PUBLIC FIRST

Innovate Cities.

Changing the Conversation

Canada's Model for 21st Century Data Governance

Changing the Conversation

Canada's Model for 21st Century Data Governance

Authors: Vinous Ali, Carly Munnelly, Rachel Wolf

Foreword

Innovate Cities was founded on the belief that data has value, and that it has the power to make the world a better place. However, to unlock the true potential data has to offer, trust, privacy and security are key, so too is public awareness and the right governance structures. That is why Innovate Cities has commissioned this independent report that takes in data stewardship models from across the globe to explore how municipalities can maximise the value of their data while maintaining the safeguards and systems necessary to build trust and confidence.

With these as its guiding principles, Innovate Cities is creating a completely unique innovation ecosystem that is unlike anything in its field - one which will serve citizens, innovators, and municipalities, while safeguarding the privacy of each.

At the core of Innovate Cities' vision is CityShield – a not-for-profit data trust that will act as a public utility providing data at cost. CityShield is a data sharing platform that empowers producers of data to share their data in an accessible and useable form, while delegating the responsibilities of strict governance to a trusted third-party fiduciary – Innovate Cities. We believe a not-for-profit data trust is an essential feature of innovation infrastructure because it will allow innovators to access data - the raw material of innovation.

The concept of a data trust is not a new one. Innovate Cities, however, offers a new and compelling evolution of the model, and a solution to the central challenge of putting data to work for universal benefit while ensuring both access to data and the protection of personal privacy.

While other data trusts make decisions about the use of data, such as who is allowed to access it and for what purpose, Innovate Cities relieves data producers of the burden of stewardship, while allowing data producers to:

- own the data;
- decide who may access the data and what can be done with it; and
- decide on the price of the data.

Across economies, there is an urgent need for every kind of data – climate data, healthcare data, infrastructure data, and innovation data, among many others. But there is also an equally important imperative to protect privacy rights at both the individual and aggregate levels. If data is maintained securely, provinces and municipalities can derive value from it to overcome some of today's fiscal challenges. CityShield satisfies both these requirements with its central commitments to privacy, transparency, diversity, equity, and inclusion, while acting as an 'honest broker' of data, rather than an owner. Bound by both fiduciary and privacy duties, and focussed on social outcomes, CityShield offers a clear advantage.

An essential element of Canada's innovation ecosystem, CityShield will help to build and grow Canada's innovation expertise in the future. CityShield not only removes the burdens of data governance and privacy compliance – it gives users control over who can access their data while promoting collaboration between innovators, arriving at optimal solutions for the present day and the future, and designing thoughtful, sustainable cities that meet the needs of citizens and the world they live in. This report examines how to make the promise of big data a reality with a clear roadmap for action.

Table of Contents

Executive Summary

The purpose of this report

Policy makers are immersed in reports about the opportunities provided by big data. Municipalities are often told, in the abstract, about the value that data can bring - in terms of better services to citizens, reduced costs, and revenue.

But all too often, the questions - about security, about who owns the data and who controls it, and about the actual value of the data that exists - prevent action. There have been smart city pioneers, such as Singapore; Barcelona; and New York. There have also been jurisdictions, including Singapore with its Trusted Data Sharing Framework; and the EU's General Data Protection Regulation (GDPR), that have created clear regulatory frameworks.

However there are many more cities, including (but not only) in Canada, that are not global behemoths and who lack the resources, and the federal and provincial framework, to proceed.

This paper uses case studies from across the world to investigate how some of the principles behind data stewardship can be used to help municipalities maximize the value of their data, while maintaining control and safeguarding their citizens. Learning from these case studies, the report makes recommendations to federal, provincial, and municipal governments so that they - and the people living within them - neither miss out on the opportunities data can bring, nor face unnecessary exposure to the risks.

The benefits of data

The private sector has embraced big data. It has created new business models, new services, and offered enormous productivity gains to sectors from tech, to agriculture, to pharmaceuticals.

The public sectors' experience has been patchier. Some cities, such as Barcelona, have embraced the opportunities of data and become truly smart cities. Some countries, such as Singapore and Estonia, now run the majority of their services digitally and with the use of data.

But the experience of Innovate Cities, which commissioned this report, is that many cities, provinces, and countries remain uncertain about the practical use of their data, and its value.

Fundamentally, public sector data can be used to:

- **Provide a revenue source or substantially reduce costs for municipalities.** The former comes from thoughtfully leasing aggregated, anonymized data to private sector companies in, for example, transportation. The latter comes from identifying large scale inefficiencies or fraud, or from removing unnecessary friction for citizens;
- **Improve public services.** The intelligent use of data has been used to improve transport networks and stop people waiting endlessly for their bus or train; identify dangerous buildings in violation of codes; and predict and plan for bed occupancy in hospitals.
- **Create a cycle of research and innovation.** For example, during Covid-19, the King's College London Zoe app used millions of submissions from volunteers to identify new symptoms of Covid, evaluate effective interventions and drive better responses from the government.

This matters to citizens. We all want better functioning public services, we want our taxes to be used to maximum effect, and we want more funding to be available for health and other services.

The need for an honest broker

Given the perplexing array of benefits promised by data, why do we not live in a completely smart world? Clearly, there are questions of capacity and application. But more than that, public servants and citizens have been understandably concerned about how data might be used, how secure it is, and the extent to which people and their elected representatives maintain control.

In Canada, Sidewalk Labs has been a cautionary tale. Its partnership with Waterfront Toronto, in 2017, planned to install sensors to record real-time data. This is an innovation that has been used, in different forms, in multiple cities around the world. However, a lack of clarity around the governance and stewardship of the data, and potential risks to individuals' privacy and security, meant the project fell apart.

Sidewalk underlines the need for clarity, transparency, and the public interest. There needs to be a data stewardship model that can realize the benefits of data for all parties, while maintaining public consent.

Data stewardship models

Data stewardship models are an attempt to resolve these questions. Many are still in early stages, and different (and often conflicting) definitions exist. Their structure comes down to four questions:

1. Who owns the data;
2. Why the data is being stewarded;
3. The responsibility, legal or contractual, of the data steward;
4. How decisions about sharing the data are made.

In this paper we looked at four common models for data stewardship: Contractual/Corporate models; Data Trusts; Data cooperatives; and Data commons. We summarize the difference briefly below. There were challenges even with this categorisation - for example some define data trusts around a very specific legal structure, while others define it in terms of its responsibilities for data security.

	Ownership	Stewardship	Responsibility	Decision-Making
Contractual/ Corporate models	Data is owned by the data providers, with the data steward tasked with sharing and stewarding the data based on the terms of a pre-agreed contract. In some instances, the data may be held and managed by the data steward depending on the terms of the contract.	Stewardship is designed to maximize revenue and benefits for the data providers.	The data steward is bound by the terms set out in their contract with the data provider.	The data steward makes decisions on how to share the data in line with the terms and conditions set out in their contract with the data provider.
Data trusts	Data is owned and held by the individual data provider, the trust acts as a broker who facilitates the sharing of the data with potential users based on a contractual agreement.	Stewardship is designed to give benefits to the data provider by sharing data securely, and give advantages to members who pay for access to the data.	The data trust is bound by the terms set out in their contract with the data provider as well as by a legal fiduciary duty to act in the beneficiaries best interest.	The trust has the fiduciary duty to make decisions about the data in the best interest of the data generators. Third parties who buy the data are bound by the rules of use set by the trust.
Data cooperatives	Data is owned by the cooperative.	The main stewardship principle is that data should be managed democratically by the individuals who generate it.	Data cooperatives are bound by the terms laid out in their contractual underpinnings and steward data accordingly.	Decision-making is delegated to the cooperative, who make decisions to share the data in the best interest of the members.
Data commons	Data in a data common is seen as a common resource, and therefore ownership remains undefined.	The main stewardship principle is common access to the data	Data commons lack any legal or contractual responsibility for stewarding the data, however in some instances access may be restricted to prevent unintended harm (e.g. poachers having access to endangered animals' movement data) through accreditation mechanisms.	Once an individual adds their data to the common they cannot control who accesses it or how it is used.

Given the vast amount of data collected by municipalities, data trusts and contractual models provide the most effective form of governance - data commons and cooperatives are primarily citizen-driven models that serve a specific group or segment of the population.

Based on our research, we think there is a good argument for a new category that combines some of the best stewardship principles and obligations from across these models. This model would require the data steward to:

1. Be a non-profit;
2. Have social outcomes in its object;
3. Facilitates the safe and controlled use of data;
4. Be bound by fiduciary duties.

To enable this new category to emerge and establish itself the right environment must be created. This report explores how this can be achieved.

What regulation and legislation does a data trust need to exist and succeed?

One of the reasons there is ambiguity surrounding the definition and models of data stewards concerns the lack of clear national legal frameworks. There are more than 2,500 data privacy laws globally, and 88% of global companies report spending more than $1 million annually on GDPR compliance.[1] This means that, for smaller municipalities, the effort:reward ratio is poor.

We have identified four clear policy and regulatory reforms to help enable effective data stewardship through a data trust model:

1. Governments must introduce a clear and succinct definition of a data trust and any related concepts. Currently, there are varying definitions of data trusts in the academic literature, which makes it impossible for the concept to be used in any regulations or legislation;
2. Governments must introduce legislation or guidance on the re-use of data without additional consent. Canadian legislation on the definition of 'socially beneficial purposes, where data can be reused without consent' is insufficiently clear;
3. The limits in regulatory provisions concerning data portability, access, and erasure must be better defined. Data trusts must be able to exercise data rights on behalf of their beneficiaries for a trust to run effectively (or at all);
4. Government must clarify whether data trusts have fiduciary duties akin to property trusts. If so, property definitions must include data.

In Canada, enormous variation across municipalities and provinces make it more challenging for trusts to operate and for municipalities to get clear and consistent advice.

[1] https://www.itgovernance.eu/blog/en/how-much-does-gdpr-compliance-cost-in-2020

Data stewardship in practice

The concept of data stewardship is novel, responding to the explosion in data collection in recent times. Case studies, particularly those with long time spans, are therefore few and far between. However, those that do exist help us understand where there are regulatory issues, and identify early successes and lessons.

In the full report we look at seven case studies of data stewards and in each case looked at its:

- Purpose;
- Structure and governance;
- Impact

Our case studies were:

1. The Biobank in the UK, which uses rich longitudinal data to provide support to researchers across the world;
2. Place in the US, which makes high quality mapping data available to underserved regions;
3. Agri-Gaia in Germany, which looks at AI applications in agriculture;
4. Salus Coop in Spain, which is a citizen-driven organization collecting the health data of participants;
5. Driver's Seat in the US, which enables gig economy workers to collect and share their own generated data;
6. Vivli in the the US, which is a data broker for the sharing of scientific clinical trial data; and
7. Brighthive in the US, which acts as a data platform for data sharing between private organizations, governments and academic institutions.

What is clear from these is that: 1. Their usage is pre-defined (by broad sector); 2. Revenue generation is not a criterion or aim; 3. Either individuals or the trust itself make the decision about its usage.

Separately, we found:

1. Numerous examples of municipalities and the public sector using their own data to unlock services and generate revenue. For example, Departments of Motor Vehicles (DMVs) across the United States routinely sell data to third parties (which, because of a lack of governance, has caused major controversy); [2]
2. Examples of private companies using open public data - for example in Ghana, a private company Esoko uses open weather and market price data to help small-scale Ghanain farmers improve their output. [3]
3. Examples of private companies providing AI services, at a fee, to the public sector to identify waste and reduce cost. For example, Aviana global has worked with the California Franchise Tax Board to reduce costs, and Faculty AI has worked with numerous UK public bodies to do the same. [4]

What we did not find was a successful marriage of the three - the use of non-profit data stewardship structures to remove some of the bureaucracy, governance, and security from municipalities' hands; the provision of revenue or cost saving to municipalities; and the potential for the private sector to use the data under carefully designed, secure conditions.

[2] https://www.caranddriver.com/features/a32035408/dmv-selling-driver-data/

[3] https://esoko.com/who-we-are/

[4] https://avianaglobal.com/portfolio/the-ai-effect-increased-revenue-and-reduced-debt/;
https://faculty.ai/ourwork/defining-ai-implementation-across-uk-government/

This is a reminder that innovation will continue to occur in this area, and that legislation and regulation needs to be simultaneously flexible enough to allow new models, while providing clarity to actors. From our review, the two consistent themes were:

1. **Despite their differences, the data trusts we looked at all had social gain at their core.** Although the purposes and sectors of the data trusts we looked at differ wildly, they shared a common aim to improve the welfare and livelihoods of the individuals and communities they serve.
2. **In the short-term, data trusts are most likely to flourish in areas where data privacy legislation is well-developed.** All of the data trusts we have looked at are all concentrated in Europe and the United States, which may reflect the existence of more developed data rights landscapes - a fundamental prerequisite to data trust activities. This finding has been reiterated by a survey of data trusts by GPAI, Aapti and ODI, in which 37 of the 45 respondents were based in Europe and North America. [5]

Roadmap for municipalities and provincial governments

There is clearly an enormous opportunity if we can get the legislation and regulation for data stewardship right. We need smaller municipalities to have the same confidence and capacity as New York, London, and Barcelona. Their citizens deserve effective services and lower taxes just as much. As this report makes clear, there is too much of a lag between the definition and structures for data stewardship and the potential for big data.

In our first section, we laid out what we think is needed from the federal government. Provincial and municipal government must also play their part. Clearly, local decision making must be respected, but we have identified three universal principles:

1. Lay the ground rules.

As this report has demonstrated there are multiple approaches to enabling the creation and use of data trusts. Before embarking on this journey Governments must answer the following questions:

- How do you want to define data trusts? Should it reflect the same legal structure as a property trust? How are fiduciary duties defined with regards to data stewardship - is this a legal obligation or simply a principle to guide behaviors? Can your definition allow innovative models?
- What regulations are currently missing? How can data privacy and data stewardship laws and regulations be improved? Is there somewhere to look to for best practice (e.g. EU GDPR, California CCPA, etc.)?
- Are there jurisdictional issues, particularly with data trusts that operate across provinces? How can these be solved through cooperation or through standardization of regulation?

2. Build public trust.

We have seen attempts to open up the use of data through the use of intermediaries or third-parties fail due to a lack of public trust and transparency. This is one of the great advantages of data stewardship - it guarantees security, and control by elected representatives. Beyond this governments should consider:

- Publication of explicit, transparent, rules of engagement.
- Citizen juries or other deliberative engagements with the public. [6]
- Consultations.

[5] https://docs.google.com/document/d/18HPZbsd9DLQp5fk7iSzS6fs-ptGiWSJrm34UdR_3aMg/edit

[6] https://www.involve.org.uk/resources/blog/project-update/delving-data-trust-decision-making

3. Invest in pilot schemes.
In our experience, use of data is often stymied because of abstracted conversations. Well-designed pilots allow you to experience and learn from experience, and gain public trust by real-life example. Pilots should have:

- Clear objectives and metrics assigned to measure their success.
- Take an interdisciplinary approach to design that brings together experts-in-the-field, policymakers and relevant delivery partners and stakeholders.
- Be committed to full transparency of findings to enable others to benefit from any learnings and to earn and grow public trust and confidence.

Fundamentally, though, municipalities should feel reassured that models do exist that can provide them the security and control they require.

Part 1: What are we trying to achieve? The purpose of a data trust.

The benefits of data

We are living in an increasingly digitized world, with vast amounts of data created, stored, and shared on a daily basis. This trajectory will continue as connected devices become more commonplace and as a society we shift to "smart cities", following the example of Barcelona,[7] where data on city life is used to provide better services and amenities and improve citizen welfare. This vast amount of data holds enormous economic and social value for governments, researchers, and private industry across the world.

For organizations, access to large scale datasets can provide the basis for innovation, allowing companies to boost productivity, reduce costs and ultimately offer a better service to their customers and clients.[8] For example, Airbus and several partners developed a data sharing platform, where aircraft manufacturers and supply chain partners can access, use and share real-time data to design aircraft parts. As a result, design processes that took several weeks could now be completed in hours, leading to cost savings and efficiency gains for aircraft manufacturers, the airlines that buy them, and ultimately the end user.[9]

Access to data is also beneficial from a research perspective, as improved access to data can help companies uncover insights that would otherwise be missed. For example, reducing barriers to access to NASA's satellite imagery data was found to have had a significant positive impact on the quantity and quality of related research over the following decade.[10]

For municipal, provincial and federal Governments, opening access to data can accelerate economic development by providing a revenue source or substantially reducing costs, leading to more efficient services, and creating a cycle of research and innovation.[11]

Case studies: the economic development benefits of unlocking Government data

Making Government data available can:

1. Provide a revenue source or substantially reduce costs for municipalities.

According to an investigation by VICE, Departments of Motor Vehicles (DMVs) across the United States routinely sell individual's personal data to third parties, including tow companies, insurance companies and even - controversially - private investigators. According to public records, several states have made tens of millions of dollars per year selling this data.[12] Safeguards to ensure this data is handled and shared appropriately would help to ensure governments can continue to raise this revenue whilst maintaining public trust and confidence.

[7] https://dl.acm.org/doi/fullHtml/10.1145/3117800
[8] https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf
[9] https://theodi.org/article/case-study-the-value-of-sharing-data-in-supply-chain-optimisation/
[10] https://www.pnas.org/content/117/38/23490
[11] https://theodi.org/article/using-open-data-for-public-services-report-2/
[12] https://www.vice.com/en/article/43kxzq/dmvs-selling-data-private-investigators-making-millions-of-dollars

The North Texas Innovation Alliance (NTXIA), a regional consortium of municipalities, agencies, companies and academic institutions, is currently considering proposals for monetising data. The NTXIA is still exploring options for how they might monetize data in practice, but in general they see data monetisation as one solution to what they see as ever increasing demand from residents for better public infrastructure while long-term projections for revenues trend downward. Many cities in Texas have struggled to raise sufficient revenue to support public services due to state law tax caps, a problem that has only been exacerbated by the pandemic.[13] One model being looked into in Texas is a "freemium system", where some data is made freely available while others are made available on a subscription basis.[14]

2. Improve public services.

In 2002, the Mayor of New York created the Mayor's Office of Data Analytics (MODA), tasked with aggregating and analyzing data from across different city agencies. One of MODA's ambitions was to use data from other city agencies to improve firefighters' ability to identify dangerous buildings during inspections. MODA was able to aggregate and analyze infrastructure data to identify buildings most at risk of having safety violations and therefore improve the efficiency of FDNY's inspection process. The first 25% of inspections had previously resulted in the discovery of 21% of the most severe violations, whereas after this analysis was done, the first 25% of inspections exposed more than 70% of buildings with the most severe violations.[15]

In Australia, the Government's research agency, the Commonwealth Scientific and Industrial Research Organization, aggregated and analyzed hospital data from across the country to generate a tool that used historical data to project expected patient admission and discharge, as well as expected medical urgency and speciality. The tool has been trialed in 30 hospitals with a 90% accuracy rate in forecasting bed demand. Estimates suggest that AUD 23 million in annual savings could be generated if the tool were rolled out across Australia.[16]

3. Create a cycle of research and innovation.

In the UK, Transport for London (TfL) began publishing increasing amounts of real-time transit data between 2007-2011, data which is now used in over 600 apps. Providing access to transport data has improved journeys and saved people time - equivalent to between £70m and £90m in savings for passengers per year. This data has also supported innovation and created jobs, with over 700 jobs at companies dependent on TfL's data availability created since TfL began opening its data.[17]

In Spain, the National Observatory of Telecommunications and the Information Society (ONTSI) produce a yearly estimate of the size of the "infomediary sector", which broadly consists of businesses that exist because of data. In their 2020 report, they found the sector to be made up of over 700 businesses valued roughly at a combined €2.5 billion, employing nearly 22,000 employees. 75% of these businesses rely on national data sources from public entities alongside private data, therefore underscoring the importance of access to open government data.[18]

[13] https://cities-today.com/cities-want-more-from-their-data-including-money/
[14] https://gcn.com/data-analytics/2021/11/can-cities-monetize-their-data/316288/
[15] http://www.spatialcomplexity.info/files/2015/06/Big-Data-in-the-Big-Apple.pdf
[16] https://www.oecd-ilibrary.org/sites/1ab27217-en/index.html?itemId=/content/component/1ab27217-en
[17] https://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf
[18] https://www.ontsi.es/sites/ontsi/files/2020-06/CharacterizationInfomedarySector2020.pdf ;
https://data.europa.eu/en/news/asedie-publishes-annual-infomediary-sector-report

In Ghana, open government data has been used by a private company - Esoko - to provide weather and market price data to small-scale Ghanian farmers to improve their output. Esoko sends SMS and voice message alerts to farmers' cell phones with updates on market prices, weather forecasts, crop price bids, and crop production protocols. Esoko also offers market evaluation and monitoring products to businesses. Esoko has reached 350,000 farmers in 10 countries and a research study has found that farmers that use Esoko received 10% more for maise and 7% more for groundnuts than farmers that did not use Esoko.[19]

In the US, opening weather data through the United States National Oceanic and Atmospheric Administration (NOAA) has led to cost savings on weather-related damage and underpinned the growth of a multi million-dollar industry of tools and apps. Impact estimates of NOAA's data include:[20]

- NOAA real-time data is critical to a private weather service industry worth over $700 million in value added annually
- Electricity generators save an estimated $166 million annually due to 24-hour temperature forecasts reliant on NOAA real-time data
- NOAA data contributes to improved forecasting, warnings and associated emergency responses resulting in $3 billion in savings in a typical hurricane season

The need for an honest broker

However, with the rise in the collection, storing and sharing of big data, concerns have been increasingly raised about the opportunity for this data to be misused and for sensitive data to enter the wrong hands.[21] The potential for misuse is particularly concerning given the speed at which this enormous amount of data has been generated, the lack of comprehensive data governance legislation around the world, and the fact that this data has been largely concentrated in the hands of a few private companies.[22]

Further, there are limitations with the current popular model of data governance which relies on consent to waive legal protection over an individual's data.[23] A Deloitte survey found that 91% of consumers accept terms and conditions without reading them when installing apps or signing on to online services.[24] This is because these terms and conditions are often long[25] and complex,[26] and the lack of alternative services makes 'opting-out' an impractical choice.[27]

As a result, there is a growing need for an honest data broker to transparently hold, transmit, and sell data while maintaining the trust of citizens and governments. That is where data stewardship comes in. Data stewardship is a governance approach that formalizes accountability for managing data on behalf of an individual or organization.[28] According to the Open Data Institute (ODI) (2019) "a steward of data … can decide who has access, under what conditions and to whose benefit."[29]

[19] https://odimpact.org/case-ghanas-esoko.html

[20] https://odimpact.org/case-united-states-noaa-opening-up-global-weather-data-in-collaboration-with-businesses.html

[21] https://static1.squarespace.com/static/5e3b09f0b754a35dcb4111ce/t/603ce3325e1da817afe6b193/1614603061204/WP+2+–+DTI+–+global+perspectives.pdf

[22] https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

[23] https://il.boell.org/en/2021/11/18/data-stewardship

[24] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-2017-global-mobile-consumer-survey-executive-summary.pdf

[25] https://www.thinkmoney.co.uk/blog/what-phones-know-about-you/

[26] https://www.visualcapitalist.com/terms-of-service-visualizing-the-length-of-internet-agreements/

[27] https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/ and https://www.thebritishacademy.ac.uk/publications/data-ownership-rights-controls-seminar-report/

[28] https://repositorium.sdum.uminho.pt/bitstream/1822/69192/1/JBEBJ20.pdf

[29] Hardinges, J, Wells, P, Blandford, A, Tennison, J, Scott, A (2019) 'Data trusts: lessons from three pilots' Open Data Institute.

For municipalities and other government bodies, using data stewardship provides a safe and secure way to maximize the potential of data they already hold. Right across the globe municipalities are time-poor with their resources stretched across multiple priority areas, the use of data stewards can help relieve some of the pressure by managing the data and helping unlock its insights.

The development of different data stewardship models is still very much in its infancy, and the Sidewalk Labs project is a good example of how data stewardship can go wrong and very quickly lose the public's trust.

Case study: Sidewalk Labs and the development of data trusts in Ontario

Sidewalk Labs entered into a partnership with Waterfront Toronto in 2017 to construct a "smart" redevelopment of a portion of Toronto's waterfront. The project planned to install sensors throughout the area, which would be used to record real-time usage data that could be analyzed to "improve urban life." This would include, for example, pedestrian, bicycle and car traffic patterns, public amenities usage data, and even tenancy lengths for nearby apartment complexes. All of this data would be held in what was called an "Urban Data Trust", which would be responsible for governing the collection and sharing of the data.

The project was met with intense criticism from the public and media, with citizens particularly concerned about issues of privacy, surveillance, and data sovereignty. Sidewalk Labs' proposed data trust ultimately failed to solve the emerging governance concerns, including the lack of clarity around the limitations for data sharing and access, the inability for other companies to participate in data collection, and the risk to individual's privacy and security.[30] Following several months of public outcry, the project was eventually canceled in 2020 (though the official reasons for canceling were said to be related to the pandemic and economic downturn).

Since then, the Government of Ontario has launched a public consultation on proposed reform to privacy protection laws. Many of the proposed reforms revolved around improving models of consent, portability and erasure, as well as enhancing enforcement powers and introducing a legislative framework for data trusts.[31] In 2021, following this consultation, the Government of Ontario set out it's proposed data privacy reforms in a white paper entitled Modernizing Privacy in Ontario.[32] The white paper includes many proposals for strengthening data privacy laws, including reforming consent, introducing a right to erasure, right to data portability, expanded application outside of commercial organizations, a rights-based approach to privacy, and stronger oversight and enforcement mechanisms.[33]

It is clear from the experience of Sidewalk Labs that in order to reap the benefits of big data – e.g. revenue, innovation, improved public services, etc. – we need a data stewardship model that all parties trust to transparently hold and share their data. The Government of Ontario has clearly recognized this and has begun the process of implementing better defined data sharing frameworks. Other governments can learn from Ontario's experience and implement these policies proactively rather than reactively.

[30] Scassa, T. (2020) Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto. Technology and Regulation.
[31] https://news.ontario.ca/en/release/57985/ontario-launches-consultations-to-strengthen-privacy-protections-of-personal-data
[32] https://www.ontariocanada.com/registry/showAttachment.do?postingId=37468&attachmentId=49462
[33] https://www.lexology.com/library/detail.aspx?g=b5927f23-3fa5-461e-9375-c64194c2f5f3

Data stewardship models

The development of data stewardship models is still in early stages, and as such, many different (and sometimes conflicting) definitions exist. Many researchers and organizations are contributing to the literature on this topic, including GPAI, the ODI, and the Ada Lovelace Institute, whose work is helping to develop distinct definitions for the different emerging stewardship models. As the body of literature grows over the coming years, we expect definitions to change and solidify, with a lead consensus emerging.

Broadly, the data stewardship models differ based on their approaches to:

1. Who owns the data;
2. Why the data is being stewarded;
3. The responsibility, legal or contractual, of the data steward;
4. How decisions about sharing the data are made.

The four main emerging stewardship models are contractual/corporate models, data trusts, data cooperatives and data commons. Each of these models is described in more detail below.

Contractual/Corporate Model

In a corporate/contractual model, data generators enter into a contract with a data-sharing venture who manage how and with whom the data is shared. This data-sharing venture may take several different corporate forms - it can be either a profit-making enterprise where profits are applied to the business or divided amongst shareholders, or a non-profit organization where any profits are, for example, applied to a not-for-profit cause related to the company's purpose. Regardless of the corporate form, the entity would operate as a data-platform owner and manager, entering contractual agreements with both the providers of data and the potential users.[34]

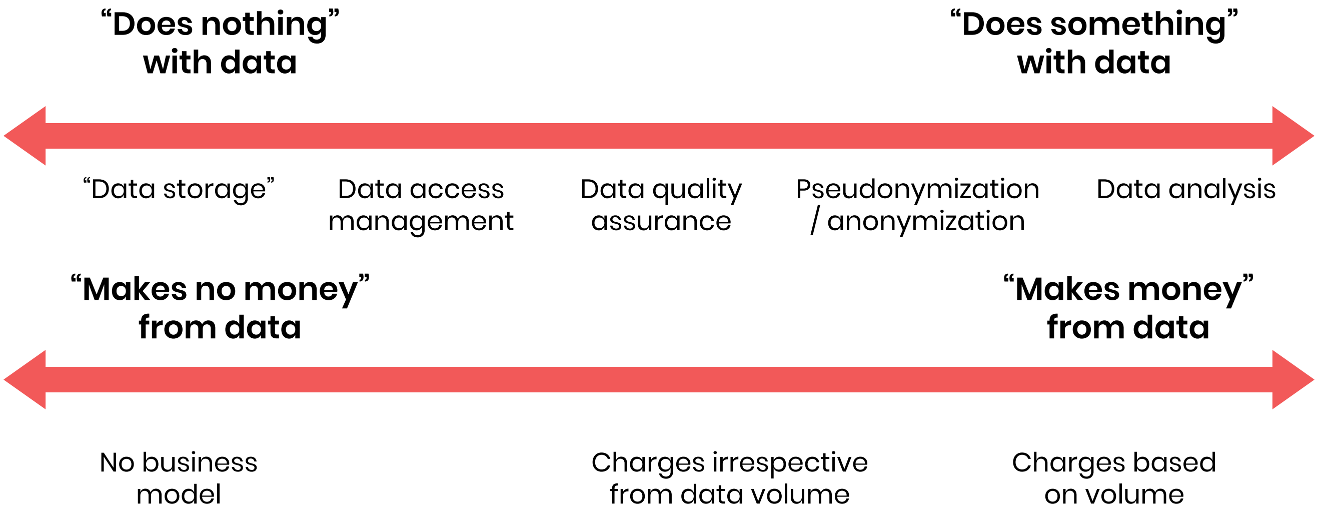
Data Trusts

Data trusts are a relatively new concept and therefore different definitions have emerged, with some defining it around a legal structure[35] and others defining it more loosely as a framework to delegate data sharing responsibility to a third party to ensure data exchanges are secure and mutually beneficial.[36] One of the most widely accepted definitions of a data trust, and the one we will use for the purposes of this report, comes from the ODI: "[a data trust is] a legal structure that provides independent stewardship of data."[37]

Trustees of a data trust take responsibility to steward data for an agreed purpose. Trustees are bound by fiduciary duties, which legally require them to act in the best interest of the data generator. The individual data generators hold and own their own data, but grant the trustees the power to make decisions about who has access to the data and for what purposes. This model therefore transfers liability for ensuring the data is shared appropriately to the trustees.

As illustrated in the figure below, the level of involvement that data trusts have in storing, analyzing or anonymising the data, as well as its charging methods, can vary from trust to trust.[38] The scope of activities that data trusts carry out will likely be refined as data stewardship models become better defined.

[34] https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/
[35] Hardinges, J, Wells, P, Blandford, A, Tennison, J, Scott, A (2019) 'Data trusts: lessons from three pilots' Open Data Institute.
[36] https://www.mmu.ac.uk/media/mmuacuk/content/documents/business-school/future-economies/Mills-2020.pdf
[37] Hardinges, J, Wells, P, Blandford, A, Tennison, J, Scott, A (2019) 'Data trusts: lessons from three pilots' Open Data Institute.
[38] https://www.stiftung-nv.de/sites/default/files/regulation_for_data_trusts_0.pdf



Data Cooperatives

Data cooperatives are mutual organizations which are owned and democratically controlled by members. Decisions around the stewardship of the data is a joint responsibility of the individuals involved and data is shared in a way that serves the best interest of the entire cooperative. There are differences in how cooperatives are governed, but most commonly, the decision-making power is delegated to the cooperative who acts in the best interest of the members. Large groups have more negotiating power over the terms and conditions of sharing their data, and therefore data cooperatives can be a useful tool for empowering individuals.[39]

Data Commons

In the data commons model, data is pooled and governed as a common resource. This model is much more informal than the other data stewardship models. Typically, decisions around data governance are developed democratically and the institutional framework is less rigid. Data commons tend to see access to data as a public good, and the primary aim is to increase data access, experimentation and interaction rather than to make money.[40] Wikipedia is an example of a data common, where individuals contribute information and data to the platform for the sake of improving access to that information and data, rather than any monetary gains.

Data commons, unlike other models of data stewardship, offers relatively unrestricted access to data and is built on the principle that all data in the common should be accessible by all stakeholders. This raises concerns around accountability and privacy - if ownership is completely decentralized then who is responsible for ensuring it is accurate, up-to-date, and held securely? Some researchers have suggested introducing a centralized body responsible for maintaining the data, which could alleviate some of the accountability and privacy concerns.[41]

[39] https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/
[40] https://www.mmu.ac.uk/media/mmuacuk/content/documents/business-school/future-economies/Mills-2020.pdf
[41] Lawrence and Laybourn-Langton (2018) 'The Digital Commonwealth: From private enclosure to collective benefit' IPPR Commission on Economic Justice.

	Ownership	Stewardship	Responsibility	Decision-Making
Contractual/ Corporate	Data is owned by the data providers, with the data steward tasked with sharing and stewarding the data based on the terms of a pre-agreed contract. In some instances, the data may be held and managed by the data steward depending on the terms of the contract.	Stewardship is designed to maximize revenue and benefits for the data providers.	The data steward is bound by the terms set out in their contract with the data provider.	The data steward makes decisions on how to share the data in line with the terms and conditions set out in their contract with the data provider.
Data trusts	Data is owned and held by the individual data provider, the trust acts as a broker who facilitates the sharing of the data with potential users based on a contractual agreement.	Stewardship is designed to give benefits to both the data provider by sharing data securely, and to members who pay for access to the data.	The data trust is bound by the terms set out in their contract with the data provider as well as by a legal fiduciary duty to act in the beneficiaries' best interest.	The trust has the fiduciary duty to make decisions about the data in the best interest of the data generators. Third parties who buy the data are bound by the rules of use set by the trust.
Data cooperatives	Data is owned by the cooperative.	The main stewardship principle is that data should be managed democratically by the individuals who generate it.	Data cooperatives are bound by the terms laid out in their contractual underpinnings and steward data accordingly.	Decision-making is delegated to the cooperative, who make decisions to share the data in the best interest of the members.
Data commons	Data in a data common is seen as a common resource, and therefore ownership remains undefined.	The main stewardship principle is common access to the data in the common.	Data commons lack any legal or contractual responsibility for stewarding the data, however in some instances access may be restricted to prevent unintended harm (e.g. poachers having access to endangered animals' movement data) through accreditation mechanisms.	Once an individual adds their data to the common they cannot control who accesses it or how it is used.

Of the four models, data commons and cooperatives are most relevant for individuals seeking to leverage power by pooling their data, whereas data trusts and contractual/corporate models are most relevant for organizations (public or private) who hold large amounts of data and want to monetize that data or open it for the purposes of research and innovation, whilst ensuring data privacy and sharing regulations are upheld. The legal structures of a data trust add further safeguards and security around the data stewardship process and, therefore, are the most attractive option for governments seeking to safely reep the benefits of the data they hold.

Based on our research, we believe there should be a fifth stewardship model, which is a subset of 'data trusts' and has the following characteristics:

1. Be a non-profit;
2. Have social outcomes in its object;
3. Facilitates the safe and controlled use of data;
4. Be bound by fiduciary duties.

An appropriate regulatory environment must be introduced to enable this new category of data trusts to emerge and grow. The rest of this report explores how this can be achieved.

Example of a new stewardship model: Innovate Cities

Innovate Cities' Data Trust, CityShield, is a data-sharing platform designed to "enable innovation among data collectors, with interactions monitored through accountability, strict governance and 'Privacy-by-Design' principles."[42] Innovate Cities defines a data trust as "a technological platform that allows participants to share their data with one another, while delegating the responsibilities and obligations of management to a trusted third-party entity."[43]

This data trust is built on the idea that improving access to data will drive innovation that creates better and safer cities. Privacy is at the center of the trust, and Innovate Cities provides the oversight and accountability to ensure data is used solely for its intended purpose.

[42] https://innovatecities.com/

[43] https://innovatecities.com/wp-content/uploads/2021/05/CityShield-Innovate-Cities-Data-Trust-FINAL.pdf

What regulation and legislation does a data trust need to exist and succeed?

Data trusts are a relatively new concept and therefore there are very few legal frameworks or related regulations that have existed long enough to be evaluated. At the time this report is being written, the Data Economy Lab is undertaking the most comprehensive review to date of existing data trust policy, regulation and legislation across the world.[44] The publication of this report will add much needed clarification to the role of governments in accelerating the development of data trusts across the world.

Regardless, preliminary work on this topic has identified a number of broad policy and regulatory advancements that could simplify and improve the landscape for data trusts in the short-term, which we explore in more detail below. For global organizations, complying with data privacy laws can be costly and time consuming. There are more than 2,500 laws governing data privacy globally and 88% of global companies report spending more than $1 million annually on GDPR compliance costs alone.[45] Given the level of complexity and uncertainty in existing data privacy legislation, it is important that any new regulations work to reduce the level of complexity in this policy space rather than add to it.[46]

We have identified four main policy and regulatory reforms which would help to enable effective data stewardship through a data trust model:

First, governments must introduce a clear and succinct definition of a data trust and any related concepts. Currently, there are varying definitions of data trusts in the academic literature, which makes it impossible for the concept to be used in any regulations or legislation.[47] Ideally, this definition would be set at the national level to ensure consistency across different municipalities and provinces.

Second, governments must introduce legislation or guidance on the re-use of data without additional consent. Unless it was originally disclosed that an individual's data would be shared via a data trust at the point of collection and consent, then sharing data requires new legal justification. One way (partially) around this is through the de-identification of data - in Canada, the Digital Charter Implementation Act (2020) allows de-identified data to be shared with specified public institutions without consent of the individual data subject so long as it is used for "socially beneficial purposes."[48] However, issues remain around what activities can be included under the umbrella of having "socially beneficial purposes".

Therefore, it would be helpful for governments to introduce further guidance on how data can be re-used and shared, as well as how trusts should develop data sharing rules which balance the benefits from sharing data against the interests of the individual data subjects who may not have consented for their data to be shared in this way.

Third, the limits in regulatory provisions concerning data portability, access and erasure must be better defined. Trustees in data trusts must be able to exercise data rights - such as portability, access and erasure - on behalf of their beneficiaries for a data trust to run successfully. This has given rise to an emerging debate about whether regulatory intervention is therefore needed to make data rights mandatable to a trustee.[49]

[44] https://thedataeconomylab.com/2021/08/10/enabling-data-sharing-for-social-benefit-through-data-trusts-a-top-down-analysis-of-legal-frameworks-for-data-trusts/
[45] https://www.pwc.com/ca/en/services/consulting/privacy/privacy-canadian-business-hub/creating-data-trust.html
[46] https://www.stiftung-nv.de/sites/default/files/regulation_for_data_trusts_0.pdf
[47] https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf
[48] https://www.cigionline.org/articles/governance-innovation-and-privacy-promise-data-trusts-and-regulatory-sandboxes/
[49] https://datatrusts.uk/blogs/international-policy-developments

Fourth, governments must clarify whether data trusts have fiduciary duties akin to property trusts. If so, property definitions must include data. Including fiduciary duties as a legal obligation in the data trust model would increase accountability, ensuring that the socially-conscious governance system is more than just talk, but rather, is embedded in its legal structure.[50] However, currently data is not included under the definition of property in trust law, therefore presenting a challenge to applying trust law to data trusts.

Many governments, across the federal, provincial and municipal levels, have begun developing the policy framework necessary for data trusts to exist and thrive - particularly by clarifying and improving data privacy legislation. We take a closer look at recent developments across the UK, EU, Canada, Quebec, Singapore and California below.

The EU has the strongest data privacy legislation of the jurisdictions we looked at, and has been used as the 'gold standard' when developing data privacy legislation worldwide. In Canada and the US there is massive variation in data privacy standards across states/provinces - Quebec and California, for example, have some of the world's most comprehensive data privacy laws, whereas those provinces and states reliant on federal legislation are much weaker. For further details on each of the governments reviewed in this report please see Annex 1.

Summary of data privacy legislation across several jurisdictions

	Main data privacy legislation	Who the legislation covers	Consumer rights covered (access, rectification, erasure, portability)	Enforcement agency	Penalties for noncompliance
UK	The Data Protection Act (2018) and UK GDPR	Any personal data controllers and personal data processors operating within the UK or offering goods or services to individuals within the UK	Yes to all	Information Commissioner	GDPR penalties, plus an unlimited fine for the new offense of intentionally or recklessly re-identifying individuals from anonymized data
EU	GDPR (2018)	Any data controllers and data processors that hold EU citizens' personal data, regardless of their physical location	Yes to all	European Data Protection Board ensures the consistent application of data protection through the EU and each member state has their own national body for enforcing GDPR	For less severe violations, fine up to 2% of annual global turnover or €10 million (whichever is greater) For more severe violations, fine up to 4% of annual global turnover or €20 million (whichever is greater)

[50] Ibid.

Canada	PIPEDA (2000) and the Digital Privacy Act (2015)	Organizations deemed to be a ‘federal work, undertaking or business’ (e.g. banks, telecommunications, etc.) and organizations who collect, use and disclose personal data in the course of commercial activity	Does include access and rectification rights but does not include portability or erasure rights	The Office of the Privacy Commissioner of Canada (OPC) does not have the power to issue fines for compliance themselves, but rather can make recommendations to judicial courts	Fine up to CAN$100,000 per violation
Quebec	Quebec Privacy Act (1993) and Bill 64 (2021)	Organizations based in or with “real and substantial” connection to Quebec	Yes to all	Quebec Information Access Commission (CAI)	Fine for administrative penalties up to $10 million or 2% of global turnover (whichever is greater) Fine for penal offenses up to $25 million or 4% of global turnover (whichever is greater)
Singapore	Personal Data Protection Act (2013) and amendments	All organizations collecting, using or disclosing personal data in Singapore, regardless of whether or not the company is registered or physically located in Singapore	Right to portability, right to access and rectification in limited cases, but not to portability or erasure	Personal Data Protection Commission	Fine of up to SG$1 million or 10% of annual turnover (whichever is greater)
California	CCPA (2018) and CPRA (2020)	California businesses of a “substantial” size that collect personal data	Yes to all	The California Privacy Protection Agency	Fine of up to $2,500 per violation for negligent violations and up to $7,500 per violation for intentional violations

Part 2: Data stewardship in practice

Summary and lessons learned

As previously discussed, data trusts are a relatively new concept and, therefore, there are very few examples of established data trusts in practice. Further, the examples that do exist tend to be relatively new, making it difficult to ascertain their impact. According to a recent survey by the ODI and Aapti Institute, 80% of data trust projects are less than 5 years old or yet to be operational.[51]

For the purposes of this report, we have identified seven data stewards with varying structures and purposes across different sectors. Where possible, we have tried to evaluate their impact, although this has not been possible for all of our case studies depending on their maturity. What is clear from these case studies is that: 1. Their usage is pre-defined (by broad sector); 2. Revenue generation is not a criterion or aim; 3. Either individuals, or the trust itself, makes decisions about how the data is used.

The main lessons learned from the experience of these data stewards are:

1. **Despite their differences, all the data trusts we looked at all had social gain at their core.** Although the purposes and sectors of the data trusts we looked at differ wildly (UK BioBank was created to improve public health outcomes, whereas Agr-Gaia was built to improve efficiency in the agricultural sector), they shared a common aim to improve the welfare and livelihoods of the individuals and communities they serve. Whether empowering self-employed individuals (Driver's Seat) or opening mapping data up to underserved regions (Place), the primary goal of these data trusts is social gain rather than profit.
2. **In the short-term, data stewards are most likely to flourish in areas where data privacy legislation is well-developed.** All of the data stewards we have looked at are all concentrated in Europe and the United States, which may reflect the existence of more developed data rights landscapes - a fundamental prerequisite to data stewardship activities. This finding has been reiterated by a survey of data trusts by GPAI, Aapti and ODI, in which 37 of the 45 respondents were based in Europe and North America.[52]

[51] https://docs.google.com/document/d/18HPZbsd9DLQp5fk7iSzS6fs-ptGiWSJrm34UdR_3aMg/edit
[52] https://docs.google.com/document/d/18HPZbsd9DLQp5fk7iSzS6fs-ptGiWSJrm34UdR_3aMg/edit

Case Studies

UK BioBank

Purpose
UK BioBank was established by the Medical Research Council and the Wellcome Trust in 2006 to steward in-depth genetic and health data from half a million participants aged 40 - 69. Since its inception, the BioBank has conducted regular assessments of its initial cohort of participants, providing rare longitudinal data which forms a major resource to modern medicine - the largest and richest dataset of its kind.[53]

Scientists from around the world are able to apply to use the UK Biobank's data on the condition that their research is health-related, in the public interest, and is published in an academic journal or open source publication site.

Structure and Governance
It is worth noting that UK BioBank is not explicitly conceived of as a data trust - instead it is a charitable company whose board of directors' act as charity trustees under UK charity law and company directors under UK company law.[54] It is funded by the UK Department of Health, the Medical Research Council, the Scottish Executive and the Wellcome Trust.

The usage of participant data instead follows a strict ethical framework established at the company's outset, and a number of sub-committees within the BioBank's Board of Directors are responsible for holding the organization accountable to its ethical principles and its mission of contributing positively to global public health.

The Board oversees the management and control of UK BioBank and answers to the Medical Research Council and the Wellcome Trust. Individual sub-committees are charged with ensuring that priorities of the BioBank are met.

- The International Scientific Advisory Board guides the BioBank's overall scientific direction and ensures that UK BioBank's activities meet the needs of the global research community and improve public health.
- The Ethics Advisory Committee is charged with identifying and advising on ethical issues that come up in the process of collecting and using data as well as conducting detailed research in order to ensure that its advice is evidence-based.
- The BioBank Access Sub-Committee makes decisions relating to scientific access to UK BioBank's data including the use of datasets for potentially contentious research.

Impact
The UK BioBank's data sets have provided the empirical basis for hundreds of research papers.

Researchers in Canada were able to use data from more than 400,000 UK BioBank participants in order to make a breakthrough discovery in the fight against sepsis - a life-threatening reaction to infection which is the cause of a staggering 11 million deaths every year.

[53] https://www.ukbiobank.ac.uk/
[54] https://theodi.org/article/data-trusts-in-2020/

By examining genetic data and cholesterol measures in the over 3000 BioBank participants who developed sepsis, academics established a strong link between a patient's level of "good" high-density lipoprotein cholesterol (known as HDP) and their ability to fight off sepsis. The research suggests that raising HDL levels during sepsis infection may be a major aid to overcoming the illness. Drugs which increase HDL in patients are already readily available and used in the hope of reducing heart attacks and strokes.

This breakthrough would not have been possible without UK BioBank, which provides both a large sample size and detailed individualized genetic and health data.[55]

Place

Purpose
Place was launched early in 2021 in the US with the goal of making high-quality primary mapping data available to regions that are currently underserved.

Mapping data is crucial for a wide range of infrastructural and digital projects, but good quality data mapping data is held only by a limited number of actors across the world.

Place instead collects, processes, aggregates and manages mapping data which it then offers to its members in return for a membership fee.

Structure and Governance
Place is explicitly conceived of as a non-profit data trust. It is currently working with GovLab and FutureState in order to define its own long-term governance structure.[56]

At present, the plan is to distinguish Place's governance model through the creation of a separate legal data trust - known as Place Trust - which will hold all Place Data and will issue licenses for the use of these data to members. Place Trust will be independent of and separate from the operational body that creates the mapping data in the first place.

Place Trust is overseen by an independent board of trustees who are to be selected for their commitment to Place's mission. By putting ownership of the data into the hands of this independent board, Place hopes to create an organization which remains beholden to the public interest rather than shareholder interests.

At present, Place is funded by Omidyar Network, Rockefeller Foundation, Microsoft Planetary Computer, the David Weekley Family Foundation and others - but the long-term goal is for Place to be entirely financially sustainable.

Membership of Place will operate according to the notion of a "club good", whereby each member agrees to a set of terms and conditions governing the use of data and the ethical framework of that use. Membership fees pay for the collection of the mapping data.

Impact
Place has yet to be fully launched.

[55] https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/our-impact/cholesterol-and-blood-poisoning-the-story-unfolds
[56] https://thisisplace.org/blog-1/introducingplace/its-really-hard

Agri-Gaia

Purpose
Rapid developments in science have meant that the analysis of data has assumed an immense significance for increasing productivity in farms around the world.

This creates issues for small and medium-sized agricultural businesses, who have limited access to open data and difficulties turning small scale self-generated data into helpful insights. Agri-Gaia is a project launched by the German Federal Ministry for Economic Affairs and Energy which seeks to provide an infrastructure through which farmers can share the masses of data they generate and then reap the rewards. In this way, Agri-Gaia helps farmers turn data they acquire into optimized algorithms that inform future work.

Agricultural businesses will be able to exchange data and algorithms on the decentralized Agri-Gaia platform. The scale of the project will be such that users can download industry-specific AI building blocks to download directly to their tools. What's more, the platform is being designed to be interoperable and standardized such that data can be exchanged regardless of the manufacturer of the equipment that originally collected it.[57]

Governance and Structure
The Agri-Gaia project is still in the development stage. Kick-started by a €12m fund from the the German Federal Ministry for Economic Affairs and Energy,[58] the project is led by a group of research and industry partners under the supervision of the German Research Center for Artificial Intelligence.

Impact
Though the Agri-Gaia project is still in its early days, one ongoing project concerns the creation of an AI-supported application to guide the timely use of fertilizers. Farmers will be able to download the application - which is being tested over the course of 2022 - to their machines.[59]

Salus Coop

Purpose
Salus Coop is a non-profit data cooperative for health research founded in Barcelona by members of the public in September 2017.

Its vision is to have an entirely citizen-driven organization collecting the health data of participants who retain full sovereignty over the data they donate. Generated data is made available for research on the condition that it be accessible at no cost and that all data is anonymized and unidentified.[60]

Structure and Governance
Individual members retain access to all data they have donated, and have a right to know how their data will be used and by whom.

Impact
Salus Coop's health and social data has been made available to Pompeu Fabra University for a project seeking to improve home care services for the elderly through innovative new methods, including algorithms that provide insight for policy-makers and care managers.[61]

[55] https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Artikel/UseCases/agri-gaia.html
[58] https://www.wevolver.com/article/agri-gaia-strong-alliance-of-companies-develops-open-ai-standard-for-agriculture-based-on-gaia-x/
[59] https://www.bosch.com/stories/agri-gaia/
[60] https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/
[61] https://www.saluscoop.org/proyectos/ephocas

Driver’s Seat

Purpose

Driver’s Seat is an app which enables gig economy workers to collect the smartphone data they generate over the course of their shifts. Participants share the data they generate with the cooperative, which aggregates data from thousands of drivers to produce useful insight which helps drivers optimize their incomes.

Driver’s Seat also sells these insights to city agencies to help inform their own transportation planning decisions. The cooperative has a stated commitment to only sell data to customers that “respect our driver-first mission”.[62]

Drivers are also able to delete their account and all associated data at any time.

Structure and Governance

Driver’s Seat is a Limited Cooperative Association in the US. This means that it is a for-profit member-owned business. More detail on their governance structure is not readily available online.

Driver’s Seat is committed to full transparency, and drivers will always know who their data is being shared with and when.

Impact

Driver’s Seat is also in its early stages, but an early release of data indicates the types of insight that the platform might generate for its drivers.

One study, based on data collected by drivers between July and October 2021 found that tip rates for drivers were marginally higher on Uber than on Lyft - 28.5% of single destination Uber trips resulted in a tip compared to 25.5% on Lyft.

Another short-form study released to the public and based on driver-pooled data from DoorDash, GrubHub, Lyft, Uber and UberEats found that 85.8% of food delivery trips result in tips compared to only 24.8% of rideshares.[63]

Vivli

Purpose

Vivli is a neutral data broker which was launched in 2016. It aims to promote and facilitate scientific sharing and reuse of clinical research data. Members can store, share, and reuse anonymized individual participant-level data from clinical trials conducted by academia, pharmaceutical companies, foundations or non-profit entities.[64]

Structure and governance

Vivli is an independent non-profit data broker, which hosts anonymized clinical trial data in its system. Data contributors and data users each pay a fee when they share or request data, which covers the operating costs for the platform.

[62] https://driversseat.co/
[63] https://blog.driversseat.co/
[64] https://vivli.org/about/overview/

For data contributors, listing and storing anonymized personal data on the platform costs $4,000. Data contributors also have the option of including an independent review panel in their membership for an additional $5,500.[65] The independent review panel is run by the Wellcome Trust, and acts as an independent, neutral party that reviews research proposals based on their merit on behalf of the data contributors.[66] Data users pay a per-diem membership fee to access the platform, with different charges levied for different data libraries.[67]

To access data, users submit a data request on the platform, which is reviewed for feasibility by the data contributor, and then reviewed by an approving entity or the independent review panel to assess the merits of the research proposal and ensure it is in line with the data contributor's data sharing agreement[68] Each data contributor sets their own data sharing policies.[69]

Impact

In 2021, Vivli's inventory of data grew to include over 6,000 clinical studies from 38 members. Researchers used data accessed from Vivli to produce 51 publications, including articles in leading scientific journals, in 2021 (up from 7 in 2020).[70] Researchers have used data from the Vivli platform to make advancements in the understanding, management and treatment of various diseases, including rheumatoid arthritis, Crohn's disease, and breast cancer.[71]

Brighthive

Purpose

Brighthive is a platform that provides the business, legal and technical framework needed for organizations to safely share their data. Their mission is to responsibly promote the use of data to increase efficacy, equity and efficiency across private organizations, academic institutions and government entities.[72]

Structure and governance

Brighthive is a data platform which enters into contracts with data generators outlining how (and with whom) their data can be used and shared. These contracts - called a Data Trust Agreement - create a bespoke data trust for each member, defined by a legal, technical and governance framework:[73]

1. **Legal framework:** members maintain ownership and control of their own data, and set the terms of use through a formalized legal framework which can be amended as the member sees fit.
2. **Technological platform:** Brighthive's technology platform ensures data is shared securely and responsibility, and makes different members' data interoperable.
3. **Governance:** Brighthive data trusts establish a governing body made up of data generator representatives to monitor the trust and oversee how the data is used. Further, each trust has an external trustee that ensures the contract is adhered to.[74]

[65] https://vivli.org/how-to-share-your-data-on-the-vivli-platform/
[66] https://vivli.org/about/independent-review-panel/
[67] https://vivli.org/resources/vivli-secure-research-environment/
[68] https://vivli.org/about/data-request-review-process/
[69] https://vivli.org/members/ourmembers/
[70] https://vivli.org/annualreport/
[71] https://vivli.org/interviews-with-researchers-about-the-impact-of-their-secondary-analysis/
[72] https://brighthive.io/about/
[73] https://medium.com/brighthive/the-use-of-data-trusts-to-improve-impact-responsibly-e04aef5f40d9
[74] https://acumenacademy.org/blog/how-brighthive-helps-organizations-answer-complex-social-questions-ethical-data-sharing/

Brighthive generates revenue through its Data Trust Agreements, with members paying an annual subscription fee to access the data on the platform. The company charges for the linking and use of data rather than for data collection itself. Brighthive's profit is derived from the value of the data trust rather than the sheer quantity of data - they only collect the data required to fulfill the data trust agreements.[75]

Impact

Although Brighthive is a relatively young company, the impact of their data platform can be found in the efficiency gains it provided to the government of Colorado. In order to remain compliant with the Workforce Innovation and Opportunity Act (WIOA), the government of Colorado must report on whether students who have attended workforce development training programmes funded by federal grants have successfully landed jobs. Before Brighthive, this was done manually - with training providers calling alumni and emailing the government with information which was then manually cleaned in spreadsheets. Brighthive initiated a data trust consisting of 5 members, including private training programs, community colleges and government departments. The data trust automated the data sharing process, making a process that previously required an entire government IT team to be done with a single server. According to a co-founder of Brighthive, this has led to huge efficiency gains and cost savings for the government of Colorado.[76]

[75] https://acumenacademy.org/blog/how-brighthive-helps-organizations-answer-complex-social-questions-ethical-data-sharing/

[76] https://acumenacademy.org/blog/how-brighthive-helps-organizations-answer-complex-social-questions-ethical-data-sharing/

Part 3: Roadmap for municipalities and provincial Governments

The opportunity

For Government at all levels - municipal, provincial and federal - data provides a huge opportunity. Data can act as a catalyst for innovation, growth and improved outcomes. This is not a new or particularly radical statement, after all data has been collected and used for its insights by policymakers, researchers and commercial entities for hundreds of years. What is different now is the sheer scale at which data is produced and collected. While Canada is starting on this journey, for example the recent work in the health space culminating in the Building Canada's Health Data Foundation, there is more to be done.[77]

Zoe App

Zoe, a UK health startup, was originally conceived to better understand - by gathering data and applying machine learning, how and why people respond differently to various kinds of food. However, at the start of the covid outbreak Zoe pivoted to start tracking coronavirus symptoms and how these symptoms changed and progressed over time.

The Zoe covid study app launched in partnership with King's College London has since had 4 million contributors globally and is the world's largest ongoing study of Covid-19.[78] With users reporting their symptoms Zoe has been able to predict the prevalence of the virus and track Covid infections first in the UK and now in other countries. Daily reporting has generated greater scientific understanding of the virus.

Zoe was granted £2m to continue its operations helping to detect local hotspots, ensuring the UK government remains one step ahead of the virus.[79] The data is anonymised and shared giving researchers an in-depth insight resulting in the publication of over 300 scientific papers identifying new symptoms and characteristics of the disease.

Government, in particular now, holds vast swathes of data that offer huge insights on how to improve services, drive efficiencies and identify issues before they become overwhelmingly challenging. But making the most of this data can be extremely time consuming and require close and careful management.

While some datasets of not personally identifiable information can be opened up for anyone to use and build on, helping to create jobs in the new digital economy, others will require closer care. Data trusts can support governments unleash the power of their data whilst maintaining the highest levels of privacy and security protections.

By entrusting data to an independent trust, governments can:

- Enable greater access to datasets by private and third sector entities who share agreed goals and purposes.
- Enable greater collaboration on grand challenges, for example improving urban spaces or supporting an aging population, by creating new products, services and insights.
- Reduce the cost and resource burden on the government by outsourcing management of data to an independent and trusted partner.
- Create new jobs and businesses by opening up more datasets for developers to experiment and innovate with.
- Create new revenue streams from data that otherwise would be unproductive.

[77] https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/pan-canadian-health-data-strategy-reports-summaries/expert-advisory-group-report-02-building-canada-health-data-foundation.html

[78] https://covid.joinzoe.com/

[79] https://www.digitalhealth.net/2020/08/zoe-covid-19-symptom-study-app-government-grant/

The steps

In order to make this vision a reality authorities will need to take the following steps:

1. Lay the ground rules
As this report has demonstrated there are multiple approaches to enabling the creation and use of data trusts. Before embarking on this journey Governments should:

- Conduct an audit of all current legislation and regulation related to data collection, management, storage and use. This will help identify where there are gaps, complexities or obstructions to the use of data trusts.
- Identify priority issues and challenges that may benefit from the use of data trusts. As we have seen, opening up data can have unintended but welcome consequences, for example creating innovative products and services from insights made available to entrepreneurs. However, as we have seen, a mission or challenge-led approach can help focus energies and ensure that there are clear shared goals at the outset.

Some questions you can ask include:

- How do you want to define data trusts? Should it reflect the same legal structure as a property trust? How are fiduciary duties defined with regards to data stewardship - is this a legal obligation or simply a principle to guide behaviors?
- What regulations are currently missing? How can data privacy and data stewardship laws and regulations be improved? Is there somewhere to look to for best practice (e.g. EU GDPR, California CCPA, etc.)?
- Are there jurisdictional issues, particularly with data trusts that operate internationally or across provinces? How can these be solved through cooperation?

2. Build public trust
We have seen attempts to open up the use of data through the use of intermediaries or third-parties fail due to a lack of public trust and transparency. Similarly, research has shown that "the erosion of public trust and confidence in data-collecting organizations and in the technologies that rely upon this data (including AI) has provoked a backlash that threatens society's ability to access and use trusted data for the public good". [80] It is never too early to engage the public on this issue to help create space for an informed and constructive discussion.

[80] https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

This can take the form of:

- Citizen juries or other deliberative engagements with the public. These are complex issues that require a base level of understanding to engage with constructively, and so the use of citizen juries, in which a small group of people, representative of the demographics of a given area, come together to deliberate on an issue, can be extremely helpful.[81] Already, citizen juries have been used by various jurisdictions to explore the use of data trusts. For example, in London as part of a larger project looking at data trusts, small workshops were held with the public to explore the concept of a data trust and the expectations, hopes and fears associated with its introduction.[82]
- Consultations: a place remains for more traditional consultations which offers an opportunity for municipalities to test ideas and ways of working as well as elicit interesting and thoughtful contributions from experts in the field.
- Transparency: Whether before, during or after the introduction of any novel mechanism for handling citizens data, transparency is key. Nervousness and distrust is natural and spotlight is the best disinfectant. Transparency at every stage can help build public trust and reassure those with concerns that their data is secure and being used in an appropriate manner.

3. Invest in pilot schemes

Moving from theory to practice can be complicated, the use of pilots can help test new ideas in a controlled manner. Piloting can help test assumptions, highlight issues that need to be addressed, and build confidence in further pilots and more extensive rollouts. Across the globe pilots are underway exploring different models of data stewardship. Each is different shaped by the local context and regulatory environment it is embedded in. While these pilots will allow other jurisdictions and interested actors to learn a lot this is no substitute for testing out ideas on “home turf”.

In order for pilots to be successful, they should:

- Have clear objectives and metrics assigned to measure their success.
- Take an interdisciplinary approach to design that brings together experts-in-the-field, policymakers and relevant delivery partners and stakeholders.
- Be committed to full transparency of findings to enable others to benefit from any learnings and to earn and grow public trust and confidence.

[81] https://www.involve.org.uk/resources/methods/citizens-jury#:~:text=A%20Citizens’%20Jury%20is%20a,of%202%20to%207%20days.

[82] https://www.involve.org.uk/resources/blog/project-update/delving-data-trust-decision-making

Annex 1: Data Privacy Legislation

UK

The UK has been a leader in the development of data trusts since the publication of its 2017 report "Growing the Artificial Intelligence Industry in the UK", which considered ways to facilitate the implementation of data trusts to advance the development of Artificial Intelligence (AI). AI requires vast amounts of data and data trusts are being considered as a mechanism for data to be shared safely in a way that meets both the data holders and the data users mutual needs and interests.[83] To accelerate the role out of data trusts, the report suggested implementing a Data Trusts Support Organization (DTSO), which would "lead on the development of tools, templates and guidance for those who want to share and use data, so data owners and consumers can come together to form data trusts as and when they wish to do so."[84] DTSOs have not yet been implemented.

Further, in 2019, the UK Government invested £700,000 in three data trust pilot schemes aimed at maximizing the use of data to improve urban space, reduce global food waste, and crack down on illegal wildlife poaching.[85] The purpose of the project was to further develop a "data trust life-cycle", detailing best practices for designing, launching, running and evaluating data trusts.[86]

Data trusts were also highlighted in the National Data Strategy (2020) as a data sharing mechanism that can "fuel growth and innovation".[87] Following this Strategy, the UK has launched a new public consultation seeking feedback on how the Government can best support the activities of data intermediaries, including data trusts, by reforming the data privacy legislation - The Data Protection Act (2018) which implements GDPR in the UK. Their proposed reforms are aimed at removing barriers in the GDPR to promote a pro-growth and trusted data regime post-Brexit, as well as broadening the role of the Information Commissioner's Office to increase regulation in the sector. The results of the consultation are due to be published Spring 2022.[88]

EU

The development of data intermediaries, including data trusts, is reliant on the existence of clear and comprehensive data rights - including definitions around what individual data rights exist and how they can be transferred to data trusts. The EU began introducing data rights in 2009 through the EU Charter of Fundamental Rights, which recognized the right to personal data protection under an individual's right to privacy.[89] Data rights were further enhanced in 2018 under the General Data Protection Regulation (GDPR), which harmonized European data protection law, strengthened individual's data protection rights, and provided data protection supervisory authorities with means of enforcement.[90] Although some evaluations have suggested GDPR still has room for improvement, it nonetheless represents the global standard in data protection legislation and has been the basis for many other international data protection regulations, including the Californian Consumer Privacy Act and the Thai Data Protection Act.[91]

[83] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf
[84] Ibid.
[85] https://www.gov.uk/government/news/digital-revolution-to-use-the-power-of-data-to-combat-illegal-wildlife-trade-and-reduce-food-waste
[86] http://theodi.org/wp-content/uploads/2019/04/ODI-Data-Trusts-A4-Report-web-version.pdf
[87] https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy#data-1-3
[88] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf
[89] https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en
[90] https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1163
[91] https://free-group.eu/2020/01/31/evaluation-of-the-general-data-protection-regulation/

In 2020, the Commission published the European Data Strategy, which aims to create a single market for data, allowing data to flow freely within the EU across different sectors.[92] Following this, the Commission has proposed The Data Governance Act, a world-leading governance framework to safeguard the activities of data intermediaries, including data trusts. The Act includes a range of safeguards aimed at ensuring neutrality and accountability of data intermediaries, including the introduction of:[93]

- Conditions around the re-use of certain public sector data, including confidential or personal data.
- A requirement for data intermediaries to notify a competent public authority as they are created, who is in turn tasked with monitoring their compliance with existing regulations.
- A framework for the voluntary registration of entities which collect and process data made available for altruistic purposes
- A framework for the establishment of a European Data Innovation Board tasked primarily with advising the European Commission in the development of policies, regulations, guidelines and legislation related to data sharing.

In addition, in late February 2022 the Commission introduced more advanced data sharing legislation through the highly anticipated Data Act.[94] The Data Act intends to harmonize the different regulations around the use of private and public sector data, encouraging data sharing and re-use.

Canada (Federal)

Data protection laws in Canada are complicated: they include a mix of general and sector-specific regulations, and exist at both the federal and provincial level. Quebec, British Columbia and Alberta each have their own provincial statutes for data privacy law which extend the scope of the legislation to organizations not otherwise governed by the federal law.

Federally, the Personal Information Protection and Electronic Documents Act 2000 (PIPEDA) regulates personal data use by the private sector and the Privacy Act regulates data use by the public sector. PIPEDA has been criticized as insufficient, most notably because it only applies to "commercial activities", leaving a gap in data privacy legislation applicable to non-commercial organizations (such as non-profits and charities).[95] The Digital Privacy Act in 2015 expanded the enforcement powers of the Information Commissioner, introduced new consent exceptions, and made it mandatory for organizations to notify the Commissioner and individuals when a data breach occurs.[96]

In 2020, the federal Government introduced Bill C-11 (the Digital Charter Implementation Act or the Consumer Privacy Protection Act) to replace the previous legislation that regulated data privacy in the private sector - the Personal Information Protection and Electronic Documents Act (PIPEDA). Bill C-11 aimed to improve data privacy, however, there were many holes in the legislation identified, including around consent, and protections for children and youth.[97] [98] After serious debate and many concerns raised about Bill C-11, including by the Privacy Commissioner of Canada, it failed to pass through Parliament. It is expected that a new version of the bill will be introduced at some point during this Parliamentary session (before October 2025).[99]

[92] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
[93] https://data.consilium.europa.eu/doc/document/ST-14606-2021-INIT/en/pdf
[94] https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113
[95] https://mcmillan.ca/insights/covid-19-realities-push-ontario-government-to-launch-public-consultation-to-improve-the-provinces-privacy-laws/
[96] https://www.insideprivacy.com/international/canada/highlights-of-the-canada-digital-privacy-act-2015/
[97] https://www.ontariocanada.com/registry/showAttachment.do?postingId=37468&attachmentId=49462
[98] https://www.pwc.com/ca/en/services/consulting/privacy/privacy-canadian-business-hub/creating-data-trust.html
[99] https://www.dlapiperdataprotection.com/index.html?t=law&c=CA

Although there have been some efforts to encourage data sharing, including cross-sectorally - for example, the Canadian Data Governance Standardization Collaborative aimed to streamline data standardization practices across civil society, industry, and academia in 2019[100] - Canadian federal laws do not currently grant personal data rights of portability and erasure. This creates challenges for data trusts, who need clarity around their decision-making ability over the data they steward.

Quebec

Due to the inadequacies of the federal act (PIPEDA) Quebec, like British Columbia and Alberta, has introduced its own data privacy laws - the Quebec Privacy Act. The Quebec Privacy Act applies to both consumer and employee personal data for organizations within Quebec that are not otherwise governed by the federal regulation, PIPEDA (which only covers commercial activities).[101]

In 2021, Quebec reformed its data privacy legislation massively with the adoption of Bill 64, which transforms the legal framework for data privacy to resemble the EU GDPR. Bill 64 will introduce several major reforms as it enters into force over the next two years, including:[102]

1. Introducing individuals' right to erasure and data portability.
2. Mandating that organizations appoint a person responsible for the protection of personal information.
3. Requiring organizations to enact appropriate contractual safeguards when transferring data outside of the province to ensure adequate protection is upheld.
4. Introducing more severe penalties for lack of compliance.
5. Changing to an "opt-in" model of consent for organizations using tracking, localization or profiling technologies.

Singapore

Singapore's overarching data privacy legislation is the Personal Data Protection Act (PDPA), which was introduced in 2013. The PDPA governs the collection, use, and sharing of personal data in addition to other sector-specific legislation (e.g. the Banking Act).[103] It is applicable to all organizations collecting, using or disclosing personal data in Singapore, regardless of whether or not the company is registered or physically located in Singapore.[104] The PDPA favors innovation by granting exemptions on a case-by-case basis for the development of new technologies.

Major reforms aimed at strengthening the PDPA were passed in 2020 and have started coming into effect since 2021. The reforms include: reforming the model of consent, making it mandatory to notify the Personal Data Protection Commission (PDPC) of data breaches, broadening the authority of the Commission, and introducing harsher criminal offenses for the mishandling of data.[105] The PDPC has also introduced a data portability obligation, in which organizations must transmit an individual's data to another organization in a machine-readable format if requested, making data sharing easier.[106]

[100] Standards Council of Canada (SCC), Canadian Data Governance Standardization Collaborative (DGSC), see https://www.scc.ca/en/flagships/data-governance
[101] https://www.dlapiperdataprotection.com/index.html?t=law&c=CA
[102] https://www.dlapiperdataprotection.com/index.html?t=law&c=CA
[103] https://www.pdpc.gov.sg/Overview-of-PDPA/The-Legislation/Personal-Data-Protection-Act
[104] https://www.dlapiperdataprotection.com/index.html?t=law&c=SG
[105] https://www.epiqglobal.com/en-gb/thinking/blog/singapore-makes-changes-to-data-privacy
[106] https://www.oecd-ilibrary.org/sites/3b6a594d-en/index.html?itemId=/content/component/3b6a594d-en#section-102

In 2019, Singapore launched the Trusted Data Sharing Framework, which established guidelines for organizations seeking to explore data sharing partnerships. The framework covers data-sharing strategies, legal and regulatory considerations, technical and organizational considerations, and operationalising data sharing. The aim is that this Framework makes data sharing easier for organizations.[107]

California

There are no comprehensive federal data privacy laws in the US, and instead a number of states have introduced their own state-level legislation. California introduced the nation's strongest data privacy legislation in 2018 (and came into effect in 2020) with its California Consumer Privacy Act (CCPA). The CCPA gives individuals the right to disclosure, erasure, access, opt-out and non-discrimination and was largely influenced by and modeled after the EU GDPR.[108]

The CCPA will be strengthened by the California Privacy Rights Act (CPRA), which was voted for in 2020 and will come into effect in 2023. The CPRA establishes an agency responsible for enforcing the data privacy rules laid out in the CCPA, limits the collection, use, retention and sharing of personal data to that which is "reasonably necessary", requiring annual cybersecurity audits for all businesses who process consumers' personal information which presents a risk to individual's privacy or security, amongst other reforms.[109]

[107] https://www.zdnet.com/article/singapore-unveils-framework-to-facilitate-trusted-data-sharing-between-organisations/
[108] https://oag.ca.gov/privacy/ccpa
[109] https://www.whitecase.com/publications/alert/dust-settles-california-privacy-rights-act-ballot-initiative-modifies-and