

Enabling data sharing for social benefit through data trusts

Apti Institute and the Open Data Institute for the Global Partnership for Artificial Intelligence 2021

Output 1: Data trusts global survey

Table of contents

Executive Summary	2
1. The emergence of data trusts	4
1.1. Data stewardship	4
1.2. Bottom-up data stewardship	7
1.3. Data trusts: an evolving conceptual framework	10
1.4. Differing interpretations of data trusts	12
1.5. Institutionalising data trusts and codifying fiduciary responsibilities	14
2. International knowledge, attitudes and practices of data trusts	19
2.1. Context	19
2.2. Awareness and understanding	20
2.3. Practices	22
2.4. Attitudes towards data trusts	26
3. Case studies	27
3.1. Criteria and selection	27
3.2 Driver's Seat	28
3.3. Open Humans	31
3.4. MIDATA	34
3.5 Insights from case studies	37
4. Key findings and takeaways	38
5. Endnotes	39
5.1 About Aapti, ODI, and GPAI	39
5.2 Authors	39
5.3 Report drafting	39
5.4 Acknowledgements	39
6. Bibliography	40
6.1 Review of literature (academic papers, blogs, comments, news reports)	40
6.2 Policy, regulation and strategy documents	48
6.3 Tools, guides and videos	49

Executive Summary

Progress in artificial intelligence (AI) requires access to data. Who shares data, for what purposes and under what conditions will therefore shape the development of AI and the challenges it is put to.

However, many organisations currently regard data as something to hoard, causing it to be inaccessible to those who could otherwise use it to create new products or insights. At the same time, **a lack of involvement of individuals and communities in shaping how data is used will deny beneficial uses of data**, due to people withdrawing their consent - in the broad sense - for its collection and sharing.

In response, **data stewardship** has emerged as a responsible, rights-preserving and participatory concept. It aims to unlock the economic and societal value of data, while upholding the rights of individuals and communities to participate in decisions relating to its collection, management and use.

In this context, **this research set out to understand global knowledge, attitudes and practices of data trusts**. It was undertaken for the Global Partnership on AI (GPAI) by Aapti and the Open Data Institute between August and October 2021, using a combination of literature review, survey and case studies. It adopted the GPAI's Data Governance Working Group interpretation of data trusts as:

“a form of data stewardship that supports data producers to pool their data (or data rights) with the aim of collectively negotiating terms of use with potential data users, through the oversight by independent trustees, with fiduciary duties, and within a framework of technical, legal and policy interventions that facilitate data use and provide strong safeguards against mis-use”.

The project's literature review (Section 2) found significant theory, interest and experimentation around new forms of 'bottom-up' data stewardship that seek to empower people to participate in the process of data collection, use and sharing. The analysis frames **data trusts as a particular, evolving form of bottom-up data stewardship**. It found emerging consensus on distinctive features of data trusts and that practitioners deploy a variety of operational strategies to realise its functions.

A survey (Section 3) was completed by 45 people building or running data trusts and similar bottom-up data stewardship initiatives, or who are working on data stewardship and related topics. Analysis of the survey (summarised in Section 5) found that:

- **today's data stewardship projects deliver many of the functions associated with data trusts**, but delivering all the functions attributed to a data trust, as per the GPAI Data Governance Working Group's interpretation, remains a challenge
- **there is general optimism about the potential of data trusts** among people

working on data stewardship.

- the **interest in data trusts as a form of data stewardship seems to be concentrated in Europe and North America.**
- There are **a number of real-world initiatives that demonstrate multiple routes to realising bottom-up data stewardship that do not follow the data-trust definition or deliver all of the functions associated with data trusts.**
- The **purpose for bottom-up data stewardship can differ significantly**, from supporting altruism to generating commercial return and this defines how models design their governance mechanisms.

The case studies (Section 4) document three bottom-up data stewardship initiatives: Driver's Seat, Open Humans and MIDATA. They represent **real-world examples of how groups can be empowered around data they've generated** and are actively making available data for broad societal benefit.

The intent is for this report to act as a reference point on the subject of data trusts for practitioners seeking inspiration, as well as policymakers, funders and other enabling actors considering how to support the field.

1. The emergence of data trusts

- This section presents a review of literature on data stewardship and the evolution of the concept of data trusts, discussing the work of Sylvie Delacroix and Neil Lawrence, Sean McDonald and other scholars, and organisations such as the Mozilla Data Futures Lab and the Ada Lovelace Institute.
- The review documents significant theory, interest and experimentation around new forms of 'bottom-up' data stewardship that seek to empower people to participate in the process of data collection, use and sharing.
- The review frames data trusts as a particular, evolving form of bottom-up data stewardship, and found rapid proliferation around the use of the term among practitioners and scholars of stewardship.
- While there is divergent opinion from around the world on how data trusts could be constructed, the stewardship community is nevertheless working to consolidate their understanding of the term.

1.1. Data stewardship

The effective collection, use and sharing of data can help address the pressing challenges of our time - from surfacing remedies for climate change¹ to improving public health².

The transformative power of data is best explained through the lens of examples such as the Human Genome Project, undertaken between 1990 and 2003.³ Led by the US Government's National Institute of Health, the Project made available data on DNA sequencing within 24 hours of its discovery. The consequent availability of that data for research and development has not only saved lives, but also generated \$796 billion in economic impact and supported over 300,000 jobs in 2010 alone.⁴ More recently, the same data sharing norms established by the Human Genome Project (the 'Bermuda principle')⁵ were adopted in the development of vaccines against SARS-COV-2.⁶ A lab in

¹ Szasz, Open Data Institute (2020), "Tackling Climate Challenges through Data Access: Microsoft and the ODI", <https://theodi.org/article/tackling-climate-change-challenges-through-data-access-microsoft-and-the-odi/>

² Harper, International Journal of Infectious Diseases (2016), "Sharing Public Health Data Saves Lives", [https://www.ijidonline.com/article/S1201-9712\(16\)31285-1/fulltext](https://www.ijidonline.com/article/S1201-9712(16)31285-1/fulltext)

³ See <https://www.genome.gov/human-genome-project/What>

⁴ Tripp and Greuber, Battelle Memorial Institute (2011), "Economic Impact of the Human Genome Project", <https://www.battelle.org/docs/default-source/misc/battelle-2011-misc-economic-impact-human-genome-project.pdf?sfvrsn=6>

⁵ See <https://dukespace.lib.duke.edu/dspace/handle/10161/7407>

⁶ First and Collins, Forbes (2021), "NIH Director Dr. Francis Collins: Connecting The Dots From The Human Genome Project To The COVID-19 Vaccine", <https://www.forbes.com/sites/billfrist/2021/01/20/nih-director-dr-francis-collins-connecting-the-dots-from-the-human-genome-project-to-the-covid-19-vaccine/?sh=738447175438> [Podcast]

China released the genome sequence of the coronavirus in January 2020, which was subsequently used by researchers around the world to develop antidotes, even without access to physical genomic samples of the virus.

However, despite the positive benefits of data, the emergence of new approaches to its collection, use and sharing - particularly driven by developments in machine learning⁷ - is underpinned by two disconcerting trends.

First, the market-driven imperatives of corporations have helped create digital enclosures⁸ that hampers the ability to use data for broad-based social benefit. Much of the current data economy is defined by a paradigm of extraction⁹, whereby the role of individuals and communities as the generators of data goes unrecognised.¹⁰ This process of has been described varyingly as the “attention economy”,¹¹ “surveillance capitalism”¹² and “computational capitalism”¹³, with corporations’ use of data existing beyond the control of those individuals and communities.¹⁴ For instance, patients signing-up to digital health applications have little knowledge of who has access to their data and how it will be used,¹⁵ just as rideshare drivers in the gig economy are excluded from the management of algorithms that govern their work.¹⁶

Second, a lack of consideration of ethics and equity, and a lack of engagement with those affected by data’s use, undermines trust in the process of data sharing. In the UK, the Government came under scrutiny for its GP Data for Planning and Research

⁷ Expert Panel, Forbes Technology Council (2019), “15 Social Challenges AI Could Help Solve”, <https://www.forbes.com/sites/forbestechcouncil/2019/09/03/15-social-challenges-ai-could-help-solve/?sh=76e9dd973533>

⁸ Andrejevic (2009), Amsterdam Law Forum, “Privacy, Exploitation and the Digital Enclosure”, https://www.researchgate.net/publication/228226821_Privacy_Exploitation_and_the_Digital_Enclosure

⁹ Morozov, E., (n.d.), Council of Europe, “Digital Intermediation of Everything: At the Intersection of Politics, Technology and Finance”, <https://rm.coe.int/digital-intermediation-of-everything-at-the-intersection-of-politics-t/168075baba>

¹⁰ Manohar, Kapoor and Ramesh (2020), Aapti Institute, “Data Stewardship: A Taxonomy”, <https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/>

¹¹ Beuno (2017), London: Rowman and Littlefield International, “The Attention Economy: Labour, Time and Power in Cognitive Capitalism”

¹² Zuboff (2018), New York: PublicAffairs, “The Age of Surveillance Capitalism: The Fight for Human Future at the New Frontier of Power”

¹³ Beller (2018), London: Pluto Press, “The Message is Murder: Substrates of computational capital”

¹⁴ Lawrence (2016), The Guardian, “Data trusts could allay our privacy fears”, <https://www.theguardian.com/media-network/2016/jun/03/data-trusts-privacy-fears-feudalism-democracy>

¹⁵ Sur (2021), Medianama, “Online medical platforms are playing fast and loose, collecting patient data”, <https://www.medianama.com/2021/09/223-india-digital-health-medical-platforms-data-consent-records/>

¹⁶ O’Connor (2016), Financial Times, “When your boss is an algorithm”, <https://www.ft.com/content/88fdc58e-754f-11e6-b60a-de4532d5ea35>

(GDPR) proposal, which would facilitate access to the health records of 55 million people.¹⁷ The proposal's introduction was postponed, having been subject to criticism for not giving patients a meaningful say in how the system should work.¹⁸ Equally, high profile data breaches such as the 2017 Equifax data breach, where an unauthorised third party gained access to data on as many as 143 million Americans, serve to erode the trust we have as consumers in the processing of data about us.¹⁹

These trends have given rise to inter-related phenomena - 'data hoarding' and 'data fearing'²⁰. 'Data hoarding' relates to a scenario where organisations restrict access to data due to misperceptions about its value to their operations or the risks associated with data sharing. The benefits of data collection and use would only be enjoyed by a few, while the negative impacts of its use would affect society as a whole.²¹ On the other end is the scenario of 'data fearing', where data might not be collected or used to the extent it could, due to concerns about the harm that it can cause people being left unaddressed. People might avoid using services, or withdraw consent for data to be collected, which means that we end up missing data and the uses of it that could support human flourishing.

The concept of data stewardship is a response to these 'data hoarding' and 'data fearing' scenarios. Data stewardship can be understood as an approach to data governance that is responsible, rights-preserving and participatory in nature²². In effect, data stewardship aims to unlock the societal value of data, while upholding the data rights of individuals and communities to participate in decisions relating to its collection, management and use.²³

The development of machine learning and artificial intelligence is contingent on the practice of responsible data stewardship which aims to enable meaningful participation

¹⁷ Vallance (2021), BBC News, "GP Data Sharing: What is it and can I opt out?", <https://www.bbc.com/news/technology-57555013>

¹⁸ Crouch (2021), Digital Health, "GP Data September implementation date is scrapped", <https://www.digitalhealth.net/2021/07/gdpr-september-implementation-date-scrapped/>

¹⁹ Forbes (2017), "Equifax Data Breach Impacts 143 Million Americans", <https://www.forbes.com/sites/leemathews/2017/09/07/equifax-data-breach-impacts-143-million-americans/#6f6ed8d3356f>

²⁰ Open Data Institute (2021), "What are data institutions and why are they important?", <https://theodi.org/article/what-are-data-institutions-and-why-are-they-important/>

²¹ Newman (n.d.), Federal Trade Commission, "How Big Tech enables harms to consumers, especially to low-income and other vulnerable sectors of the population", https://www.ftc.gov/system/files/documents/public_comments/2014/08/00015-92370.pdf

²² Ada Lovelace Institute (2021), "Disambiguating data stewardship", <https://www.adalovelaceinstitute.org/blog/disambiguating-data-stewardship/>

²³ Manohar (2019), Aapti Institute, "Responsible data sharing for public good: Theoretical bases and policy tools", <https://thedataeconomylab.com/2020/07/31/data-sharing-for-public-good-theoretical-bases-and-policy-tools/>

in data governance. Without it, practitioners may find themselves unable to access data required to infer patterns, inform analytics²⁴ and develop new algorithms. Stewardship is critical for building trust in the creation and use of AI as it involves people and communities in questions on the use and value of data. Therefore, those working with AI have a responsibility to imagine, test and implement new approaches to stewarding data that unlock the societal value of data while upholding the rights of individuals and communities, and respect community rights' and interests.²⁵

1.2. Bottom-up data stewardship

Corporations, governments and civil society organisations find themselves occupying powerful positions in determining how data is put to use. However, within most current mechanisms for data collection, use and sharing, the involvement of individuals and communities is non-existent.

In response, a more empowering paradigm of 'bottom-up data stewardship' has emerged. It builds on the ideals of data stewardship, recognising individuals and communities as more than recipients of information - or mere providers of consent - about how data about them is used²⁶, and seeks to empower them to participate in the process of data collection, use and sharing.

An ecosystem of research and practice has emerged around this concept of 'bottom-up data stewardship'. The MyData Global community, for example, is set out "to empower individuals by improving their right to self-determination regarding their personal data"²⁷. The Mozilla Data Futures Lab was launched in 2021 to support experimentation around "new approaches to data stewardship that give greater control and agency to people"²⁸. The Ada Lovelace Institute advocates for 'participatory data stewardship', where people whose data is used or about which data decisions are taken are meaningfully involved.²⁹ Aapti Institute's work at the Data Economy Lab³⁰ aims to empower individuals and communities to play a bigger part in data governance, and it has documented numerous examples of this in practice.³¹

²⁴ Digital Curation Centre, Trilateral Research, University of Edinburgh (2020), "Role of Data in AI", <https://gpai.ai/projects/data-governance/role-of-data-in-ai.pdf>

²⁵ Sridharan, Kapoor & Manohar (2021), "Health data stewardship: Learning from use cases", <https://thedataeconomylab.com/2021/09/29/health-data-stewardship-learning-from-use-cases/>

²⁶ Ada Lovelace Institute (2021), "Participatory data stewardship: A framework for involving people in the use of data", <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>

²⁷ MyData (n.d.). Retrieved from <https://mydata.org/>

²⁸ Mozilla (n.d.). Retrieved from <https://foundation.mozilla.org/en/data-futures-lab/>

²⁹ Ada Lovelace Institute (2021). Retrieved from <https://www.adalovelaceinstitute.org/event/exploring-participatory-mechanisms-data-stewardship-report-launch/>

³⁰ Aapti Institute (2021). Retrieved from <https://thedataeconomylab.com/>

³¹ Aapti Institute (2021). Retrieved from <https://thedataeconomylab.com/tracking-stewardship/>

Research suggests that the bottom-up data stewardship initiatives emerging from this ecosystem can be functionally very different, especially in terms of the types of involvement they afford to individuals and communities³². Some initiatives - such as Bitsabout.me³³ - enable people to make granular, individual decisions about who has access to data about them, for what purposes and in exchange for what;³⁴ other initiatives - such as Salus.coop³⁵ and others described below - enable people to participate in collective decision-making as part of a community;³⁶ and a few initiatives - such as Jumbo³⁷ - enable people to delegate another party to mediate data collection and use.³⁸ Research on bottom-up data stewardship has also demonstrated the array of other ways such initiatives can differ.³⁹ Aapti's Stewardship Mapper⁴⁰, which draws on interviews and analysis of 100+ bottom-up data stewardship initiatives, describes nine categories of 'design choices' practitioners can make in constituting them, from business models to technical features.

In particular, a number of promising initiatives have emerged to enable groups to generate or repurpose data about them, and exert collective control over it for a common purpose. For instance:

- Variant Bio⁴¹ works with historically marginalised populations to facilitate people-driven therapeutics. Communities are engaged prior to the beginning of research projects; their data is then collected and used within a framework that focalises community concerns.
- Driver's Seat⁴² is an independent, driver-owned cooperative where members' data is used to derive insights that help them optimise their performance.

³² Hardinges and Keller (2021), The Open Data Institute, "What are "bottom-up" data institutions and how do they empower people?"

<https://theodi.org/article/what-are-bottom-up-data-institutions-and-how-do-they-empower-people/>

³³ BitsaboutMe (2021). Retrieved from <https://bitsabout.me/en/>

³⁴ Digime (2021). Retrieved from <https://digi.me/> ; Schluss Foundation (2021). Retrieved from <https://schluss.org/>

³⁵ Salus Coop (2021). Retrieved from <https://www.saluscoop.org>

³⁶ LunaPBC (2021). Retrieved from <https://www.lunadna.com/> ; Open Data Manchester (n.d.). Retrieved from <https://www.opendatamanchester.org.uk/>

³⁷ Dumbo (2021). Retrieved from: <https://www.withjumbo.com/>

³⁸ Ciitizen Corporation (2021). Retrieved from <https://www.ciitizen.com/> ; UTS-CRiCOS (2021). Retrieved from <https://www.ciitizen.com/>

³⁹ Sridharan, Kapoor & Manohar (2021), "Health data stewardship: Learning from use cases", <https://thedataeconomylab.com/2021/09/29/health-data-stewardship-learning-from-use-cases/>

⁴⁰ Aapti Institute (2021). Retrieved from <https://thedataeconomylab.com/mindmap/>

⁴¹ Variant Bio (2021). Retrieved from <https://www.variantbio.com/>

⁴² Driver's Seat Cooperative LCA (2021). Retrieved from <https://driversseat.co/>

- Swash⁴³ enables users to control what data is collected about their browsing habits, as well as aggregate and sell access to this data to generate financial return.
- OpenHumans⁴⁴ empowers individuals and communities to explore and share their personal data for the purposes of education, health and research.
- MIDATA⁴⁵ enables users to contribute to medical research and clinical studies by granting selective access to their personal data.
- Gyeonggi Data Dividend⁴⁶ ensures that any financial profits generated by selling access to data about transactions using the local currency are returned to citizens in the form of a dividend.

Viljoen's 2020 paper articulates the rationale for reorienting power relationships within the digital economy in favour of communities and to enable them to exercise meaningful control over their data⁴⁷. The paper argues that the process of data collection, use and sharing requires reworking on account of its social effects, whereby "personal choices over data sharing should reflect the effects this choice has on others, not only because of the political and moral benefits of considering others, but also because under current conditions of datafication".⁴⁸ It also builds on Elinor Ostrom's ground-breaking research on the governance of lakes, forests and other common pool resources, which demonstrates how communities can forge institutional frameworks to govern their use in a sustainable and mutually beneficial manner. The resultant theory of self-regulation by communities - termed Ostrom's design principles - holds invaluable insights for assigning collective rights over data.⁴⁹

Overall, bottom-up data stewardship represents an opportunity to subvert existing patterns of storing and sharing data⁵⁰, empowering groups to play an active role in deciding how and the purposes for which data can be used.⁵¹

⁴³ Swashapp.io (2021). Retrieved from <https://swashapp.io/>

⁴⁴ Open Humans Foundation (n.d.). Retrieved from <https://www.openhumans.org/>

⁴⁵ MIDATA (n.d.). Retrieved from <https://www.midata.coop/en/home/>

⁴⁶ Gyeonggi Do (2021), Gyeonggi Province Becomes First Local Autonomy in World to Implement a Data Dividend. Retrieved from: <https://english.gg.go.kr/blog/daily-news/gyeonggi-province-becomes-the-first-municipality-in-the-world-to-implement-a-data-dividend-governor-lee-jae-myung-says-it-is-the-beginning-sign-of-the-era-of-data-sovereignty/>

⁴⁷ Viljoen, S., (2020) Yale Law Journal, Forthcoming. "Democratic Data: A Relational Theory For Data Governance", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727562

⁴⁸ Ibid

⁴⁹ Coyle (2020), Ada Lovelace Institute, "Common governance of data: Appropriate models of collective and individual rights", <https://www.adalovelaceinstitute.org/blog/common-governance-of-data/>

⁵⁰ Sundarajan (2020), Aapti Institute, "Role of data stewards in enhancing accountability", [Role of data stewards in enhancing accountability](#)

⁵¹ Manohar, Kapoor and Ramesh (2019), Aapti Institute, "Data stewardship: A Taxonomy", <https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/>

Data justice: A social justice agenda for the digital age

Growing datafication is a significant feature of contemporaneous capitalism, such that human and economic development have come to be governed by digital footprints that people leave in the wake of their interactions with technology. Consequently, the ways in which data is processed by corporations and governments affect not only the organisation of information, but also people's access to services and ultimately, their autonomy itself.

Scholars such as Linnet Taylor⁵² have draw attention to the "structural discrimination" inherent to intensifying datafication, such that institutions of the state (through population databases and surveillance) and corporations (as dominant entities with accumulated data and processing abilities) function to amplify exclusion and disempowerment of individuals and communities. Elsewhere, Lina Dencik and Anne Kaun⁵³ illustrate the debilitating impact of datafication on the welfare state, leaving citizens with limited bargaining power and agency to control the use of their data.

In such a milieu, reconstituting the conventional agenda of social justice becomes crucial to forge ethical pathways to regulate datafication. Data justice is an expression of this impulse, making "fairness in the way people are made visible, represented and treated as a result of their production of digital data" as crucial considerations that should guide policy and regulation on data.

1.3. Data trusts: an evolving conceptual framework

The concept of 'data trusts' has evolved from this backdrop of bottom-up data stewardship. The idea of extracting value from data and restructuring its distribution through the use of data trusts was posited by Professor Neil Lawrence in 2016,⁵⁴ whereby data trusts could act as "power brokers" to mediate the use of data for public benefit, without compromising the rights of data subjects to whom the data relates.

⁵² Taylor (2017), Sage Journals, "What is data justice? The case for connecting digital rights and freedoms globally", <https://journals.sagepub.com/doi/10.1177/2053951717736335>

⁵³ Dencik and Kaun (2020), Global Perspectives - University of California Press, "Datafication and the Welfare State", <https://online.ucpress.edu/gp/article-abstract/1/1/12912/110743/Datafication-and-the-Welfare-State?redirectedFrom=fulltext>

⁵⁴ Lawrence (2016), The Guardian, "Data trusts could allay our privacy fears", <https://www.theguardian.com/media-network/2016/jun/03/data-trusts-privacy-fears-feudalism-democracy>

Subsequently, Delacroix and Lawrence articulated ‘bottom-up data trusts’ in 2019⁵⁵ as a tool of collective engagement used by communities to decide on how their data is used and shared by third parties. They described how trustees could be bound by fiduciary obligations of undivided loyalty and care towards its beneficiaries, defining the terms for purpose-led data sharing. It set out use cases to outline the advantages of the approach in different contexts. For example, in the context of social media and financial information, data trusts could have a role to play in negotiating terms on behalf of data subjects to ensure data could be made available for research and public policy purposes.

The Data Trusts Initiative builds on this conception of data trusts. It describes them as a “mechanism for individuals to pool their data rights into an organisation”⁵⁶, with i) independent stewardship of the pooled rights, ii) fiduciary responsibilities, iii) operations guided by a framework of institutional safeguards and iv) the facilitating of collective action.⁵⁷ This conception of the data trust seeks to rebalance the respective control that corporations and individuals have over personal data, and provide a legal mechanism to empower data subjects to choose to appoint others to make those decisions on their behalf. As a Mozilla Fellow, Anouk Ruhaak is also working on scenarios where multiple people ‘hand over their data assets or data rights to a trustee’, such as data donation platforms that allow users of web browsers to donate data on their usage of different services.⁵⁸

Similarly, Aapti’s interpretation of a data trust refers to “a legal arrangement wherein a person authorises an individual or entity to manage certain property for the benefit of a third party or for certain defined purposes”.⁵⁹ In the context of the data economy, the data (or rights over it) constitutes the property⁶⁰ that will be managed by the trust and the trustee (authorised representative individual or entity) is bound by fiduciary

⁵⁵ Lawrence and Delacroix, International Data Privacy Law (2019), “Bottom-up data Trusts: disturbing the ‘one size fits all’ approach to data governance”, <https://academic.oup.com/idpl/article/9/4/236/5579842>

⁵⁶ Data Trusts Initiative (2021), “Data trusts: international perspectives on the development of data institutions”, <https://static1.squarespace.com/static/5e3b09f0b754a35dcb4111ce/t/603ce3325e1da817afe6b193/1614603061204/WP+2+-+DTI+-+global+perspectives.pdf>

⁵⁷ Ibid.

⁵⁸ Ruhaak (2019), Mozilla Foundation, “Data trusts: Why, what and how”, <https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34>

⁵⁹ Manohar (2019), Aapti Institute, “Trust Law, Fiduciaries and Data Trusts”, https://thedataeconomylab.com/wp-content/uploads/2020/10/DataTrustsPpr_SM.pdf

⁶⁰ Certain scholars have demonstrated that data lacks the requisite quality to be considered and treated as “property” in law. Refer Professor McFarlane’s work for more information - <https://www.law.ox.ac.uk/research-and-subject-groups/property-law/blog/2019/10/data-trusts-and-defining-property>

obligations to act in the best interests of its beneficiaries and according to the defined purposes.

Another major work outlining the potential function of data trusts is that of McDonald and Wylie⁶¹. Their work describes the potential for data trusts in consumer protection and fiduciary governance for the data economy.⁶² It argues that the data trust, a legal arrangement where a trustee is appointed with fiduciary obligations towards a specified beneficiary, is suited to creating predictable data supply chains with increased accountability.

As a model of bottom-up data stewardship, data trusts represent a compelling instrument to unlock data for public benefit uses within a framework of fiduciary duties, such that data sharing decisions are compliant with the interests and rights of communities.⁶³ As the Global Partnership for AI has itself described, they offer the potential to “expand access to data for innovation while putting citizen interests at the heart of stewardship”.⁶⁴

1.4. Differing interpretations of data trusts

There has been a lack of consistent, global interpretation of data trusts among scholars, policymakers and other actors, reflecting the nascent nature of the research and practice of bottom-up data stewardship.

India’s proposed framework for the governance of non-personal data recommends the appointment of data trustees as representatives to steward community data and channel its use for socially beneficial purposes such as entrepreneurship, innovation, research and policymaking.⁶⁵ Similarly, Ontario state authorities in Canada are exploring legal mechanisms to establish data trusts that would enable ‘privacy-protective data sharing’.⁶⁶

⁶¹ McDonald, S., Wylie, B. (2018), CIGI, “What is a Data Trust”,
<https://www.cigionline.org/articles/what-data-trust/>

⁶² McDonald, S., Wylie, B. (2018), CIGI, “What is a Data Trust”,
<https://www.cigionline.org/articles/what-data-trust/>

⁶³ Sadowski, Viljoen and Whittaker (2021), Nature, “Everyone should decide how their data is used - not just tech companies”,
<https://media.nature.com/original/magazine-assets/d41586-021-01812-3/d41586-021-01812-3.pdf>

⁶⁴ The GPAI Data Working Group (2021), “Understanding Data Trusts”,
<https://ceimia.org/wp-content/uploads/2021/07/2021-07-09-GPAI-summary-understanding-data-trusts-updated.docx.pdf>

⁶⁵ Kris Gopalakrishnan, et al., (2020), Ministry of Electronics and Information Technology (Govt. of India), “Report by the Committee of Experts on Non-personal Data Governance Framework”,
https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

⁶⁶ Newsroom (2020), Government of Ontario, “Ontario Launches Consultations to Strengthen Privacy Protections of Personal Data”,

The European Commission's proposed Data Governance Act, 2020 outlines a framework for "data intermediaries" - entities which provide "data sharing services" that "contribute to the efficient pooling of data as well as to the facilitation of bilateral data sharing".⁶⁷ Although the Act does not explicitly identify a specific category or type of data intermediary, scholars have put forth data trusts as possible mechanisms to promote enfranchisement and meaningful realisation of data rights of subjects within the EU.⁶⁸

In the UK, a 2018 report by Hall and Pesenti called for the creation of data trusts as a "trusted and proven framework" to increase the availability and use of data for growing the domestic AI industry.⁶⁹ Other research subsequently interpreted data trusts as 'providing independent fiduciary stewardship of data', building on the work of Porcaro and others that had imagined organisations "[putting] their user-data in some form of irrevocable, spendthrift-esque 'data trust', which would then be managed by a third-party trustee (a nonprofit, for instance)". Using this interpretation, the ODI undertook pilot projects to generate insights on the potential application of data trusts in the contexts of food waste management, wildlife poaching and urban mobility.⁷⁰ It has also observed UK Biobank, OpenCorporates and Oversight Board as examples of independent, fiduciary stewardship of data applied in practice⁷¹.

This interpretation is adopted elsewhere. Based in the US, PLACE describes itself as a data trust for creating, storing and accessing mapping data, governed by independent trustees drawn from different geographies and sectors⁷². Johns Hopkins Medicine has similarly been described as having 'a data trust administrator', responsible for retaining patient privacy while enabling medical records to be used to improve care and facilitate research⁷³. Toronto's experience with a proposal to set up a 'civic data trust' is another

<https://news.ontario.ca/en/release/57985/ontario-launches-consultations-to-strengthen-privacy-protections-of-personal-data>

⁶⁷ European Commission (2020), "European Data Governance (Data Governance Act)",

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767>

⁶⁸ Data Trusts Initiative (n.d.), "Understanding the Data Governance Act: In conversation with Sylvie Delacroix, Ben McFarlane and Paul Nemitz",

<https://datatrusts.uk/blogs/understanding-the-data-governance-act-in-conversation-with-sylvie-delacroix-ben-mcfarlane-and-paul-nemitz>

⁶⁹ Hall and Pesenti (2017), "Growing the Artificial Intelligence Industry in the UK",

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

⁷⁰ Open Data Institute (2019), "Data trusts: lessons from three pilots",

<https://theodi.org/article/odi-data-trusts-report/>

⁷¹ Hardinges (2020), The Open Data Institute, "Data trusts in 2020"

<https://theodi.org/article/data-trusts-in-2020/>

⁷² Verhulst et al (n.d.), PLACE, "Establishing a data trust: From concept to Reality",

<https://www.thisisplace.org/blog-1/introducingplace/establishing-a-data-trust>

⁷³ Dell Technologies (2019), "Lessons from a user-trusted data trust",

<https://www.delltechnologies.com/en-us/perspectives/lessons-from-a-user-trusted-data-trust/>

example that worked to this interpretation.⁷⁴ In 2019, Sidewalk Labs proposed setting up a civic data trust to govern data generated by sensors and cameras around the neighbourhood that would be developed. Ultimately, the project was abandoned in 2020 citing unprecedented economic uncertainty but perhaps mostly due to local opposition to the development plans.

This research works to a definition based on the Global Partnership for AI's consensus statement, whereby a data trust is "a form of data stewardship that supports data producers to pool their data (or data rights) with the aim of collectively negotiating terms of use with potential data users, through the oversight by independent trustees, with fiduciary duties, and within a framework of technical, legal and policy interventions that facilitate data use and provide strong safeguards against mis-use". This interpretation is aligned with the concept as put forward by Lawrence and Delacroix in their paper on 'bottom-up data trusts' and as adopted by the Data Trusts Initiative. It marries the conception of groups of individuals coming together to contribute data within the framework of independent fiduciaries duties.

1.5. Institutionalising data trusts and codifying fiduciary responsibilities

There has been significant discussion around the challenge of institutionalising the data (or data rights) to be held and managed by data trusts, and codifying the fiduciary responsibilities of trustees.

Establishing data trusts involves the pooling of data, or data rights, and the exertion of control over these data (rights) by a trustee. Data rights enshrined within legislation are therefore a prerequisite for their development.⁷⁵ Over the past decade, we have seen the introduction of significant new data protection laws globally that represent the basis of a data rights framework. Jurisdictions in Canada have implemented data protection legislations at the federal⁷⁶ and provincial levels,⁷⁷ with Ghana close behind in enacting its Data Protection Act, 2012.⁷⁸ The most significant step towards articulation of data rights is the EU's General Data Protection Regulation, 2016⁷⁹ which sought to impose

⁷⁴ Tusikov (2019), Centre for Free Expression, "Urban Data" and "Civic Data Trusts" in Smart Cities", <https://cfe.ryerson.ca/blog/2019/08/%E2%80%9Curban-data%E2%80%9D-%E2%80%9Ccivic-data-trusts%E2%80%9D-smart-city>

⁷⁵ Data Trusts Initiative (2021), "Data trusts: international perspectives on the development of data institutions", <https://static1.squarespace.com/static/5e3b09f0b754a35dcb4111ce/t/603ce3325e1da817afe6b193/1614603061204/WP+2+-+DTI+-+global+perspectives.pdf>

⁷⁶ The Personal Information Protection and Electronic Documents Act, 2004

⁷⁷ To date, Alberta, British Columbia and Quebec have provincial laws to govern processing of personal information. Additionally, Ontario, Newfoundland and Labrador, Nova Scotia and New Brunswick have enacted independent legislations on health information processing.

⁷⁸ The Data Protection Act, 2012 (Act 843)

⁷⁹ Regulation (EU) 2016/679 - General Data Protection Regulation

controls on data processing, rooted in principles of individual harm, rights and privacy. This has spurred a flurry of personal data protection regulations that have been introduced or implemented in multiple jurisdictions outside Europe, such as India,⁸⁰ Kenya,⁸¹ Brazil,⁸² South Africa,⁸³ among others.

However, despite alignment between some regions, there remain significant variations in the data rights afforded by different jurisdictions. In addition to these variances, it is important to note that the manner in which specific rights - such as the right to data portability, which is crucial to support bottom-up data stewardship - are enforced differ across legal regimes. Further complications arise for jurisdictions that do not have an operative data protection legislation and consequently, provide little clarity on the data rights of citizens.⁸⁴ Therefore, the feasibility of data trusts will be a function of the extent and nature of data rights afforded by the relevant legal jurisdiction, and the way those rights are enforced. This dynamic is explored further in legal research undertaken by Aapti in parallel to the research described by this report.

Enabling data sharing for social benefit through data trusts: Legal review

In addition to the present research, Aapti was commissioned by GPAI to examine the existing and necessary legal mechanisms required to develop data trusts. To do so, the researchers undertook a rigorous process of comparative legal analysis across 11 jurisdictions to draw out variations in data protection laws and rights, data sharing frameworks and fiduciary obligations - all of which constitute essential legislative underpinnings of a data trust.

The resultant comparative analysis throws up several key insights that demonstrate disparity in maturity of legal landscapes for data trusts around the globe, and point to the need for administrative and legislative investments in data governance in several countries. Further, it was found that for legal systems which do not embed fiduciary duties matching common law structures, there may be a need to explore diverse structures for enabling human-centric data governance. Key takeaways from this research have been summarised below:

- **Disparity across nations and lack of digital infrastructure:** Given the diversity - both economic and political - of the jurisdictions analysed, it was

⁸⁰ The Personal Data Protection Bill, 2019

⁸¹ The Data Protection Act, 2019

⁸² Lei Geral de Proteção de Dados, 2019

⁸³ Protection of Personal Information Act, 2013 (came into force, partly, in 2020)

⁸⁴ For more information on the status of data protection and privacy legislations across the world, refer to the UNCTAD's remarkable tracker available at <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>

found that the maturity in articulating rights over data varied significantly. This includes regulatory measures like standardisation of data formats or sharing purpose, enabling interoperability, or introducing digital public infrastructure.

- **Personal data rights and building for autonomy:** Even within some of the countries with more robust digital infrastructures, the absence of certain personal data rights - such as access, portability and erasure - pose challenges in creating a sustainable data trust ecosystem. For instance, Canada, Australia, and South Korea, while faring well on digital infrastructure, have yet to recognise clear data portability rights.
- **Legislative implementation and regulatory oversight:** Legal concepts which may be common across jurisdictions are not always implemented uniformly. For instance, the extent of adoption of trusts in Kenya and South Africa - countries with common law origins - is not as crystalised as compared to jurisdictions such as England, which extensively use trusts for a variety of commercial and non-commercial purposes.
- **A 'data trust conundrum' for stewardship:** It is evident that the conception of data trusts is most fundamentally rooted in English trust law. Based on this analysis, even in jurisdictions that have common law influence and recognise trusts, the evolution of its concepts have not mirrored the English experience.

In the past, the role of fiduciaries in the context of data stewardship has generally been limited to ensuring compliance - that is determining whether grounds for processing of data are in accordance with applicable regulations.⁸⁵ However, this conception of 'trustees' holding broader and more explicit fiduciary duties that compel them to act in the best interests of data producers has started to be discussed within the data governance ecosystem. This renewed approach to the duties of "trustees" has its roots in the stewardship of common pool resources, such as Scotland's ports trusts.⁸⁶ With myriad activities - from fisheries management to renewable energy generation - being undertaken, the ports are managed through a trust framework set through an act of law of the Parliament. These trust ports are governed by a stakeholder-representative board comprising users of the port, members of the local community, government agencies and follows a democratic model of decision-making. The trust framework helps balance various perspectives and interests while generating valuable income for regional and national economies; at the heart of the trust port model is the fiduciary

⁸⁵ Bailey and Goyal (2019), Data Governance Network, "Fiduciary relationships as a means to protect privacy: Examining the use of the fiduciary concept in the draft Personal Data Protection Bill, 2018", https://datagovernance.org/files/research/NIPFP_Rishab_Trishree_fiduciaries_-_Paper_4.pdf

⁸⁶ Transport Scotland (n.d.). Retrieved from <https://www.transport.gov.scot/transport-network/ports-and-harbours/port-governance/>

duty of loyalty and care which mandates trustees to act in the best interest of its stakeholders.⁸⁷

Sean McDonald⁸⁸ has expanded on the application of explicit, contextual fiduciary duties to the governance of data, describing data trusts as a tool to hold companies accountable for their decisions and the promises made to users. In this view, data trusts offer a credible legal container for articulating fiduciary accountability and establishing processes to enable the right of redress.

The concept of information fiduciaries⁸⁹ proposed by Balkin also provides a seminal analysis of fiduciary duties. It has, however, been subject to criticism,⁹⁰ chiefly in that the approach fails to take cognisance of the entrenched business models that drive diverging interests between end users and data processors. According to this criticism, users are not adequately equipped by platforms and data processors in order to express and act on their interests, which is not changed by the imposition of fiduciary duties on data processors. Also, as fiduciary obligations are normally settled by courts, the cost of solving each dispute on violation of fiduciary obligations may prove to be too much for users as well as the legal system to bear.⁹¹

Divergent views have also been expressed as to the legal forms most suitable for data trusts to take to 'house' the data (rights) contributed by data subjects, and to codify the fiduciary responsibilities of its trustees.

In the UK, the ODI worked with a legal consortium on its initial pilots who argued that the mechanism of trust law was 'inappropriate' for constructing data trusts, largely on the basis that data cannot be made the property of a trust under existing law⁹². This statement was subsequently challenged by the legal community. In October 2019, Professor Ben McFarlane of University of Oxford questioned this finding, suggesting that people's rights over data, such as those conferred by the General Data Protection Regulation, rather than data itself, could be made the property of a legal trust and

⁸⁷ Kapoor and Ramesh (2019), The Data Economy Lab, "Principles for Revenue Models for Data Stewardship",

<https://thedataeconomylab.com/2020/07/31/principles-for-revenue-models-of-data-stewardship/>

⁸⁸ McDonald, S. (2019), The Fiduciary Supply Chain: Models for Platform Governance,

<https://www.cigionline.org/articles/fiduciary-supply-chain/>

⁸⁹ Balkin, J. (2016), UC Davis Law Review, Vol 49, No.4, "Information Fiduciaries and the First Amendment", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3700087

⁹⁰ Pozen, D. and Khan, L. (2019) "A Skeptical View of Information Fiduciaries", Harvard Law Review, 2019, <https://harvardlawreview.org/2019/12/a-skeptical-view-of-information-fiduciaries/>

⁹¹ Ibid

⁹² Open Data Institute (2019), "Data trusts: lessons from three pilots", <https://docs.google.com/document/d/118RqyUAWP3WlyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit>

asserted collectively by its trustees.⁹³ A similar argument has been made by Sylvie Delacroix and Neil Lawrence, who have suggested that while there are challenges, they ‘do not constitute reasons to doubt that data rights can be held under a legal Trust’. A paper published in the National University of Singapore Faculty of Law’s journal similarly found that ‘the traditional trust, the historical creation of English Equity jurisprudence and now found around the world, is a perfectly sensible vehicle for the management of data’.⁹⁴

The use of alternate legal forms to construct data trusts may be appropriate in jurisdictions that do not follow the common law tradition. For example, Germany does not have a trust law framework, but nonetheless has institutional forms such as the Sparkassen (cooperative or not-for-profit banks) that carry fiduciary obligations ascribed to the common law trust. And in Quebec, a civil law jurisdiction, the Quebec trust enables data rights to be pooled and administered by a trustee, which has sparked widespread excitement around the opportunity.⁹⁵ These examples indicate the difference in legal structures being experimented with in different legal jurisdictions in order to facilitate the development of data trusts.

⁹³ McFarlane (2019), University of Oxford - Faculty of Law, “Data trusts and defining property”, <https://www.law.ox.ac.uk/research-and-subject-groups/property-law/blog/2019/10/data-trusts-and-defining-property>

⁹⁴ Lu Jia Jun, et.al. (2019), NUS Law Working Paper No. 2019/019, “The basics of private and public data trusts”, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3458192##

⁹⁵ Hulin (2021), Data Trusts Initiative, “How can civil law jurisdictions support data trusts? The Quebec example”, <https://datatrusts.uk/blogs/how-can-civil-law-jurisdictions-support-data-trusts-the-quebec-example>

2. International knowledge, attitudes and practices of data trusts

- This section describes the results of a survey to understand knowledge (awareness and understanding), attitudes (perspective) and practices (implementation) of data trusts around the world.
- Respondents included those operating data trusts or similar bottom-up data stewardship initiatives, and organisations working on the topic of data stewardship.
- The survey: surfaced awareness of and optimism towards the concept of data trusts; found general agreement with the definition put forward by GPAI; found no projects delivering all the functions of real-world data trusts through one vehicle but a number of similar bottom-up data stewardship initiatives that were delivering many of the functions; and encountered a variety of legal forms and technologies used to construct the different initiatives.

2.1. Context

This section describes the results of a survey undertaken to understand the knowledge about, attitudes towards and practices of data trusts from around the world.

The survey was divided into four sections. The first covered basic information about the respondent: their country of work, role and organisation name. The second sought their current understanding of data trusts, and included questions about the definition of data trusts and other forms of data stewardship. The survey then split into two parts, with one for completion by practitioners (people building or running data trusts, or similar bottom-up data stewardship initiatives) and the other for experts working on data stewardship and related topics. Practitioners were asked questions about how their initiatives work and whether they identified themselves as data trusts. Experts were asked about their perceptions of the current state of data trusts, and their thoughts about the future of the movement. In general, the survey combined structured and open-ended questions designed to enable quantitative analysis as well as space for respondents to elaborate their ideas.

The survey was initially distributed by Aapti and ODI to around 100 practitioners and experts. It was also disseminated through social media platforms and newsletters, the GPAI Working Group, and snowballed through respondents sharing the survey themselves. At the close of the survey there were 45 responses. There was an even split between practitioners and experts, 22 and 23 respectively. Responses were heavily-weighted towards Europe (and in particular from the UK), with nearly $\frac{3}{4}$ of all the

respondents working there. The sectors that respondents described working from were technology and data, health and research.

2.2. Awareness and understanding

The concept of data stewardship was a familiar topic for the respondents of the survey, as you might expect from practitioners and experts. Respondents were slightly less familiar with the idea of data trusts.

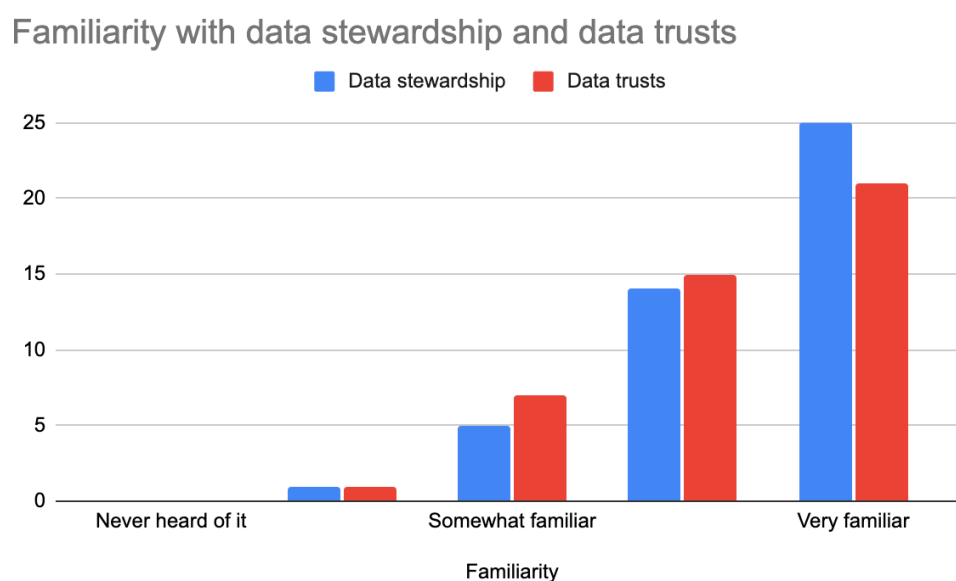


Figure 1: Respondents familiarity with data stewardship and data trusts

Beyond data trusts, our respondents were aware of a variety of other forms of data stewardship - the most familiar being data trusts, data commons, and data exchanges. The familiarity with data trusts was unsurprising given the focus of this project. Respondents also shared examples of data stewardship that were not listed as prompts, such as statistics agencies and national bodies.

As the literature review details, there are varying interpretations of data trusts. In this case, 82% of respondents agreed with the GPAI definition of data trusts as “a form of data stewardship that allow data producers to pool their data (or data rights) and facilitate collective negotiation of terms of use with potential data users, working through independent trustees who are bound by strong fiduciary duties, within a framework of technical, legal and policy interventions that facilitate data use and provide strong safeguards against mis-use”. The respondents who did not agree with the definition offered a variety of interesting ideas and opinions.

A number of respondents disagreed with the language used in the GPAI definition. In one case, the respondent felt that the who the trustees have a fiduciary duty to was missing, preferring the language of trustees "*acting on behalf of the data subjects*". The respondent also described the need for data trusts to be working towards a specific purpose and felt this should be reflected in the definition. Other respondents questioned the use of the term "*data producer*" and requested that the legal mechanism intended to create the fiduciary duties should be documented.

A further contention made by respondents was whether the definition reflected a feasible reality. In particular, respondents questioned whether the imposition of fiduciary duties and technical, legal and policy safeguards was an ideal, and too high a bar to be met. One respondent preferred to talk about 'data intermediaries' and another described there not to be a need for an independent trustee. These responses perhaps represent something deeper than varying ideas for the definition of data trusts, and instead reflecting the existence and need for various forms of bottom-up data stewardship.

When asked to list examples of data trusts, respondents responded in three ways:

- They gave examples of similar bottom-up data stewardship initiatives (the difference between these and data trusts is discussed below).
- They gave theoretical examples of data trusts.
- They stated that they were not aware of any practical examples of data trusts.

A number of examples given by respondents - including Swash⁹⁶ and MIDATA⁹⁷ - did not appear to have all of the functions of a data trust set out by the GPAI definition. In particular, none appeared to have independent trustees with fiduciary responsibilities. Other examples were not yet active, like the Liverpool Civic Data Cooperative⁹⁸. Other examples were theoretical or abstract, such as 'a Health Bank', a theoretical data trust of residents of a housing block and mentions of the examples put forward by Sylvie Delacroix and Neil Lawrence's paper⁹⁹.

⁹⁶ Swash, (2021). Retrieved from: <https://swashapp.io/>

⁹⁷ MIDATA, (2021). Retrieved from: <https://www.midata.coop/en/home/>

⁹⁸ Liverpool Civic Data Cooperative, (2020). Retrieved from: <https://www.liverpoolcityregion-ca.gov.uk/liverpool-city-region-combined-authority-announces-proposal-s-for-5-3m-funding-for-data-driven-health-improvements/>

⁹⁹ Lawrence, N. & Delacroix, S. (2019, October 1). Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*. Retrieved October 26, 2021, from <https://academic.oup.com/idpl/article/9/4/236/5579842>.

2.3. Practices

To find out about the practices of data trusts, part of the survey was targeted specifically to practitioners who are building or running data trusts.

These questions focused on understanding how each initiative identified itself, their function and purpose, and included practical questions on their approach to data stewardship. There were 22 respondents to this part of the survey.

In terms of identifying their initiatives, around a third of practitioners did not identify as data trusts. Within these respondents, we observed two main types of organisations. The first consist of organisations like Schluss & digi.me, who see themselves as the creators of infrastructure onto which data trusts can be built - for example by creating the technology for users to collect their data in a pod or vault. The second was formed of organisations with similar aims to the concept of data trusts, as put forward by GPAI, but function differently. For example, Swash described having “built-in trustless structures” rather than fiduciary responsibilities to achieve the outcome of empowering people with their data, and CSIRO detailed how they work to enable the aggregation of data for the common good but unlike data trusts, participating actors control what happens to the data rather than trustees.

The survey asked practitioners to select which of the 6 functions, as per the GPAI definition of data trusts, their initiatives undertake, as follows:

1. Provide a platform for people to pool data.
2. Provide a platform for people to establish desirable terms and conditions of data use.
3. Negotiate use of the data in accordance with agreed terms and conditions, facilitating safe and controlled data use.
4. Appoint expert trustees to take responsibility for the stewardship of the data.
5. Create a regime of strong fiduciary responsibilities to bind the trustees to act in the interests of the data trust’s members.
6. Establish safeguards and oversight mechanisms to prevent data misuse and to take remedial action in the event of the trust’s terms and conditions being breached.

On average, the respondents stated that they had between four and five of the functions, with responses ranging from just one to the full six.

Functions of data trusts like initiatives in practice

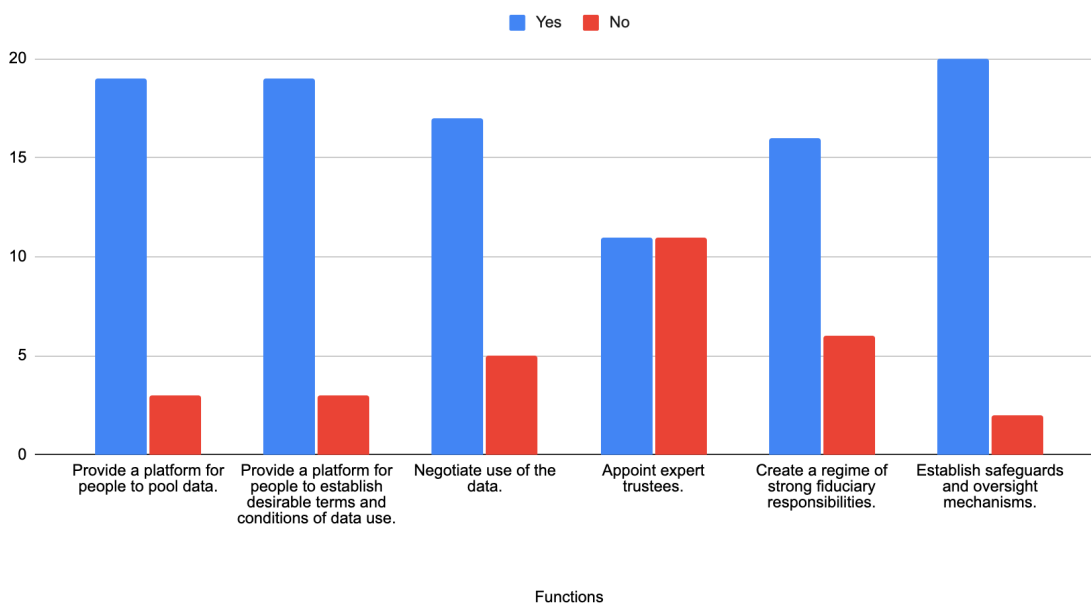


Figure 2: “Which of the following functions does your initiatives have?” - question to data stewardship practitioners.

There was a relatively even distribution across five of the six roles, with one function outlying: *Appoint expert trustees to take responsibility for the stewardship of the data*. This may reflect that the requirement to appoint expert trustees to oversee the stewardship of data may be one of the more difficult functions to achieve in practice. Conversely, providing a platform to pool data, and establishing safeguards over that data, appear some of the more common and achievable functions in practice.

Seven respondents said their initiatives had all six of these functions, appearing to determine them as data trusts. However, of these seven: the name of one of these was not given (N/A); two were not yet active (PLACE, Donate your Data Foundation); two instead develop technology for data trusts (Sightline Innovation, PolyPoly); and one preferred not to define as a data trust (DataYogi). This left one for us to examine further (Worker Info Exchange).

Worker Info Exchange is a non profit organisation dedicated to helping workers access and gain insight from data collected from them at work.¹⁰⁰ The project is relatively early stage, but is active and currently stewards data on behalf of gig workers. On their website, there is clear evidence they are providing a platform to pool data and establish desirable terms, they are negotiating safe usage of their users data, and establishing

¹⁰⁰ Worker Info Exchange, (2021). Retrieved from: <https://www.workerinfoexchange.org/>

safeguards to prevent data misuse. However, it is unclear as to whether there are any data trustees and whether those trustees have fiduciary responsibilities to act in the best interests of their members. The lack of evidence means that for the purposes of this report, Worker Info Exchange is not considered a data trust. Despite not meeting the GPAI definition, Worker Info Exchange is an excellent example of bottom-up data stewardship in practice¹⁰¹.

The survey also asked experts about the frequency that they had seen the six functions of data trusts in their work.

Frequency of functions of data trusts as seen by experts

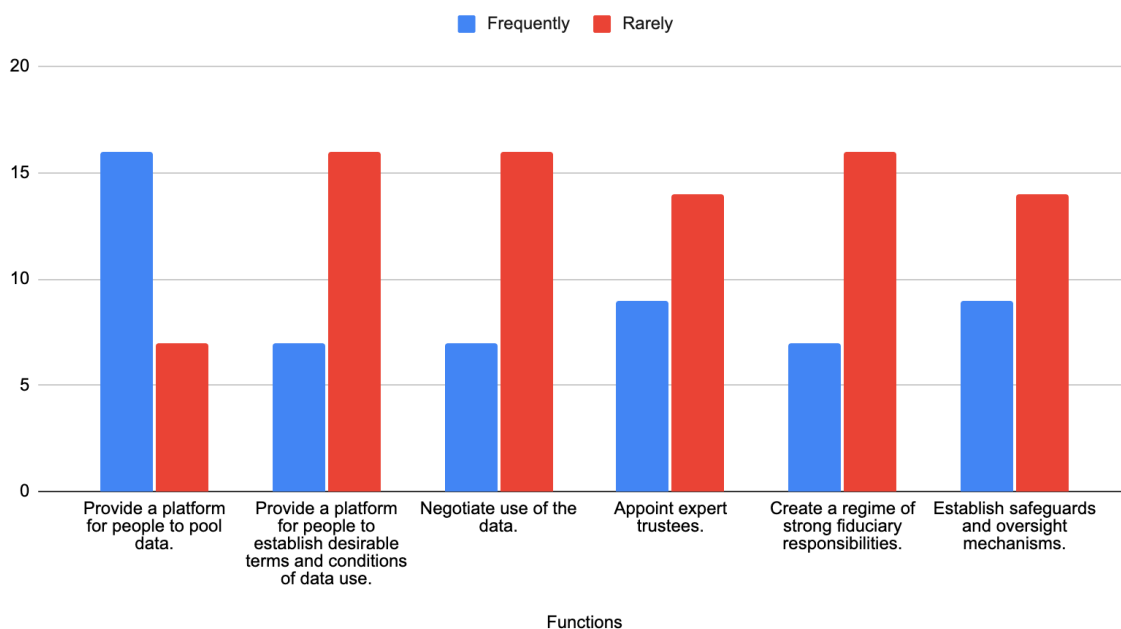


Figure 3: “In the data stewardship initiatives you have come across in your work, please select the frequency that you see the following features” - question to data stewardship experts.

Two thirds of the experts had frequently seen initiatives *providing a platform for individuals and groups to pool their data*. The other 5 functions were rarely observed in a practical setting.

The survey explored how bottom-up data stewardship initiatives function, asking respondents to describe legal forms, technical systems and business models.

¹⁰¹ This project reached out to Worker Info Exchange to respond to this analysis, Worker Info Exchange agreed with our description of the initiative not currently meeting the GPAI definition, but described how the team was currently exploring whether becoming a data trust of this nature was a feasible undertaking.

Data trusts have been conceptualised as being constructed using trust or common law, but only one respondent specified that they had given thought to this (they had undertaken some research to understand how trust law could be applied in their context). Respondents described a wide range of legal approaches used to underpin their initiatives. Limited companies, foundations and nonprofits were the most frequent legal forms, with constructs such as cooperatives also mentioned. A number of respondents alluded to the fact that different legal forms were suitable in different circumstances.

As well as the legal form of their initiative, respondents also commented on the wider legal conditions around them. Respondents cited the General Data Protection Regulation as a legislation or law that has been ‘specifically helpful or harmful to your initiative’. Other similar laws, such as the CDPR in California and the “Law for a digital republic in France” were also mentioned.

The technologies adopted by the respondents were also varied. Some organisations described using blockchain, while the development of ‘pods’ or ‘vaults’ for personal data were also described. Blockchain was described as being useful by initiatives working towards creating a ‘trustless’ structure, in which the distributed technology was seen as replacing the need for trustees to make decisions. The ‘pods’ and ‘vaults’ were described by initiatives seeking to help users to have more control over data about them on an individual, rather than group, basis. There was also discussion of privacy enhancing technologies to ensure the safety of users' data, as well as Application Programming Interfaces to facilitate access to the data being brought together by the initiatives.

Respondents also shared information about their business models. Most of the respondents described being funded by non-earned revenue streams, such as philanthropic funding, public funding or private investment. The predominance of unearned revenue reflects the relative infancy of bottom-up data stewardship initiatives and finding them early in their quest for sustainability¹⁰². A quarter of respondents described earned revenue streams, including membership fees for the users of the initiative, selling access to data to third parties and analysing and packaging the data into insights.

¹⁰² Dodds, L., Szász, D., Keller, J., Snaith, B. and Duarte, S. (2020, April). Designing sustainable data institutions. *Open Data Institute*. Retrieved 26 October 2021, from: <https://theodi.org/article/designing-sustainable-data-institutions-paper/>.

In terms of the scale and maturity of the initiatives, 50% of the respondents had 1,000 members or less, and 80% of the initiatives are less than 5 years old (or are yet to be operational).

2.4. Attitudes towards data trusts

The survey also sought to understand the respondents' attitudes and perspectives on data trusts.

A number of survey respondents, both experts and practitioners, were positive about the potential of data trusts. By increasing the control people have over data about them, a number of respondents thought that data trusts could help to rebalance the power asymmetry in the data economy. One respondent stated that through data trusts, people's data could be better safeguarded from private interests, avoiding 'abusive data relationships'. Respondents felt data trusts could also work towards a future where there was increased access to and usage of data for public benefit. They described how data trusts could be designed to tackle "problems in which individuals are interested in combining their data to get a broader analysis of a specific challenge, or where individuals benefit from pooling their data".

Some respondents described the lack of practical examples of data trusts holding the concept back. There was a strong desire, verging on impatience, among some of the experts to see an operational data trust, even to the point where some described that the concept had become 'somewhat of a fantasy'. Others were concerned that the focus on an approach to bottom-up data stewardship that only exists as a concept may detract from other similar - but importantly, operational - approaches (such as those featured as case studies in this report). Some respondents were also concerned about the potential uptake of data trusts as and when they became functional, feeling that the demand for them originates largely from experts and reaching a critical mass of users would require a dramatic shift in culture, understanding and skills across the data economy.

3. Case studies

- This section consists of three case studies of bottom-up data stewardship initiatives: Driver's Seat, Open Humans and MIDATA.
- They represent real-world examples of how groups can be empowered around data they've generated, and are actively making available data for broad societal benefit.
- They have been selected on the basis of their community-centrism and maturity, and the studies unpack the purposes for their formation, their stakeholders served, their legal structure, their internal data governance principles and the technical safeguards used to mediate access to data.

3.1. Criteria and selection

This section of the report consists of three case studies of bottom-up data stewardship initiatives. They are intended to shed light on the practical considerations involved in designing such an initiative, and to surface perspectives from practitioners operating them.

The case studies rely on secondary research in the form of Aapti and the ODI's internal interview notes, analysis and videos. The cases were selected using the two criteria:

1. **Maturity** - There are many interesting proposals and theoretical models for data trusts and bottom-up data stewardship initiatives. However, for the purposes of the case studies, we skewed towards those that are operational (i.e. actively supporting data to 'flow' between actors).
2. **Community-centrism** - Empowering people to exercise meaningful control over data takes many forms. We have chosen cases that adopt a collective, participatory approach to data stewardship, such that their members dictate how their data is used, by whom and for what purposes.

Ideally, Aapti and the ODI would have liked to feature initiatives that responded to the survey as case studies. However, the researchers found that [Driver's Seat](#), [Open Humans](#) and [MIDATA](#) not just satisfy the above mentioned criteria more so than respondents, but are perhaps some of the most promising examples of bottom-up data stewardship in general.

As well as showing the variety of design choices available, the case studies, albeit anecdotal, highlight the virtues of bottom-up stewardship.

Aspect of analysis	 Driver's Seat	 OPEN HUMANS	 MIDATA
Purpose	Empower gig workers to take control of their data and derive monetary value through it	Facilitate community data governance to produce open, participant-centric research in health	Enable members to gain control over data and amass health data for use in biomedical research
Structure	Limited cooperative association; funded by pvt investments, grant money; revenue from monetization of aggregated insights	Non-profit organization; funded and sustained through grant money; does not generate any revenue	Non-profit cooperative; private investments and grant money source of funding; does not generate any revenue
Stakeholders	Gig workers are the data generators; data users include local government and transportation agencies	Individual members of OH comprise data generators; academic institutions, independent researchers are data users	Members and MIDATA account holders are data generators; pharmaceutical companies and research orgs. are data users
Governance principles	Data sharing and selling decisions governed through cooperative board; members can revoke consent for data use through email notices	Data sharing decisions are authorized by individuals and varies on a project-to-project basis	Data sharing authorized through internal ethics review board and general assembly of members through democratic votes
Privacy controls	Member data is anonymized, and only aggregated datasets and insights sold to data users/buyers	Data may or may not be anonymized/pseudonymized before sharing, varies according to project requirements	Data may or may not be anonymized/pseudonymized before sharing, varies according to project requirements

Figure 4: Snapshot of insights from case studies (Source: Aapti/ODI analysis)

3.2 Driver's Seat

Overview

Driver's Seat enables workers in the gig economy to gain control over their data and access analytics that help them earn more from their labour.

Founded in 2019 in Colorado, USA, the platform was developed over conversations with Uber and Lyft drivers whose work was mediated by the 'extraction, processing and delivering of data' via algorithms that tended to deny workers of their agency and a voice in negotiating their working conditions.¹⁰³ Driver's Seat also aids in the monetisation of driver's mobility data by selling access to public authorities and local governments that use these insights in policy making and resource allocation.¹⁰⁴ Consequently, the entity unlocks societal value by directing data use towards public benefit, while simultaneously compensating the members of its cooperative for their role in generating this data.

¹⁰³ Witt, Hays [Aapti Institute] (2021). "Data Economy Lab | Tracking Stewardship: Driver's Seat - Empowering gig workers through data" [Video]. Youtube. <https://youtu.be/a-l8tfeoB3g>

¹⁰⁴ Ibid.

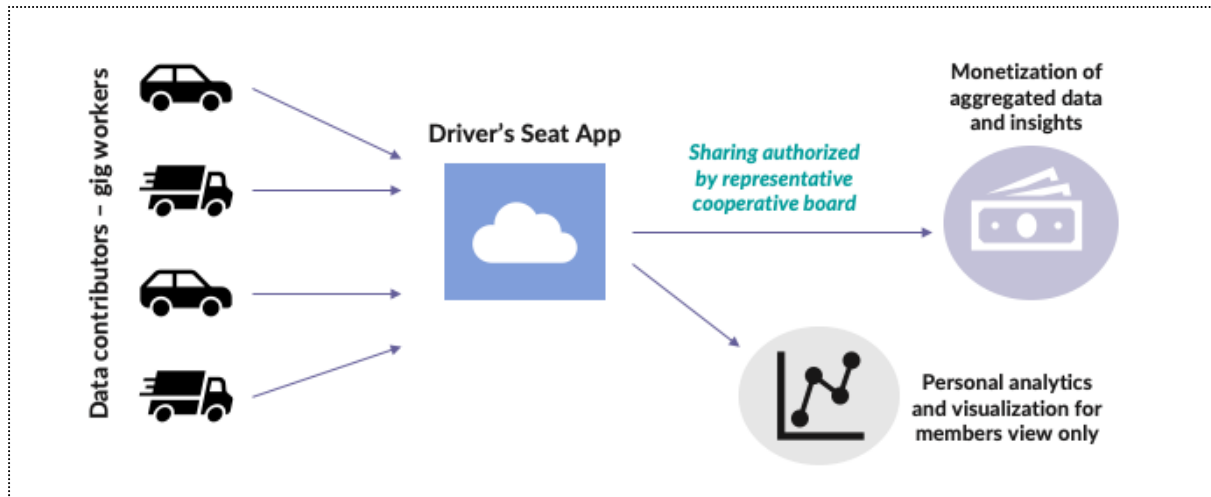


Figure 5: Structure of Driver's Seat (**Source:** Aapti/ODI analysis)

Purpose

The core stated purpose of Driver's Seat is to enable participation of gig workers in the governance of their data and derive monetary/instrumental value that could potentially reduce the precarity associated with platform-mediated labour. Additionally, the platform presents opportunities for monetisation of aggregated mobility data and its insights by selling it to public transportation agencies.¹⁰⁵ This information provides crucial inputs for policy making in areas such as pollution control and congestion management. In fact, municipal authorities are afforded visibility into rideshare operational patterns and analytics that are otherwise challenging to access or comprehend.

Structure

Driver's Seat is registered as a for-profit, limited cooperative association in Colorado, USA. The primary beneficiaries of this platform are rideshare and delivery workers who have signed up to become members of the cooperative. It adopts a delegated cooperative structure that is governed by a representative board, with at least 51% of the board members drawn from the larger community of workers.¹⁰⁶ This structure is quite unlike MIDATA, considered below, which instead adopts a 'one member, one vote', approach to internal governance.

Rideshare and delivery workers who are a part of Driver's Seat share data with the application, which functions as a data storage and analytics platform. Insights and visualisations created on the platform not only enhance the agency of gig workers by placing data in their control, but also provide crucial information that enables them to optimise for higher wages and better working conditions. Moreover, Driver's Seat enables this data to be shared with local government transit operators, promising to

¹⁰⁵ Dickey (2020), TechCrunch, "Coop helps Uber, Lyft drivers to use data to maximise earnings", <https://techcrunch.com/2020/02/06/co-op-helps-uber-lyft-drivers-use-data-to-maximize-earnings/>

¹⁰⁶ Stewardship Navigator (2021), Aapti Institute, <https://airtable.com/shrH2lvivQ0ughB94/tblHfyAY7elk1plux>

improve the reach and effectiveness of citizen service delivery and experience.¹⁰⁷

Membership fee, grant money and private investments are the intended funding streams for Driver's Seat. However, monetisation of insights and aggregated data sold to public authorities is currently its sole source of revenue.¹⁰⁸

Stakeholders

The primary stakeholders served by Driver's Seat are gig workers: rideshare and delivery drivers employed by Uber, Lyft, DoorDash, Uber Eats, Amazon Flex and Postmates operating within the US. These workers also constitute the data generators whose mobility information and personal data is managed by the platform.

Data users include local government and transportation agencies that use analytics and aggregated mobility data supplied by Driver's Seat as a part of urban policy and planning.¹⁰⁹

The representative board governing the cooperative is the designated decision-making body that authorises all data sharing and selling activities.

Governance principles

The data stewarded by Driver's Seat includes members' personal data as well as anonymised mobility information. The data rights are vested in the cooperative board, with individual members retaining only the right to revoke consent for data use.

Significantly, data sharing and selling decisions are authorised through the cooperative board in which gig workers hold at least 51% representation and voting rights. Further, the members are also entitled to a minimum 51% of the share of profits generated by Driver's Seat. The twin features of representation and share in profits are mandated by the Colorado Uniform Limited Cooperative Association Act, 2012.¹¹⁰

Members of the cooperative participate in data decisions by electing representatives to the board. Individual consent for data access, processing and sharing is obtained at the point of on-boarding to the Driver's Seat application. Lastly, members can revoke consent for any of these functions and request deletion of their data via email.¹¹¹

Privacy controls

Driver's Seat anonymises the data contributed by drivers and shares only aggregated datasets and insights with its data users once approved by the board.¹¹² Personalised

¹⁰⁷ Ibid.

¹⁰⁸ Ibid.

¹⁰⁹ Witt, Hays [Aapti Institute] (2021). "Data Economy Lab | Tracking Stewardship: Driver's Seat - Empowering gig workers through data" [Video]. Youtube. <https://youtu.be/a-l8tfeOB3g>

¹¹⁰ To better understand limited cooperative associations, visit https://www.sos.state.co.us/pubs/business/news/2012/20120402_ULCAA_Dean.html

¹¹¹ Stewardship Navigator (2021), Aapti Institute, <https://airtable.com/shrH2lvivQ0ughB94/tblHfyAY7elk1plux>

¹¹² Ibid.

analytics and visualisation derived from individual members' data is not shared with third parties.

Analysis

Monetisation of aggregated data and insights by Driver's Seat performs two functions - one, it contributes to the income of gig workers who are profit-participants in the entity and two, creates a viable source of revenue that contributes to the financial sustainability of Driver's Seat. Delegated representation afforded through the entity's cooperative board upholds participation of data generators in decision-making as a core of its operating principles.

The experience of Driver's Seat holds interesting insights for data trusts and data trust-like initiatives that hope to facilitate effective purpose-driven data sharing. Additionally, the social value element can be fulfilled by identifying appropriate stakeholders, in this case public transit agencies, who stand to benefit from the use of the data stewarded by the initiative. Delegated representation and decision-making through the cooperative board could potentially reduce the burden on beneficiaries to evaluate granular considerations on data sharing, while simultaneously availing valuable advisory services from the board.

3.3. Open Humans

Overview

Open Humans is a not-for-profit that allows individuals and communities to donate personal data for use in research, education and health projects.

Established in 2015 in the US, the platform helps individuals access and understand their personal data through an Open Humans account and donate it for projects that align with their values or goals. The entity facilitates public benefit data sharing while providing granular and dynamic controls, creating opportunities for bottom-up decision-making and data governance.

Open Humans differs from Driver's Seat on two fronts: Open Humans supports granular decision-making functions by allowing each member to approve or disapprove use of their data in a specific project, and the incentive to participate is altruistic as the entity does not stand to make any profits or generate revenue from its sharing activities.



Figure 6: Structure of Open Humans (**Source:** Open Humans website)

Purpose

Open Humans is designed to empower individuals and communities around their personal information, by combining technology and community governance of data to advance an open, participant-centric approach to human subjects research.¹¹³ It has built a platform for collaboration between communities and researchers, directing data to projects or purposes that fulfill the data donors' personal expectations. The entity also facilitates citizen science and subject participation in health research. Open Humans currently has 12,364 members.¹¹⁴

Structure

Open Humans is registered as a 501(c)(3) nonprofit organization in the US¹¹⁵. In addition to providing a platform for collaboration and data sharing, Open Humans allows individuals to run analytics on their personal data and understand it through free and open source "notebooks" that can be embedded in one's browser.¹¹⁶ The entity is funded through grants from Robert Wood Johnson Foundation, Knight Foundation and Shuttleworth Foundation.¹¹⁷

Stakeholders

Individuals who have signed up as members of the Open Humans community are the data producers, while academic institutions, citizen scientists and researchers comprise the data requestors and users. Data sharing decisions are made by individuals who can

¹¹³ Ball, Mad [Aapti Institute] (2021). "Data Economy Lab | Tracking stewardship: Open Humans - Empowering citizens, patients and researchers through data" [Video]. Youtube. <https://youtu.be/L9GHP-u0gK4>

¹¹⁴ As on 20 September, 2021, reported on the organisation's website.

¹¹⁵ Open Humans Foundation (n.d.). Retrieved from <http://openhumansfoundation.org/>

¹¹⁶ Ball, Mad [Aapti Institute] (2021). "Data Economy Lab | Tracking stewardship: Open Humans - Empowering citizens, patients and researchers through data" [Video]. Youtube. <https://youtu.be/L9GHP-u0gK4>

¹¹⁷ As disclosed in the organisation's website on 20 September, 2021.

agree to sharing their personal data, particularly health data, for a specific project. Therefore, consent is specific to a project and can be revoked at any point.¹¹⁸

Governance principles

The members of Open Humans are the primary decision-makers. They retain full control over their personal data and authorise its use through granular controls on a project-to-project basis.¹¹⁹ Individuals and communities willing to donate their data for research are encouraged to examine the purpose and value of a particular project before agreeing to share their data.

Similarly, any member of Open Humans - irrespective of their academic background - can create projects/research studies about a specific theme. This is best demonstrated through the many citizen science projects hosted on the platform.¹²⁰ All projects undergo a community review process, performing a role akin to the ethics or internal review board within academia. The review is an open and public process in which every member can participate. Communities can de-platform a project as a part of the review process.¹²¹

Privacy controls

Data may or may not be anonymised or pseudonymised before it is shared; it is dependent on the nature of research involved.¹²² Member data is centrally stored on the Open Humans platform and a copy in shareable formats is made available to third parties who have been authorized to access the data.¹²³ Analytics, insights and visualisation accessed by individuals through personalised “notebooks” and such tools hosted on the platform are not shared with third parties, unless otherwise consented to by an individual.

Analysis

Open Humans presents a unique use case of a bottom-up data steward that is explicitly concerned with facilitating data sharing for social benefit. Individuals and communities are invited to become a part of the entity and participate actively in research projects that appeal to them personally, exploring themes that are otherwise often marginalised within conventional academic discourse. For instance, a project on the dynamics and perceptions of the neovagina is currently being hosted on the Open Humans platform. It allows members to deliberate on the research methodology and frame questions that should be addressed as a part of this study. In essence, the platform embeds health

¹¹⁸ Stewardship Navigator (2021), Aapti Institute, <https://airtable.com/shrH2lvivQ0ughB94/tblHfyAY7elk1plux>

¹¹⁹ Ibid.

¹²⁰ Ball, Mad [Aapti Institute] (2021). “Data Economy Lab | Tracking stewardship: Open Humans - Empowering citizens, patients and researchers through data” [Video]. Youtube. <https://youtu.be/L9GHP-u0gK4>

¹²¹ Ibid.

¹²² Stewardship Navigator (2021), Aapti Institute, <https://airtable.com/shrH2lvivQ0ughB94/tblHfyAY7elk1plux>

¹²³ Ibid.

research within a strong framework of citizen-driven science and community interests.

3.4. MIDATA

Overview

MIDATA is a member-owned cooperative that provides an open source technical platform for account holders to store their personal data and share it with researchers and service providers.

Developed by ETH Zurich and Bern University of Applied Sciences, MIDATA Switzerland was established in 2015 and the entity supports the creation of other regional or national cooperatives that use MIDATA's technical infrastructure.¹²⁴ MIDATA account holders can control who has access to their data and direct its use in specific clinical studies.

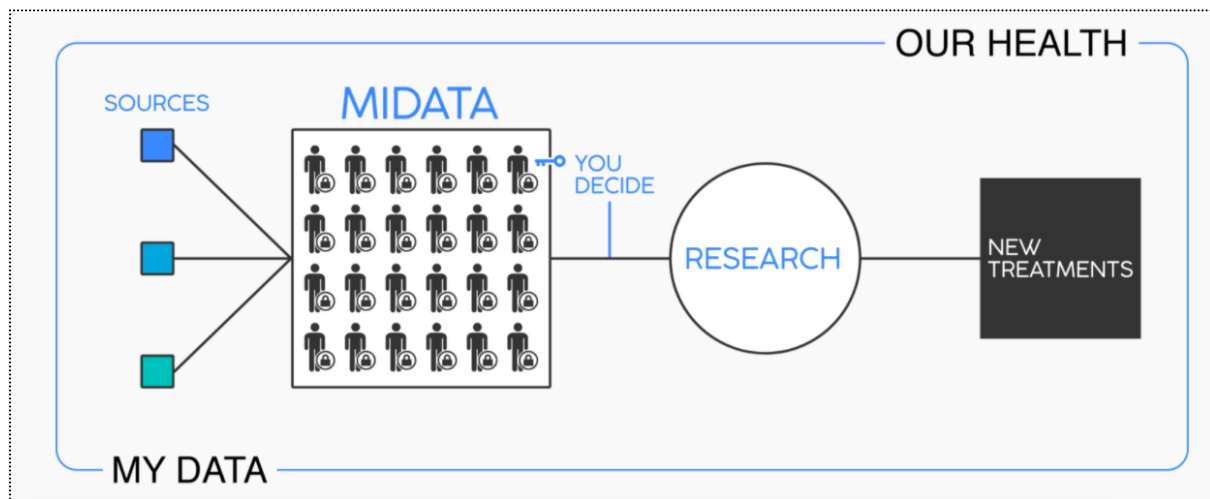


Figure 7: Structure of MIDATA (Source: MIDATA website)

Purpose

The cooperative was established to fulfill two objectives: to enable citizens to gain control over health information and to amass valuable aggregated health data for use in medical research.¹²⁵ MIDATA allows individuals to engage with health research projects and determine if they wish to contribute their data to a certain study. Thus, it facilitates active participation of data subjects in medical research.¹²⁶

Structure

MIDATA is a non-profit cooperative, registered under Article 828 of the Swiss Code of Obligations.¹²⁷ Its primary function includes the development, deployment and maintenance of a common technical infrastructure that allows MIDATA data account

¹²⁴ Ibid.

¹²⁵ Ibid.

¹²⁶ MIDATA (2021). Retrieved from <https://www.MIDATA.coop/en/partners/>

¹²⁷ Corporate law of Switzerland is primarily contained within the Swiss Code of Obligations, from Article 552 to 1186. Cooperatives such as MIDATA are governed as per provisions under Article 828 to 926.

holders and members to store their personal data. Significantly, members of the cooperative govern the use of data through a system of rights vested within Article 4 of MIDATA's Articles of Association.¹²⁸

Members of the cooperative are required to pay a one-time fee of CHF 40 to cover administration and operational costs incurred for managing the cooperative.¹²⁹ Private investments, including grant money, seem to constitute its primary source of funding.

Stakeholders

Individuals who hold an account on the MIDATA platform and members of the MIDATA cooperative are the primary data producers in this context. The nature of data stored and managed through the MIDATA platform is primarily personal data, including sensitive health information such as genomic data and medical records. They consent to share their data on a per-project basis with pharmaceutical companies, research institutions and other interested third parties that make up the category of data requestors.¹³⁰

Governance principles

Data sharing decisions undergo two levels of review and authorisation. First, every proposal containing a request for data is reviewed by MIDATA's Data Ethics Review Board.¹³¹ The board may choose to admit a proposal, depending on the nature and purpose of research involved. Creating value for MIDATA's members and data account holders whilst contributing to society's knowledge on a particular health condition is a crucial consideration in the process of review. Second, the proposal that has been vetted by the Board is then sent to the general assembly of members for further authorisation. The general assembly follows a 'one member, one vote' model typical of cooperatives and a simple majority vote is necessary to gain approval for a proposed project.¹³²

¹²⁸ The original MIDATA Article of Association, published in German, can be accessed here - https://www.MIDATA.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626.pdf ; an unofficial English translation is also available on the website, viewable on https://www.MIDATA.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626_EN.pdf

¹²⁹ MIDATA (2021). Retrieved from <https://www.MIDATA.coop/en/faq/>

¹³⁰ Stewardship Navigator (2021), Aapti Institute, <https://airtable.com/shrH2lvivQ0ughB94/tblHfyAY7elk1plux>

¹³¹ Ibid.

¹³² Hafen, Ernst [Aapti Institute] (2020). "Data Economy Lab | Tracking stewardship: MIDATA - Unlocking value and control over our health data" [Video]. Youtube. <https://youtu.be/MfnDX-Sswr4>

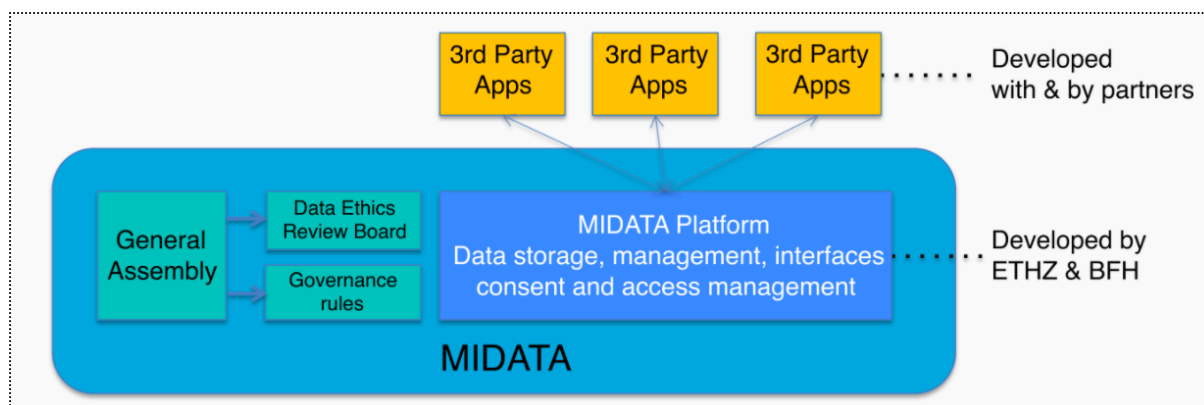


Figure 8: Governance mechanism of MIDATA (**Source:** MIDATA website)

Individual members and data account holders on the platform may consent to independently and exclusively share their personal data (or a specific subset therein) with other account holders, the cooperative itself or third parties requesting data.¹³³ Consent collection is digitally-mediated and obtained through the MIDATA platform. Data producers (i.e. individuals) reserve the right to revoke consent for data use through the process of research and beyond.¹³⁴

Privacy controls

All data on the MIDATA platform is stored centrally on servers in Switzerland and follow multi-level encryption and 'a perfect forward secrecy protocol'.¹³⁵ Individual data may or may not be anonymised or pseudonymised prior to sharing, varying on the proposal under consideration and the nature of data required to undertake research. Third party data requestors are granted access to granular datasets of individuals who consent to sharing data for a specific project, after obtaining necessary authorisation for the project from the Data Ethics Review Board and the general assembly.¹³⁶

Analysis

MIDATA is founded on traditional cooperative principles that have been transposed onto the current data-driven world. It creates an environment for cooperative members to pool their data and use it in pursuit of their collective desires. Sophisticated standards for health data interoperability followed by MIDATA combined with a direct democracy approach to data sharing help achieve transparency and utmost technical safeguards for responsible bottom-up stewardship. Authorisation layers in the nature of MIDATA's Data Ethics Review Board help ensure compliance with community interests and that the members of the cooperative are involved at every step of the data value chain. Lastly, the not-for-profit nature of MIDATA avoids any conflict of interests that

¹³³ Stewardship Navigator (2021), Aapti Institute, <https://airtable.com/shrH2lvivQ0ughB94/tblHfyAY7elk1plux>

¹³⁴ Ibid.

¹³⁵ Ibid.

¹³⁶ Ibid.

may arise when financial imperatives are posed against public good solutions.¹³⁷

3.5 Insights from case studies

The experiences of Driver's Seat, Open Humans and MIDATA chronicled in this section surface valuable reference points on the subject of data trusts and data trust-like initiatives. Although there exists significant differences in their structure - MIDATA and Driver's Seat are data cooperatives while Open Humans is a non-profit - the three use-cases nonetheless deliberately focalise community empowerment as the underlying purpose of their initiatives.

By examining real-world case studies, the research demonstrates the multiplicity of avenues and design choices that are available to builders of data trusts and data trust-like initiatives to actualise bottom-up mechanisms for stewardship. Further, it illustrates how initiatives managing the use of data can embed participation of data generators as a cornerstone of their governance principles. This marks a crucial departure from the current disempowering paradigm of data sharing that is opaque,¹³⁸ extractive¹³⁹ and marginalises the role of individuals and communities in data decisions.¹⁴⁰

These case studies highlight the virtues of bottom-up stewardship. Data sharing decisions within the use-cases considered go hand-in-hand with respect for agential rights of individuals and creation of social benefit through the use of data. Responsible data stewardship makes available data for social benefit in ways that are democratic and privacy-preserving, while balancing complex considerations of market incentives and public welfare.¹⁴¹ Thus, upcoming data trusts and data trust-like initiatives should embody participatory mechanisms for stewardship that can engineer effective outcomes for all stakeholders - communities, private entities and the public at large.

¹³⁷ Kapoor, Aapti Institute (2021), "Rethinking data monetisation", <https://thedataeconomylab.com/2021/06/14/rethinking-data-monetisation/>

¹³⁸ Engler, Brookings (2021), "Tech cannot be governed without access to its data", <https://www.brookings.edu/blog/techtank/2020/09/10/tech-cannot-be-governed-without-access-to-its-data/>

¹³⁹ Sadowski, The Reboot (2021), "The Internet of Landlords Makes Renters of Us All", <https://thereboot.com/the-internet-of-landlords-makes-renters-of-us-all/>

¹⁴⁰ Medina, The Conversation (2021), "NHS data gathering: government plans to collect and share health records are hugely concerning – here's why", <https://theconversation.com/nhs-data-gathering-government-plans-to-collect-and-share-health-records-are-hugely-concerning-heres-why-162699>

¹⁴¹ Tenisson, et al., Open Data Institute and Bennett Institute for Public Policy (2020), "The value of data: Policy implications", https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_Policy_Implications_Report_26_Feb_ok4noWn.pdf

4. Key findings and takeaways

The research generated the following key findings and takeaways on the global state of data trusts:

1. **While there is emerging consensus around what functions a data trust should deliver, there remain questions about the specific operational strategies which can deliver these functions in practice.**¹⁴² However, the research encountered a plurality of bottom-up data stewardship initiatives that enable groups to engage in data sharing for social benefit and embody features that are nonetheless similar and attributed to data trusts.
2. **There is general optimism about the potential of data trusts among people working on data stewardship.** Both practitioner and expert respondents to the survey described a positive outlook on data trusts as an approach to data stewardship. Indeed, many were eager - and in some cases impatient - to see real-world examples to begin to test the considerable theory behind them.
3. **The interest in data trusts as a form of data stewardship seems to be driven from Europe and North America.** The response to the survey - with 37 of the 45 respondents based in Europe and North America - suggests a relative maturity in terms of imagining new forms of data stewardship. This may be due to the existence of data protection regulations that afford data rights, such as the right to access and portability, which are prerequisites to actualise bottom-up initiatives (such as the ones featured here as case studies). This means the maturity of the data rights landscape needs to be borne in mind while recommended data trusts in different jurisdictions.
4. **The purpose for bottom-up data stewardship can differ significantly.** The examples examined in this research are markedly different in their purpose, and subsequently their legal forms, governance processes and business model. Initiatives such as MIDATA and Open Humans are driven by altruistic motivations for data sharing. On the other hand, initiatives such as Driver's Seat, Swash and Digi.me seek to financially compensate those that have contributed data by charging interested parties for access.
5. **There are a number of real-world initiatives that demonstrate multiple routes to realising bottom-up data stewardship.** The survey findings and case studies exhibit a diversity of initiatives united in their efforts to empower individuals and communities to steward data. This highlights that there is no 'one-size-fits-all' framework for operationalising participatory forms of data stewardship.

¹⁴² Data Governance Working Group (2021), Global Partnership for Artificial Intelligence, "Understanding data trusts", <https://ceimia.org/wp-content/uploads/2021/07/2021-07-09-GPAI-summary-understanding-data-trusts-updated.docx.pdf>

5. Endnotes

5.1 About Aapti, ODI, and GPAI

Aapti Institute is a public research firm that works at the intersection of technology and society, building policy-relevant and actionable insights on the digital economy. It was founded in 2019 in Bangalore, India. Through its Data Economy Lab, a flagship effort to rebalance power in the digital economy, Aapti supports research, conversation and experimentation around the practice of data stewardship.

The **Open Data Institute** works to make data work for everyone by working with businesses and governments to build an open, trustworthy data ecosystem. It is independent, nonprofit and nonpartisan, founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. From its headquarters in London and via its global network of startups, members and nodes, the ODI offers training, research and strategic advice for organisations looking to explore the possibilities of data.

The **Global Partnership on Artificial Intelligence (GPAI)** is a multi-stakeholder initiative which aims to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities. Built around a shared commitment to the OECD Recommendation on Artificial Intelligence, GPAI brings together engaged minds and expertise from science, industry, civil society, governments, international organisations and academia to foster international cooperation.

5.2 Authors

This report was written by Astha Kapoor & Soujanya Sridharan from Aapti Institute and Jack Hardinges and Joe Massey from the Open Data Institute. The report was written in collaboration with the GPAI Data Working Group, whose insight and expertise helped to shape the direction, content and focus of this report.

5.3 Report drafting

This report was written in the autumn of 2021, with the research taking place over the summer. The survey was developed in July and distributed over the month of August which was followed by analysis and drafting of the report in September. The first draft of the report was reviewed by GPAI in late September.

5.4 Acknowledgements

We would like to thank GPAI for giving us the opportunity and funding to conduct this research and write this report, and for supporting the research with their knowledge and passion. We also thank the respondents who gave their time to answer the survey - their insights form the basis of this report.

6. Bibliography

6.1 Review of literature (academic papers, blogs, comments, news reports)

1. AI Council & Ada Lovelace Institute (2021, March 4). "Disambiguating data stewardship." *Ada Lovelace Institute*. Retrieved October 6, 2021, from <https://www.adalovelaceinstitute.org/blog/disambiguating-data-stewardship/>.
2. AI Council & Ada Lovelace Institute (2021, March 4). "Exploring legal mechanisms for data stewardship". *Ada Lovelace Institute*. Retrieved October 6, 2021, from <https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/>.
3. Andrejevic, M. (2009, September). "Privacy, Exploitation and the Digital Enclosure." *Amsterdam Law Forum*, 1 (4). Retrieved October 6, 2021, from https://www.researchgate.net/publication/228226821_Privacy_Exploitation_and_the_Digital_Enclosure.
4. Artyushina, A. (2020). "Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto". *Telematics and Informatics*, Volume 55. Retrieved October 6, 2021, from <https://www.sciencedirect.com/science/article/pii/S0736585320301155>.
5. Artyushina, A. (2021, June 10). "The future of Data Trusts and the Global Race to dominate AI". *Bennett Institute for Public Policy*. Retrieved October 6, 2021, from <https://www.bennettinstitute.cam.ac.uk/blog/data-trusts1/>.
6. Bailey, R., & Goyal, T. (2019). "Fiduciary relationships as a means to protect privacy: Examining the use of the fiduciary concept in the draft Personal Data Protection Bill, 2018". *Data Governance Network Working Paper 04*. Retrieved October 6, 2021, from https://datagovernance.org/files/research/NIPFP_Rishab_Trishree_fiduciaries_-_Paper_4.pdf.
7. Balkin, J. M. (2020, November 18). "The fiduciary model of privacy". *SSRN*. Retrieved October 6, 2021, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3700087.
8. Barth, B. (2018, August 18). "The fight against Google's smart city". *The Washington Post*. Retrieved October 6, 2021, from <https://www.washingtonpost.com/news/theworldpost/wp/2018/08/08/sidewalk-abs/>.
9. Baruh, L., & Popescu, M. (2017). "Big data analytics and the limits of privacy self-management". *New Media & Society*, 19(4), 579–596. Retrieved October 6, 2021, from <https://doi.org/10.1177/1461444815614001>.

10. Beller, J. (2018). "The Message is Murder: Substrates of computational capital". *London: Pluto Press*. Retrieved October 6, 2021 from <https://doi.org/10.2307/j.ctt1x07z9t>.
11. Beuno, C. (2016, October 1). "The Attention Economy: Labour, Time and Power in Cognitive Capitalism". *London: Rowman and Littlefield International*.
12. Blankertz, A. (2020, February). "Designing Data Trusts: Why We Need to Test Consumer Data Trusts Now". *Stiftung Neue Verantwortung*. Retrieved October 6, 2021, from https://www.stiftung-nv.de/sites/default/files/designing_data_trusts_e.pdf.
13. Christopher, B. (2019, September 16). "Rentier capitalism: The UK case". *Bennett Institute for Public Policy*. Retrieved October 6, 2021, from <https://www.bennettinstitute.cam.ac.uk/blog/rentier-capitalism-uk-case/>.
14. Coyle, D. (2020, October 30). "Common governance of data: appropriate models for collective and individual rights". *Ada Lovelace Institute*. Retrieved October 6, 2021, from <https://www.adalovelaceinstitute.org/blog/common-governance-of-data/>.
15. Crouch, H. (2021, July 20). "GDPR September implementation date is scrapped". *Digital Health*. Retrieved October 6, 2021, from <https://www.digitalhealth.net/2021/07/gdpr-september-implementation-date-scrapped/>.
16. Data Governance Working Group (2021). "Understanding data trusts". *Global Partnership for Artificial Intelligence*. Retrieved October 6, 2021, from <https://ceimia.org/wp-content/uploads/2021/07/2021-07-09-GPAI-summary-understanding-data-trusts-updated.docx.pdf>.
17. Deming, D. (2021, February 19). "Balancing privacy with data sharing for the public good". *The New York Times*. Retrieved October 6, 2021, from <https://www.nytimes.com/2021/02/19/business/privacy-open-data-public.html>.
18. Dencik, L., & Kaun, A. (2020, June 23). "Datafication and the Welfare State". *University of California Press*. Retrieved October 6, 2021, from <https://online.ucpress.edu/gp/article-abstract/1/1/12912/110743/Datafication-and-the-Welfare-State?redirectedFrom=fulltext>.
19. Dickey, M. R. (2020, February 6). "Co-op helps Uber, Lyft drivers use data to maximize earnings". *TechCrunch*. Retrieved October 6, 2021, from <https://techcrunch.com/2020/02/06/co-op-helps-uber-lyft-drivers-use-data-to-maximize-earnings/>.
20. Dodds, L., Szász, D., Keller, J., Snaith, B., & Duarte, S. (2020, April). "Designing sustainable data institutions". *The Open Data Institute*. Retrieved October 6, 2021, from: <https://theodi.org/article/designing-sustainable-data-institutions-paper/>.

21. Element AI & Nesta (2019) "Data Trusts: A new tool for data governance". *Element AI*. Retrieved October 6, 2021, from https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf
22. Engler, A. (2020, September 10). "Tech cannot be governed without access to its data". *Brookings*. Retrieved October 6, 2021, from <https://www.brookings.edu/blog/techtank/2020/09/10/tech-cannot-be-governed-without-access-to-its-data/>.
23. Forbes Technology Council (2019, September 3). "15 Social Challenges AI Could Help Solve". *Forbes*. Retrieved October 6, 2021, from <https://www.forbes.com/sites/forbestechcouncil/2019/09/03/15-social-challenges-ai-could-help-solve/?sh=76e9dd973533>.
24. Frist, B. (2021, January 20). "NIH Director Dr. Francis Collins: Connecting The Dots From The Human Genome Project To The COVID-19 Vaccine". *Forbes*. Retrieved October 6, 2021, from <https://www.forbes.com/sites/billfrist/2021/01/20/nih-director-dr-francis-collins-connecting-the-dots-from-the-human-genome-project-to-the-covid-19-vaccine/?sh=738447175438>.
25. Graham, M. (2019, October 2). "Lessons From a User-Trusted Data Trust". *DELL Technologies*. Retrieved October 6, 2021, from <https://www.delltechnologies.com/en-us/perspectives/lessons-from-a-user-trusted-data-trust/>.
26. Grueber, M., & Tripp, S. (2011, May). "Economic Impact of the Human Genome Project". *Battelle Memorial Institute*. Retrieved June 6, 2021, from <https://www.battelle.org/docs/default-source/misc/battelle-2011-misc-economic-impact-human-genome-project.pdf?sfvrsn=6>.
27. Gyeonggi-Do (2020, February 20). "Gyeonggi Province Becomes First Local Autonomy in World to Implement a Data Dividend; Governor Lee Jae-myung Says It "Heralds an Era of Data Sovereignty"". *Gyeonggi-Do*. Retrieved October 6, 2021, from <https://english.gg.go.kr/blog/daily-news/gyeonggi-province-becomes-the-first-municipality-in-the-world-to-implement-a-data-dividend-governor-lee-jae-myung-says-it-is-the-beginning-sign-of-the-era-of-data-sovereignty/>.
28. Hardinges, J. (2020, May 17). "Data trusts in 2020". *The Open Data Institute*. Retrieved October 6, 2021, from <https://theodi.org/article/data-trusts-in-2020/>.
29. Hardinges, J., & Keller, J. R. (2021, January 29). "What are data institutions and why are they important?" *The Open Data Institute*. Retrieved October 6, 2021, from <https://theodi.org/article/what-are-data-institutions-and-why-are-they-important/>.

30. Hardinges, J., Wells, P., Blanford, A., Tennison, J., & Scott, A. (2019, April). "Data trusts: lessons from three pilots". *The Open Data Institute*. Retrieved October 6, 2021, from <https://theodi.org/article/odi-data-trusts-report/>.
31. Harper, D. (2016). "Sharing Public Health Data saves lives". *International Journal of Infectious Diseases*, 53, 24–25. Retrieved October 6, 2021, from <https://doi.org/10.1016/j.ijid.2016.11.067>.
32. Hulin, A.-S. (2021, July 14). "How can civil law jurisdictions support data trusts? The Quebec Example". *The Data Trusts Initiative*. Retrieved October 6, 2021, from <https://datatrusts.uk/blogs/how-can-civil-law-jurisdictions-support-data-trusts-the-quebec-example>.
33. Joshi, D. (2020, August 11). "Non-Personal Data: Examining Data Trusts?" *Centre for Law & Policy Research*. Retrieved October 6, 2021, from <https://clpr.org.in/blog/non-personal-data-what-is-data-trusts/>.
34. Kapoor, A. (2021, February). "Collective bargaining on digital platforms and data stewardship". *Friedrich-Ebert-Stiftung*. Retrieved October 6, 2021, from <http://library.fes.de/pdf-files/bueros/singapur/17381.pdf>.
35. Kapoor, A. (2021, June 14). "Rethinking data monetisation". *The Data Economy Lab*. Retrieved October 6, 2021, from <https://thedataeconomylab.com/2021/06/14/rethinking-data-monetisation/>.
36. Keller, J., & Hardinges, J. (2021, June 25). "What are bottom-up data institutions and how do they empower people?" *The Open Data Institute*. Retrieved October 6, 2021, from <https://theodi.org/article/what-are-bottom-up-data-institutions-and-how-do-they-empower-people/>.
37. Kop, M. (2021, April 3). "The Right to Process Data for Machine Learning Purposes in the EU". *Jolt Digest*. Retrieved October 6, 2021, from <https://jolt.law.harvard.edu/digest/the-right-to-process-data-for-machine-learning-purposes-in-the-eu>.
38. Lau Jia Jun, J., Penner, J. E., & Wong, B. (2019, September 23). "The basics of private and public data trusts". *SSRN*. Retrieved October 6, 2021, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3458192#.
39. Lawrence, N. (2015, March 5). "Beware the rise of the digital oligarchy". *The Guardian*. Retrieved October 6, 2021, from <https://www.theguardian.com/media-network/2015/mar/05/digital-oligarchy-algorithms-personal-data>.
40. Lawrence, N. (2016, June 3). "Data Trusts could allay our privacy fears". *The Guardian*. Retrieved October 6, 2021, from <https://www.theguardian.com/media-network/2016/jun/03/data-trusts-privacy-fears-feudalism-democracy>.

41. Lawrence, N., & Delacroix, S. (2019, October 1). "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance". *International Data Privacy Law*. Retrieved October 6, 2021, from <https://academic.oup.com/idpl/article/9/4/236/5579842>.
42. Manohar, S. (2019). "Trust law, fiduciaries, and data trusts". *The Data Economy Lab*. Retrieved October 6, 2021, from https://thedataeconomylab.com/wp-content/uploads/2020/10/DataTrustsPpr_S M.pdf.
43. Manohar, S. (2020, July 31). "Data Sharing for Public Good: Theoretical Bases and Policy Tools". *The Data Economy Lab*. Retrieved October 6, 2021, from <https://thedataeconomylab.com/2020/07/31/data-sharing-for-public-good-theoretical-bases-and-policy-tools/>.
44. Manohar, S., Ramesh, A., & Kapoor, A. (2020, June 24th). "Data Stewardship – A Taxonomy." *The Data Economy Lab*. Retrieved October 6, 2021, from <https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/>.
45. Mathews, L. (2017, September 7). "Equifax Data Breach Impacts 143 Million Americans". *Forbes*. Retrieved October 6, 2021, from <https://www.forbes.com/sites/leemathews/2017/09/07/equifax-data-breach-impacts-143-million-americans/#6f6ed8d3356f>.
46. Maxson Jones, K., Ankeny, R., & Cook-Deegan, R. (2013) "The Bermuda Principles". *Duke University Center for Public Genomics*. Retrieved October 6, 2021, from <https://dukespace.lib.duke.edu/dspace/handle/10161/7407>.
47. McDonald, S. (2019, March 5). "Reclaiming data trusts". *Centre for International Governance Innovation*. Retrieved October 6, 2021, from <https://www.cigionline.org/articles/reclaiming-data-trusts/>.
48. McDonald, S. (2019, October 28). "The Fiduciary Supply Chain". *Centre for International Governance Innovation*. Retrieved October 6, 2021, from <https://www.cigionline.org/articles/fiduciary-supply-chain/>.
49. McFarlane, B. (2020, August 16). "Data Trusts and defining property". *Oxford Law Faculty*. Retrieved October 6, 2021, from <https://www.law.ox.ac.uk/research-and-subject-groups/property-law/blog/2019/10/data-trusts-and-defining-property>.
50. Medina, I. (2021, June 18). "NHS Data Gathering: Government plans to collect and share health records are hugely concerning – here's why". *The Conversation*. Retrieved October 6, 2021, from <https://theconversation.com/nhs-data-gathering-government-plans-to-collect-and-share-health-records-are-hugely-concerning-heres-why-162699>.
51. MIT Technology Review Insights (2020, March 26). "The global AI agenda: Promise, reality, and a future of data sharing". *MIT Intelligence Review*. Retrieved October 6, 2021, from

- <https://www.technologyreview.com/2020/03/26/950287/the-global-ai-agenda-promise-reality-and-a-future-of-data-sharing/>.
52. Montgomery, J. (2021). "International perspectives on data institutions: lessons for data trusts". *The Data Trusts Initiative*. Retrieved October 6, 2021, from <https://datatrusts.uk/blogs/international-perspectives-on-data-institutions-lessons-for-data-trusts>.
 53. Montgomery, J. (2021). "Understanding the Data Governance Act: in conversation with Sylvie Delacroix, Ben McFarlane and Paul Nemitz". *The Data Trusts Initiative*. Retrieved October 6, 2021, from <https://datatrusts.uk/blogs/understanding-the-data-governance-act-in-conversation-with-sylvie-delacroix-ben-mcfarlane-and-paul-nemitz>.
 54. Morozov, E. (2017, October 19-20). "Digital intermediation of everything: at the intersection of politics, technology and finance." *Council of Europe Portal*. Retrieved October 6, 2021, from <https://rm.coe.int/digital-intermediation-of-everything-at-the-intersection-of-politics-t/168075baba>.
 55. Newman, N. (n.d.). "How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population". *Federal Trade Commission*. Retrieved October 6, 2021, from https://www.ftc.gov/system/files/documents/public_comments/2014/08/00015-92370.pdf.
 56. Niklas, J. (n.d.). "Digital Rights are Human Rights". *Digital Freedom Fund*. Retrieved October 6, 2021, from <https://digitalfreedomfund.org/article-22-the-right-to-social-security/>.
 57. Ostrom, E. (2009, December 8). "Beyond Markets and States: Polycentric Governance of Complex Economic Systems". *Nobel Prize Lecture*. Retrieved October 6, 2021, from https://www.nobelprize.org/uploads/2018/06/ostrom_lecture.pdf.
 58. O'Connor, S. (2016, September 8). "When your boss is an algorithm". *Financial Times*. Retrieved October 6, 2021, from <https://www.ft.com/content/88fdc58e-754f-11e6-b60a-de4532d5ea35>.
 59. Patel, R. (2021, September 7). "Participatory Data Stewardship". *Ada Lovelace Institute*. Retrieved October 6, 2021, from <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>.
 60. Patnaik, A. (2021, February 15). "Rethinking Personal Data Regulation in India". *The New Indian Express*. Retrieved October 6, 2021, from <https://www.newindianexpress.com/opinions/2021/feb/15/rethinking-personal-data-regulation-in-india-2264123.html>.
 61. Pozen, D. E., & Khan, L. M. (2019, December 10). "A skeptical view of information fiduciaries". *Harvard Law Review*. Retrieved October 6, 2021, from

- <https://harvardlawreview.org/2019/12/a-skeptical-view-of-information-fiduciaries/>.
62. Ramesh, A., & Kapoor, A. (2020, July 31). "Principles for Revenue Models of Data Stewardship". *The Data Economy Lab*. Retrieved October 6, 2021, from <https://thedataeconomylab.com/2020/07/31/principles-for-revenue-models-of-data-stewardship/>.
 63. Reddy T, P. (2021, July 16). "Counterpoint: Solving India's judicial backlog requires a nuanced conversation". *Scroll.in*. Retrieved October 6, 2021, from <https://scroll.in/article/1000329/counterpoint-solving-indias-judicial-backlog-requires-a-nuanced-conversation>.
 64. Ruhaak, A. (2019, November 13). "Data trusts: Why, what and how?" *Medium*. Retrieved October 6, 2021, from <https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34>.
 65. Sadowski, J. (2021, March 15). "The internet of landlords makes renters of us all". *The Reboot*. Retrieved October 6, 2021, from <https://thereboot.com/the-internet-of-landlords-makes-renters-of-us-all/>.
 66. Sadowski, J., Viljoen, S., & Whittaker, M. (2021). "Everyone should decide how their digital data are used — not just tech companies". *Nature*, 595(7866), 169–171. Retrieved October 6, 2021, from <https://doi.org/10.1038/d41586-021-01812-3>.
 67. Scott, K. (2018, April). "Data for public benefit - understanding patient data". *Understanding Patient Data*. Retrieved October 6, 2021, from https://understandingpatientdata.org.uk/sites/default/files/2018-04/Data%20for%20public%20good_0.pdf.
 68. Sridharan, S., Manohar, S., & Kapoor, A. (2021, September 29). "Health Data Stewardship: Learning from use-cases". *The Data Economy Lab*. Retrieved October 26, 2021, from <https://thedataeconomylab.com/2021/09/29/health-data-stewardship-learning-from-use-cases/>.
 69. Stucke, M. E. (2018, March 17). "Here Are All the Reasons It's a Bad Idea to Let a Few Tech Companies Monopolize Our Data". *Harvard Business Review*. Retrieved October 6, 2021, from <https://hbr.org/2018/03/here-are-all-the-reasons-its-a-bad-idea-to-let-a-few-tech-companies-monopolize-our-data>.
 70. Sundarajan, P. (2020, August 21). "Role of data stewards in enhancing accountability". *The Data Economy Lab*. Retrieved October 6, 2021, from <https://thedataeconomylab.com/2020/08/21/role-of-data-stewards-in-enhancing-accountability/>.

71. Sur (2021). "Online medical platforms are playing fast and loose, collecting patient data", *Medianama*. Retrieved October 6, 2021, from <https://www.medianama.com/2021/09/223-india-digital-health-medical-platform-s-data-consent-records/>. [Paywalled]
72. Szász, D. (2020, October 26). "Tackling climate change challenges through data access – Microsoft and the ODI". *The Open Data Institute*. Retrieved October 6, 2021, from <https://theodi.org/article/tackling-climate-change-challenges-through-data-access-microsoft-and-the-odi/>.
73. Taylor, L. (2017, November 1). "What is Data Justice? the case for connecting Digital Rights and Freedoms globally ". *SAGE Journals*. Retrieved October 6, 2021, from <https://journals.sagepub.com/doi/10.1177/2053951717736335>.
74. Tennison, J., et. al., (2020). "Value of Data Report - Bennett Institute for Public Policy" . *Bennett Institute*. Retrieved October 6, 2021, from https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_Policy_Implications_Report_26_Feb_ok4noWn.pdf.
75. Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016, December 1). "Data colonialism through accumulation by dispossession: New metaphors for daily data". *Environment and Planning D-Society & Space*, 34(6), 990-1006. Retrieved October 6, 2021, from <https://escholarship.org/uc/item/5bf9164g>.
76. Tisne, M. (2020, July 14). "The Data Delusion: Protecting individual data isn't enough when the harm is collective". *Luminate*. Retrieved October 6, 2021, from https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/the_data_delusion_formatted-v3.pdf.
77. Transport Scotland (n.d.). Retrieved October 6, 2021 from <https://www.transport.gov.scot/transport-network/ports-and-harbours/port-governance/>.
78. Tusikov, N. (2019, August 6). ""Urban Data" & "Civic Data Trusts" in the Smart City". *Centre for Free Expression*. Retrieved October 6, 2021, from [https://cfe.ryerson.ca/blog/2019/08/"urban-data"-civic-data-trusts-smart-city](https://cfe.ryerson.ca/blog/2019/08/).
79. Vallance, C. (2021, June 6). "GP data sharing: What is it and can I opt out?". *BBC*. Retrieved October 6, 2021, from <https://www.bbc.co.uk/news/technology-57555013>.
80. Various authors (2021, January). "The Digital New Deal: Visions of Justice in a Post-Covid World". *Just Net Coalition and IT for Change*. Retrieved October 6, 2021, from <https://itforchange.net/digital-new-deal/pdf/>.
81. Verhulst, S., Ramesh, A., Young, A., Rabley, P., & Keefe, C. (2021, June 4). "Establishing a data trust". *PLACE*. Retrieved October 6, 2021, from <https://www.thisisplace.org/blog-1/introducingplace/establishing-a-data-trust>.

82. Vial, G. (2020, December 8). "The data problem stalling AI". *MIT Sloan Management Review*. Retrieved October 6, 2021, from <https://sloanreview.mit.edu/article/the-data-problem-stalling-ai/>.
83. Viljoen, S. (2020, November 11). "A Relational Theory of Data Governance". *Yale Law Journal*. Retrieved October 6, 2021, from, <https://ssrn.com/abstract=3727562>. [Forthcoming]
84. Wilson, D. S. (2016, October 29). "The Tragedy of the Commons: How Elinor Ostrom Solved One of Life's Greatest Dilemma"s. *Economics*. Retrieved October 6, 2021, from <https://economics.com/tragedy-of-the-commons-elinor-ostrom/>.
85. Wylie, B., & McDonald, S. (2018, October 9). "What is a Data Trust?" *The Centre for International Governance Innovation*. Retrieved October 6, 2021, from <https://www.cigionline.org/articles/what-data-trust/>.
86. Zarkadakis, G. (2020, November 10). ""Data Trusts" Could Be the Key to Better AI". *Harvard Business Review*. Retrieved October 6, 2021, from <https://hbr.org/2020/11/data-trusts-could-be-the-key-to-better-ai>.
87. Zuboff, S. (2019, Jan 15). "The Age of Surveillance Capitalism: The Fight for Human Future at the New Frontier of Power". *New York: Public Affairs*.

6.2 Policy, regulation and strategy documents

Legislation/policy/ strategy documents	Jurisdiction	Link
Proposals to modernize Protection of Personal Information and Electronic Documents Act, 2020	Canada	https://news.ontario.ca/en/release/57985/ontario-launches-consultations-to-strengthen-privacy-protections-of-personal-data
Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act), 2020	European Union	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767
Regulation (EU) 2016/679 of the European Parliament and of the Council - General Data Protection Regulation, 2016	European Union	https://eur-lex.europa.eu/eli/reg/2016/679/oj
Entering the new paradigm of artificial intelligence and series (strategy document), 2019	European Union	https://rm.coe.int/eurimages-entering-the-new-paradigm-051219/1680995331

Report by the Committee of Experts on Non-personal Data Governance Framework, 2020	India	https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf
Growing Artificial Intelligence Industry in the UK, 2017 (strategy document)	United Kingdom	https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf
Investigation of Competition in Digital Markets: Department of Justice - Subcommittee on Antitrust, Commercial and Administrative Law, 2020	United States of America	https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519

6.3 Tools, guides and videos

Name	Author/publisher	Type	Link
Stewardship mapper	Aapti Institute	Tool	https://thedataeconomylab.com/mindmap/
Tracking stewardship	Aapti Institute	Video repository	https://thedataeconomylab.com/videos/
Stewardship Navigator	Aapti Institute	Database	https://airtable.com/shrH2IvivQ0ughB94/tblHfyAY7elk1plux
A Human Rights-based Approach to Data	Office of the United Nations High Commissioner for Human Rights	Guide	https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf
Data protection and privacy legislations worldwide	United Nations Conference on Trade and Development	Tool	https://unctad.org/page/data-protection-and-privacy-legislation-worldwide