Data Governance, Part 2: Models and Implementation

December 2023 Research Report

Katy McKinney-Bock, revised February 20, 2024

1 Introduction

In *Data Governance: Part 1*, I set a broader context for why an exploration of data governance models is important, and why currently there is work and focus going into developing data governance practices. Data governance frameworks over the last ten years have focused on themes of trust, protecting user and citizen rights, and use of data for the public interest (Marcucci et al., 2023). This is a reaction to increasing realization of the value in the data individuals produce while interacting in a digital environment, as well as the extractive practices that are being used to leverage that value, often out of the hands of the people who generated the data.

Part 2 of this data governance report pursues two goals (which turns out to be quite ambitious): first, to identify and provide a description of recognized models of data governance that could, in principle, become adopted and put into practice. The second goal was to begin to evaluate the implementation cost for each model, addressing the question: *what would it take to implement this model*?

For the first goal, I review existing typologies of data governance models, including data trusts, data collaboratives, and data collectives. For each model, I provide a description of the model, an example or set of examples from case studies in these resources which illustrate the model in practice, and I summarize any broader research on this model. I leverage resources from my previous reports, such as the Aapti Institute's Data Stewardship Navigator and the Ada Lovelace Institute's work on governance and stewardship (Ada Lovelace Institute, 2020; Soni, 2021a), as well as GovLab's database of over 200 data collaboratives, the Data Collaborative Explorer, to select governance models and case studies, (GovLab, n.d.-c).

In the second section of this report (which is more speculative), I begin to ask questions about how challenging it is to implement these governance models. I include discussion of an evaluation framework that was developed to improve trust in data institutions through better governance practices, published in October 2023 by the GPAI (The Global Partnership on Artificial Intelligence, 2023b). Then, I begin to pose some increasingly more specific questions around what it means to be "lightweight", as possible first steps toward an analytical framework or approach for evaluating the ease of implementation, though with more questions than answers, at this point.

In the next report, I will explore *context*. Some of the resources cited in this report discuss the context-sensitive nature of data governance, and the inability to create either a 'one-size-fits-all' model for governance, or even reproduce a governance model easily in a new context. Questions in the next report will turn to work by D'Ignazio and Klein (2020) on data in context as well as the concept of conceptual integrity (originally a theory of privacy) as applied to the data 'food chain' in Nissenbaum (2019).

2 Responses to standard data governance models

The "status quo" data governance model presents a number of challenges that are addressed by work on alternative data governance models. In short, this model uses contracts like terms of service that have been described as 'take it or leave it' or 'one-way' approaches to governance of data, which leaves the user/data subject disempowered, and lets companies capture and solely benefit from the value of the data. Here is a list of problem statements that grounds research into *alternative* models of data governance (see also Part 1 of this data governance report):

1. **Extraction:** "The market-driven imperatives of corporations have helped create digital enclosures that hampers the ability to use data for broad-based social benefit." (Aapti Institute et al., 2021, p. 13)
2. **Undermined Trust:** "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." (Aapti Institute et al., 2021, p. 13)
3. **Power Imbalances:** "Today's data environment is characterized by structural power imbalances. Those with access to large pools of data – often data about individuals – can leverage the value of aggregated data to create products and services that are foundational to many daily activities." (Ada Lovelace Institute, 2021a, p. 32)
4. **Fragmented/Siloed Ecology:** "Different frameworks within and across countries, regions, sectors, and organizations have resulted in a patchwork of policies, frameworks, and practices, leading to a fragmented ecology that poses certain challenges to the evolution of a common framework." (Marcucci et al., 2023, p. 2)
5. **Free Market:** "We tend to approach data governance with the idea that the current model - which aims for a worldwide free market for data, with some restrictions in a number of countries based on the notion of personal data - is the only possible normative starting point." (Lopez Solano et al., 2022, p. 18)

These discussions about how alternative models of data governance can help to re-empower people with their data are happening broadly, across and within sectors. Within agriculture, a review of data governance practices was recently published by Development Gateway, called *Farmer-Centric Data Governance: Towards a New Paradigm*. It reviews six alternative models of data governance that would work to bring the benefits of data back to farmers, based on the models proposed below: data collaboratives, data cooperatives, data commons, data marketplaces, data trusts, and indigenous data sovereignty (Development Gateway, 2023).

3 Developing alternative models of data governance

There are several papers that attempt to both define and address emerging, or alternative, models of data governance. Most of these papers set out a typology of alternative governance models, though there is not a well-established practice that would support a clearly delineated typology. The models are often theoretical and not yet established in practice - and definitions are partially overlapping, or have characteristics that could both be leveraged in a single governance approach. Mozilla Insights et al. (2020) note that the definitions of these models are "not mutually exclusive and often seek to address entirely different objectives." Lopez Solano et al. (2022) discuss six alternative models of governance which are "mainly proofs of concept or pilots that continue to be developed". Marcucci et al (2023) don't even provide a typology of governance models, but rather they analyze properties of current governance models from the ground up without presenting a resultant typology.

I review the following papers, which contain typologies that have often shared recognition of governance models (though not all of which share the goal of developing a typology), and I overlap these papers to provide a similar overview of each model that is discussed - though this is an approximation and not a clear-cut alignment of the data governance models. The Data Stewardship Navigator presents nine models; Lopez Solano et al. (2022) present six models, the Ada Lovelace Institute examines legal mechanisms for two governance models in depth, and Mozilla presents seven models of governance. For each source, I pull from the case studies that these organizations have made available, including a database of over 110 examples of data stewardship models from Mozilla, a database of data collectives from a GovLab collaboration, and case studies provided in The Data Economy Lab's Data Stewardship Navigator.

In practice, I suspect that it would be necessary to leverage elements from multiple models in order to define an organization's governance framework. It is not clear to me that any of the frameworks provide a method to select the model elements that meet a particular set of needs or context or context. Some models are more 'high level' descriptions of data institutions than others; for example, a data collaborative could take the form of a cooperative.

Table 1: Data Governance Models

Governing data and artificial intelligence for all: Models for sustainable and just data governance (Lopez Solano et al., 2022)	The Data Stewardship Navigator (Aapti Institute, n.d.; Soni, 2021a)	Exploring Legal Mechanisms for Data Stewardship (Ada Lovelace Institute, 2021a)	Shifting Power through Data Governance (Mozilla Insights et al., 2020)	Data trusts: Lessons from Three Pilots (Open Data Institute, 2019) - for "approaches often discussed in the context of data trusts" - but see also the [Map of Data Access](#)	Farmer-Centric Data Governance: Towards a New Paradigm (Development Gateway, 2023)
Data collaborative	Data collaborative		Data collaborative	Research partnerships (subset of data collaborative)	Data collaborative
Public data trust	Data trust	Data trust	Data trust	Data trust	Data trust
Data cooperative	Data cooperative	Data cooperative	Data cooperative	Data cooperative	Data cooperative
	Data exchange				
Personal Data sovereignty	Data marketplace		Data marketplace		Data marketplace / fiduciaries
(Personal data sovereignty)	Personal data store			Personal data store	
Indigenous data sovereignty			Indigenous data sovereignty		Indigenous data sovereignty
Data (semi) commons			Data commons	Data commons	
	Account aggregator				

	Data repository				
			Data fiduciary		(Data marketplace/fiduciaries)
	Ecosystem enabler				

What is important to note is that the models can be interchangeably labeled both governance and stewardship models, depending on the source. In September, I reviewed definitions of data stewardship, and in November I reviewed definitions of data governance. In this report, I will combine models of data governance with models of data stewardship, with an understanding that we are looking at information that would help practitioners to develop and employ these models in real life - which will involve both the notion of governance as a set of practices and policies that formally manage data assets (see 4.1, November), and an understanding that stewardship involves and supersedes governance by going beyond compliance (see 4.2, September) and involving concepts of "long-term care" of data, empowerment, participation, control, and preservation of rights around data (Ada Lovelace Institute, 2021b; Gorr & Zawacki, 2020; GovLab, n.d.-c; Knight, 2023; Soni, 2021b; Wilkinson et al., 2016; Wilson & Thereaux, 2020, among many others).

4 Governance models

4.1 Data Collaborative

4.1.1 Description

GovLab defines data collaboratives as a model "in which participants from different sectors - in particular companies - exchange their data to create public value" (GovLab, n.d.-c). Data collaboratives may have multiple arrangements, but they all share the property of pooling data for a set of stakeholders. Typically they have taken the form of a public-private partnership, but, as both Aapti Institute and GovLab note, this is not a requirement. Data may be hosted by one of the members of the collaborative, who stewards the data, though different arrangements are possible (Aapti Institute, n.d.).

Verhulst (2015) identifies six type of data collaboratives as the dominant models of practice, which was updated by GovLab's Data Collaboratives project in 2019 as the following (S. G. Verhulst et al., 2019):

- Open Access:
    - Public Interfaces
    - Data Pooling
    - Prizes or Challenges
- Restricted Access:
    - Trusted Intermediary

- Research and Analysis Partnerships
- Intelligence Generation

Open access collaboratives involve fewer restrictions on access; public interfaces, such as data platforms or APIs, allow others to access data directly for, e.g., purposes of building applications. Prizes/challenges make data available to applicants or contestants, usually for the purpose of innovation. Data pooling involves shared data resources to pool data collaboratively or create a database with shared data/resources, and these can be public/open or private (GovLab, n.d.-d; S. G. Verhulst, 2015). Restricted access collaboratives involve sharing data with a limited number of partners, such as among research universities or among companies. Intelligence generation involves sharing data that contains information about markets, customers, or other business intelligence trends.

Given the variety of collaborative models, Verhulst et al. (2019) define a set of key variables to represent the nature of different collaboratives: data accessibility, data attributes, collaboration dynamics, and scope. These four properties help determine the approach and methods to governance. *Accessibility* involves open/restricted access, as well as location of the data as online or on-site. *Attributes* of the data involve how long it is made available, whether the data is preprocessed, aggregated and to what extent, who is providing the data/how many providers, and how many datasets are part of the collaborative. *Collaboration dynamics* look at both the engagement around data use (more independent to more cooperative, or directed); the flow of data (unidirectional flow of data, or multidirectional), relationship, and stakeholder sectors. The *scope* of the data collaborative is defined by purpose (how specific/flexible the data use is) and timeframe (how long access to data will exist).

Both Aapti Institute and GovLab identify purposes of data collaboratives to create public good or public value, to desilo data, and to overcome the challenge of data reuse. Of the models discussed in this report, there is the most variance in methods of governance within a collaborative.

4.1.2 Examples

GovLab's website on Data Collaboratives contains a tool with over 150 examples of data collaboratives across the typology they set out. From these case studies, I selected three illustrative examples:

Community Services Industry Alliance (Australia); *Trusted Intermediary* collaborative (GovLab, n.d.-a):

The Community Services Industry Alliance (CSIA) is a program that holds data about community services. The CSIA, started in 2014, is a data collaborative that involves both private industry and the government, as well as a research university, with the goal to "strengthen the community services ecosystem" (Community Services Data Alliance, 2023). Data is collected from industry service providers, and shared with their research partner, the Griffith University Regional Innovation Data Lab, for analysis and insight. Organizations can sign up via an "Expression of Interest" form to contribute their data to CSIA if it relates to a community service area, such as workforce development or homelessness. The CSIA website doesn't have published information about their specific governance practices, but in their FAQ they commit to keeping organizations' data "safe" through the practices and data management experience at Griffith University.

JoinData (Netherlands); *Trusted Intermediary* collaborative (GovLab, n.d.-f):

JoinData is a collaborative that uses a cooperative model of governance. It is a nonprofit organization with a data platform focused around agricultural innovation. The purpose of the data platform is to both provide data for innovation, but also allow farmers to be "in charge" of their data, with agreements around data use, privacy, and safety. Currently, there are use cases for dairy, pig-farming, and arable sectors; and poultry and horticulture sectors are under development. The data platform itself has a set of permissions that individual farmers can set to share the data with other parties. "With the clear dashboard of My JoinData you can manage your own permissions and data and decide which parties can access your data" ("Frequently Asked Questions," n.d.), and the dashboard also has a timeframe and use limitations that the farmer can set.

JoinData has a cooperative governance structure, with a general assembly and a supervisory board. The cooperative has an independent audit committee that monitors data sharing. JoinData itself acts as a trusted intermediary between farmers and data users/organizations, with a focus on data security and a tool that allows individuals to share their data.

Intel's Big Data for Precision Farming (California); *Research Partnership* collaborative (GovLab, n.d.-e):

Intel is working with California universities to place sensors to monitor irrigation techniques as well as measuring snowpack in the Sierra Nevada mountain range (Intel Tackles Water Supply Problems With Big Data, 2014). This collaborative involved a public-private partnership in order to better understand irrigation as well as the food supply – big problems, which a better understanding involves large data generation and storage that is often outside the reach of research universities.

4.1.3 Other Available Research/Resources

The Data Collaborative Explorer (GovLab, n.d.-b) maintains a list of data collaboratives identified by GovLab. The list can be sorted by type, sector, data type, and region, and the three case studies above were located using this tool. Additionally, the Aapti Institute's Data Economy Lab has the Data Stewardship Navigator, which contains case studies and resources about data collaboratives (Aapti Institute, n.d.).

Research on data collaboratives is early, and there are some reports that explore both data governance practices as well as existing models of data collaboratives. For example, Ruijer (2021) establishes governance practices specific to data collaboratives, and through a 'living lab' model looked at governance practices for an "lead-organization led" approach, as opposed to a participant-governed approach. They identified some governance challenges, including clashing institutional logic around how to define public problems and outcomes, as well as measured fear and hesitation around sharing data across organizations.

Verhulst et al (2019) explores the typology of data collaboratives and develops a set of recommendations for implementing collaboratives, including identifying a need for a "scoping methodology" for assessing the variables around a given context, as well as the recommendation that a data steward role be employed within existing organizations for reviewing opportunities for data collaboratives. The third recommendation is the exploration of 'new' intermediaries that can be available to enable data collaboratives, noting that data brokers and third-party analytics providers are currently active in this space.

Verhulst et al (2023) argues for both the importance of data collaboratives, but also for the importance of developing a framework of "mutual commitment" that would enable the formation of data collaboratives in important moments, such as a natural disaster or pandemic. A *Mutual Commitment Framework* is a framework that is intended to set up agreements prior to the moment of need for data in, e.g., a crisis. The MCF is used to quickly set up a data collaborative, by sharing the needs and priorities of the stakeholders in advance.

4.2 Data Trust

4.2.1 Description

A data trust is a model of governance for data sharing that authorizes a designated Trustee or group of Trustees to make decisions around data for the benefit of a group of stakeholders, such as the individual data subjects. The formation of a data trust defines which data rights are being given to the trustee, and the trustee has a *fiduciary duty* to those stakeholders (Ada Lovelace Institute, 2021a; Hardinges, 2018). A data trust can be an independent organization, but it does not have to be (McDonald & Wylie, 2018). The ODI explains that fiduciary duty in the context of data involves "stewarding data with impartiality, prudence, transparency, and undivided loyalty" (Hardinges, 2018). In a trust, decisions around both access to data and data use can be answered, as well as an accountability mechanism for not upholding those decisions.

Definitions vary around trusts; some require data rights to be claimed under trust law, while other definitions don't require a legal definition of trust. Aapti Institute and the ODI note that there are multiple definitions of data trust. The ODI adopts a definition of data trust as: "A data trust is a legal structure that provides independent, fiduciary stewardship of data", based on work by "particularly Lilian Edwards, Sean McDonald, Keith Porcaro and David and Richard Winickoff" (The Open Data Institute, 2020). Ada Lovelace Institute's report notes that legal data trusts can only be formed in areas with trust law, though contract law or agencies can be used in similar ways to create a trust-like arrangement (Ada Lovelace Institute, 2021a). The Global Partnership for AI reached a consensus definition of trust that doesn't require trust law, though retains the concept of fiduciary responsibility and pooled data: "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"(Aapti Institute et al., 2021). All three of these approaches to data trusts recognize that the legal implementation of a data trust may vary based on the context.

Practically speaking, both the ODI's (2019) report on data trust pilots and the Global Partnership on Artificial Intelligence (2022) report on data trusts for climate work have had trouble identifying practical, repeatable examples of data trusts.

The ODI states: “There is not yet an easily repeatable approach to building a data trust”, noting that the context where people are trying to share data varies so much, that it will take time and effort each time to “understand the dynamics in each context, to build working relationships, and, even with a good delivery guide to build something lasting and useful” (Open Data Institute, 2019, p. 13).

In a follow-up report in 2020, the ODI responds to a frequently asked question: “Show me the data trusts.” In this section, Hardinges says that there are many examples of trust-like data sharing arrangements, with the specific properties:

- “Examples of one party authorizing another to make decisions about data on their behalf”
- “Examples of data-holding organizations going out of their way to become beholden to the advice or decision-making authority of independent interests”
- “Analogous attempts [to bottom-up data trusts] to support groups of individuals to contribute data to an entity that stewards it on their collective behalf”

However, Hardinges does not claim that these are data trusts. Rather, he mentions them as they share properties with data trusts and lays out the explicit claim that data trusts have two additional properties: “contextual fiduciary duties” and “independence.” Then, Hardinges provides examples of trusts that have fiduciary duties and independence: the UK Biobank, OpenCorporates, Truata, as well as examples from Facebook, Microsoft, and Sidewalk Labs.

GPAI, in their climate report in February 2022, sets out a criteria for defining a data trust (The Global Partnership on Artificial Intelligence, 2022, p. 10), paraphrased here:

1. ‘Terms of data use and constitution of a trust’
2. ‘Appointed expert trustees’
3. ‘A regime of fiduciary responsibilities’
4. “Negotiate the use of trust assets”
5. ‘Safeguards and oversight’

They immediately note that there were effectively zero examples that fit the entire criteria in their 2021 interim report: “Despite encountering several other real-world bottom-up data stewardship initiatives, it found no examples of data trusts that were able to deliver all of the functions listed above” (The Global Partnership on Artificial Intelligence, 2022, p. 11); a key issue seems to be the role of trustees as a data steward - the 2021 report found, in a survey of practitioners, only half of those surveyed provided a “yes” response to the question of whether an expert trustee had been appointed (Aapti Institute et al., 2021, p. 25).

Given the wide variety of definitions and practices around data trusts, it is perhaps unsurprising that not many examples fulfill emergent criteria of what a data trust should consist of. The differences seem to be both in the operationalizing of legal structures, as well as the level of independence of the pooled data and who is appointed to manage it.

While practices are clearly emerging for data trusts, and the research areas are quite active, there are a few concepts that ground discussion of data trusts that arguably go beyond the immediate specifics of this governance model. To wrap up, I present three themes: (i) thoughts from the GPAI report that the “evolution of data trusts” is emerging from approaches to data stewardship that take a bottom-up approach, such as that of Delacroix and Lawrence (2019), (ii) the lifecycle of a data trust, and (iii) the structural power imbalances in the current data environment. I will examine these three discussions briefly, in order to point to the possibility that 'data trust' is an emerging concept to address problems broader than, perhaps, what a legal trust would be required to address.

Bottom-up Data Trusts

Delacroix and Lawrence (2019) discuss a model of a data ecosystem that consists of many 'bottom-up data trusts' to replace the status quo; they argue that data ownership will be insufficient to solve some of the current issues with individuals retaining control over their data, as well as "collective empowerment aspirations" (Delacroix & Lawrence, 2019, p. 237).

Current challenges they mention are: "a lack of legal mechanisms that plausibly (and collectively) empower data subjects", "recent regulatory endeavors [they mention GDPR] to curb contractual freedom cannot by themselves reverse the power-asymmetry between data controllers/businesses and data subjects/consumers", and "the above risks [the two challenges just quoted] are also difficult to grasp and inherently subjective. The current 'one size fits all' approach does not allow individuals to choose among different approaches to data governance, …, which reflect both their subjective attitude toward risk and moral and political aspirations" (Delacroix & Lawrence, 2019, p. 240).

The solution they propose is the creation of an ecosystem of many different data trusts, allowing individuals to pool their rights with trusts that are aligned with their preferences. They are referring to the legal concept of data trust, with fiduciary responsibility. They discuss the requirement that, for this to be successful, individuals need to be able to 'shop' (data portability and erasure as conditions) to select trusts, and that data security be in place, as well as there being a simple process for creation of new trusts. They argue that this ecosystem would allow for data subject empowerment as well as rebalancing the power of the status quo models of governance right now.

Lifecycle of a Data Trust

The ODI has piloted the data trust model, and has an in-depth body of knowledge on this emerging governance model. In their 2019 report, they point to the lifecycle of a data trust, noting that their research indicates that "each data trust will need its own, individually designed, legal structure" (Open Data Institute, 2019, p. 13). They don't even recommend a set of templates be developed, given the context-sensitivity of each data situation. In this sense, repeatability of data trusts is a challenge - which stands in contrast to the conditions placed on an ecosystem of data trusts by Delacroix and Lawrence (2019) that they be easily implementable.

Despite this, the ODI report does propose a lifecycle of the data trust, to recognize some general processes that occur when developing and setting up a trust. There are six phases to this lifecycle:

1. Scope
2. Co-design
3. Launch
4. Operate
5. Evaluate

6. Retire

   (Open Data Institute, 2019)

This is a useful framework for understanding how their pilots were carried out, and arguably could be a broader description of processes around setting up any governance model, beyond trusts.

Power imbalances

In the literature on data trusts, there are some great problem statements about the status quo, which I will include here. Ada Lovelace Institute, in their (2021) report on data trusts and other legal mechanisms for governance, talk about trusts being a response to the limitations of consent-based models: "The limitations of consent as a model for data governance have already been well-characterized" (Ada Lovelace Institute, 2021a, p. 57). These include lengthy, hard to understand terms, the inability to review data agreements, etc. Critically, in this model, awareness of the governance policies only arises after 'misuse' has been realized, and accountability is lacking.

While it is important to note that they were made in the context of discussion of the data trust model, I believe these statements could be solved by more than one of the governance models discussed in this report. Data trusts are an emerging model to address problems which may be broader than, perhaps, what a legal trust would be required to address – in other words, perhaps the data trust model is a sufficient, but not necessary, condition to resolve the power imbalances that currently exist ("overkill").

Given this, in addition to the observation by GPAI that the European Data Governance Act in 2020 discusses the term *data intermediaries* rather than explicitly specifying a trust (Aapti Institute et al., 2021), any discussion of data trust should be very specific about what is intended. As of this literature review, I have not found a clear, current meta-analysis of this in the literature of existing data trusts. Some open questions remain: what elements of a data trust could be modeled via other legal mechanisms (Ada Lovelace as well) to continue capturing the participatory nature and benefits? Which arguments for/against trusts can be taken more broadly to other governance models?

4.2.2 Examples

PLACE (from the Data Stewardship Navigator): PLACE is a nonprofit organization with a focus on mapping data that contains high-resolution images of the earth's surface, governed by a board of trustees that oversees the trust. PLACE had support setting themselves up with a trust governance model, from GovLab and FutureState. In a series of five blog posts, they documented their process (Establishing a Data Trust, n.d.).

> "We are creating a legal trust – the PLACE Trust, as a permanent legal data trust based in the United Kingdom, which will hold all PLACE data and licenses, received from governments, in perpetuity. All data produced in partnership with PLACE belongs to the government of each country. PLACE receives from each government an irrevocable, perpetual, royalty free license

to a copy of all data and its use by PLACE members through the PLACE Trust. That Trust will issue licenses for use of this data by our members."

Their blog posts document the principles behind their trust, such as Ostrom's governing principles of the commons applied to data, and the documented need for the unified mapping dataset they steward. They also document the challenges of setting up a data trust, as the model was (and is) not well-established with models and expertise.

4.2.3 Other Available Research/Resources

There is a pilot program in the UK, the [Data Trusts Initiative](), in partnership with the University of Cambridge and University of Birmingham, that provides both financial support and mentorship for data trust pilot projects (Data Trusts, n.d.).

There have been significant resources put towards exploring data trusts in recent years. Here is a short list of references that I reviewed, which demonstrate recent research (see also the bibliography of this report):

Aapti Institute. (n.d.). *Data Trust*. Stewardship Navigator - The Data Economy Lab. Retrieved January 2, 2024, from https://tool.thedataeconomylab.com/data-models/10

Aapti Institute, The Open Data Institute, & The Global Partnership on Artificial Intelligence. (2021). *Enabling data sharing for social benefit through data trusts: An Interim Report for the 2021 GPAI Paris Summit*.

Ada Lovelace Institute. (2021). *Exploring legal mechanisms for data stewardship*. https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

Brewer, S., Pearson, S., Maull, R., Godsiff, P., Frey, J. G., Zisman, A., Parr, G., McMillan, A., Cameron, S., Blackmore, H., Manning, L., & Bidaut, L. (2021). A trust framework for digital food systems. *Nature Food*, *2*(8), Article 8. https://doi.org/10.1038/s43016-021-00346-1

Data trusts. (n.d.). Data Trusts Initiative. Retrieved January 2, 2024, from https://datatrusts.uk

Data Trusts and Defining Property | Oxford Law Blogs. (2019, October 29). https://blogs.law.ox.ac.uk/research-and-subject-groups/property-law/blog/2019/10/data-trusts-and-defining-property

Data trusts: Lessons from three pilots (report). (2019, April 14). The ODI. https://theodi.org/news-and-events/blog/odi-data-trusts-report/

Defining a "data trust." (2018, October 19). The ODI. https://theodi.org/news-and-events/blog/defining-a-data-trust/

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data Trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4), 236–252. https://doi.org/10.1093/idpl/ipz014

Edwards, L. (2004). *The Problem with Privacy* (SSRN Scholarly Paper 1857536). https://papers.ssrn.com/abstract=1857536

Establishing a Data Trust. (n.d.-a). PLACE. Retrieved January 3, 2024, from https://thisisplace.org/blog-1/introducingplace/establishing-a-data-trust

Establishing a Data Trust: It's Really Hard. (n.d.-b). PLACE. Retrieved January 3, 2024, from https://thisisplace.org/blog-1/introducingplace/its-really-hard

Farmer-Centric Data Governance: Towards A New Paradigm – Development Gateway: An IREX Venture – Data and digital solutions for international development. (n.d.). *Development Gateway: An IREX Venture*. Retrieved January 2, 2024, from https://developmentgateway.org/publication_landing/farmer-centric-data-governance-towards-a-new-paradigm/

Food Data Trust: A framework for information sharing | Food Standards Agency. (n.d.). https://doi.org/10.5281/zenodo.4575565

Hardinges, J. (2018, July 10). *What is a data trust?* The Open Data Institute. https://theodi.org/insights/explainers/what-is-a-data-trust/

Lau Jia Jun, J., Penner, J. E., & Wong, B. (2019). *The Basics of Private and Public Data Trusts* (SSRN Scholarly Paper 3458192). https://doi.org/10.2139/ssrn.3458192

McDonald, S. (2016, October 25). The Civic Trust. *Medium*. https://medium.com/@digitalpublic/the-civic-trust-e674f9aeab43

McDonald, S. M., & Wylie, B. (2018, October 9). What Is a Data Trust? *Centre for International Governance Innovation*. https://www.cigionline.org/articles/what-data-trust/

O'Hara, K. (n.d.). *Data Trusts Ethics, Architecture and Governance for Trustworthy Data Stewardship*.

O'Hara, K. (2020). Data Trusts. *European Data Protection Law Review*, *6*(4), 484–491. https://doi.org/10.21552/edpl/2020/4/4

OPEN Data trusts in 2020. (n.d.). Google Docs. Retrieved January 2, 2024, from https://docs.google.com/document/d/19W8ljnWRf2EnwRS6zMM5bF5fiuhdqg8TQrhXJxFC8Mw/edit?usp=embed_facebook

Penner, J. E. (2022). *The Law of Trusts* (Twelfth Edition, New to this Edition:, Twelfth Edition, New to this Edition:). Oxford University Press.

PLACE. (n.d.). PLACE. Retrieved January 2, 2024, from https://thisisplace.org

The Global Partnership on Artificial Intelligence. (2022). *Enabling Data Sharing for Social Benefit Through Data Trusts: Data Trusts in Climate; An interim report*.

The Open Data Institute. (2020, March 17). *Data trusts in 2020*. The ODI. https://theodi.org/news-and-events/blog/data-trusts-in-2020/

4.3 Data Cooperative

4.3.1 *Description*

Data cooperatives arise when a group of people come together under shared interests to collectively govern their data. Compared with data trusts, where there is a designated trustee(s) that manages the data and negotiates its use on behalf of the trust, in a cooperative model, the individual stakeholders jointly make decisions around the governance of their data using democratic processes (Aapti Institute, n.d.; Ada Lovelace Institute, 2021a; Lopez Solano et al., 2022). Both Ada Lovelace Institute (2021a) and Lopez Solano et at. (2022) state that the data subjects are the individuals who form the cooperative and retain decision-making power over the data pool. The Data Stewardship Navigator notes that cooperatives may be created explicitly around data pooling, but that existing cooperative structures could also engage in 'digitizing processes' to begin to steward data under an existing cooperative model (Aapti Institute, n.d.). Compared with data collaboratives, which could have a variety of arrangements for exchanges or pooling of data across organizations, data cooperatives involve development of a collective governance structure.

The Data Stewardship Navigator describes the following properties of co-ops (Aapti Institute, n.d.):

1. Members are all considered owners.
2. Collective Bargaining
3. Democratic Control
4. Interest/Clause Alignment
5. Autonomous

Legal implementation of cooperative models may vary, much like data trusts. The Ada Lovelace Institute (2021a) points out that in the UK, the law doesn't expressly define a cooperative. As a unifying statement, Ada Lovelace instead uses the International Cooperative Alliance's definition of a coop: "autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically-controlled enterprise" (Ada Lovelace Institute, 2021a, pp. 49–51).

In terms of a response to standard models, data cooperatives help empower individual data subjects to use the data for their collective benefit. Ada Lovelace points out that this is in contrast to use for social benefit "at large". Ada Lovelace also notes that data commons and data cooperatives are similar; one difference that the report postulates is that a cooperative implies that individuals are bringing in and withdrawing data/information from the cooperative as they decide to join or leave; a data commons would not change or be dependent on members' contributing data or removing data in the same way (Ada Lovelace Institute, 2021a, p. 54). Another difference is that data commons is a much more theoretical prospect, and has not been put into practice as frequently.

4.3.2 Examples

Drivers' Seat Coop (Data Stewardship Navigator, also Ada Lovelace): This data cooperative was cited as a case study by both Ada Lovelace Institute (2021a) as well as in the Data Stewardship Navigator (Aapti Institute, n.d.). Driver's Seat is a cooperative of gig workers that work for ridesharing companies.

Workers join Driver's Seat, contribute their driving data, and the coop's app provides them with insights about their driving. The members' data is also aggregated and sold to governments for transportation planning. Ada Lovelace notes one potential issue with this model, which allows drivers to collaborate to understand their data together, but they are also competing with each other in the economy for clients (Ada Lovelace Institute, 2021a).

MIDATA (Data Stewardship Navigator): The Data Stewardship Navigator identifies MIDATA as an example of a cooperative in the health industry. MIDATA stores personal data, and lets each individual user choose to supply data and participate in research studies. The software platform, the Open MIDATA Server is open source and freely available. The larger cooperative supports development of regional or special-purpose coops based on the MIDATA platform/using the same software.

4.4 Data Commons

4.4.1 Description

A data commons treats data as a shared common resource, and is generally aimed at keeping public goods available. However, definitions of what a data commons are, in addition to what kind of governance model a commons is, was probably the most variant model I came across:

- "As this report is focused on existing examples of practice, in this respect it is difficult to identify actual paradigms of data commons" (Ada Lovelace Institute, 2021a, p. 54)
- "Data commons is the one [approach to governance] with the most existing real world applications" (Mozilla Insights et al., 2020, p. 11)

Looking a bit more broadly to *knowledge commons*, I found a useful definition that could be applied to data: "The basic characteristic that distinguishes commons from noncommons is institutionalized sharing of resources among members of a community" (Workshop on Governing Knowledge Commons et al., 2014, citing Madison, Frischmann & Strandburg 2010: 841). With a contrasting perspective, Ada Lovelace Institute observes that "management of a commons is typically **informal**, via agreed institutions and social norms" (Ada Lovelace Institute, 2021a, emphasis my own).

Implementation as well as definition is variable across sources, and Mozilla Insights et al. (2020) notes that governance structures for realizing a data commons vary. Some of this may be due to the fact that data commons is primarily a theoretical concept; every mention of data commons is grounded in a discussion of Elinor Ostrom's work on governing the commons (see the September research report). Lopez Solano et al. (2022) note that data commons is "powerful imagery", but "yet to be realized even in relation to areas where it has been discussed" (i.e. research and healthcare). In addition to being highly theoretical, Lopez Solano et al. (2022) ground the lack of a "single notion" of data commons in Ostrom's work that cautions against "institutionalizing commons-oriented approaches", as there is no one-size fits all approach that will generate the best outcomes across all conditions, or *contexts*.

Mozilla Insights et al.'s (2020) claim that data commons has several real world applications points to open access movements, pooled research data, and institutions such as Wikipedia as examples of commons, as well as 'digital data commons' that arise from "technical tools and services" as well as cloud computing.

4.4.2 Examples

Case studies were difficult, given that half of the primary sources recognize that data commons haven't yet been realized, while another set of references claims it is frequent. Both sets of sources ground a model of data commons in Ostrom's work, and both refer to the same set of examples (below) as possible examples of commons, so these examples serve as a starting point for understanding approaches to a data commons(Ada Lovelace Institute, 2021a; Mozilla Insights et al., 2020):

- Wikidata
- Wikipedia
- OpenStreetMap

4.5 Data Exchange

4.5.1 Description

The Data Stewardship Navigator is the only source that defines a data exchange as a form of data stewardship or governance. A data exchange is a platform for sharing data, that involves "authorizing access" and "creating interoperability" within a system of data.

4.5.2 Examples

The two examples provided by the Data Economy Lab are:

- Estonia Data Exchange Layer (e-Estonia)
- Indian Urban Data Exchange

Though there are other commercial examples in practice that could be referenced: AWS has a data exchange for accessing third party data, the AWS Data Exchange, the Google Analytics Hub, etc. However, given the focus of alternative models of data governance on moving away from the status quo, and toward governance/stewardship models that empower data subjects, the data exchange does not seem to be a model that is specifically intended or commonly used for this purpose. The government exchanges listed above are closer to a purpose of enabling public data use for good, and data exchanges are not uncommon - but careful theory and analysis of this model as purposefully empowering data subjects has not arisen.

4.6 Data Marketplace (sub: Personal Data Marketplace)

4.6.1 Description

A data marketplace connects an individual's data to a marketplace of data buyers that allows selling or sharing of data. The Data Stewardship Navigator notes that marketplaces supposedly enables more individual control over their data and who they choose to sell it to, nothing that there are models that link a personal data store to a marketplace ("hybridization" of a personal data store with a data marketplace (Aapti Institute, n.d.)). Discussions of data marketplaces involved the concept of "personal data sovereignty" (Lopez Solano et al., 2022) and the value that individuals can both decide when to monetize their data, as well as retain user control over their own data; blockchain is mentioned across the publications in about data marketplaces.

4.6.2 Examples

Case studies presented by the Stewardship Navigator as well as Mozilla Insights are:

- Streamr (both Mozilla and the Stewardship Navigator): Open source blockchain technology; decentralized peer-to-peer network, uses data coin to monetize data.
- Streamlytics: A marketplace that allows individuals to sell their data to Streamlytics, who then sells it to organizations that train AI; according to their webpage, they collect data from individuals' usage on platforms such as Netflix, YouTube, locations on Google Maps, purchasing data on Amazon, etc.
- Bitsaboutme: 'E-bay for personal data' prior to October 2023, but the services have now been stopped. The goal of this institution was to create a place to store personal data and share when needed. However, now Bitsaboutme focuses on data from shopping, providing incentives for users to upload their receipts for food to create a dataset about grocery use (see: https://bitsabout.me/en/howto-en/).
- oneTRANSPORT Data Marketplace (see also DataCollaboratives.org)

4.7 Personal Data Store

4.7.1 Description

A personal data store is a model of governance that enables users to store their own data, and share it when needed with third parties (Open Data Institute, 2019). The data subject/owner retains control over their data. in a sense this is more focused on the technical implementation, and there have been several projects in the last few years that create personal data stores; however, as Bolychevsky & Worthington (2018) discuss, personal data stores "have not seen significant mass-market penetration". In looking at personal data stores, available solutions seem to currently focus more on technical implementation than other data governance models, like data trusts - which are more theoretical, and in practice focus more relationally on who and how to steward data or on the legal mechanisms for establishing stewardship relationships rather than the technical implementation of how.

4.7.2 Examples

digi.ME (from Data Stewardship Navigator): digi.ME is a health records personal data store that can collect all your health records and have them on hand when you need them. It is supported by the Dutch government, which gives Dutch citizens use of digi.ME for free. They call the personal data store a 'data vault', and state their high priority around data security.

Solid: Solid is an open source project proposed by Tim Berners Lee and led by his company, inrupt (Berners-Lee, 2019). Solid was established in order to re-empower individuals to have control over their data: "Solid changes the current model where users have to hand over personal data to digital giants in exchange for perceived value" (Berners-Lee, 2019), which addresses and tries to interrupt the status quo governance model, much as the other alternative governance models do. Solid's contribution is a "data pod", or personal data store, that allows the user to set data sharing permissions - so that the user can decide what to share with web apps, for example.

4.8 Indigenous Data Governance / Indigenous Data Sovereignty

4.8.1 Description

Indigenous data governance models are being developed by indigenous communities to shift control over data from others back into indigenous communities. Mozilla Insights et al. (2020) provides a definition of governance: "in this context, data stewardship entails governance on behalf of (and by a community) throughout the whole data lifecycle, including to determine what data should be 'open' or 'closed' to protect community security or intellectual property" (Mozilla Insights et al., 2020). There are several examples of governance frameworks provided in the Mozilla report, and Lopez Solano et al. (2022) describe the Māori model of governance.

4.8.2 Examples

Te Mana Raraunga (Lopez Solano et al., 2022; Te Mana Raraunga, 2018): Te Mana Raraunga is the Māori Data Sovereignty Network; a highly developed data governance framework that maintains authority for the Māori over their data and empowers Māori communities with decision-making over their data.

Global Indigenous Data Alliance (Mozilla Insights et al., 2020): Developed the CARE principles, going beyond FAIR principles for open science, ensuring that indigenous people's rights/interests are protected.

Canada First Nations (Mozilla Insights et al., 2020): OCAP data principles for use of data assert the rights for First Nations communities to have control over their data collection, ownership, and use (OCAP = ownership, control, access, and possession).

British Columbia First Nations' Data Governance Initiative (BCFNDGI) (Mozilla Insights et al., 2020): BCFNDGI is a governance framework that covers data ownership, access, possession, and the collection/management/reporting on their communities' data. With the stated goal of "lead(ing) their citizens through processes to define and measure what matters to them; this is self-determination – this is self-governance." They recognize that autonomy will happen through recognition of data sovereignty, as well as capacity building.

4.9 Account Aggregator

4.9.1 Description

The Data Stewardship Navigator includes Account Aggregators as a stewardship/governance model. Primarily in the financial sector in India, account aggregators act as an intermediary between a "financial information provider" (e.g. a bank) to a financial information user (e.g. a lender) on behalf of an individual user. Consent comes from the individual, and the account aggregator then requests the information and shares it between the two financial parties. Account aggregators are regulated by the Reserve bank of India.

4.9.2 Examples

Visions (Data Stewardship Navigator): The Data Stewardship Navigator presents Visions as a company with a structure similar to that of an account aggregator. Visions describes their product as a "Personal Data Intermediary" that works with an individual to set preferences for the use of their data that can then be shared with other organizations.

4.9.3 Other Available Research/Resources

Sahamati: Sahamati is an alliance of account aggregators that is working on best practices and adoption of the account aggregation model.

4.10 Data Repository

4.10.1 Description

The Stewardship Navigator also names a data repository as a model of stewardship/governance, describing the model as an archive that enables data use for secondary purposes. As noted, data repositories are frequently used by researchers or academics to archive their data.

4.10.2 Examples

- Findata (Data Stewardship Navigator): https://findata.fi/en/
- SPARC (from NIH registry): https://sparc.science/
- Kids First Data Resource Portal (from NIH registry): https://portal.kidsfirstdrc.org/login

4.10.3 Other Available Research/Resources

- Re3data: This is a large registry of research data repositories, which can be searched to find data repositories.
- The NIH also maintains a registry of scientific data repositories that are approved for NIH research.

5 How challenging is it to implement these models?

Part 2 of this data governance report had two goals: first, to research and provide a description of recognized models of data governance that could, in principle, become adopted and put into practice. The second goal was to begin to evaluate the implementation cost for each model, addressing the question: *what would it take to implement this model*?

The first question has shown incredible variance in the goals and concerns that emerging models address, different aspects/levels of governance, and variance in whether they are more theoretical or practical in nature. It seems that many of these models overlap in scope or could be hybridized - i.e. someone could create a data trust that enables indigenous data sovereignty, with a data repository that lets individual users set certain permissions on the use of their data even while the trustee negotiates the general use of the repository - all addressing different levels of concern and needs around data governance.

The complexity of describing the typology of emergent governance models makes the second goal much more challenging. However, the second goal is still an important one to continue evaluating, especially when considering a Better Deal for Data, which makes a commitment to developing a lightweight set of commitments that will embody best practices in governance, which should importantly rule out any models of data governance that contain extractive practices and which do not leave the data subject in 'control' over how their data is used (see model commitments in our upcoming white paper (Tech Matters, 2024)).

As part of the second goal, I was challenged to consider a hypothesis that there may be a single spectrum along which the governance models could be ordered, in terms of how easily each model could be implemented. This, in turn, will help inform which (set of) model(s) may be best suited for implementation when attempting to conform to best practices in governance. In this section, attempting to address this challenge, I include discussion of an evaluation framework that was developed to improve trust in data institutions through better governance practices, written recently by the GPAI (The Global Partnership on Artificial Intelligence, 2023b). Then, I begin to pose some increasingly more specific questions around what it means to be "lightweight", as possible first steps toward an analytical framework or approach for evaluating the ease of implementation, though with more questions than answers.

5.1 Data institutions & trust

When thinking about forming a data institution like an independent data trust, the October 2023 report by GPAI has developed the "Trustworthy Data Institutional Framework" with a set of practical tools for data governance. Their Governance Framework 2.0 takes the FAIR and CARE principles and extends them into a systematized framework that addresses four areas of data governance, with an acronym to represent principles of each: data decisions (LEAP), data activities (TASQ), data value (FAIR or QRES), and data sharing (CROP or CARE). FAIR are the familiar principles from Go-FAIR (Findability, Accessibility, Interoperability, and Reliability), and CARE refers to the CARE principles of indigenous data by the Global Indigenous Data Alliance (see the GPAI report for more information on this framework, The Global Partnership on Artificial Intelligence, 2023b).

Importantly, the report also creates an assessment framework for 'trustworthiness', based on definitions of 'technical', 'ethical', and 'interaction' aspects of trustworthiness. This governance framework is based on a data commons approach, using the indicators below, which they state are based on Ostrom's work on governing the commons (The Global Partnership on Artificial Intelligence, 2023b, p. 12):

Indicators of Trustworthiness:

- Shared Resources: "The notion of shared resources here is rooted in the open science perspective and refers to open infrastructures, which could include virtual or physical

equipment, sets of instruments, knowledge-based resources, and open computational and data manipulation service infrastructures, …"
- Communities: "This indicator refers to the dynamics around data stakeholders; with value, benefits, and risks of data-driven innovation being distributed equally amongst the communities."
- Rules: "Rules indicate the standards by which an organization governs itself."
- Governance: "This indicator refers to data governance of wider ecosystems."

The full assessment tool can be found in Table 7 on p. 19 of the report. Each indicator variable is assessed on a scale of 0 to 9, where 0-1 is "Unaware" and 8-9 is "Mastering"; the resulting measurements are taken to be the *maturity* of the trustworthiness of the institution, with the most mature organizations scoring highly (8-9).

Along with development of this assessment, a pilot project focused on climate-induced migration in the Lake Chad Basin, in Cameroon, was conducted. Gaps and challenges within the local data ecosystem were identified, which focused around the ability to collect data, build local data infrastructure, and work on local participation of communities. They proposed a next step from the pilot as the development of improving trustworthiness in data governance, valuing "sharing and openness at the center of data exchange" (The Global Partnership on Artificial Intelligence, 2023a) - the basis for the framework proposed above.

Key readings:

- The Global Partnership on Artificial Intelligence. (2023). *Trustworthy Data Institutional Framework: A practical tool to improve trustworthiness in data ecosystems.* Global Partnership on AI.
- The Global Partnership on Artificial Intelligence. (2023). *Designing Trustworthy Data Institutions: Scanning the Local Data Ecosystem in Climate-Induced Migration in Lake Chad Basin—Pilot Study in Cameroon.* Global Partnership on AI.

5.2 What do we mean when we say "lightweight"? First Steps towards an analytical framework/approach

One of the primary goals of a Better Deal for Data is to identify a minimum set of commitments that would build immediate trust between data providers and data users or, perhaps, within a data ecosystem. Outcomes of this would be the ability of two parties to more rapidly and efficiently communicate about expectations to enable data sharing and use, while protecting data subjects' rights and providing compensation/spreading the economic value of data to all involved individuals, as possible.

The commitments are a high-level presentation, and we aim to develop each commitment into a more specific set of principles that can be tested across a wide variety of use cases. The details would then be captured in a set of legally enforceable documents that would be easily accessible and implementable - modeled after the ease of a Creative Commons license. Broadly speaking, a common language/understanding around what best data practices are should easily emerge from reading the commitments - and other frameworks (such as the one above) should be straightforwardly able to assess whether they adhere to the “Better Deal for Data” or not.

However, part of this will depend on what we mean by “lightweight”. It was challenging to assess a given governance model for ease of implementation - in particular, as they were often meeting different challenges and, at times, speaking at different levels of implementation (technology-centered, vs. decision-making or organization-centered).

Here, I will try to sketch out a criteria for what we mean when we say “lightweight”, and consider what next steps we might take to continue building onto this - so that we can most efficiently meet our goals:

Lightweight Approach:

- Low cost
- Low barrier to implementation using clear, established methods
- Structured for immediate implementation (weeks, not years), with an ongoing commitment to longer term development.
- No need to set up an entirely new institution - can be done with existing parties.
- Expresses clearly an adherence to processes that are long-term and responsible

Moreover, there are some questions that come to mind, which I don’t see addressed in the challenges recognized by the literature I read on these governance models. For example: has this model been done before (or is it entirely theoretical, more like the data commons model as per Ada Lovelace’s interpretation/contra Mozilla Insights’ interpretation)? Is there existing technology that can be used to implement the model, or does new technology need to be leveraged (for example, the Solid specification to separate data from the web, into a Pod that is managed by an individual and made available to web apps as needed)? Is there an existing legal framework that is easily copied (like data trusts under trust law)? Does it have external dependencies - i.e., does the model only exist when there is an ecosystem of governance models to interoperate with (‘bottom-up’ data trust ecosystem, a la Delacroix and Lawrence 2018)? When I see challenges like “practical uptake has been slow” for personal data stores due to individuals’ lack of incentive to participate and adopt these platforms (Fallatah et al., 2023), it seems to me that the governance model of leaving personal data to individuals’ control is not sufficient driver of the models’ success in any given context. Empowerment may be implied, but if individuals have blockers (including a lack of interest) in leveraging and participating in the systems, the model may well not be a practical solution in many contexts.

I suspect that many of these decisions and questions need to be addressed upfront, with careful engagement between data stewards, processors, and users, with the communities that originate the data. This, in a sense, may be independent of the selection of any given governance model. I suspect that frameworks such as Verhulst et al.’s (2023) mutual commitment framework are trying to address these upfront decisions that would enable the next kind of discussion of how governance would work. Getting discussion of governance off-the-ground may require a model that is a bit more fundamental than the governance models discussed in this paper.

When you look at the data collaborative vs data trust comparison, the focus is on establishing who the decision-makers are for data sharing and use - where the accountability lies. When you focus on personal data stores and marketplaces, there is an assumption that most of the decision-making aspects of governance lies with the individual user ('control') and then the focus is on technical solutions for creating ways to interact and share out that data (Solid, blockchain, etc.) as well as on security of the information. When you look at data sovereignty, it is about *who decides who decides* - empowering indigenous and other communities to create governance frameworks that work for their data and context.

When considering how to assess the ease with which one could implement a model of governance, context is important, but also practical considerations of implementation. Why not put all the data into a data repository? This is an incomplete model when thinking about who gives permissions to pull data from the repo. Why not set up a trust for every context? A trustee may not be available (much less an expert trustee).

I think what is needed as an immediate next step, is a discussion of the process behind setting up governance systems - mapping community intent to available resources and appropriate example models. Finding ways to facilitate discussions of governance both while thinking about initial data collection as well as found data/data discovery, will help organizations that haven't delved into this field of practice be more cognizant of what is needed. We need to translate the BD4D commitments not into a set of approved governance models, but into a set of straightforward processes for establishing the context and setting of what characteristics of governance are needed for which data.

6 Conclusion

This report provides thorough coverage of existing typologies of data governance and stewardship models, providing key references and examples in the existing literature for emerging governance models and best practices. Secondly, I raise questions about challenges that remain unaddressed by the use of a typology to inform and govern decision-making around data. I include a brief overview of an evaluation framework for "trust" in data, and I pose questions around what it means to be "lightweight", as possible first steps toward an analytical framework or approach for evaluating the ease of implementation. While there are more questions than answers at this point, what is emerging is a need to move beyond an understanding of "models" and implementation of such models, to explore what elements of practice are critical for building trust that are feasible and lightweight to implement.

In the next report, I will explore different notions of *context* around data, with one primary question: how can the Better Deal for Data begin to test feasibility of a general set of commitments, while taking context (community, geographical, different types of data) into account?

This work is supported by a subaward from OpenTEAM as an initiative of Wolfe's Neck Center for Agriculture and the Environment, specifically funded by the U.S. Department of Agriculture under agreement number NR233A750004G032. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of any funder. In addition, any reference to specific brands or types of products or services does not constitute or imply an endorsement.

Please send feedback via email to katy@techmatters.org.

References

Aapti Institute. (n.d.). *Data Collaborative*. The Data Stewardship Navigator. Retrieved December 20, 2023, from https://tool.thedataeconomylab.com/our-data-models

Aapti Institute. (n.d.). *Stewardship Navigator—The Data Economy Lab*. Retrieved September 26, 2023, from https://tool.thedataeconomylab.com/

Aapti Institute, The Open Data Institute, & The Global Partnership on Artificial Intelligence. (2021). *Enabling data sharing for social benefit through data trusts: An Interim Report for the 2021 GPAI Paris Summit*.

Ada Lovelace Institute. (2020). *Rethinking data and rebalancing digital power*. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/11/Ada-Lovelace-Institute-Rethinking-data-and-rebalancing-digital-power-FINAL.pdf

Ada Lovelace Institute. (2021a). *Exploring legal mechanisms for data stewardship*. https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

Ada Lovelace Institute. (2021b). *Participatory data stewardship*. https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/

Berners-Lee, T. (2019, January 2). One Small Step for the Web…. *Medium*. https://medium.com/@timberners_lee/one-small-step-for-the-web-87f92217d085

Bolychevsky, I., & Worthington, S. (2018, October 4). Are Personal Data Stores about to become the NEXT BIG THING? *Medium*. https://medium.com/@shevski/are-personal-data-stores-about-to-become-the-next-big-thing-b767295ed842

Community Services Data Alliance. (2023). *Annual Review 2023*. https://csialtd.com.au/wp-content/uploads/2023/11/CSIA-AnnualPlan-2023-final-web-1.pdf

Data trusts. (n.d.). Data Trusts Initiative. Retrieved January 2, 2024, from https://datatrusts.uk

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data Trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4), 236–252. https://doi.org/10.1093/idpl/ipz014

Development Gateway. (2023). Farmer-Centric Data Governance: Towards a New Paradigm. *Development Gateway: An IREX Venture*. https://developmentgateway.org/blog/farmer-centric-data-governance-models-protecting-farmers-and-food-systems-today-and-tomorrow/

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press. https://mitpress.mit.edu/9780262547185/data-feminism/

Establishing a Data Trust. (n.d.). PLACE. Retrieved January 3, 2024, from https://thisisplace.org/blog-1/introducingplace/establishing-a-data-trust

Fallatah, K. U., Barhamgi, M., & Perera, C. (2023). Personal Data Stores (PDS): A Review. *Sensors (Basel, Switzerland)*, *23*(3), 1477. https://doi.org/10.3390/s23031477

Frequently Asked Questions. (n.d.). *JoinData*. Retrieved January 2, 2024, from https://join-data.nl/en/faq/

Gorr, K., & Zawacki, K. (2020, February 17). Data Stewardship—What is it and why does it matter? *Mozilla Foundation*. https://foundation.mozilla.org/en/blog/data-stewardship-what-it-and-why-does-it-matter/

GovLab. (n.d.-a). *Community Services Data Alliance (CSDA)*. Retrieved January 2, 2024, from https://datacollaboratives.org/cases/community-services-data-alliance-csda.html

GovLab. (n.d.-b). *Data Collaborative Explorer*. Retrieved December 11, 2023, from https://datacollaboratives.org/explorer.html

GovLab. (n.d.-c). *Data Collaboratives*. Retrieved December 8, 2023, from https://datacollaboratives.org/

GovLab. (n.d.-d). *Data Collaboratives—Phase 1: Demand*. Retrieved December 21, 2023, from https://datacollaboratives.org/canvas.html

GovLab. (n.d.-e). *Intel's Big Data for Precision Farming*. Retrieved January 2, 2024, from https://datacollaboratives.org/cases/intels-big-data-for-precision-farming.html

GovLab. (n.d.-f). *JoinData*. Retrieved January 2, 2024, from https://datacollaboratives.org/cases/joindata.html

Hardinges, J. (2018, July 10). *What is a data trust?* The Open Data Institute. https://theodi.org/insights/explainers/what-is-a-data-trust/

Intel Tackles Water Supply Problems With Big Data. (2014, June 9). Data Center Knowledge | News and Analysis for the Data Center Industry. https://www.datacenterknowledge.com/archives/2014/06/09/intel-tackles-water-supply-problems-big-data

Knight, M. (2023, June 20). What Is Data Governance? Definition, Types, Uses. *DATAVERSITY*. https://www.dataversity.net/what-is-data-governance/

Lopez Solano, J., de Souza, S., Martin, A., & Taylor, Linnet. (2022). Governing data and artificial intelligence for all: Models for sustainable and just data governance. *European Parliament*. https://doi.org/10.2861/915401

Marcucci, S., Alarcón, N. G., Verhulst, S. G., & Wüllhorst, E. (2023). Informing the Global Data Future: Benchmarking Data Governance Frameworks. *Data & Policy*, *5*, e30. https://doi.org/10.1017/dap.2023.24

McDonald, S. M., & Wylie, B. (2018, October 9). What Is a Data Trust? *Centre for International Governance Innovation*. https://www.cigionline.org/articles/what-data-trust/

Mozilla Insights, Jonathan van Geuns, & Ana Brandusescu. (2020). *Shifting Power Through Data Governance*.

Nissenbaum, H. (2019). Contextual Integrity Up and Down the Data Food Chain. *Theoretical Inquiries in Law*, *20*(1), 221–256. https://doi.org/10.1515/til-2019-0008

Open Data Institute. (2019). *Data trusts: Lessons from three pilots (report)*. https://theodi.org/news-and-events/blog/odi-data-trusts-report/

Ruijer, E. (2021). Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly*, *38*(4), 101612. https://doi.org/10.1016/j.giq.2021.101612

Soni, S. (2021a, September 28). Building the Stewardship Navigator: Our Approach and Methodology. *The Data Economy Lab*. https://thedataeconomylab.com/2021/09/28/building-the-stewardship-navigator-our-approach-methodology/

Soni, S. (2021b, November 22). Empowering communities with data stewardship. *The Data Economy Lab*. https://thedataeconomylab.com/2021/11/22/empowering-communities-with-data-stewardship/

Te Mana Raraunga. (2018). *Principles of Māori Data Sovereignty*.

Tech Matters. (2024). *A Better Deal for Data: The Case for Model Data Commitments and Agreements*. Manuscript. https://drive.google.com/file/d/13upeQ50ZGblzEUsRA6vWm5nrOVEAFMOm/view?usp=sharing

The Global Partnership on Artificial Intelligence. (2022). *Enabling Data Sharing for Social Benefit Through Data Trusts: Data Trusts in Climate; An interim report*.

The Global Partnership on Artificial Intelligence. (2023a). Designing Trustworthy Data Institutions: Scanning the Local Data Ecosystem in Climate-Induced Migration in Lake Chad Basin—Pilot Study in Cameroon. *Global Partnership on AI*.

The Global Partnership on Artificial Intelligence. (2023b). Trustworthy Data Institutional Framework: A practical tool to improve trustworthiness in data ecosystems. *Global Partnership on AI*.

The Open Data Institute. (2020, March 17). *Data trusts in 2020*. The ODI. https://theodi.org/news-and-events/blog/data-trusts-in-2020/

Verhulst, S. G. (2015, January 30). Mapping the Next Frontier of Open Data: Corporate Data Sharing. *Internet Monitor 2014: Data and Privacy*. https://medium.com/internet-monitor-2014-data-and-privacy/mapping-the-next-frontier-of-open-data-corporate-data-sharing-73b2143878d2

Verhulst, S. G., Young, A., Winowatan, M., & Zahuranec, A. J. (2019, October). *Data Collaboratives: A Descriptive Analysis and Typology of Existing Practices*. http://localhost:4000/existing-practices.html

Verhulst, S., Schroeder, A., & Hoffman, W. (2023). The Time is Now: Establishing a Mutual Commitment Framework (MCF) to Accelerate Data Collaboratives. *United Nations University Centre for Policy Research Working Papers*. https://unu.edu/publication/time-now-establishing-mutual-commitment-framework-mcf-accelerate-data-collaboratives

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), Article 1. https://doi.org/10.1038/sdata.2016.18

Wilson, R., & Thereaux, O. (2020). *Designing Trustworthy Data Institutions*. The Open Data Institute.

Workshop on Governing Knowledge Commons, Tom Dedeurwaerdere, Brett Frischmann, Charlotte Hess, David Lametti, Michael Madison, Charles Schweik, & Katherine Strandburg. (2014, August 15). *An Introduction to Knowledge Commons*.