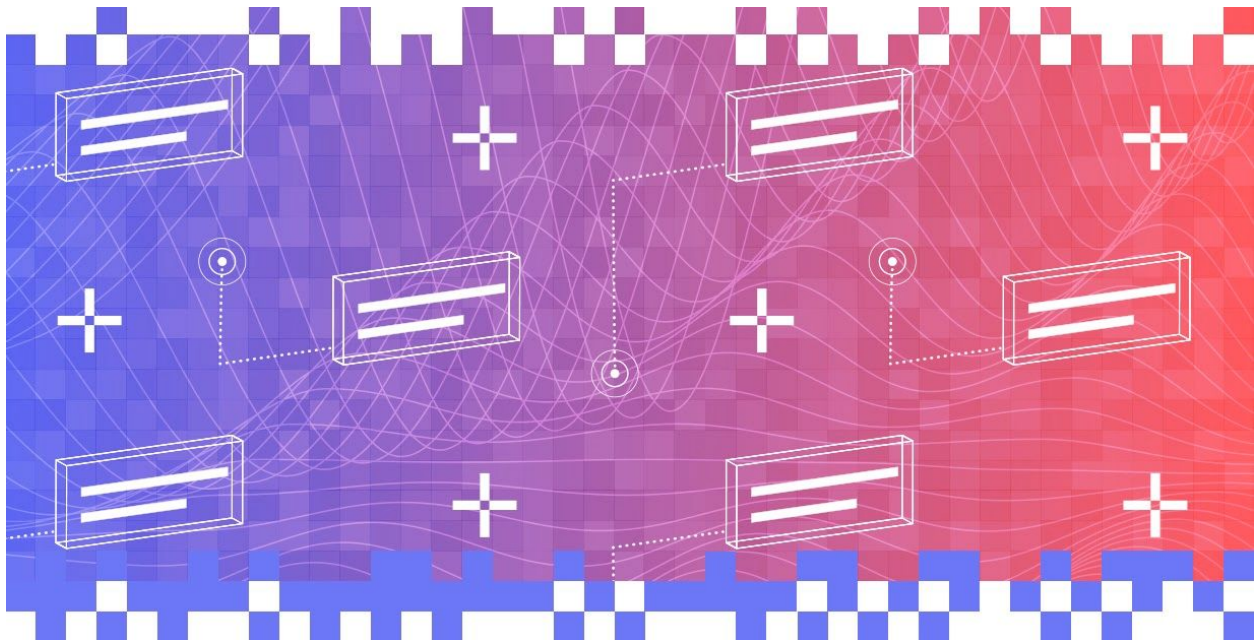


Shifting Power Through Data Governance



Authors: Mozilla Insights with Jonathan van Geuns and Ana Brandusescu

September 2020

This analysis is licensed under [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/)

Table of contents

Preface	3
Introduction	4
Key takeaways	7
7 data governance approaches explored	8
1. What is a data cooperative?	9
2. What is a data commons?	10
3. What is a data collaborative?	12
4. What is a data trust?	13
5. What is a data fiduciary?	15
6. What is Indigenous data sovereignty?	16
7. What is a data marketplace?	18
Looking into data futures	20
Frequently asked questions (FAQs)	21
Acknowledgements	22

Preface

Ideas about new forms of data governance are on the rise, suggesting alternatives to big tech monopolies. Could this help us *reimagine, reconstitute, and rebalance* skewed power dynamics?

How data is controlled and governed is central to Mozilla's understanding of internet health and trustworthy AI. We are committed to working with others to reimagine data governance as a building block toward a [trustworthy AI ecosystem](#).

The Data Futures research presented here is an open study by Mozilla Insights that examines approaches to data stewardship and alternative data governance (in theory and in practice).

This research will help inform decisions to support new technologies and infrastructure, as well as support further cross-disciplinary thinking as part of Mozilla's new Data Futures Lab. The Data Futures Lab will resource projects and foster networks across the ecosystem of actors seeking to realize data governance alternatives which redirect agency, ownership, value, and power back to the people from whom these elements are commonly extracted. We believe that informed and collaborative experimentation in the data governance field, motivated beyond the consolidation of power and profit, should be a priority and at the center of our work in this space. The more people deploy trustworthy models to govern and handle data in the future, the bigger impact it will have on the health of the internet and the AI field we are trying to change.

To kick things off with the Data Futures Lab, we will initiate convenings and we'll fund initiatives to design, build, test, or scale future products that can unleash the creative, financial, and political potential of more equitable data futures.

As part of this broader effort, [this research series](#) will be expanded periodically with more publications throughout 2020. We are committed to working in the open, collaborating with others, and sharing our findings widely to advance and complement a growing community of thought.

Introduction

Something is happening in the world of data. Yes, more data is being collected, analyzed, used, and reused at practically every juncture between life and the internet. At the same time, concerns over personal privacy and the surveillance economy are increasing too. As new realms of possibility open up for use of data, including for artificial intelligence (AI), are we approaching a time where fundamental critiques of data extraction also begin to sprout new innovations for how we collect, share, and generate purposeful value from data?

The short answer is yes. But if the dominant paradigm is one where companies like Facebook offer free services to gather data about every facet of human activity to sell ads against, what are the actual emerging alternatives? Who could even compete with such a market share advantage to empower the individuals, communities, and societies from which so much data stems? Over a decade of scholarly writing and research into data governance law and theory show that different pathways for 'stewarding' data within trusts, commons, collectives, collaboratives, fiduciaries, and more, are indeed opening up. These various approaches each have their imperfections, but at least in theory there is hope for alternatives to help redefine how we conceive of data to address imbalances of power between data holders and data subjects.

Our values

The fluid and overlapping definitions of data stewardship evident in literature and theory inevitably carry over into how we consider examples of them in real life. In practice, innovators of all kinds and in different contexts are still at the beginning of figuring out what works, for whom, and with what data governance approach. Do the many initiatives we see today for rethinking data in agriculture or health or mobility constitute a credible challenge to the entire surveillance economy? No. Does technology to manage personal data or to distribute benefits from datasets more evenly offer a glimpse of the future? Perhaps. What we do see is a clear rise in the creativity and ingenuity of display to solve real world challenges helped by data in different ways. In the past year alone, countless responses to the COVID-19 global pandemic show just how far we have come in considering data as a key to solving big problems, for better or worse.

By collaboratively studying and eventually supporting attempts to reimagine, reconstitute, and rebalance data governance, we believe we may help influence a paradigm shift in a positive direction. The values that guide us are embedded in our

frameworks for [internet health](#), [Trustworthy AI](#), and the [Mozilla Manifesto](#). We are critical of tech monopolies, critical of tech solutionism, and deeply concerned with privacy, transparency, and accountability. And we are just as concerned when it comes to surveillance and (ab)use of data by the public sector. The learnings we wish to summarize for you are sandwiched between theory and practice. Our goal is to better understand what it takes to build bridges from conceptual frameworks into actual innovations that offer better prospects for humanity in all its diversity of race, gender, geography, ability, and more. We believe in unlocking opportunities for people to gain greater insight and meaningful agency over the different kinds of data that is generated about them.

Our approach

With external researchers Jonathan van Geuns and Ana Brandusescu we have assembled [a literature catalog of more than 250 resources](#) that discuss different facets of data stewardship and alternative forms of data governance. From here, we have siphoned a list of the most commonly discussed data stewardship approaches and defined them. What you will encounter is a general overview accompanied by a few examples and a link to relevant literature. This is not a detailed topography nor legal analysis. It will serve as a foundation from which to proceed with more research and conversations with allies and builders, including many experts who helped review this work. We consider 'stewardship' to imply guidance toward a societal goal, while 'governance' refers to a process for making decisions and exercising power (we look to [Aapti Institute](#) and [Ada Lovelace Institute](#) for more refined scrutiny of the term 'stewardship').

In the absence of broad consensus around definitions and categories, especially in different languages and across different legal, political, and economic realities, we have worked with a team of regional researchers from Africa, Central and Eastern Europe, South East Asia, and Latin America to assess what types of initiatives, companies, and organizations exist around the world that employ exemplary, new approaches to data governance. Together, we have compiled a non-exhaustive [database](#) of more than 110 initiatives that we plan to continue expanding. [Public input is welcomed!](#) We hoped to uncover global interest in data stewardship, but [our analysis](#) concludes that there is a pronounced geographic imbalance as to where alternative models (as described below) are hatched and to what extent they are currently on the radar of even seasoned open data and digital rights professionals and scholars worldwide.

What's next?

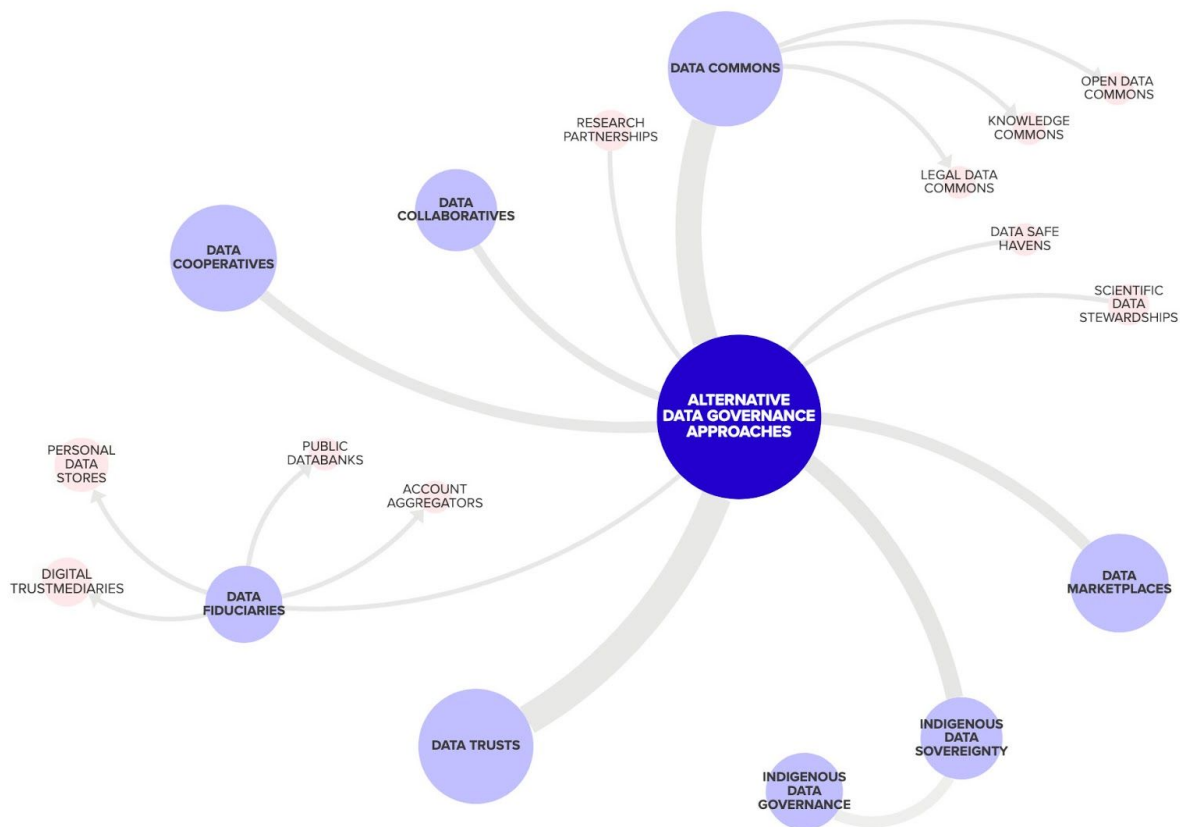
Beyond a handful of countries in Europe, Canada, and the United States we did not find evidence that these theories hold high currency in other regions, or in other major languages than English. There are exceptions, and indeed many interesting ones. But as we continue to consider how new data governance approaches may offer true relief from the myriad of harms people do (and will) experience from highly centralized and discriminatory data power worldwide, we can only emphasize that far more diverse input, co-creation, and research is necessary for the evolution of this field. For our part, we will continue an open minded but critical interrogation of data stewardship (in theory and in practice) throughout the next several months. We'll also be examining the sprouting 'ecosystem' of organizations, companies, and software developers that support innovation in this field. If you wish to be in touch with us, let us know.

As Mozilla, we have a particular interest in understanding how software and technology can be developed to benefit consumers while upholding values of equality and human rights. In better understanding stewardship, we wish to think strategically and collaboratively about our own role in the technology ecosystem to support bright ideas and real alternatives to the status quo.

Key takeaways

- We see genuine potential in different contexts for data stewardship to help open new paths away from the dominant approaches employed by big tech.
- This is an emerging field. There are relatively few examples of scalable initiatives for more than a handful of data governance approaches that are frequently repeated.
- Different data stewardship and governance approaches are not all mutually exclusive. They can be mixed and matched to accomplish a variety of goals.
- We see clear geographic and linguistic imbalances in the discourse surrounding data stewardship, alternative data governance approaches, and related legislation.
- There is a budding ecosystem of thought leadership and support for inspiring, incubating, and testing new data governance approaches in practice.

7 data governance approaches explored



Of the seven data stewardship and governance approaches described below, the first six are the most commonly mentioned in more than 250 literature resources reviewed. The final one is honorable mentions, but each exemplifies an interesting aspect of data governance. For each, we examine how they can be vehicles of empowerment for data subjects like you and me.

Before you dive into this list of definitions, please know that the different approaches are not mutually exclusive and often seek to address entirely different objectives. In other words, if you were to create a new data stewardship initiative yourself, you might employ a number of different approaches in combination with one another. Alternatively, it's possible you may have trouble discerning the difference between some approaches with very similar purposes. As mentioned in the executive summary,

this is an emerging field and these descriptions are based on theoretical explorations, rather than clear cut models with definitions everyone adheres to.

What you will find here is a simple overview of each approach, including at least one initiative, organization, or company that exemplifies it. Our aim is to make it easier to navigate the field and recognize key terms used by scholars and technologists. For each, we also link to a list of literary references on which our understanding is based. In sum, you can browse our [catalog of literature](#), see what definitions we logged by others on this [board](#), review our [database of projects](#), and [submit your own examples](#) to participate.



What is a data cooperative?

#co-owned #purposeful #value

A data cooperative is a legal construct to facilitate the collaborative pooling of data by individuals or organizations for the economic, social, or cultural benefit of the group. The entity that holds the data is often co-owned and democratically controlled by its members. Worldwide, dozens of platform cooperatives have been formed in recent years in direct opposition to major Silicon Valley platforms (say, for instance, [FairBnB](#) as an alternative to AirBnB) but only a subset of them are also data cooperatives. One example is [Driver's Seat](#), a cooperative of on-demand drivers who gather their own combined driving data in an app to gain insights that are usually kept secret by employers like Uber. When Driver's Seat sells mobility data to city agencies they share profits with drivers. In this way, cooperatives can shift power to data subjects, who typically have very few rights in mainstream ventures. From personal concerns (like privacy or labor rights) to societal concerns (like gentrification or market monopolies) cooperatives clearly offer opportunities for channeling discontent into alternatives that serve the platform economy. [Resonate](#), for example, is a cooperative online music network where listeners, musicians, labels, and staff [share profits](#) from streaming and downloads using blockchain technology for music [metadata, licensing, and payments](#). Instead of secrecy and proprietary control, innovators of data cooperatives aim for clarity and openness around the value of data along with new paths for [communal governance](#). That doesn't mean that cooperatives are always governed the same way. Since long before the internet, many countries have known cooperatives to be effective,

for instance in the banking or agricultural sectors. But jurisdictions everywhere differ on what kind of cooperatives can be created, for what purpose, and which stakeholders. Software plays an integral role too, with many data collaboratives founded around apps for data collection or analysis to either serve the purpose of one cooperative or to enable others to meet theirs. In fact, offering [alternatives to corporate cloud infrastructure](#) to cooperatives is one purpose a number of software developers appear to be drawn to as cooperatives themselves: [Collective Tools](#), [FairApps](#), and [CommonsCloud](#) are just a few. Data cooperatives are governed by legal or fiduciary obligations agreed to by all, in order to benefit members individually, and empower members collectively. They have the potential to help redress power inequities posed by 'big data' though democratizing governance of data at scale certainly presents challenges too.

Reaping the benefits of patient insights as a cooperative

[Savvy Coop](#) in the United States is a platform for patients to share their personal insights and experiences with companies and get paid fairly. Members are co-owners of the enterprise with voting power, and they choose themselves with whom to share information. Their mission is to improve health care and patient services through inclusive and collaborative design (breaking with a tradition in the health sector of developing products without speaking to patients first). As the steward of the cooperative, Savvy is legally required to operate in the best interests of patients. Further, they say that the majority of their profits are shared directly with members.

[Literature reviewed for this section](#) →



What is a data commons?

#shared #open #resource

In a data commons, data is pooled and shared as a common resource. This approach can address power imbalances by democratizing access to and availability of data. Often, a data commons is accompanied by a high degree of community ownership and leadership and has a public good cause. Of all of the approaches mentioned in this

study, data commons is the one with the most existing real world applications. As a term, it has been a seminal part of open internet discourse since the 90s. However, data commons can be created with very different rules and governance structures, not to mention different kinds of data. So, there is more than just one kind. [Wikipedia](#) and [Wikidata](#) are data commons and so is [OpenStreetMap](#). In science communities, research data is often pooled to increase the impact of data held by individual contributors. Current discussions in literature around data commons revolve around building on the learnings of [open access movements](#) and applying them to new and more ambitious forms of data management. The approach is certainly useful for envisioning alternatives to power inequities that arise from corporate or state control of data in different contexts. A prerequisite is that they be stewarded responsibly, for which many refer to [eight governing principles](#) outlined in a seminal work by Elinor Ostrom from 1990. In practice, this could mean that a commons has conflict resolution mechanisms. Or that a commons ensures that the integrity of data is high or the confidentiality of data subjects is protected. Or that there be clear licensing rules to open or restrict access to knowledge or software code. It probably goes without saying that technical tools and services (and often cloud computing) always form part of a digital data commons. When it comes to devising future data ecosystems that boost competition and enable innovation by diverse players, interoperability of such technical tools and datasets is an important consideration. Being able to apply data in more than one context, or by more than one entity, is part of an expansive vision for unlocking the power of collectively owned data by and for a community.

Raising ambitions for data commons at the city level

[DECODE](#) (Decentralized Citizen Owned Data Ecosystem) is a consortium of 15 European partners, including the cities of Barcelona and Amsterdam, who work to make more communal data available to innovators and civic groups to meet local needs. DECODE launched [four pilot projects](#) between 2017 and 2019 for privacy-preserving technologies (including distributed ledgers) to share data generated and gathered by citizens for communal use. One project was on participatory politics, another on environmental data, a third was on age/identity verification, and a fourth was a login system for hyperlocal networks (as an alternative to bypass Facebook). They see city governments as “custodians of digital rights.”

[Literature reviewed for this section](#) →



What is a data collaborative?

#coordination #information #partnerships

A data collaborative is often what you have when private sector data is combined to help inform public sector decisions. At least this is one aspect. Data in a collaborative could be shared strictly between partners, with an independent third party who manages access to the data, or publicly online. If you feel confused about overlapping approaches at this point, keep in mind that the 7 approaches explored here are simply frequently mentioned in literature. These are not definitive categories from which to understand all data governance. What has long been known as the ‘open data movement’ has often centered on ‘opening’ government data to the public. With data collaboratives, on the other hand, the usual idea is to take data that is proprietary or siloed and make it available to inform research or policy. At the city level, for instance, collaboratives of [mobility data](#) from private sources (say, from rideshare companies) could be used to help inform urban planning. At the global level, [humanitarian data](#) can help United Nations agencies respond to emergencies like the COVID-19 pandemic. Data collaboratives are not entirely uncommon, but it’s hard not to see untapped potential for more. As part of an [extensive research project](#), GovLab has assembled a [database of 200 data collaboratives](#). Among the listed projects is [Global Fishing Watch](#). They combine satellite vessel tracking data from six countries to create an online map that tracks fishing activity for better ocean management. Such big data analysis of ships and ports is also [done by others, for profit](#). But baked into the notion of data collaboratives, as explored here, is that they act as responsible data stewards to empower their members or the general public to solve societal problems. Why would companies want to participate? It could be for “reciprocity, revenue, research, regulatory compliance, reputation, or responsibility,” suggests GovLab in their research. Meanwhile, even proponents agree that a raft of privacy concerns and questions about consent and ‘big data bias’ arise from repurposing data collected in a corporate context. That is why further study and analysis is necessary. Technical platforms for coordination and exchange of data between providers and users are a necessity, as are interoperable data standards and frameworks. [SharedStreets](#) is an example of a non-profit provider of open source software for the “seamless exchange of transport data.” [BrightHive](#) is an example of a for-profit startup that offers consulting on legal frameworks and software for data collaboratives.

Visualizing the effects of deforestation from above

[Global Forest Watch](#) combines datasets about forests with satellite imagery of Earth to visually demonstrate the effects of deforestation from above. It's a partnership of dozens of companies, organizations, and research institutes convened by the World Resources Institute. Since 2014, the platform has helped thousands of people monitor and manage forests and stop illegal deforestation and fires. Since 2019, they have also operated a "pro" service for businesses concerned about deforestation in their supply chain. Their online maps currently show that the equivalent of a football pitch of primary forest is lost every six seconds.

[Literature reviewed for this section](#) →



What is a data trust?

#framework #rights #theory

A data trust is a legal relationship where a trustee stewards data rights in the sole interests of a beneficiary or a group of beneficiaries. When a person or group hands over their data to a trustee, it means the trustee has a fiduciary duty to act according to predefined terms and conditions and never in their own self interest. Data can be pooled from different sources, and a trustee can negotiate access by others on behalf of the collective. As a legal framework, trust law only exists in some parts of the world (including, the UK, US, and Canada) but fiduciary duties often still exist outside of trust law jurisdictions, for instance when a legal representative handles the estate of a deceased person on behalf of a group of specific beneficiaries. The idea of 'a trust' for data is exciting to many scholars, because it could help establish a clear 'fiduciary' responsibility for companies or organizations that handle data. It's an instrument that "uses private law infrastructure without being overly dependent on government action," [writes Sean McDonald](#). Aspirationally, civic data trusts could help shift power to protect people from vulnerabilities resulting from personal data with a clearly stated purpose to benefit a certain community. As a counter example, a company like Facebook protects the interests of its shareholders above those of the people whose data they collect. There have been experiments with data trusts in both the public and private sectors, but this is a truly nascent approach so examples are sparse. Trustlike frameworks have

occasionally been used for [medical research](#) to handle patient data privately. One famously controversial proposal for an '[urban data trust](#)' as part of Alphabet's Sidewalk Labs in Toronto, Canada was ultimately abandoned. From 2017-2020 a number of landmark scholarly texts have helped popularize the term, but definitions are still not uniformly agreed on. Occasionally, they even veer toward more general concepts of data stewardship, so even if something is called a trust it might not actually be one. Meanwhile, significant questions remain about which laws are compatible where (for either 'data' or 'rights to data') and how to govern or hold trusts truly accountable. Conceptually, would there be many trusts that compete with each other? How would a trust exercise collective bargaining power? Could data collectors themselves ever be reliable trustees? These are ongoing theoretical discussions. How do you distinguish a data trust from a different type of trust that also holds data? A growing list of non-profit organizations and government research groups are invested in answering such questions and understanding the potential of trusts, including the [Open Data Institute \(ODI\)](#) with lessons to share from [three pilot projects](#) in 2019.

Staying true to a mission of corporate transparency

[OpenCorporates](#) is the largest open database of companies in the world. They openly publish the financial and legal data of more than 1.8 million companies, selling expanded data access to paying subscribers. A [trust](#) oversees the management of OpenCorporates as a public-benefit business with a mission centered on corporate transparency. It does not neatly fit the definition of a data trust above, but it is a related example, for lack of many others.

[Literature reviewed for this section](#) →



What is a data fiduciary?

#individual #control #intermediaries

A data fiduciary is an intermediary between individuals and data collectors, which takes many experimental forms. These range from 'information fiduciaries' that focus on use of fiduciary law to require a 'duty of care' toward data subjects, to various intermediary technologies or storage solutions that act as a buffer between people's personal data and any commercial entities or online interactions. Doctors and lawyers are examples of professionals that (in many parts of the world) are bound by fiduciary obligations to serve patients or clients. In the digital realm, a data fiduciary could be an entity that protects a client's interests above all and facilitates security and control over personal data. The concept of an 'information fiduciary' was [explored by Jack M. Balkin and Jonathan Zittrain](#) in 2015/16 sparking [continued discourse](#) over the conflicts of interest inherent to a digital surveillance economy and how to generate new pathways for trust. Under the fiduciary umbrella, some scholars would also include advanced tech tools that act as trusted intermediaries (or 'trustmediaries') like [personal AIs](#), [identity layers](#), or [personal data pods](#), and [cloudlets](#). These are all approaches that recognize the significant power asymmetries that occur online between consumers and those who have access and control over personal data. Personal data stores typically focus on storing the data of a single individual and negotiating access to it by others. One of several examples is [Digi.me](#), a "private sharing app" that allows users to upload and store their data, and decide who it will be shared with. [Solid](#) (by Tim Berners Lee) is developing a platform for linked-data applications that aspires to offer full control of personal data via peer-to-peer storage. (Public) data banks are similarly focused on individual data, imagining that consumers will one day manage their data via 'data bank accounts' that just like financial banks, can have different business models (e.g. some may also invest revenues for common good causes). [DigiSahamati Foundation](#) in India advocates for a type of financial data fiduciary called an Account Aggregator, that can earn a license from India's central bank to help individuals collate digital records from different sources in a format that can be easily and securely accessed by lenders, banks, or various automated financial services. It's presented as an alternative to emailing screenshots of electronic records, as well as a potential opportunity for 1 billion unbanked people in India to collate information that gives greater access to financial services. Theoretically, fiduciaries like all of the above could help rebalance the relationship between individuals and data holders, including commercial platforms. Critics are doubtful whether data fiduciary solutions present a realistic path to

structural change, even if they could empower individuals to have more control over access to their personal data and enhance accountability through duties of care. Plenty of data fiduciary initiatives see resolving governance questions between users and platforms primarily as a business opportunity.

An app to tap for greater consumer privacy

[JumboPrivacy](#) is a new Android and iOS app that aims to help people “take control” of their privacy. The app blocks data trackers on the phone, makes it easier to alter Facebook privacy settings, and tells Google and Amazon to delete logged voice data. It also offers to automatically archive older tweets from Twitter and store them offline via the app. JumboPrivacy has a “pay what you think is fair” subscription fee and said they reached [5,000 paying customers](#) in August. In June, the company said it had [raised US\\$8 million](#) in investments, [including](#) from venture capital groups and individuals (among them VPs from Microsoft and Google).

[Literature reviewed for this section](#) →



What is Indigenous data sovereignty?

#collective #determination #ownership

Indigenous data governance is about shifting access and control over data away from governments and other institutions and directly to Indigenous Peoples. This includes information about territories, natural resources, and people, as well as about collectively owned knowledge and intellectual property. Considering how often the withholding of information has been used as a vector of subjugation and control worldwide, it's a data governance approach that illustrates how important data sovereignty can be to self-determination and justice. This is one reason the approach has found its way into general discourse on data governance. Around the world, many Indigenous Peoples and tribes have sought to [gain the right](#) to have unrestricted access to data and information about themselves and their communities, as well as to independently govern the collection, ownership, and use of their own data. 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. Perspectively, Indigenous data governance creates opportunities for more sustainable, consistent, accurate, and relevant data management for the benefit of communities themselves. Some scholars have written on [differences in epistemologies](#) between different Indigenous communities (and differences to mainstream epistemologies) as being a factor in how virtual spaces like websites, social media platforms, or shared data are perceived conceptually. In 2020, the [Global Indigenous Data Alliance](#) has faulted mainstream open data discourse, including the well known “[FAIR data principles](#)” (Findable, Accessible, Interoperable, Reusable) of ignoring historical contexts and power differentials and propose additional safeguards in their “[CARE Principles for Indigenous Data Governance](#)” (Collective Benefit, Authority to Control, Responsibility, Ethics). In Canada, the First Nations Information Governance Centre applies their “OCAP data principles” (Ownership, Control, Access, and Possession) to determine how First Nations data is researched, collected, protected, used, or shared. In the Canadian province of British Columbia, the [First Nations' Data Governance Initiative](#) (BCFNDGI) has created a [Strategic Framework](#) that formalizes a tripartite governance approach to First Nations' data management and establishes a common understanding of the initiative as being led and managed by First Nations. Data sovereignty initiatives are part of transforming relationships with local and national authorities to participate directly in community investment decisions, as well as in ending exploitative and inaccurate research practices that can have detrimental effects. There are learnings in this field that can advantagefully be applied to many decisions in the tech sector. Drawing on in depth conversations with Indigenous people in Aotearoa, Australia, North America, and the Pacific, Jason Edward Lewis has written a [position paper](#) for those who want to design and create AI where 'ethical' is defined as aligning with Indigenous perspectives.

Explaining cultural knowledge rights in a digital format

[Traditional Knowledge \(TK\) Labels](#) is an initiative of [Local Contexts](#), a group that supports Indigenous Peoples in managing their IP and cultural heritage in the digital world. These black and white graphic labels can be applied to “cultural heritage that is digitally circulating,” for instance, to show whether something is sacred or is only intended for seasonal use. The labels are reminiscent of Creative Commons licenses that explain how information can be shared, but are created specifically for Indigenous communities to help educate and inform others about how they wish cultural knowledge to be used or to whom exactly material belongs. The [U.S. Library of](#)

[Congress](#) uses TK labels on digitized Native American audio recordings as part of a collaborative project for historical preservation and access.

[Literature reviewed for this section](#) →



What is a data marketplace?

#commodity #ownership #transferable

A data marketplace is an approach that enables consumers to sell or trade their personal data for services or other benefits. It's common knowledge by now that data is "valuable" but it's also an open question what kind of price tag you can put on data (or access to processing the data), particularly when transactions tend to be business-to-business. Quite a few new platforms are positioning themselves as new intermediaries of the data economy, seeking trust from individuals and cultivating a new willingness to share or transfer personal data. The questions posed through this approach are multiple. Why shouldn't people be paid for data that big tech profits from? How could you develop interoperable systems to transfer data between platforms? Could it lead to more transparency around processing and algorithms? Critiques of this approach can be scathing, suggesting that data marketplaces incentivize data collection that should not happen in the first place. Bypassing data brokers that operate in the shadows may sound appealing as a business model, but critics argue that it promotes a commodification of personal data that is core to what is already wrong with the surveillance economy. In theory, it is imagined that customers of data marketplaces could pool their data for collective bargaining (similar to credit unions where people pool their finances for the collective benefit of members). Members could own and manage such "data unions" and at the same time earn profits from aggregated data. The [Streamr](#) marketplace, for example, allows business and individuals to bundle and sell real-time data with the promise of earning revenue in an ethical way. [The Data Union](#) in the US is another initiative that describes itself as an activist entity to push for personal data ownership as a mechanism to drive change. Again, this approach bears similarities to others and is not necessarily exclusively different.

Enhancing the value and access to transportation data

[oneTRANSPORT Data Marketplace](#) service in the UK is a platform for public and private sector organizations to upload mobility data for discovery and monetization. Data holders manage their own data through the platform where it can be accessed by local government and transport authorities for research and planning. oneTRANSPORT offers access to analytics tools as an incentive for using the platform and encourages making datasets more accessible.

[Literature reviewed for this section](#) →

Looking into data futures

This is an exciting moment to be thinking about and working on emerging questions around alternative data governance and data stewardship. When we embarked on a literature review to better understand common alternative data governance approaches described above, we hoped to be able to identify definitions and categorizations that would help to guide our review of global initiatives. But the newness of this field is evidenced in the fluid and overlapping definitions of approaches that simply aren't easily mapped to real life. We found that there is a gap between theoretical discussions and practice, and that there is still much to explore in thinking about how to bridge this gap and work with others toward exploring alternative data futures.

Research by Mozilla Insights will continue through 2020 with additional publications (see [overview](#)) accompanied by roundtable conversations and interviews with other researchers, supporting organizations, technical service providers, and entrepreneurs. With input from this broader ecosystem, our hope is to identify more ways for ideas to be applied in practice.

To name a few areas of interest to Mozilla, it could be in the realm of community-driven approaches to multilingual voice technology; collective governance models for mobility data to support labor rights; and new techniques to expand consumer rights to data privacy.

The questions we are asking about who needs data stewardship (and where), how to break silos of anglocentric discourse, and how to mitigate [significant risks](#) — these are questions we ask from a perspective of urgency. We don't know whether alternative data governance is the answer, but we know that data collection and ambitions for use of data are growing at a faster pace than systems of accountability and trust that could help protect lives and communities.

Shifting power through data governance is something that deserves a try.

Frequently asked questions (FAQs)

- How can power be shifted through data stewardship? For instance, if a trusted entity uses legal safeguards to uphold the rights and interests of individuals or groups.
- Does 'data governance' have to do with governments? No, it's a term that refers to what rules are applied to data by whomever has data to manage and control.
- What kind of data do you mean? It depends on context. We're interested in personal data, but also agricultural, mobility, and all other kinds. Whatever can unlock more equity.
- How important are community centric approaches for managing data? Very. Data can be leveraged in different contexts for community empowerment and collective needs.
- Is this all about selling personal data? Not to us. We want to empower users with alternatives to the commoditization of data by big tech, data brokers, and others.
- How does a person get more control? For instance, by deciding who can use or share data or by having the technical means to give or revoke access to different entities.
- What difference can a clear purpose make? It can facilitate informed consent around data use or sharing. Plus digital infrastructure can be designed with intention.

Acknowledgements

Mozilla Insights project team

[J. Bob Alotta](#) - VP, Global Programs

[Kasia Odrozek](#) - Project lead and strategy

[Solana Larsen](#) - Editing and strategy

[Stefan Baack](#) - Research

[Eeva Moore](#) - Outreach

With thanks to all contributors

Afef Abrougui, Nabeel Ahmed, Beatriz Botero Arcila, Tetyana Bohdanova, Ana Brandusescu, Tim Davies, Sylvie Delacroix, Alix Dunn, Jonathan van Geuns, Kristina Gorr, Chris Hartgerink, Astha Kapoor, Moses Karanja, Max Kortlander, Dr. Srivatsa Krishna, Danny Lämmerhirt, Raegan MacDonald, Madeleine Maxwell, Claude Migisha, Marilia Monteiro, Edafe Onerhime, Natalie Pang Lee San, Aidan Peppin, Abigail Phillips, Keith Porcaro, Anouk Ruhaak, Stephanie Russo Carroll, Nathan Schneider, Socrates Schouten, Taís de Souza Lessa, Mark Surman, Sander van der Waal, Peter Wells, Amelia Winger-Bearskin, Richard Whitt.