case Was Innovated

Notes

aapti institute 2019

Aapti is a research institution generating public, policy-relevant, actionable and accessible knowledge on the frontiers of tech and society, about our networked lives, to support the creation of a fair, free, and equitable society.

AUTHORS

Siddharth Manohar, Astha Kapoor, Aditi Ramesh


contact@aapti.in


www.aapti.in


medium.com/@aapti