DIGITAL DATA PARTNERSHIPS: BUILDING THE FOUNDATIONS FOR COLLABORATIVE DATA GOVERNANCE IN THE PUBLIC INTEREST

Sarah Gagnon-Turcotte
Miranda Sculthorp
Steven Coutts

February 2021

With the support of:

SYNAPSE C

Laboratoire d'innovation urbaine de Montréal

OpenNorth

ACKNOWLEDGEMENTS

This report was developed with the support of two organizations committed to data partnership initiatives in Québec: Synapse C and the Montréal Urban Innovation Lab. We would also like to thank everyone who agreed to be interviewed for this project for their time and generosity. Finally, thanks to our colleagues at Open North, including Megan Wylie and Lauriane Gorce, for their support on this project.

ABOUT OUR PARTNERS

Synapse C

Synapse C aims to develop a data culture within the arts and culture community. To this end, the non-profit organization develops and shares expertise in creating value from data for the arts and culture sector in Québec and Canada, while working with the cultural, academic, and entrepreneurial sectors to become an international reference in this space.

Recently, Synapse C has been working to define best practices in data collection and pooling. Through leading initiatives and groups in the pooling of their data, Synapse C has helped share and analyze several datasets while ensuring that they are safely hosted. To date, more than 60 cultural sector organizations have benefited from their work, mainly in Québec. In keeping with its values of transparency and collaboration, Synapse C ensures that its findings are shared with as many organizations as possible.

With its experience in leading and supporting data partnerships in the arts and culture sector, Synapse C wants to explore and contribute to collaborative data governance models that serve the public interest. To this end, it has partnered with Open North to further explore the governance of data partnerships.

Montréal Urban Innovation Lab

The Montréal Urban Innovation Lab of the City of Montréal (LIUM) fosters and supports the emergence of innovative solutions from all walks of life. LIUM is a space within the City dedicated to innovation, a free space where we explore, experiment, and imagine a future that has met the challenges of today. It is a space in which Montréalers, businesses, municipal employees, and partners are invited to co-create a more people-focused, creative, open and efficient city.

LIUM is the unit in charge of implementing the wide-ranging *Montréal in Common* program, a winning project of the Government of Canada's Smart Cities Challenge. Bringing together more than 22 partners with a desire to rethink the city, its 13 projects will help drive the development and testing of innovative solutions in three main areas: (1) *food*, (2) *mobility*, and (3) *data and regulatory experimentation*. The third theme expresses the desire to use data sharing and collaborative governance as levers to implement all the projects, develop a better understanding of needs, measure impacts, and make more informed decisions. The role of Open North is to help support and oversee this process.

The actions to be deployed by 2024 as part of *Montréal in Common* will rely on innovation and new technologies to improve the quality of urban life in all its aspects: service efficiency, rich human relations, a healthy and stimulating environment, and a place where everyone feels good and belongs.

ABOUT US

Open North

Founded in 2011, Open North is a Montréal-based non-profit organization that has its roots in the open data and civic technology space. Today, its interdisciplinary team works with a wide variety of innovative public administrations and community stakeholders in key areas of data and technology management and governance.

Our applied research, capacity building, and advisory services are driven by our values of transparency, autonomy, and responsibility. Our mission is to empower communities to reinvent how they use and manage data and technology.

Open North has developed expertise in operationalizing data governance in smart cities as well as data pooling and sharing in various sectors (including culture, health and mobility). Open North is currently playing a key role in establishing collaborative data governance in Montréal as part of the *Montréal in Common* program.

Projects under this program will address multiple data governance issues: data collection, access, and sharing; privacy; consent; data ownership, control, and security; data openness; compliance with existing legislation, etc. The objective is to define clear governance frameworks that enable data to be stewarded as a commons.

This research report on data partnerships aims to establish a solid foundation for Open North's actions under its mandate in *Montréal in Common*, as well as consolidate useful findings that can support the work of organizations like Synapse C that seek to accelerate the implementation of projects to create value from data in Québec.

ACKNOWLEDGEMENTS

Research and Writing

Sarah Gagnon-Turcotte, Director, Applied Research Lab, Open North

Miranda Sculthorp, Senior Research Analyst, Applied Research Lab, Open North

Steve Coutts, Research Analyst, Applied Research Lab, Open North

Document Layout

Tatev Yesayan, Design and Communications Consultant

Recommended citation:

Gagnon-Turcotte, Sarah, Miranda Sculthorp and Steven Coutts (2021). *Digital Data Partnerships: Building the Foundations for Collaborative Data Governance in the Public Interest.* Open North.

CONTENTS

FIGURES, TABLES AND BOXES

Figures

Tables

Boxes


BUREAUX
A LOUER
(514)
282-1155
CANPRO
MASSAGE
DUNNS
AVIS
PARC
Sainte-Catherine
CLUB

PREFACE

Motivated by the desire to explore the best data management and governance practices, the Montréal Urban Innovation Lab (LIUM) and Synapse C, commissioned Open North to research terms and conditions of data sharing.

Over the past decade, the many applications of data have shown how data can contribute to decision-making processes for municipal organizations and, more broadly, to societal transformations. These recent developments require us to question the concepts of data sharing, governance, and management, as well as establish common concepts that facilitate collaboration, build resilience and agility, and support responsible ethics. As data oligopolies emerge in some sectors, LIUM, Synapse C and Open North, along with their data partners, want to position data as a shared asset and use it to drive our collective development. Making data a shared asset requires the development of new data use frameworks that respect human rights, as well as the intellectual property of contributors.

Starting with these observations, the study was motivated by several shared goals:

- Strengthen the capacity of actors in the data ecosystem to monitor changes in data usage and the impact of new regulations;
- Improve knowledge of these issues and explore and review existing models with a view to defining the mechanisms specific to the case of Montréal;
- Initiate a dialogue with as many people as possible to develop a framework for seamless collaboration with well-defined decision-making processes.

The Open North team has successfully translated these particularly complex goals in this document. The risks and challenges associated with data governance are indeed worthy of attention. The high potential for individual and collective impact, including from the professional and organizational spheres, must be addressed if we are to create a real space for civic innovation, citizen engagement, decision-making, and multi-stakeholder collaboration.

Over the past two years, Synapse C has worked to identify best practices in collective data use, and subsequently disseminate them to as many organizations as possible. In 2011, the City of Montréal opened up many of its datasets to the public, on which the LIUM team has been working since 2015 in order to realize maximum social and economic value, while respecting the human and civic responsibilities they support. Our two teams are particularly motivated and excited to see this research become a key document to be added to the toolkit of any organization that wishes to use data from an individual and collective perspective.

Diane De Courcy
Executive Director, Synapse C

Stéphane Guidoin
Director, LIUM

EXECUTIVE SUMMARY

In the age of digital transformation and artificial intelligence, data—whether it be open data or big data—and the issues data raise are in the spotlight. Spurred by the potential of shared data, a growing number of public, private, and civil society stakeholders are interested in sharing digital data with third parties to achieve public policy objectives and resolve complex social issues. New forms of inter-organizational cooperation that aim to share, combine, cross-reference, and leverage datasets are emerging every day.

However, these digital data partnerships require time, effort, resources, and sustained collaboration. Their success also depends on strong data governance that protects the public and maintains its trust.

Organizations that wish to engage in data partnerships will find in this report a discussion of the different success and activation factors for digital data partnerships, along with practical information that will guide them in building shared data governance that is collaborative, responsible, effective, and accountable.

Digital data partnerships: definitions and concepts

It quickly became apparent during our research that there were no turnkey data partnership or governance models that could be easily applied, duplicated and scaled. Despite the interest in notions such as data trusts and digital commons, practical, mature, and documented experiences are still few and far between.

Organizational factors, as well as the social and political contexts in which digital data partnerships operate have a significant influence on the data governance frameworks used by organizations to share and exchange their data. Indeed, the role of collaboration is critical to the success of data partnerships, despite the diverse configurations of private, public, and civil society actors that are involved, the types of data they value, and the objectives they pursue. Moreover, recognizing citizen concerns about new technological advances and the absence of a proven regulatory framework to protect their rights, it seems that data partnerships that focus on the common good enjoy greater legitimacy and capacity for action.

These findings are documented in the first chapter of our report, and they continue to serve as a guide for the analysis and explorations of best practices in data governance which follow.

Key components of data governance

As digital data partnerships multiply and evidence of their potential grows, more and more stakeholders in sectors other than information technology are becoming interested in their potential. Chapter 2 therefore provides a practical knowledge base in data governance by defining common terms and concepts.

Simply put, data governance determines who makes decisions, how decisions are made, and how decision-makers are held accountable for the collection, use, sharing, or control of an organization or group's data.

To help break down this definition, we use a conceptual framework developed by Abraham, Schneider, and vom Brocke (2019) in order to identify the building blocks of data governance. This framework is primarily descriptive, rather than prescriptive. It highlights the influence of antecedents (such as legislative frameworks or organizational culture) on governance and identifies three key elements that define the scope of data governance, that is, the level of organizational governance, the characteristics of (shared) data, and the domain scope, which in turn influence the concrete mechanisms through which governance is operationalized on a daily basis.

Three guiding principles for data governance in the public interest

Chapter 3 is devoted to an in-depth discussion of several categories of data governance mechanisms. These include informed consent; anonymization; risk assessment; data quality, standardization, and interoperability; access management; compliance monitoring; and the auditability of decisions. The mechanisms that can be deployed are as numerous as the issues they seek to address. Their selection must consider not only the antecedents, but also the context of each partnership and the scope of the established governance.

To help organizations shape their governance choices toward morally and socially desirable ends, we have structured this chapter around three key guiding principles. We thus propose that the governance frameworks developed by digital data partnerships, be guided by the following principles:

- **Responsible: realizing value from data in a responsible and ethical manner**
- **Effective: managing data effectively and consistently**
- **Accountable: assessing compliance and impact on an ongoing basis**

Montréal perspectives

In the final chapter of the report, we present the results of discussions with representatives of Montréal organizations involved in data sharing initiatives. The real-world experiences and perspectives of these stakeholders, who are active in the arts and culture sector or the *Montréal in Common* smart city program, have played an important role in the creation of this report. For one, they confirmed the influence of organizational factors as an enabling condition and success factor for data partnerships.

The interviews also highlighted a number of barriers to participation, including: the role of data culture within the organization, the degree of adherence to the data sharing initiative, a lack of organizational capacity, and data production costs. They also highlighted the importance of securing third-party support (e.g. legal experts, government initiatives, public funds) in response to the complexity of issues raised as well as technical requirements of the project.

Despite these major obstacles, we found the participants to be genuinely interested in exploring and developing alternative models of data governance that are tailored to their needs and ambitions.

Conclusion

This report was written with the intention of making a useful contribution to the existing theoretical and practical corpus on data governance. We hope that it will also support, in a concrete way, the movement toward shared and pooled data in Québec, where for some years now, key players such as Synapse C and the Montréal Urban Innovation Lab have been exploring and experimenting with new approaches to data governance.

For those who are interested in exploring and experimenting new digital data partnerships, we conclude by summarizing some of the key lessons learned during the making of this report:

- **Recognize that the public interest is defined and negotiated by citizens;**
- **Invest time in your collaboration and experimentation processes;**
- **Create data governance that is tailored to your needs;**
- **Document your impact and share your successes.**

INTRODUCTION

Ever since the advent of the computer, the production of digital data has been growing exponentially. As a source of information, innovation, and competitive advantage, data is proving to be an increasingly important means for organizations to increase their economic and social impact. In all sectors, both public and private organizations are exploring new ways to use data to realize their full potential. The possibilities for combining and sharing data seem almost endless.

Therefore, it is no surprise to see an increasing appetite for **digital data partnerships**. Such initiatives, which bring together many organizations, including public agencies, to join forces to collect, exchange, combine, and share their data, are multiplying worldwide. However, far from providing benefits alone, data use and sharing also raise important issues and risks that these new types of partnerships will inevitably face.

These numerous and complex challenges often relate to: privacy, informed consent, responsible and ethical data use, privatization and access, biased or discriminatory algorithmic decision-making, and citizen participation in decision-making. In Canada and internationally, the modernization of the legal framework required to respond to these challenges has been slow to respond to growing public concerns.

This legal vacuum has gradually fostered the emergence of digital **data governance** as a means to better frame data use and sharing. This concept, derived from the field of enterprise data management, is now used by a multitude of practitioners and researchers to define frameworks and principles capable of rebuilding public trust. These insights have led to the development of various models of data governance such as data trusts, data cooperatives, and digital commons.

However, examples of these governance models being put into practice are still rare. The factors that make digital data partnerships successful, like the fundamental principles of responsibility, effectiveness, and accountability in data governance, remain relatively undocumented. Therefore, identifying innovative practices and techniques for using, combining, and sharing data responsibly and ethically in new collaboration models is essential.

This report aims to support the creation and success of digital data partnerships intended to serve the public interest, while proposing concrete recommendations for establishing the required data governance mechanisms. It is part of a broader movement to share data in Montréal where, in recent years, key players have committed to exploring and experimenting with new approaches to data governance.

As part of our research, we first identified a range of data partnership models discussed in the literature

This report aims to support the creation and success of digital data partnerships intended to serve the public interest, while proposing concrete recommendations for establishing the required data governance mechanisms. It is part of a broader movement to share data in Montréal where, in recent years, key players have committed to exploring and experimenting with new approaches to data governance.

and sought to identify their governance mechanisms, success factors, risks encountered, and obstacles faced. We reviewed over 100 articles and reports from grey and academic literature on data governance, data sharing, data pooling, and inter-organizational governance. We then completed this analysis by conducting a series of interviews with experts from Montréal's arts and culture community and representatives of organizations involved in the *Montréal in Common* smart city program. We hope that our work will feed into new experiments and research that can eventually serve to validate our results.

This report has three main chapters. The first delves into the theoretical foundations of data governance, by presenting the definitions and concepts at the heart of digital data partnerships. It will familiarize the novice reader with a technical vocabulary, while also providing an overview of the main components of data governance. The second chapter provides tangible examples of how data governance mechanisms embody within digital data partnerships the key principles of responsibility, effectiveness, and accountability. The final chapter summarizes the lessons learned from our discussions with stakeholders involved in data sharing initiatives and offers a concrete perspective on the challenges they face. In our conclusion, we offer recommendations for organizations that are considering embarking on the foundation of new digital data partnerships.

CHAPTER 1

DIGITAL DATA PARTNERSHIPS: DEFINITIONS AND CONCEPTS

NEW TYPES OF DIGITAL DATA PARTNERSHIPS

In its simplest form, data sharing is an exchange of data between entities for a specific purpose (Thuermer et al., 2019). For decades, such exchanges have been taking place in various forms in the public, private, and academic sectors.

For many organizations, data sharing is an essential operational function. For example, government departments and agencies share data to facilitate internal planning and improve service delivery. In the private sector, business models based on data aggregation, sharing, brokering, and analysis have increasingly emerged in recent years (D'Addario et al., 2020). In the academic sector, research centres have a long history of coming together to achieve new results by aggregating and analyzing large anonymous datasets, such as genomic datasets (Byrd et al., 2020). Nowadays, however, digital data is being shared between different organizations at an unprecedented scale, and increasingly between public and private partners (Verhulst et al., 2019).

The number of **digital data partnerships** has grown substantially in recent years, and indeed, this trend is not surprising. The modern global economy is increasingly dependent on the flow of data between individuals and organizations. Data is being used in a growing range of social and economic activities, and many entities are now producing and sharing data. New forms of collaboration between non-profit organizations, private companies, academic research centres, and public administrations are emerging to leverage these new data flows.

We define a digital data partnership as any initiative where two or more organizations align around a common goal and parties engage in the sharing of data to realize its value.

BOX 1: DATA PARTNERSHIP EXAMPLES

PULSAR

Implemented by Université Laval and Alliance santé Québec, PULSAR is a collaborative space for sustainable health research and innovation, the first of its kind in Québec. As both a virtual and real space, PULSAR brings together actors from all walks of life looking at health research differently to significantly and sustainably improve the population's health and well-being. The project directly involves citizens in the process, inviting members of the public to join the platform and participate in sustainable health studies led by the network partners. The data generated by the research will ultimately feed into a digital sustainable health data bank to create a valuable information resource for studying health in all its dimensions.

IDAHO HEALTH DATA EXCHANGE

The Idaho Health Data Exchange (IHDE), a non-profit organization, is the State of Idaho's health information exchange organization. To achieve its goals, the IHDE works with a wide range of stakeholders and is actively building a state-of-the-art technology infrastructure to provide access to reliable data and information, combining traditional health care data with other data sources. The exchange of health information allows doctors, nurses, laboratories, and other health care providers to safely and quickly access their patients' electronic health information and thereby improve the speed, quality, safety, and cost of patient care.

APIDAE TOURISME

Apidae Tourisme is a network of tourism stakeholders created in 2004 in the Rhône-Alpes region in France. The platform is used to collaboratively manage tourism information across all the territories covered by the project. The network members, who produce and share their data on the platform, are also the primary users of the data. The platform also captures, stores, and uses tourism information to inform customers about the network members' destination offerings. The Apidae network now includes 23 French departments, 1 overseas community, and more than 23,800 platform users.

YORKINFO PARTNERSHIP

The YorkInfo Partnership describes itself as a "government market" for data and analytics professionals. It coordinates the sharing of geographic information system (GIS) data—including aerial photography, water and waste infrastructures, and road networks—between nine local municipalities and one regional municipality. The data can be accessed by all partners through an online portal, and supports planning and development, emergency services, social services, and economic development throughout the region.

TUI'KN PARTNERSHIP *STRENGTH IN NUMBERS PROJECT*

As part of the Tui'kn partnership, the First Nations of Nova Scotia are working with provincial and federal partners to improve their access to reliable health information through the Strength in Numbers Project. This initiative has led to the creation of the Nova Scotia First Nations Client Linkage Registry, a registry of the First Nations population in Nova Scotia directly linked to provincial health data. This allows First Nations to better track a set of health indicators for their population. One of the corner-stones of this project is a data-sharing agreement between the First Nations and the Government of Nova Scotia.

GLOBAL FISHING WATCH

Global Fishing Watch is a collaborative effort between SkyTruth, Oceana, and Google, to map and measure fishing activity worldwide using data from the automatic identification system, a vessel tracking system used by large fishing vessels. A map of the data is available to anyone with Internet access. It allows users to track when and where commercial fishing is taking place around the world. Governments can use this map to do such things as identify and take action against vessels that are not authorized to fish in their waters or that are illegally fishing in protected areas.

Our research to date has identified many initiatives where data is shared, linked, or pooled as part of a multi-stakeholder partnership. The Data Collaboratives Explorer site maintained by The GovLab, for example, lists more than 500 data sharing initiatives, including several dozen multi-stakeholder initiatives, illustrating the multitude of forms digital data partnerships can take, the different sectors in which they are found, the different types of data they use, and the diverse goals that guide them.

BOX 2: REASONS FOR SHARING DATA

The Data Sharing Toolkit identifies five main reasons why organizations are interested in sharing their data in a partnership (Smart Dubai and Nesta, 2020, p.17):

- Discovery of new insights and identification of the key questions that need to be addressed
- Unlocking new sources of value and innovation by opening up data to third parties
- Providing a more complete and accurate picture of complex issues for rapid decision-making
- Increased prediction capability and forecasting
- Optimized process efficiency and coordination

Several organizations, including the Open Data Institute (UK), The GovLab (US), Nesta (UK), and, more recently, Mozilla's Data Futures Lab have committed significant resources to document digital data partnership models and further their adoption.

The most commonly identified models include data collaboratives, data cooperatives, data commons, data trusts, and the concept of personal data sovereignty.

Since digital data partnerships are an emerging phenomenon, there is still no established definition for many of these terms. They are sometimes used interchangeably in the literature, making them difficult to distinguish. For example, as Bass and Old (2020, p.11) point out, "data commons may include a trust or a co-operative-like structure, or the term 'data trust' may be used to refer to something resembling a commons model to others."

As we will see in this chapter, this confusion between the different governance models is due in part to the fact that each data partnership initiative is unique. Each partnership originates in a particular dynamic and context, determined by the interactions between the various actors involved, their expectations and levels of expertise, their motivations and relationships, and the laws and rules that govern them.

Indeed, we found considerable differences between the conceptual and idealized definitions used to describe data governance models and their real-world implementation.

Moreover, despite efforts to date to distinguish different data governance models with greater precision, our research has shown significant gaps in the documentation of compelling, mature, and successful examples (Coutts and Gagnon-Turcotte, 2020). These new approaches are still emerging and therefore require careful analysis.

Based on these considerations, this report does not seek to promote or define any particular digital data partnership model. Instead, we have focused on the

BOX 3: ALTERNATIVE DATA GOVERNANCE MODELS

DATA COLLABORATIVES

The GovLab defines data collaboratives as forms of partnerships that bring together private companies, research institutions, and government agencies that are designed to combine data and generate public value (Verhulst and Sangokoya, 2015).

DATA COOPERATIVES

Data cooperatives are similar to traditional cooperatives. In that sense, they are a group of people who come together to achieve common goals in a joint organization. Data cooperatives can be defined as mutual "organisation[s] owned and democratically controlled by members, who delegate control over data about them" (Hardinges et al., 2019, p. 9).

DIGITAL COMMONS

Digital commons, on the other hand, are initiatives where individuals or organizations share data as a common resource and collectively set the rules governing access (Bass and Old, 2020).

DATA TRUSTS

Data trusts are defined as legal structures mandated to provide independent stewardship of data for the benefit of their trustees (Hardinges et al., 2019). This model has received a lot of attention in recent years. For example, the Open Data Institute began exploring data trusts with the British government in 2018. What distinguishes data trusts from other models is that the trust is an intermediary distinct from the members of the data sharing initiative.

PERSONAL DATA SOVEREIGNTY

The main feature of personal data sovereignty is that individuals have direct control over their personal information. New digital platforms and initiatives now provide individuals "the means [...] to control, use and share their data – and re-users with whom data subjects decide to share their data" (Micheli et al., 2020, p. 9). In particular, the personal data sovereignty movement has been strengthened by the right to data portability under Article 2, Directive 20, of the General Data Protection Regulation (GDPR) adopted by the European Union.

fundamentals of data governance, as it is implemented through partnerships in general. In adopting this approach, we provide a more operational perspective focused on data governance that can be adapted and scaled, regardless of the preferred form and structure.

Before addressing data governance directly, however, we feel it is essential to discuss two factors that, despite the diversity of digital data partnerships, have quickly emerged as elements critical to their success: collaboration and the pursuit of the public interest.

FOUNDATIONS OF A SUCCESSFUL DIGITAL DATA PARTNERSHIP

In our research, while attempting to identify the key characteristics of the governance of data partnerships, we also identified many factors which influence their success (for more on success factors, see Chapter 4 which highlights success factors identified during interviews with Montréal organizations). However, two factors stood out as essential foundations for digital data partnerships in the current context: **creating a climate conducive to trust and collaboration and the pursuit of a positive social impact in the public interest.**

The collaborative process

The role of collaboration in a data partnership may at first seem obvious: any group or partnership needs to collaborate in order to achieve their common goals. However, the challenges faced by participants in a digital data partnership are such that the role of collaboration becomes essential. These challenges include an unclear and rapidly evolving legal framework, the difficulty of anticipating potentially negative impacts stemming from new ways of using data, the high level of technical expertise required (especially for partnership formed in non-technology sectors or industries), and public scrutiny. In such a high-risk environment, **trust and collaboration** are among the key characteristics of interactions in successful digital data partnership initiatives.

As Ansell and Gash (2007) demonstrate in their *collaborative process*, trust and collaboration create a continuous and virtuous cycle. Trust fosters engagement in the process, which contributes to developing a shared vision of issues and objectives, which then promotes dialogue and the achievement of 'quick wins'. These achievements in turn demonstrate the value of the collaborative process and serve to build trust among the partners.

Barriers to participation

Before stakeholders engage in a collaborative process, the relationships between them can be characterized by certain dynamics, whose tensions are sometimes described in the literature as competition, loss of control, and divergent objectives.

Competition

Organizations may be reluctant to share data that they believe provides a competitive advantage to other parties (Klievink et al., 2018). For example, two companies may invest considerable time and resources in creating databases of information about their respective customers, including information on demographics, purchase histories, and purchase preferences. From the perspective of maintaining the status quo, companies would see these datasets as representing a unique competitive advantage. However, a digital data partnership is unlikely to occur if both companies consider themselves engaged in a zero-sum game. It is more likely to succeed if both companies have entered into the partnership believing that they can realize greater value by sharing their data, for example, through "generat[ing] additional revenue and reach[ing] a broader market" (D'Addario et al., 2020, p. 14).

Loss of control

Similar to the concept of competition, resource theory holds that to create a competitive advantage, companies must possess a precious and scarce resource, while limiting the use of this resource to others through imitation, transfer, or substitution (Wade and Hulland, 2004). Sharing data in a partnership means relinquishing some control over a fundamental organizational resource (Klievink et al., 2018), leaving stakeholders feeling vulnerable. In a collaborative environment, members of a partnership must agree to let their data assets be used in processes they do not fully control. However, all partners stand to gain: in allowing partners a certain degree of autonomy on how they use the data of others, parties will succeed at paving the way for the creation of new information, services, or products (Klievink et al., 2018).

Divergent objectives

Even though organizations have different objectives (Vangen and Huxham, 2012), they can still create synergies by combining their data resources and capabilities. Organizations may agree on a shared vision, but differ on exactly how to achieve it. For example, in the pursuit of the vision of creating equitable and sustainable local food systems, one organization may focus on providing food bank services, while another may focus on reducing food waste in supply chains (Bolychevsky et al., 2019). A divergence in the preferred method may hinder the creation of a digital data partnership.

Overall, these dynamics can be overcome when the parties are firmly committed to a collaborative process and to defining and adhering to a set of common objectives.

Alignment around a shared vision

Indeed, Ansell and Gash (2007) identify the development of a shared vision among stakeholders as one of the critical steps in the collaborative process. This implies that the parties agree on a definition of the problem to be resolved, align themselves around one or more common objectives, and agree on the knowledge and means necessary to resolve the problem. These bases must be established before identifying and securing the capacity or resources required for an organization to engage in a digital data partnership.

The partners must develop a clear idea of the problem they are trying to resolve, precisely determine the value and usefulness of data for solving this problem, and determine the specific benefit that a digital data partnership can offer. At this stage, the parties must seek to answer questions such as: *Why this project? What are our objectives? Why these*

data? What will they be used for? A shared vision must emerge from this exercise, in light of which the partners can ultimately measure their progress. Otherwise, the partnership risks seeing a costly drift in objectives or having its partners waver in their participation in the initiative over time.

Applying concepts from the collaborative process of Ansell and Gash (2007) to a concrete example of data partnership in the Netherlands, Klievink et al. (2018) add that a positive feedback loop between trust and collaboration fosters institutionalization of the latter. This collaboration makes it easier for partnerships to ride out difficult periods or overcome tensions between parties, including competitive dynamics, when they emerge. They also note that a history of collaboration between parties can help support the emergence and success of digital data partnerships.

Use cases

It should be noted that rallying around a shared vision can be a challenge, especially when there are many players involved with diverse interests, contributing to the wide diversity of data that can be shared. In the context of data partnerships, especially those focused on innovation, where the parties are interested in finding new ways to generate value from data that have never been shared before, this can be even more difficult.

In these circumstances, identifying the **use cases** that warrant data sharing greatly facilitates the development of a shared vision. Use cases clarify the objectives the parties wish to achieve and then identify the data that must be shared and how it will be used. This approach facilitates the identification of barriers and risks to sharing, and contributes to a shared understanding of the issues at hand.

Considering the significant resources, especially technical resources, that may be invested in a data partnership before any results are obtained, it is worthwhile to adopt iterative and agile approaches focused on testing and experimentation. These approaches are widely used in the information technology sector, as they encourage short development cycles and prototyping in order to discover whether a type of data use is viable and desirable. Exploring and identifying use cases as a basis for the collaborative process is synonymous with iterative and experimental approaches.

This philosophy holds promise in digital data partnerships because it encourages the parties to adapt and innovate in the face of new and ever-changing data and digital environments. Furthermore, prototyping is a good way to build trust around small successes that can then be scaled up. In this sense, experimentation is an essential component of the *collaborative process*, as it is described by Ansell and Gash (2007).

BOX 4: DISCOVERING DATA VALUE THROUGH USE CASES

The term use case refers to a way of using software, or in this case data, to achieve a specific utility for the stakeholders involved. Identifying specific use cases is critical to creating value from data in a digital data partnership.

A single dataset can give rise to multiple use cases. Each use case defines the scope of the data's use for achieving its purpose. The functional requirements can then be described, and the governance mechanisms required by the system identified. Use cases are generally built from an end-user perspective (definition based on Jacobson et al., 2011).

To develop use cases, it is useful to start by identifying and documenting available and potentially useful data, for example through a data audit or data mapping.

The pursuit of the public interest

Our presentation of digital data partnerships has thus far focused on several key aspects: data sharing and value creation involving different stakeholders, the development of a collaborative dynamic, the alignment around a set of common objectives, and the articulation of use cases. The question then becomes, what kinds of objectives can be pursued through data partnerships? While many different kinds of objectives can be pursued, we see the most potential for positive social impact when digital data partnerships pursue goals in the public interest.

Much more than a simple good practice, we find the pursuit of the public interest is necessary for successful digital data partnerships. As we know, appropriating the value of public and personal data for commercial and profit purposes is increasingly poorly perceived by the public and can lead to serious issues of legitimacy (Artyushina, 2020). Project credibility and relationships with local partners and beneficiaries can also be undermined when stakeholders feel they are being exploited for their data, or when projects have negative or unforeseen consequences. The consequences for organizations engaged in an initiative that does not succeed at gaining the public's trust can be serious and can affect funding as well as create legal liability issues. Anchoring digital data partnerships in the public interest, is thus increasingly widespread.

Indeed, a quick review of recent data governance discourse and projects reveals different ways in which the public interest is currently embodied in digital data partnership initiatives. First, most stakeholders acknowledge that there are ethical reasons for protecting data and ensuring they are not misused – reasons that go beyond mere compliance with existing laws and regulations. For example, according to one of the most reputable guides in the field the *Data Management Body of Knowledge* (DAMA-DMBOK) (Earley et al., 2017, p. 49), data management must be guided by ethical principles including respect for individuals, charity, and justice.

Second, more and more initiatives now embody the concept of data for the public good. This notion recognizes that public institutions and governments contribute to the production of data on the public – for instance on data demographics, infrastructure, mobility or health – which ought to be used for the public's benefit (European Commission, 2020, p. 8). Simply put, data has a public value — or in the words of the City of Montréal Digital Data Charter:

"Organizations collect data that are public in scope or that serve the public interest on behalf of the citizens. These data represent a shared asset and a common good. Therefore, it is the city's duty to allow each individual to benefit from the value of these data by making it available to residents."

Other schools of thought go further and consider data to be a public good in itself, or a common resource that must not only serve the public, but be managed collectively as well. This idea of stewarding data as a common resource is largely inspired by the theories of Elinor Ostrom, Nobel laureate in Economics, in her work on the management of commons (Ostrom, 1990).

So then, in concrete terms, how can a digital data partnership ensure that its objectives are indeed anchored in the pursuit of the common good?
In response to this question, we identified three complementary pillars of action: the creation of tangible benefits for the public, the engagement of citizens, and the adherence to good data governance standards.

> **Organizations collect data that are public in scope or that serve the public interest on behalf of the citizens. These data represent a shared asset and a common good. Therefore, it is the city's duty to allow each individual to benefit from the value of these data by making it available to residents.**
>
> **- Digital Data Charter of the City of Montréal (2020)**

Generate tangible benefits for the public

A digital data partnership in the public interest must first ensure that its use of public and personal data generates real benefits for society. These can be direct benefits for individuals or broader positive social impacts. For instance, at a series of workshops held in communities in England, three organizations (Involve, Understanding Patient Data, and the Carnegie UK Trust) studied the definition and assessment of benefits that can be generated from the sharing of personal information among public service providers (Scott, 2018). The results of their study identify a number of criteria for assessing the impacts of data sharing initiatives: the number of people who will benefit, the severity of the need (for example, improving conditions for vulnerable populations such as individuals experiencing homelessness), whether key social issues (such as social isolation) are addressed, and the duration of long-term impacts on individuals and the services provided. These are just some of the criteria that apply to the public sector.

In addition to assessing the tangible benefits of their initiatives, members of a digital data partnership may reflect on how these benefits can be distributed equitably among members of the public, for example, in accordance with the principle of justice, as identified in the DAMA-DMBOK (Earley et al., 2017, p. 52).

Engage the public in a meaningful way

Citizens have expectations, aspirations, and even demands about how their (personal) data and data that resides in the public domain should be collected, used, and managed. The public's contribution to these matters is therefore essential for ensuring the legitimacy and success of digital data partnerships.

On the one hand, governments (at all levels) play a key role in this area of intervention, as they have a responsibility to initiate and lead public discussions and debates on data and their responsible use through public consultations or other consultation forums. Governments also play a critical role in developing and building the skills and knowledge that citizens and stakeholders need to participate in discussions and decision-making in data partnerships (see Box 5).

Generally speaking, however, any digital data partnership, even one established among private organizations[1], has an interest in implementing robust consultation processes that allows its stakeholders, as well as citizens, to have an input in the initiative and in particular, the way that data is collected and used. To build legitimacy, these approaches must engage a wide variety of voices and perspectives, which implies the implementation of mobilization strategies adapted to different populations, especially those typically underrepresented in consultation processes. Importantly, citizens and stakeholders should be involved and engaged early on, well before any data collection takes place.

One way to encourage citizen participation is to create explicit communication channels or permanent bodies, such as citizen advisory committees, where the public is encouraged to participate in the governance of the initiative itself. These different modes of citizen participation help maintain the public's trust on an ongoing basis.

1 A private organization can be a non-profit organization, an association, or a cooperative.

Adhere to strong principles that guide data governance

Finally, in order to achieve socially responsible outcomes, digital data partnerships must ensure that the data they use are managed responsibly, effectively and with strong accountability mechanisms in place. This is where data governance comes into play.

Data governance is a framework for thinking about data management practices and the decision-making processes that underpin them. Data governance exists to ensure that the way in which data are collected, processed, accessed, used, stored, shared, etc. ultimately serves to achieve the partnership's shared objectives. It also serves to ensure that the parties' data processing complies with legislation and broader data governance principles.

The remainder of this report will be entirely dedicated to exploring data governance and its role in digital data partnerships. In Chapter 2, we break down the basic elements of data governance, while contributing to the creation of a shared vocabulary on the subject. Then, in Chapter 3, we examine in greater detail the concrete governance mechanisms which enable us to achieve the principles of responsibility, effectiveness, and accountability. These principles are central to any partnership which aims to pursue goals in the public interest.

BOX 5: DATA LITERACY

Recognizing that data plays an increasingly central role in decision-making processes, to empower citizens to participate fully in our society, many authors (including Ridsdale et al., 2015; Bhargava and D'Ignazio, 2015; Calzada Prado and Marzal, 2013; Wolff et al., 2016) stress the importance of building the general public's data literacy skills. Wolff et al. (2016, p. 23) define this concept as "the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. [Data literacy] is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyse, visualise, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process."

BOX 6: THE STEPS OF THE DATA SHARING PROCESS

As we have seen, building trust and collaboration and setting clear objectives based on the pursuit of goals in the public interest, will make a digital data partnership more likely to succeed. However, we are aware that any project will require its partners to go through many other steps, such as identifying a business model that will ensure the partnership's financial viability or deciding on the technologies to be deployed.

Here is an example of the key decision points identified by The Royal Academy of Engineering (2019) to develop a data sharing initiative:

1. Define the opportunity
2. Identify the scope of data to be shared and how it will be used
3. Develop the business model that allows value to be generated and shared
4. Develop the model for data sharing and the partnership
5. Ensure that the right people are involved with appropriate skills and expertise
6. Identify the constraints, including legal and regulatory requirements, on how data is shared and used, and how these should be addressed
7. Identify the architectures and technologies needed to enable data sharing
8. Develop the mechanisms for good governance and oversight, to enable trusted data sharing

The remainder of this report focuses on Step 8: building data governance that can enable partners to achieve their objectives.

CHAPTER 2

THE KEY COMPONENTS OF DATA GOVERNANCE

The more data an organization collects, the more it needs to manage the use of the data on an ongoing, systematic basis. This need is even more critical when the data are held or used by multiple organizations and then shared and aggregated to derive value. **Data governance** must, therefore, be at the heart of any digital data partnership initiative.

WHAT IS DATA GOVERNANCE? A CONCEPTUAL FRAMEWORK

Since data governance and its fields of application are subject to various disciplinary interpretations, there are multiple ways to conceptualize it. It is useful to look to the field of information technology (IT) for definitions on data governance. According to Weill, IT *management* is about what decisions are made, while IT *governance* is about who makes those decisions and how they are held accountable (Weill, 2004). This view distinguishes two levels of abstraction: the "what" (management) and the "how" (governance).

Data governance must not be confused with data *management*. For example, DAMA-DMBOK (Earley et al., 2017, p. 67) describes data governance as "the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets."

In short, data governance determines who makes decisions, how the decisions are made, and how the decision-makers are held accountable for the collection, use, sharing, or control of an organization or group's data.

In short, data governance determines who makes decisions, how the decisions are made, and how the decision-makers are held accountable for the collection, use, sharing, or control of an organization or group's data.

This clarity concerning the responsibilities and decision-making processes associated with any particular dataset is critical to the success of any digital data partnership. A data governance framework can also help enshrine and assess compliance and adherence of the project according to the principles of responsibility, effectiveness, and accountability, which ultimately serve to ensure the common good is pursued and protected.

To facilitate understanding data governance and its operationalization, we have used a conceptual framework developed by Abraham, Schneider, and vom Brocke (2019) in our analysis. This conceptual framework, described in Box 7, facilitates identifying key components of data governance and their interrelationships. This enables a better understanding of the role and impacts of data governance in digital data partnerships.

It is important to note that this framework only describes these elements as they appear in the literature; it is not a prescriptive framework that indicates how data governance should work or which values should guide it. Since the framework is intended to be descriptive rather than prescriptive, this report also discusses the importance of a shared vision, principles, and objectives to digital data partnerships.

BOX 7: UNDERSTANDING DATA GOVERNANCE THROUGH A CONCEPTUAL FRAMEWORK

The following conceptual framework, developed by Abraham, Schneider, and vom Brocke (Figure 1) synthesizes the key components of data governance based on a review of the literature published on data governance over the past two decades:

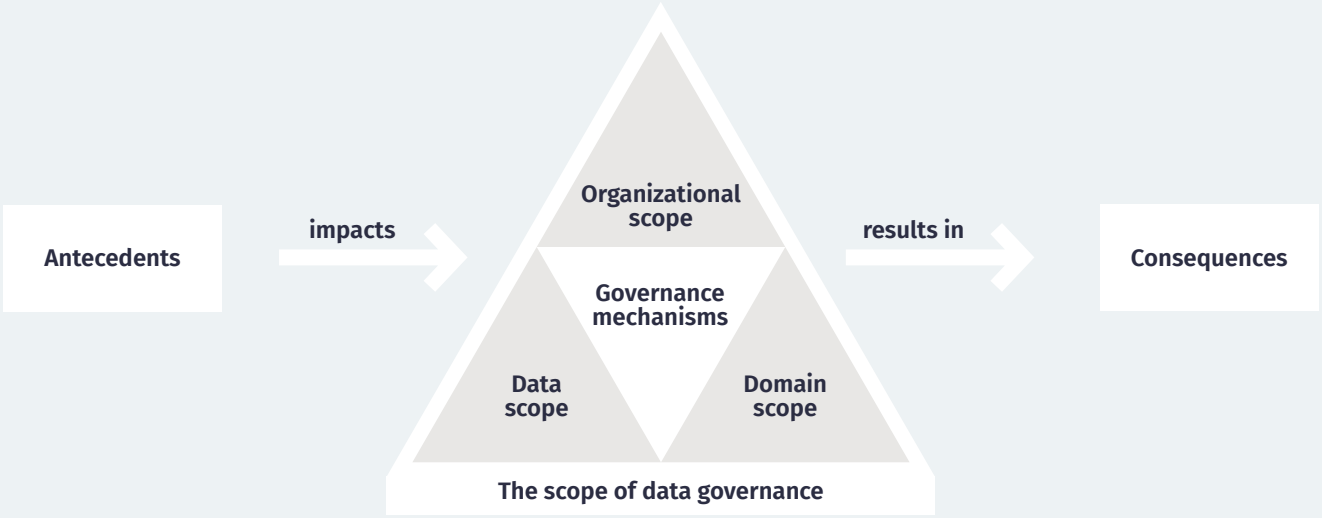


FIGURE 1: DATA GOVERNANCE FUNDAMENTALS

Retrieved from the *International Journal of Information Management*, vol. 49, Rene Abraham, Johannes Schneider, and Jan vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda" p. 428, (2019), reproduced with permission from Elsevier.

They conclude that such governance is expressed through a complex set of explicit or implicit **mechanisms** that can take many forms, including policies, procedures, practices, etc. that govern the collection, use, sharing, and control of data.

These mechanisms exist to ensure the operationalization of data governance. They are shaped by various **antecedents** at the political, legal, regulatory, organizational, or even cultural levels.

These mechanisms also depend on the **scope** of the data governance, which has three dimensions: its **domain scope**, the **organizational scope, and the data scope**.

Finally, data governance choices have measurable **consequences**. These may include improving the short-term operational efficiencies of an organization or group, mitigating certain risks (such as privacy breaches), or, over the longer term, increasing public trust in data governance.

ANTECEDENTS

The first element of data governance consists of various antecedents that determine and influence data governance. According to Abraham, Schneider, and vom Brocke (2019), there are both pre-existing internal (organizational) and external (mainly regulatory) antecedents.

Internal antecedents, which are directly related to the mode of operation and priorities of the organizations governing the data, may be strategic, organizational, or cultural. For example, an organization's profitability objectives, degree of centralization, and type of leadership can all influence the preferred governance framework. The presence of silos, the organizational culture (driven by innovation or not, for example), or the level of buy-in from management are also considered to be internal antecedents. These organizational aspects are important, but we will not discuss them in detail in this report. They must nevertheless be kept in mind by those preparing to initiate a digital data partnership.

The **external antecedents** are primarily existing laws and regulations, as well as existing norms and standards. Both the legal framework and the standards can vary between regions and industries. Moreover, the type of data, the jurisdiction or the domain scope of the governance may determine that one law applies instead of another. For instance, the applicable legal system may designate the application of certain types of mechanisms, to ensure compliance. Overall, special attention must be paid to these conditions from the outset when developing a digital data partnership.

The legal context

In Canada, many laws set out antecedents for digital data partnerships. In this section, we will briefly highlight those we find most relevant to data partnerships, including those governing personal data and intellectual property.

Privacy protection

The *Personal Information Protection and Electronic Documents Act* (PIPEDA) is the main source of law governing personal data at the federal level. It applies to any organization that collects, uses, and disseminates personal information while conducting commercial activities (see Table 1). It sets out the rules (including their exceptions) that organizations must follow to use consumers' personal information.

Several provinces, including Alberta, British Columbia, and Québec, have provincial privacy laws similar to the corresponding federal legislation (Office of the Privacy Commissioner of Canada, 2017).

Québec has two primary laws which describe how personal data must be processed. Public sector organizations are subject to the *Act respecting access to documents held by public bodies and the protection of personal information*, while private sector organizations are subject to the *Act respecting the protection of personal information in the private sector*.

In June 2020, the Québec government introduced Bill 64, an *Act to modernize legislative provisions as regards the protection of personal informationl*. to strengthen the existing legal framework. This bill seeks to modernize the legislative framework governing personal information protection in both the private and public sectors, so that it becomes more responsive to today's technological realities and is better aligned with international legislative frameworks. The proposed legislative reform is based on two main principles: giving citizens control over their personal information and making organizations accountable for the data they collect and use (Du Perron, 2020a).

Bill 64, if enacted, will introduce a significant change in the way personal data is defined. Indeed, this new bill introduces an innovative distinction between "de-identified" and "anonymization". Bill 64 considers personal information to be de-identified when it

TABLE 1: LEGAL OBLIGATIONS UNDER THE *PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS ACT*	
1. Accountability	An organization is responsible for personal information under its control. It must appoint someone to be accountable for its compliance with these fair information principles.
2. Identifying Purposes	The purposes for which the personal information is being collected must be identified by the organization before or at the time of collection.
3. Consent	The knowledge and consent of the individual are required for the collection, use, or disclosure of personal information, except where inappropriate.
4. Limiting Collection	The collection of personal information must be limited to that which is needed for the purposes identified by the organization. Information must be collected by fair and lawful means.
5. Limiting Use, Disclosure, and Retention	Unless the individual consents otherwise or it is required by law, personal information can only be used or disclosed for the purposes for which it was collected. Personal information must only be kept as long as required to serve those purposes.
6. Accuracy	Personal information must be as accurate, complete, and up-to-date as possible in order to properly satisfy the purposes for which it is to be used.
7. Safeguards	Personal information must be protected by appropriate security relative to the sensitivity of the information.
8. Openness	An organization must make detailed information about its policies and practices relating to the management of personal information publicly and readily available.
9. Individual Access	Upon request, an individual must be informed of the existence, use, and disclosure of their personal information and be given access to that information. An individual shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate.
10. Challenging Compliance	An individual shall be able to challenge an organization's compliance with the above principles. Their challenge should be addressed to the person accountable for the organization's compliance with PIPEDA, usually their Chief Privacy Officer.

Reproduced from "PIPEDA Fair Information Principles" by the Office of the Privacy Commissioner of Canada (2019).

no longer allows a person to be *directly* identified (Du Perron, 2020b). However, de-identified data still carries the risk of identifying an individual *indirectly* through advanced data analysis techniques that are widely available today. De-identified data therefore remains subject to legislation (Du Perron, 2020b; Rocher, Hendrickx and de Montjoye, 2019). Personal information is considered *anonymized* when the masking of the data no longer allows for the direct or indirect identification of an individual, and this alteration of data is considered *irreversible* (Bill 64, article 111). Thus, Bill 64 provides that only anonymized data can be exempted from the right to the protection of personal information, a situation similar to that provided for in EU Regulation 2016/679 (the General Data Protection Regulation [GDPR]) (Du Perron, 2020b).

The data governance legislative landscape will continue to evolve significantly in the coming years, and not just in Québec. In response to stricter privacy legislation introduced in other jurisdictions, notably in Europe, the Canadian federal government is currently reviewing its own legislation with the aim of better responding to the opportunities and challenges of our digital society.

It's worth noting that international laws may also impact digital data partnerships. For example, GDPR is extraterritorial in scope in that it applies to non-EU organizations that offer goods or services or collect data on European Union residents (not only its citizens). For example, a Canadian university that recruits international students from the EU may be subject to the GDPR if it processes their personal information (Information and Privacy Commissioner of Ontario, 2018). Legal changes in Canada and abroad could have a significant impact on organizations wishing to realize value from their data.

Intellectual property

Data has become an essential source of value in the new digital economy. Data partnerships will almost certainly face the question of who owns the data and what this ownership entails (The British Academy and The Royal Society, 2018). Property rights to data, as Teresa Scassa (2018a, p. 2) notes, "provide a powerful basis for control." Just as an organization may wish to own its data for commercialization purposes, a government may assert its property rights to data to earn revenue or, conversely, make it available as open data. We are also witnessing a growing number of voices in favour of new digital rights that recognize citizens' property rights to their personal information (Bass et al., 2018; Mozilla Insights, 2020).

While there are several sources of property rights under Canadian law, including the *Copyright Act*, the *Patent Act*, and the *Trademarks Act*, it is not clear whether data, in general, are subject to property rights. Data is different from other types of assets, which impact their ability to be possessed. First, they are non-competitive, which means distributing and sharing data does not decrease the quantity of data available. The original creator of a given piece of data can give an exact copy to another party without losing any part of the original. Second, some data may deal with several people at one time. Consider, for example, genetic information that concerns not only an individual but their entire family as well, making it difficult for an individual to claim exclusive ownership. In Canada, the courts have ruled that individuals can have the right to access and correct their personal information but that these rights are not recognized as a property right.

Current copyright law distinguishes between "facts" or "ideas" in the abstract (which are not protected) and

BOX 8: DATA OWNER VS. DATA HOLDER

As the Office québécois de la langue française (OQLF) explains, the term *data* or *file owner* (in French « propriétaire d'une donnée ou d'un fichier ») generally refers to a named individual who is responsible for managing and protecting one or more computer files and who has the authority to make any decisions concerning that file or files, with a view to ensuring their integrity and confidentiality. Therefore, the term *owner* in this context does not necessarily refer to the person who holds the property rights to the data or file(s) under the current legal framework.

For this reason, the OQLF goes on to say that the use of the terms data or *file owner* is sometimes disputed and that the term *file holder* or *custodian* (in French « détenteur de fichier ») is also used. Despite these reservations, the terms *data* or *file owner* are firmly established in the field of information security.

To adequately distinguish the two concepts, we will use the term "data holder" to refer to individuals who are responsible for managing the data and "data owner" for those who hold the property rights to the data.

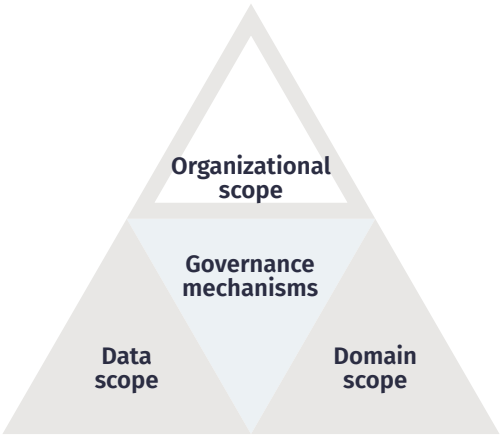
original expression of ideas (which are protected). Thus, a compilation of facts, including a dataset to which new information is continually added, may not be covered by copyright protection, even though it requires curation. When data are inferred or derived by analyzing and processing large amounts of data, it depends on whether the data are considered facts (which are not protected) or original expressions of ideas (which are protected) (Scassa 2018a). Therefore, the data is closely linked to the systems that produce it, making any generalizations on the issue of data ownership difficult (Scassa, 2018a).

In conclusion, given the complexity of the current legal landscape and the fact that the rules and their interpretations in the digital sector are still changing, partners in a data partnership should rely on experts to verify compliance with existing legislation or to draft clear agreements where data ownership and control are clearly specified.

THE SCOPE OF DATA GOVERNANCE

Data governance is defined by a scope consisting of three elements, according to Abraham, Schneider, and vom Brocke (2019) (see Figure 1): (1) the organizational scope, (2) the data scope, and (3) the domain scope.

The organizational scope: Who participates in digital data partnerships?



EXCERPT FIGURE 1

The **organizational scope** is the unit of reference to which data governance applies. The unit of reference can be a single organization, multiple organizations, as in digital data partnerships, or even an ecosystem. For this reason, Abraham, Schneider, and vom Brocke distinguish between the intra-organizational (single entity) and inter-organizational (partnership and ecosystem) levels. Therefore, the organizational scope is determined by the stakeholders involved which, in turn, influences how the governance is organized, the interactions it generates, and the preferred data governance mechanisms.

It is useful to consider the roles played by the different stakeholders in a digital data partnership. The role that each partner plays may depend on the partners' objectives, the identified needs, and

BOX 9: KEY STAKEHOLDERS AND THEIR ROLES

Data partnerships are organized around a variety of relationships and digital data flows between different actors. Therefore, defining the roles of the actors involved in these initiatives is a useful complement to the data governance framework developed by Abraham, Schneider, and vom Brocke. An overview of the key roles that can be played in a partnership is presented below (OECD, 2019; IMDA and PDPC, 2019).

- **Data holders.** They have the expertise needed to decide how the data are used and shared. They are sometimes called "data owners" even though they have no legal property rights to the data they hold. These data holders choose to share their data or enrich them with other parties.
- **Data users.** These actors are the recipients of the data that are the subject of the partnership. They generally strive to generate value by analyzing and processing shared data to transform it into useful information.
- **Data intermediaries.** These entities provide the means and assistance needed to share data. Their responsibilities may include providing expertise in data preparation and analysis, legal assistance, risk management, initiative funding, and even capacity-building activities.
- **Governing bodies.** One or more bodies may be empowered to exercise leadership, coordination, oversight, and compliance functions for the established initiative or data governance framework.
- **Regulatory authorities and bodies.** These actors influence data partnerships by creating laws and standards for using and protecting data and by publishing guidelines or codes of practice. Such authorities may include governments, associations, and institutions.
- **Beneficiaries.** These are entities that derive benefits or advantages from the data partnership. For example, the beneficiaries may include citizens, communities, or groups.

Note that these roles are not mutually exclusive and may overlap.

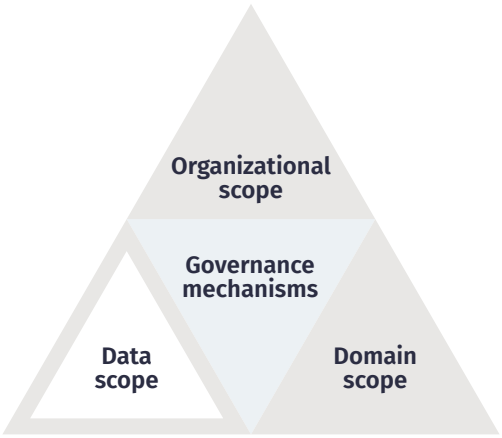
capacities (see Box 9). For example, an organization may decide to join a partnership to share important databases with others. Another organization may join the partnership, not sharing any data itself, but rather providing support through offering the technical capabilities needed to analyze and process data.

The organizational level at which data governance is traditionally exercised is ***intra*-organizational**, meaning limited to a single organization. For example, one common objective of data governance within an enterprise is to increase operational efficiency by defining and implementing principles, policies, and practices (all governance mechanisms) around data quality, sharing and storage. Such governance can help the organization operate more efficiently by improving coherence between its units, reducing redundancies, and improving data searches and access to results. A single organization can bring together all types of actors under one roof: data holders, users, beneficiaries, intermediaries, and governing bodies.

The level of data governance in digital data partnerships is then ***inter*-organizational**. This type of collaboration requires structural mechanisms that align stakeholders to achieve common objectives and structure their participation, while procedural mechanisms standardize and coordinate data management practices at the group level, improve access to data, or reduce risks. The objectives of digital data partnerships include developing innovative uses for information and solving problems through interdisciplinary or cross-sectoral collaboration and thinking.

Data governance can even extend to the **ecosystem level**, as in the case of municipalities engaged in various smart city projects. These cities seek to establish global governance frameworks that integrate decision-making processes that can influence the collection, use, and sharing of data from their administrations, citizens, partners, and active stakeholders throughout their territory. The establishment of coalitions (Cities Coalition for Digital Rights), the adoption of declarations (Montréal Declaration for a Responsible Development of Artificial Intelligence) or charters setting out founding principles (metropolitan data charter of the City of Nantes, France) or the establishment of transnational ties with active movements in other jurisdictions are all excellent examples of ecosystem-based governance mechanisms.

The data scope: What kinds of data are shared?



EXCERPT FIGURE 1

Data governance should be tailored to the **data scope** to which it applies. According to Abraham, Schneider, and vom Brocke, the data type needs to be precisely determined since it largely conditions the data governance mechanisms and even the applicable regulatory framework. The specificities of the data, such as their origin, can raise completely different issues.

However, there is no dominant taxonomy that identifies specific data types or classifies their

BOX 10: WHAT ARE DIGITAL DATA?

Data represent facts in the form of text, numbers, images, sound, or video (Earley et al., 2017, p. 19). When we talk about digital data, we are often referring to data that are encoded in a format that allows them to be processed by computers (Office québécois de la langue française, 2004). Note that data are not just facts about the world but interpretations of facts that require context to be meaningful (Earley et al., 2017, p. 19). The same underlying information can be represented in different ways, depending on the purpose of its use.

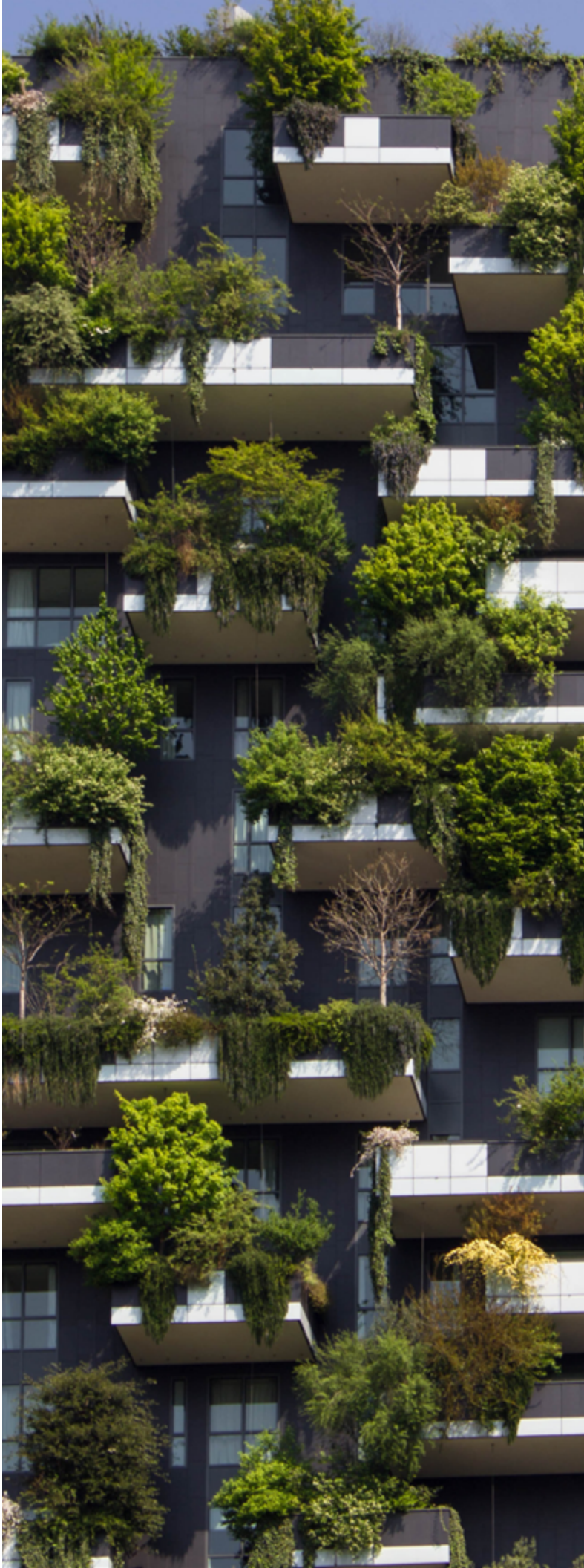
characteristics. For example, Abraham, Schneider, and vom Brocke only distinguish between *traditional data* (such as transactional data and reference data) and *big data* (such as Web and social media data). While conventional data governance focuses on ensuring consistent use across the organization, big data governance focuses on supporting innovation and reducing risks (Abraham, Schneider, and vom Brocke, 2019, p. 431).

Data can be categorized in several other ways: by their domain (personal, private, public), their openness (closed, shared or open), their source (freely provided, observed), their function (master, reference, metadata, transactional), their degree of sensitivity (de-identified, anonymized), or even their subject matter (mobility data, social data). Therefore, in the following sections, we further explore some of these categories to help organizations joining a digital data partnership better identify the characteristics of the data they use and understand the issues they raise.

The data domain: personal, private, or public

Data can be considered as belonging to one of the following three domains, which determine the rules (i.e., laws and regulations) governing their use (OECD, 2019):

- The **personal domain** encompasses all personal data "relating to an identified or identifiable individual for which data subjects have privacy interests." This area is generally governed by data privacy regulatory frameworks;
- The **private domain**, which includes all proprietary data generally protected by intellectual property rights (including copyrights and trade secrets) or access and control privileges (such as those provided for in contract law and criminal law applied to cybercrime), and whose economic interest generally prohibits their sharing;



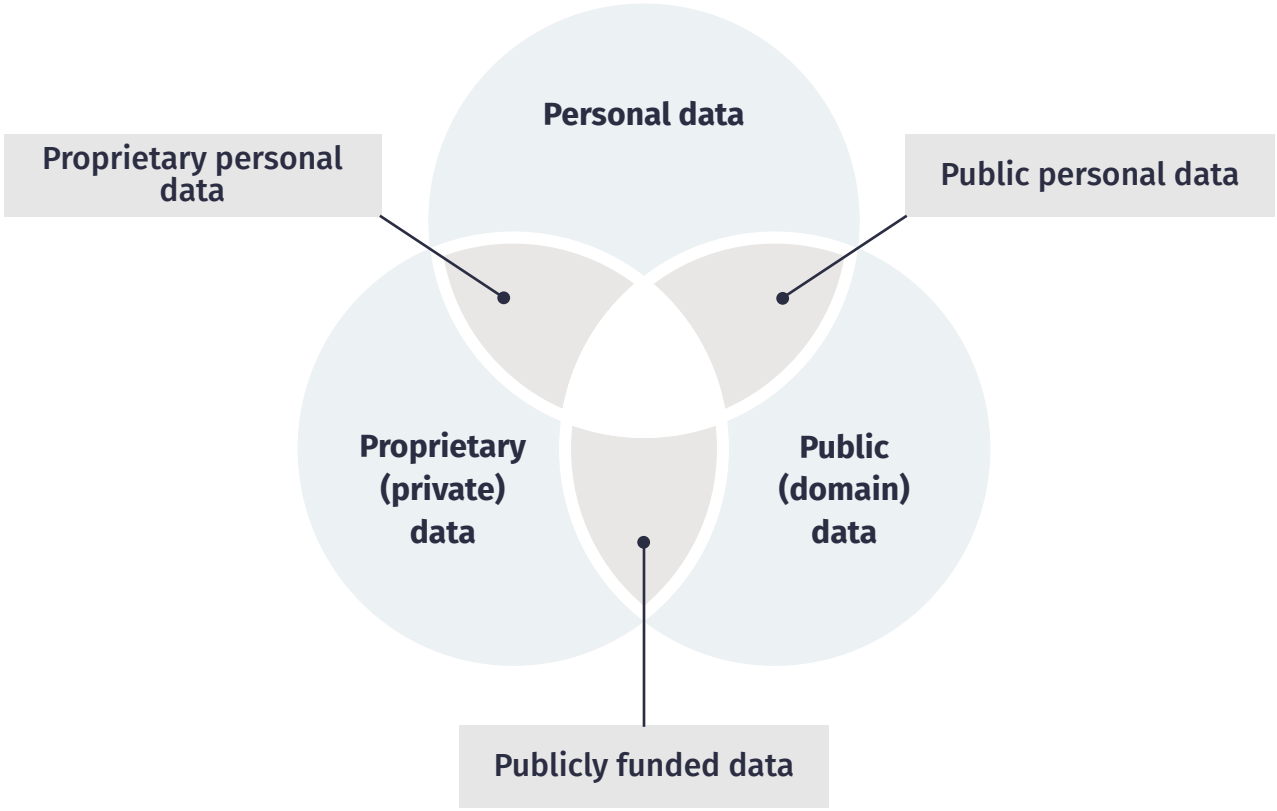


FIGURE 2: DATA DOMAINS

Reproduced from "Enhancing Access to and Sharing of Data" (OECD, 2019). The reproduction of this figure has been authorized.

- The **public domain**, which covers all data not protected by intellectual property rights or any other rights having similar effects. The term "public domain" refers not just to data which is free of copyright protection, but also implies that the data is freely accessible and reusable (public domain data is generally considered to be open data).

It is important to understand that the "public" and "private" qualifiers refer to the data, not to the type of organization that produces or uses them. For example, many private sector organizations are increasingly interested in publishing their data as open data, i.e., in the public domain. Similarly, a public sector organization can produce data from the private domain for its own use (OECD, 2019).

These areas may overlap, as shown in Figure 2. *Publicly funded data* lie at the intersection of the public and private domains and could be created, for example, through public-private partnerships. These overlaps may be caused by potentially conflicting views and interests of parties seeking to create value with their data by sharing it in a partnership (OECD, 2019).

Generally speaking, any dataset must always be carefully examined within its context to determine the legislative and regulatory frameworks that apply. The legal frameworks governing these areas vary, depending on the partnership's applicable jurisdiction, the characteristics of the data in question, and the purpose of their use. These regulations are generally based on privacy and intellectual property laws, which we discussed in the previous section on antecedents.

BOX 11: PERSONAL DATA

Personal data and privacy are inextricably linked. Privacy is considered a cornerstone of our freedom and a constitutionally protected right under the *Canadian Charter of Rights and Freedoms*. The individual right to privacy has been interpreted in Canadian jurisprudence as including protections for "personal privacy, territorial privacy and informational privacy" (*R. v. Jarvis*, 2019 SCC 10, [2019] 1 S.C.R. 488). Information privacy is most relevant to personal data, as it refers to the individual right to control to whom, how much, and for what purpose personal information is disclosed.

Under this regulatory framework, personal information is information about an individual that can be used to identify that individual (Québec Access to Information Commissioner, n.d.). PIPEDA (SC 2000, c 5, s 2[1]) states that "personal information includes any factual or subjective information, recorded or not, about an identifiable individual" which includes:

- age, name, ID numbers, income, ethnic origin, or blood type;
- opinions, evaluations, comments, social status, or disciplinary actions;
- employee files, credit records, loan records, medical records, existence of a dispute between a consumer and a merchant, intentions (for example, to acquire goods or services, or change jobs).

It is worthwhile to note that some data, such as personal data, can easily allow direct identification of individuals, sometimes based on a single piece of data. However, even seemingly non-personal data, can allow the indirect identification of individuals (for example through the correlation of two datasets) and can thus have an impact on privacy.

Generally speaking, personal data enjoy greater protection than data representing environmental or other non-human subjects, as they are the only data that can lead to several privacy violations and prejudice that may result from various inappropriate practices (see Table 2 on page 55).

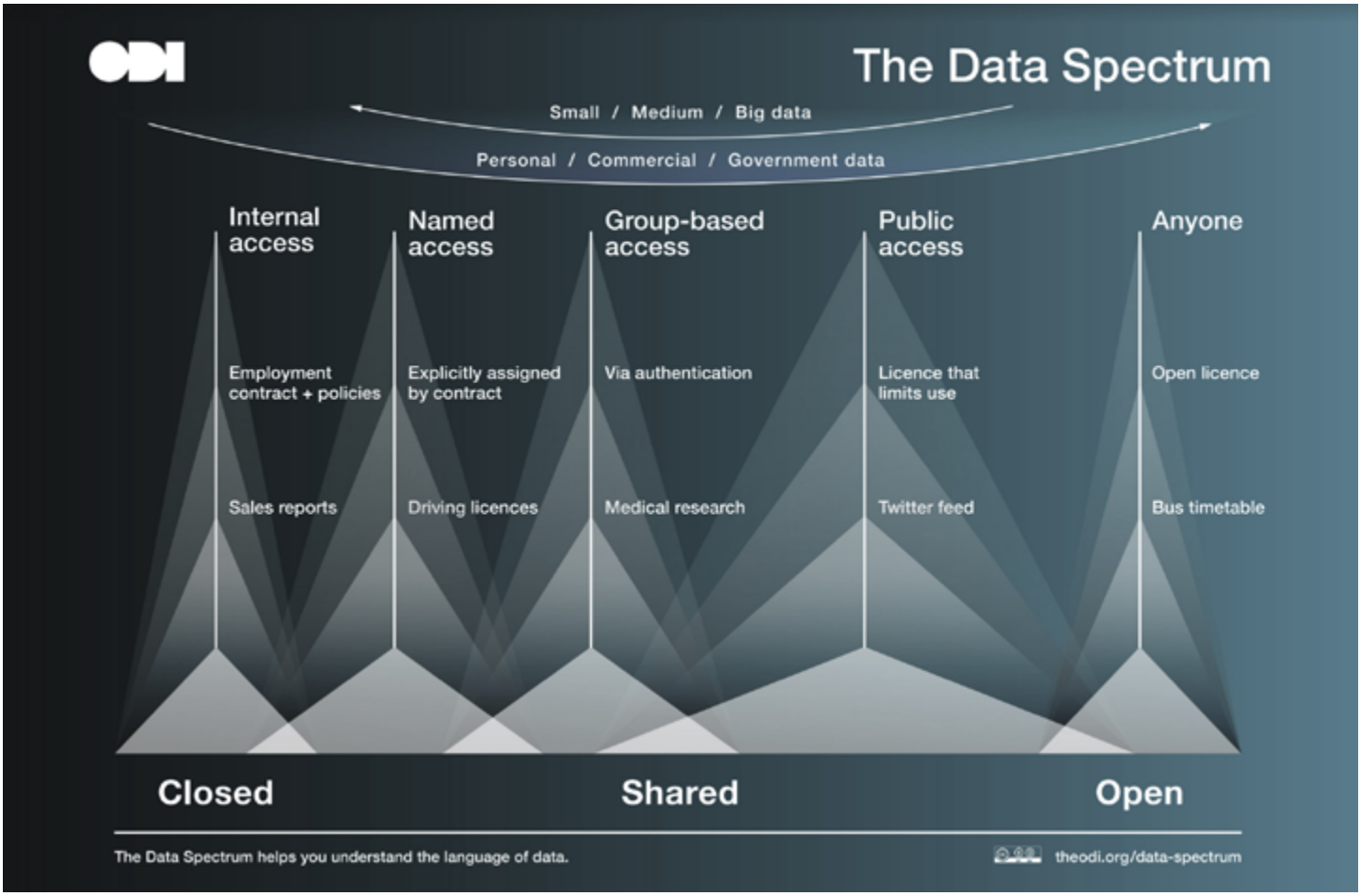


FIGURE 3: THE DATA SPECTRUM

Reproduced from "The Data Spectrum" by the Open Data Institute (n.d.). This figure is subject to a CC BY 4.0 license. It has been modified.

Degree of data openness: closed, shared, or open

Another characteristic of data is its degree of openness. The Open Data Institute (ODI) places the different degrees of a dataset's openness on a scale from *closed* to *shared* to *open* (Figure 3). This degree of openness corresponds to an organization's preferred level of data access. Organizations have the option of granting some rights of use, while reserving others (for example, through a contract or a Creative Commons licence). The legislative framework that applies to a particular data domain can also determine whether a given dataset can be accessed and used.

On the **Closed** side of the spectrum are the data that organizations collect and use internally and which external parties cannot access. Examples of closed data include sensitive information about employees, finances, operations, or trade secrets. On the **Open** side of the spectrum are freely available data, which include data in the public domain or under open licence. Between the **Closed** and **Open** ends of the spectrum is **Shared** data. Unlike open data, these data are available in a *controlled* or *limited* way.

Data sources

When a digital data partnership considers sharing *personal* data, it must examine the data's **source**. As Abrams (2014, p. 1) points out, in our current digital environment, "more and more data originates from observations that are less obvious to the individual and are a product of processing itself." Therefore, the way to address data privacy issues may depend on the individual's level of awareness at the time of data collection.

A data taxonomy is useful for linking a data source to an individual's awareness level. According to Abrams (2014, p. 5), the four main categories are as follows:

- **Freely provided data**, which an individual actively and deliberately shares (for example, by creating a profile on a social network or providing credit card information for online shopping);
- **Observed data**, which are about activities that are captured and recorded. Mobile phone geolocation data and data that describe user behaviours on the Web are some examples;
- **Derived data**, which are generated from other data. They become new data items for a particular individual. One example of derived data is the credit score calculated using an individual's financial history;
- **Inferred data** are generated by analysis or linked to data about an individual. An example of inferred data is an individual's credit score obtained from their observed payment history.

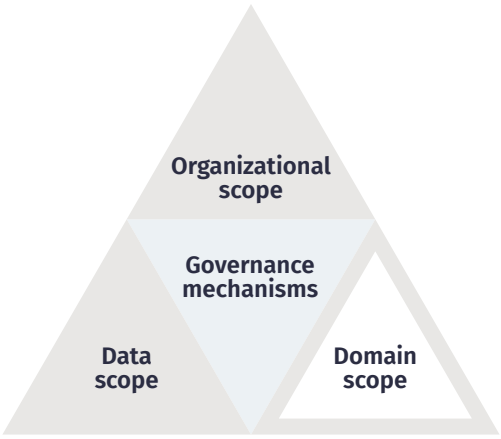
In the context of digital data partnerships, it may be relevant to add a category, as proposed by the OECD (2019):

- **Acquired data** (also known as purchased data or licensed data) are obtained from third parties (such as data brokers) through commercial licensing agreements or non-commercial means (for example, through open government initiatives). As a result, contractual and legal obligations may influence the reuse and sharing of this data.

When someone subscribes for a service, they generally provide their data freely. When they agree to the terms of service, it is accepted that they voluntarily authorize the collection and processing of their personal data. In the case of the other four categories, individuals have little or no opportunity to provide meaningful consent because they may not know that data about them is being collected.

In specific circumstances where consent cannot be obtained, but there is a strong public interest in data collection and use, alternatives to consent or additional measures may be necessary (Jones et al., 2017a). In general, this taxonomy of data types provides a useful starting point for determining not only what data can be shared, but also the data governance mechanisms that will facilitate the flow of data between parties while ensuring that individuals are protected and informed (Abrams, 2014).

The domain scope



EXCERPT FIGURE 1

The final element that helps define the scope of data governance is the **domain scope** (Abraham, Schneider, and vom Brocke, 2019, pp. 431–32). On a day-to-day basis, data management concerns different fields of application. Some are intended to ensure data quality, others exist to set up the digital infrastructures required to store the data, share it, etc. Data governance therefore deals with these different fields of application, in order to achieve its objectives.

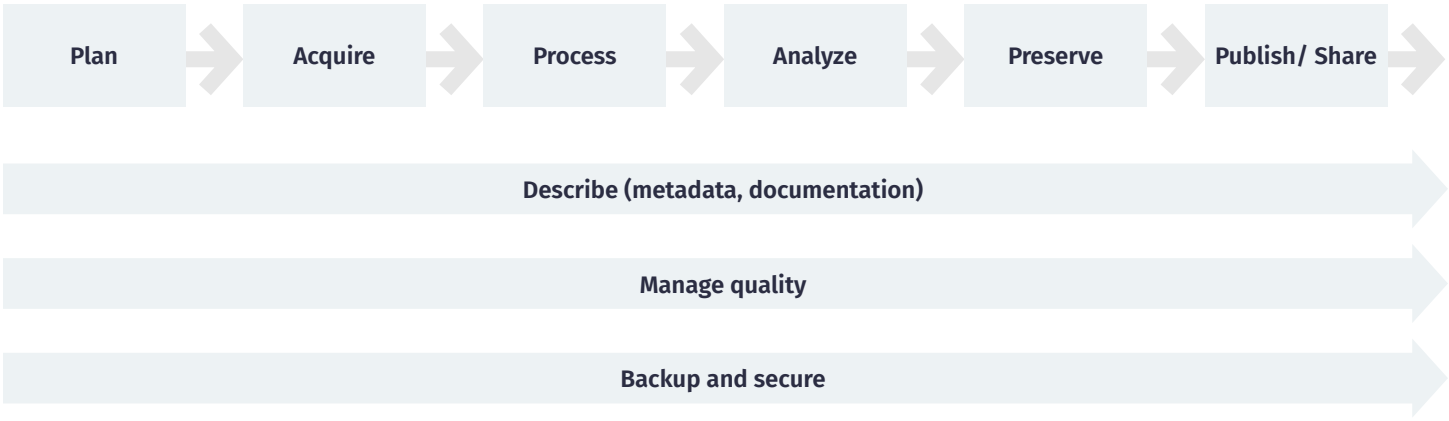
Abraham, Schneider, and vom Brocke (2019) identify six broad fields of application:

- **Data quality:** implementing quality management techniques to measure, assess and improve data quality so they can be used as intended in any given context;
- **Data security:** controlling internal and external access, protecting privacy, and ensuring data authenticity, availability, confidentiality, integrity, and reliability;
- **Data architecture:** developing and maintaining the organizational data management model and plan as well as policies, guidelines, and standards to be followed;
- **Metadata:** documenting and classifying data, flows, and models, as well as other information essential to understanding the data and systems through which they are created, maintained, and accessed;
- **Data storage infrastructure:** providing hardware and software capabilities to meet functionality, reliability, capacity, and other needs;
- **Data lifecycle:** establishing processes and procedures that determine what happens to data, from their collection to their deletion.

The data lifecycle is interesting to consider here, because it can also be used as a data management framework. The lifecycle is typically presented as a diagram that helps visualize the different phases data go through and the transformations they undergo throughout their lifespan. The other fields of application can all be incorporated into this diagram to visualize and understand the various aspects that data governance deals with.

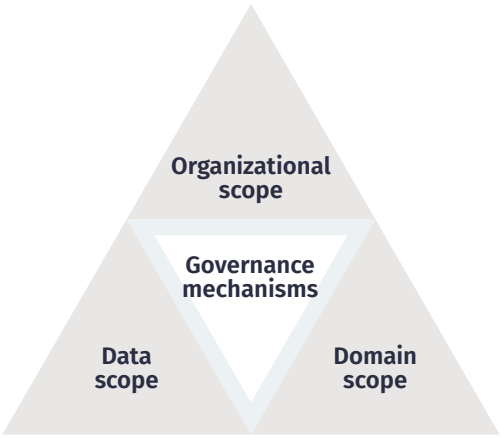
The figure below shows the lifecycle model for science data used by the *United States Geological Survey* and developed by Faundeen et al. (2014, p. 2). This diagram is a good example of a lifecycle that integrates the key fields of application of data governance.

FIGURE 4: SCIENTIFIC DATA LIFECYCLE MODEL



Reproduced from The United States Geological Survey Science Data Lifecycle Model by Faundeen et al. (2014, p. 2).

DATA GOVERNANCE MECHANISMS



EXCERPT FIGURE 1

The organizational scope, domain scope, and data scope together form the scope of data governance. This scope defines and, in turn, determines the data governance mechanisms that must be deployed.

According to Abraham, Schneider, and vom Brocke, governance mechanisms are the f*undamental dimension* of data governance. The authors group them into three categories (2019, pp. 428–430):

- **Structural mechanisms** determine the decision-making bodies, hierarchical structures, and responsibilities in place. They define roles and responsibilities and assign decision-making authority;
- **Procedural mechanisms** govern policies, standards, processes, procedures, contracts, performance measures, compliance monitoring, data strategies, and problem management. This ensures that data are recorded accurately, stored securely, used efficiently, and shared appropriately;
- **Relational mechanisms** are the various practices that facilitate collaboration among stakeholders. They include communication, training, coordination, and decision-making.

These mechanisms operationalize governance and put it into practice in the day-to-day activities of organizations participating in digital data partnerships. For this reason, the next chapter will be fully devoted to exploring concrete examples of data governance mechanisms.

We will examine the key principles and mechanisms that the parties will seek to develop and implement in a digital data partnership. We will then present insights gained through interviews with individuals in the Montréal ecosystem to better understand the data governance risks, tensions, and issues they are currently facing in their work and within their organizations.


HealthMap | Novel Coronavirus (COVID-19)
1,436,794
North Atlantic Ocean
South Atlantic Ocean
Indian Ocean
South Pacific Ocean

CHAPTER 3

THREE GUIDING PRINCIPLES FOR DATA GOVERNANCE IN THE PUBLIC INTEREST AND THEIR PRACTICAL IMPLEMENTATION

As we have seen, there are multiple interrelated aspects of data governance. The prevalent legal framework, the organizations involved, their objectives, their organizational culture, the type of data they share, etc., will all influence the governance framework they will collectively develop to lead a successful digital data partnership. But what exactly is data governance in concrete terms?

This chapter will explore several broad categories of the governance mechanisms that embed data governance on a day-to-day basis.

Before we explore how to operationalize governance in more detail, we must first highlight three key principles under which we have decided to group governance mechanisms. These principles enable organizations to better direct their governance choices toward morally and socially desirable outcomes. They are responsibility, effectiveness, and accountability.

These **principles** are translated into objectives whose impact will influence data governance as a whole.

- **Responsible: Realizing value from data in a responsible and ethical manner**
- **Effective: Managing data effectively and consistently**
- **Accountable: Assessing compliance and impact on an ongoing basis**

These principles are translated into objectives whose impact will influence data governance as a whole.

Responsible: **Realizing value from data in a responsible and ethical manner**

Effective: **Managing data effectively and consistently**

Accountable: **Assessing compliance and impact on an ongoing basis**

RESPONSIBLE

First, to maintain public trust and the legitimacy of their initiatives, digital data partnerships must ensure that all necessary measures are in place **to realize value from data in a responsible and ethical manner**. This starts with the need to comply to legislation and protect the rights of citizens. This principle also recognizes that in our current digital context, where laws do not always keep pace with data collection and the emergence of new technologies, organizations have an additional duty to remain vigilant, assess potential risks, and minimize adverse impacts.

EFFECTIVE

In a spirit of inclusiveness and public empowerment, digital data partnerships must establish governance mechanisms that promote **effective and consistent data management**. The quest for data quality, interoperability, and accessibility makes it easier to share, link, and combine data, which fosters innovation and creates new opportunities that can benefit the public.

ACCOUNTABLE

Finally, accountability is an essential feature of digital data partnerships in the public interest. This principle entails establishing mechanisms to **assess compliance and the impacts** of its decisions throughout the data lifecycle

We believe these three key principles offer a holistic approach to data governance because they recognize that putting in place data governance, particularly data governance which aims to embody ethical practices, requires time and practice.

This chapter is divided into three sections, each dealing with one of these principles. Each section starts by placing the principle in context, explaining its importance and desired objective, and describing the main risks, issues, and challenges associated with it. We then illustrate how these objectives can be achieved through various data governance mechanisms.

We will provide examples from the three categories of mechanisms: structural, procedural, and relational. As we will see, these categories are not mutually exclusive. Several mechanisms may overlap or complement each other to achieve the same objective. Moreover, our selection of data governance mechanisms does not claim to be exhaustive. Different mechanisms may be equally useful or effective, depending on the digital data partnership's circumstances. However, our selections aim to address the main concerns raised by citizens and organizations, as well as those identified in the literature.

Finally, it is important to note that we will not discuss the legal aspects of the proposed mechanisms. As we indicated in the previous chapter, the regulations and laws applicable to digital data partnerships vary by industry, data characteristics, and domain scope. As we will see, some of the mechanism categories described in the following sections are mandated by law, while others are voluntary. For example, informed consent can be a strict and formal requirement when a private company collects personal data. Nevertheless, even without such a requirement, the collection of non-personal data can benefit from informed consent since even data that do not represent individuals can be used to make decisions that affect people's lives (Earley et al., 2017). Compliance with the law is an essential foundation for any responsible and ethical data partnership.

RESPONSIBLE: REALIZING VALUE FROM DATA IN A RESPONSIBLE AND ETHICAL MANNER

Governments, private organizations, and citizens collectively produce and use an increasing amount of digital data in the hopes of identifying new sources of value creation and generating new knowledge to inform decision-making. Personal data is a common commodity, the amounts of open public data are growing, and big data is being used across all sectors of the economy. This data is often subject to advanced automated and algorithmic analysis, and the results are increasingly feeding into public decision-making processes. Data and algorithms offer vast and exciting opportunities to generate information that can improve citizens' lives. However, these opportunities also come with significant ethical challenges.

The harm that can result from negligent data use is well documented and includes the invasion of privacy, loss of individual autonomy and agency, and the presence of interpretive bias and discrimination (Shamsi and Khojaye, 2018; Järvinen et al., 2014; Peña Gangadharan and Niklas, 2019).

The consequences of such harm on individuals are real and can be physical, emotional, or financial. When identifiable data are disclosed in sensitive contexts, it can trigger violence, discrimination, or exclusion. Entire groups can also be harmed (even without individuals being identified) when discriminatory policies are implemented as the result of poor quality data or incorrectly perceived associations in data (The Engine Room, 2016).

Daniel Solove (2006) has identified the potential harms that can affect individuals, groups, or communities if their privacy is breached. Based on this author's work, the following list (see Table 2) shows the magnitude of the risks involved in privacy breaches, as well as the need to implement measures to process and handle data safely, following sound ethical principles and in accordance with legislative frameworks.

The principle of accountability therefore exists to ensure that all data is used in a way that prevents such harm while anticipating and minimizing the risk of initially unforeseen adverse impacts. Of course, responsible and ethical use of data begins first and foremost with compliance with the law and the protection of citizens' rights.

Here we propose to explore five types of governance mechanisms that aim to enshrine responsibility in data governance: the adoption of a **declaration of principles** to define the project's vision and ethical values; **consent**, to recognize the individual's right to decide whether to share their data; **recourse options** and **anonymization**, to protect the rights and privacy of individuals, including privacy by design, and; **risk assessment**, to address legal uncertainty.

TABLE 2: A TAXONOMY OF PRIVACY BREACHES, BY DANIEL SOLOVE (2006)

Domain	Privacy breach	Description
Information collection	Surveillance	Watching, listening to, or recording of an individual's activities
	Interrogation	Various forms of questioning or probing for information
Information processing	Aggregation	The combination of various pieces of data about a person
	Identification	Linking information to particular individuals
	Insecurity	Carelessness in protecting stored information from leaks and improper access
	Secondary use	Use of information collected for one purpose for a different purpose without the data subject's consent
	Exclusion	Failure to allow the data subject to know about the data that others have about her and participate in its handling and use, including being barred from being able to access and correct errors in that data
Information dissemination	Breach of confidentiality	Breaking a promise to keep a person's information confidential
	Disclosure	Revelation of information about a person that impacts the way others judge her character
	Exposure	Revealing another's nudity, grief, or bodily functions
	Increased accessibility	Amplifying the accessibility of information
	Blackmail	Threat to disclose personal information
	Appropriation	The use of the data subject's identity to serve the aims and interests of another
	Distortion	Dissemination of false or misleading information about individuals
Invasion	Intrusion	Invasive acts that disturb one's tranquility or solitude
	Decisional interference	Incursion into the data subject's decisions regarding her private affairs

Reproduced from "A Taxonomy of Privacy" by Daniel Solove (2006).

Governance mechanisms

Declaration of principles

The adoption of a set of principles by stakeholders involved in a digital data partnership, whether in the form of a declaration, manifesto, a project charter or a data sharing agreement, is an effective way to articulate and demonstrate adherence to a set of values or positions on data governance (Coutts and Gagnon-Turcotte, 2020).

For example, even though they have no legal status or force, statements of principle such as Canada's Digital Charter or the recent Digital Data Charter of the City of Montréal demonstrate a public commitment to defend citizens' existing digital rights, as well as introduce new ones. In the field of data, several important declarations guide the actions of digital stakeholders, relating to open data in particular (for instance, the international Open Data Charter) and artificial intelligence (see the Montréal Declaration for a Responsible Development of Artificial Intelligence). Adherence to such texts or the adoption of an internal charter—or principles and values, at the very least—can reinforce project legitimacy and trust between partners, as well as promote the harmonization of their interests.

In addition, developing such statements of principles through stakeholder engagement or even public participation exercises, encourages parties to consider the needs and concerns of a wider public, which can lead to greater social acceptance of the project.

However, it must be emphasized that voluntary compliance with data governance values and standards may be insufficient without integration into a broader regulatory and governance framework (Bennett and Raab, 2018).

Informed consent

Informed consent is a key mechanism for protecting privacy (Policy and Research Group of the Office of the Privacy Commissioner of Canada, 2016). Indeed, "consent functions as a way for individuals to protect their privacy by exercising control over their personal information—what personal information organizations can collect, how they can use it, and to whom they can disclose it" (Policy and Research Group, Office of the Privacy Commissioner of Canada, 2016). The pursuit of informed consent also recognizes individuals' right to decide whether to share data and what data they choose to share.

In order to provide informed consent, individuals must understand the nature, purpose, and consequences of sharing their personal information. According to the *Guidelines for obtaining meaningful consent* published by the Office of the Privacy Commissioner of Canada, organizations seeking informed consent must ensure that individuals understand the key elements that affect their decision. Organizations must, therefore, provide the following information to the individual making a decision regarding their consent:

- Personal information to be collected;
- Third parties to whom the personal information will be disclosed;
- Purposes for which personal information will be collected, used, or disclosed;
- Risk of harm and other consequences.

These elements are a fundamental informational requirement that organizations must meet in order to obtain informed consent when collecting personal data.

However, informed consent poses several challenges. On the one hand, it can be difficult to implement in practice, as many people agree to privacy policies without reading them (Scassa, 2018b). Then, even if a

BOX 12 : PRIVACY PROTECTION

The Office québécois de la langue française (1999) defines privacy protection as the implementation of a set of administrative, technical, and physical measures designed to prevent intrusions into the privacy of individuals or the private affairs of individuals and organizations, which arise specifically from the collection, processing, dissemination, and disclosure of information relating to these individuals or organizations.

person consents to a particular use of their data, it is almost impossible to go back to them after the fact to obtain consent for the 'reuse' of data for purposes other than the original collection, a situation which may arise in a digital data partnership (Curty and Qin, 2014). Furthermore, the idea of limiting data use to a specific purpose is now being challenged by machine learning, which functions by analyzing as much data as possible for purposes that evolve as the data are processed (Gellert, 2016).

The mechanisms for obtaining consent to reuse data have recently garnered attention from authors, particularly those working in the health research sector. Here, we find growing interest in new models, namely dynamic consent and meta-consent.

Dynamic consent allows for the obtainment of consent in multiple phases of data collection and processing (Budin-Ljøsne et al., 2017; Kaye et al., 2015). The dynamic consent process generally offers granular options at different "points of contact." These points of contact permit the use (or reuse) of the same set of personal information with the individuals' informed consent whenever the reasons for the collection, use, or disclosure of these data change

(Budin-Ljøsne et al., 2017). In their report *Trusted Data Sharing Framework*, the Infocomm Media Development Authority of Singapore (IMDA) and Personal Data Protection Commission (PDPC) (2019) suggest that when an organization intends to share data for a purpose other than that for which the consent was obtained, it should inform the individuals in question and highlight any new risks resulting from the secondary data sharing. Of course, individuals must be able to withdraw their consent if they do not agree with secondary data sharing.

Meta-consent, on the other hand, is an approach in which the participants under study are asked to share their preferences "regarding the type(s) and frequency of consent decisions—giving them putative control over precisely how consent will continue to be sought from them on an individualized basis" (Sheehan et al., 2019, p. 227). They may decide, for example, that they prefer to have their data used solely for non-commercial purposes, or even that they prefer not to be contacted at all. However, it remains to be seen whether these new approaches to consent management offer tangible benefits compared to the general consent practices familiar to researchers (Sheehan et al., 2019).

BOX 13 : REUSING DATA

Curty and Qin (2014, p. 1) define data reuse as the "re-analysis of a dataset or a combination of different datasets for the purpose of answering the original research questions with a new method of analysis, or answering new questions based on old data that was not necessarily the focus of the original data collection."

Recourse options

While informed consent is a basic condition for collecting and using personal data while observing an individual's right to privacy, the individual in question may still wish at some point to restrict access to their data, especially if they feel their data has been misused. Providing individuals with recourse options is therefore an important aspect of using data responsibly and ethically. By providing channels individuals can use to submit and resolve complaints or issues concerning their data, a digital data partnership shows it respects the individual's data and their right to determine how it must be used.

Recourse mechanisms can take many forms, ranging from the withdrawal of individual consent, the compensation for misuse of data, and the modification of personal data.

Since the adoption of the European Union's General Data Protection Regulation (GDPR), the recognition of individual digital rights has been gaining ground and the establishment of recourse options in the area of personal information is increasingly recognized as a responsible data governance practice. Article 21 of the GDPR gave individuals the right to object at any time to the processing of their personal data. It effectively enables them to make the organization stop or prohibit the handling of their personal information. Under the GDPR, individuals can file a complaint with their national privacy commission if they believe their data rights have been violated. They can also obtain compensation if a company or organization fails to comply with the GDPR (EU Regulation 2016/679 of the European Parliament and Council, 2016).

Anonymization

In addition to obtaining informed consent and providing recourse options, digital data partnerships should apply a set of procedural mechanisms that protect data confidentiality. One of the techniques most often described in the data governance literature is anonymization. Anonymization is a process that aims to minimize the risk of an individual being identified through data (Elliot et al., 2020, p. 10). Indeed, anonymization techniques can target different levels of data identification. ISO/IEC 19441 (developed to ensure data interoperability and portability in cloud services) identifies five categories for this purpose:

- **Identied data:** Data that can be unambiguously associated with a specific person because personal identifiable information is observable in the information;
- **Pseudonymized data:** Data in which all identifiers are replaced by aliases, where the attribution function ensures that replacements cannot be reversed through reasonable effort by an individual other than the one making them;
- **Unlinked pseudonymized data:** Data in which all identifiers are deleted or replaced with aliases, where the attribution function is erased or irreversible so that links cannot be restored through reasonable effort, including by the entity that performed the operation;
- **Anonymized data:** Unlinked data whose attributes are modified (for example, by randomization or generalization of their values) so that the data alone, or in combination with other data, do not directly or indirectly identify any individual with a reasonable degree of assurance;
- **Aggregate data:** Statistical data that does not contain individual-level entries and is combined from information about enough different persons that individual-level attributes are not identifiable.

As Elliot et al. (2020) point out, the purpose of anonymization is merely to make re-identification more difficult; it is not a foolproof solution. On this matter, the authors note the importance of carefully examining the environment in which data is shared or disseminated (Elliot et al., 2020). This is very important in our current digital environment, as we are witnessing developments in data analysis and artificial intelligence that can link seemingly non-personal information to identified or identifiable individuals with increasing ease (OECD, 2019).

For example, a 2019 study showed that anonymized data can be re-identified and successfully associated with an identifiable individual at a rate of 99.98% using 15 demographic factors (Rocher, Hendrickx, and de Montjoye, 2019). A previous study showed that in a dataset of 1.5 million people collected over 6 months using triangulated location points from mobile towers, 95% of individuals could be uniquely identified based on only four time-stamped geolocation points (de Montjoye et al., 2013).

Ultimately, while the anonymization technique can ultimately enhance privacy by removing identifiable elements from a dataset, significant privacy risks remain even after data are anonymized. The application of anonymization techniques is therefore only one step in the privacy process: other data protection mechanisms must be integrated throughout the data lifecycle (to review the concept of the data lifecycle, see the Chapter 3 Effective: Managing data effectively and consistently).

Privacy by design

One of the approaches to protecting privacy that applies to the entire data lifecycle is privacy by design, which adheres to the notion that "privacy assurance must ideally become an organization's default mode of operation" (Cavoukian, 2009, p. 1). The application of privacy by design principles (see Box 14) aims to cut "across the entire structure of a business or organization, end-to-end, including its information technology, business practices and processes, physical design and networked infrastructure" (Cavoukian and Dixon, 2013, p. 6).

In general, obtaining informed consent and applying anonymization techniques may be considered procedural data governance mechanisms that can protect the confidentiality of personal information. However, these procedures are insufficient on their own. Once one understands that anonymization is not an infallible solution, there arises the need for additional mechanisms to manage risks and provide recourse options in the event of a privacy breach.

BOX 14 : PRINCIPLES OF PRIVACY BY DESIGN

According to Cavoukian and Dixon (2013), the seven principles of privacy by design are:

1. Proactive not Reactive; Preventative not Remedial
2. Privacy as the Default Setting
3. Privacy Embedded into Design
4. Full Functionality — Positive-Sum, not Zero-Sum
5. End-to-End Security — Full Lifecycle Protection
6. Visibility and Transparency — Keep it Open
7. 7. Respect for User Privacy — Keep it User-Centric

Risk assessment

Data protection risk assessment uses calculated and contextual risk analysis tools and methodologies to assess and manage the risks associated with data processing activities planned by digital data partnerships. The purpose of these tools is to calibrate and operationalize organizations' legal obligations, such as privacy protection, that are contained in laws and regulations, based on the actual risks and benefits of the proposed use of data (Centre for Information Policy Leadership, 2014).

Risk assessment seeks to identify upstream threats to personal data or harm that may result from data processing, and trace their causes. According to the Centre for Information Policy Leadership (2014, p. 4), "the question should be whether there is a significant likelihood that an identified threat could lead to a recognised harm with a significant degree of seriousness."

Like privacy by design, the risk assessment approach aims to integrate data protection into an organization's *mode of operation*. This means that the organization adopts practices in which risk management is considered an integral part of the decision-making process, not as a separate technical or legal constraint (OECD, 2019).

Before data collected through a digital data partnership is processed, a risk analysis can help identify any potential harm that could result from the disclosure of personal information. In this process, one would consider questions such as (The Engine Room, 2016, p. 92):

- Are there any individuals or groups who might have an interest in discovering or revealing the identities of the data subjects?
- Could cross-referencing with other available datasets reveal the identities of the subjects of the data being disseminated?
- What would be the consequences of deanonymizing the data on the individuals or groups in question?

In addition to anticipating the risks associated with data processing, the "categorisation of data based on type, value, sensitivity and criticality to the organisation" are essential for risk management (IMDA and PDPC, 2019, p. 45). This data categorization enables the identification of different risk levels and the planning of appropriate security measures. As noted in the section of this report that deals with the data scope, different data types have different legal, regulatory, jurisdictional, and even contractual requirements that must be reviewed, managed, verified, and considered in risk assessment (IMDA and PDPC, 2019, p. 45).

In short, risk assessment and risk management require the establishment of specific structures and processes. One of the most frequently cited risk assessment and management frameworks in the data governance literature is the Fives Safes Framework, originally developed in 2003 by the United Kingdom's Office for National Statistics (see Box 15).

BOX 15 : FIVE SAFES FRAMEWORK

The Five Safes Framework takes a multifaceted approach to managing disclosure risk. Each 'Safe' corresponds to an aspect of the risk of disclosure. The framework asks specific questions to facilitate the assessment and qualitative description of each aspect of risk (or safety). This allows data custodians to implement the appropriate controls, not only over the data, but also over how the data are accessed. This framework is designed to facilitate the safe dissemination of data and avoid excessive regulation (Ritchie, 2017). The five elements of the framework are the following (Ritchie, 2017):

- Safe projects: Is this use of the data appropriate?
- Safe people: Can the researchers be trusted to use it in an appropriate manner?
- Safe settings: Does the access facility limit unauthorised use?
- Safe data: Is there a disclosure risk in the data itself?
- Safe outputs: Are the statistical results non-disclosive?

EFFECTIVE: MANAGING DATA EFFECTIVELY AND CONSISTENTLY

Despite available opportunities, many organizations remain reluctant to participate in digital data partnerships. In Chapter 1, we saw that competitive dynamics, fear of losing control, and competing objectives can create barriers to collaboration. However, one of the barriers to participation most frequently identified by organizations, including those we interviewed for this report, is the cost of participating in such initiatives.

Indeed, the collection, processing, and analysis of large amounts of data can quickly become costly, especially for small organizations with limited technical capability. **The adoption of the effectiveness principle aims to raise stakeholder awareness of the benefits of effective and consistent data management.** As we will see in this chapter, clearly applied mechanisms can be used to structure, share, link, combine, and analyze data more easily. By limiting data preparation and processing efforts, while maximizing the analyses and usable information that are generated, effectiveness increases potential benefits and reduces barriers to participation. Effectiveness is also important from the citizen point of view. After all, citizens are also data users, and access to good quality data can foster innovation and strengthen project transparency and accountability.

In practical terms, the principle of effectiveness aims to foster the collection and production of quality, interoperable data that can be easily accessed by legitimate stakeholders and thereby simplify data processing while reducing the risk of incorrect or discriminatory interpretations.

A lack of data processing rigour poses significant challenges for a digital data partnership, including:

- **Poor quality data:** Inaccurate, incomplete, or obsolete data come with risks because data can be misunderstood and misused, creating serious biases and harm for individuals (Earley et al., 2017);

In practical terms, the principle of effectiveness aims to foster the collection and production of quality, interoperable data that can be easily accessed by legitimate stakeholders and thereby simplify data processing while reducing the risk of incorrect or discriminatory interpretations.

- **Incompatible datasets:** Organizations all have their own ways of using and structuring data, and they often use different softwares to do so. This means it can be difficult to link computer systems to share data (Information Commissioner's Office, 2019). This quality of systems is known as interoperability, defined by IEEE (1990) as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged." At the data level, interoperability is the capacity of two or more datasets to be linked, combined, and processed (Janssen et al., 2014). Low interoperability means the data cannot be combined effectively to conduct useful analyses;
- **Unauthorized access to data:** In order to share data within a partnership, data access privileges granted to the different parties must be clearly defined. Without clear limits and safeguards for protecting access to information, unauthorized entities may access and misuse the data.

To facilitate data sharing within an ecosystem, whether at the territorial, sector, or partnership level, data quality, interoperability, and accessibility are thus important considerations. Indeed, these are levers for unlocking the potential of data that is held or produced collectively. To mitigate risks and minimize potential issues, digital data partnerships must therefore establish clear and consistent mechanisms and ensure that each party understands its role and responsibilities in implementing them.

Governance mechanisms

The data lifecycle

As we saw in Chapter 2, the data lifecycle is a useful management framework that makes it easy to visualize the various stages that data go through in a project, from their collection to their destruction. The lifecycle can be used for early planning and then to structure the data governance on an ongoing basis, to ensure that the required mechanisms are in place at each stage. It also provides a way to reflect on the effectiveness of these mechanisms in a systematic approach that considers their interactions. Therefore, it is important to assess and re-evaluate quality, interoperability, and access management at each stage of the data lifecycle.

Data quality

In the data governance literature, decisions to implement techniques that measure, assess, and improve data quality receive considerable attention and are generally viewed as essential to data governance (Earley et al., 2017; Khatri and Brown, 2010).

This concern for data quality is even more prevalent in literature that focuses specifically on data reuse and sharing (Yoon, 2017; Peer, Green, and Stephenson, 2014; Sposito, 2017). Before using data from other parties (in a data partnership), users must be assured that the data is of high quality. Users generally assess data relevance, reliability, and validity when deciding whether or not to use data from others (Faniel and Jacobsen, 2010). In other words, data quality is an important factor because of its relationship to trust.

Data quality is measured using a variety of criteria, which may vary depending on the context. The DAMA-DMBOK (Earley et al., 2017, pp. 458–459) identifies dozens of aspects of data quality, as well as several ways to group and think about them. According to Earley et al. (2017), the most important aspects of data quality are precision, completeness, consistency, integrity, plausibility, timeliness, uniqueness, and validity (pp. 458–459) (see Box 16).

In practice, to ensure data quality, organizations that share data must establish clear standards tailored to their context and use cases. It is preferable to select and implement these mechanisms collectively, to ensure that all partners adhere to existing rules and take part in holding actors to account in the event of adverse outcomes.

A variety of quality assurance mechanisms can be implemented to achieve the desired outcomes, and each of the three types of mechanisms (structural,

BOX 16 : ASPECTS OF DATA QUALITY

These definitions relating to data quality are proposed by DAMA-DMBOK (Earley et al., 2017, pp. 458–459).

Completeness: The proportion of data stored against the potential for 100%

Uniqueness: No entity instance (thing) will be recorded more than once based upon how that thing is identified

Timeliness: The degree to which data represent reality from the required point in time

Validity: Data is valid if it conforms to the syntax (format, type, range) of its definition

Accuracy: The degree to which data correctly describes the 'real world' object or event being described

Consistency: The absence of difference, when comparing two or more representations of a thing against a definition

procedural and relational) can serve a purpose here. Some useful examples of data quality mechanisms include (Information Commissioner's Office, 2019, p. 28):

- Periodic sampling (can to help ensure that shared data is accurate);
- Templates that outline how data should be recoded (can be used to ensure that data is recorded consistently and coherently);
- Agreement upon shared retention periods and deletion methods (can help to ensure the timeliness of data);
- Training for the people and organisations involved (can be an effective way to reduce data processing errors).

Metadata management

Metadata is data that provides information on other data (Riley, 2017). While this is a relatively simple concept, it plays an important role in data partnerships, because within or between organizations, no individual can have complete knowledge of all data produced or used. As Sebastian-Coleman (2018) explains, "Without reliable Metadata, an organization does not know what data it has, what the data represents, where it originates, how it moves through systems, who has access to it, or what it means for the data to be of high quality. Without Metadata, an organization cannot manage its data as an asset" (pp. 142–143).

In practice, metadata consists of definitions of the different elements of data, including information on their origin and lineage. Metadata may document, for example, the data integration and processing techniques that have been applied to a dataset as it moves through a system. Ideally, metadata gives data users the answers to the following questions (Opendatasoft, 2020, p. 5):

- **What:** What is the dataset about?
- **Who:** Who is the source?
- **Why:** Why does it exist?
- **How:** How should it be used? What rules apply?
- **When:** To what timeframe does it belong?
- **Where:** In what jurisdiction is it located?

As the volume of data used and shared by a partnership grows, it becomes increasingly essential to implement metadata management strategies. Most systems that process data generate metadata automatically, but these systems risk generating metadata that is inconsistent or inaccurate. A deliberate management strategy, however, gives partners a way to agree beforehand on the information they need to know about the data they are using. It thus provides a foundation for them to implement the practices needed to ensure that the information is collated and maintained adequately.

For example, when personal data on gender, ethnicity, or sexual orientation is involved, it is important to ensure that definitions and categorization rules are in place, that they are inclusive, and that all partners have agreed to them (see Box 17 for an example of an inclusive metadata creation process). Periodic quality inspections and audits can also be conducted to ensure that the metadata is up to date.

In addition to improving the understanding of data within organizations, good-quality, well-managed metadata maximizes discoverability and interoperability, by allowing users to easily locate and retrieve datasets (Opendatasoft, 2020). The availability of metadata is also an important aspect of data traceability, a topic that we discuss in more detail in Chapter 3 Accountable: Assessing compliance and impact on an ongoing basis.

BOX 17 : CREATING INCLUSIVE METADATA

CultureBrew.Art (CBA) is a digital platform that promotes and fosters intersectional interculturalism across the Canadian performing arts and media sector. To this end, its primary tool is a database of Indigenous and racialized artists, that companies and agencies can access as subscribers.

The CBA database can be queried using the categories of gender, racial/ethnic heritage, language, artistic discipline, and other areas identified by research data collected through community consultations (Visceral Visions, 2020). The definitions for each of these categories have been collectively determined to ensure their accuracy.

Standards and interoperability

In order for a digital data partnership to combine, compare, or link datasets from a variety of sources, the datasets need to be portable[2], standardized, and interoperable.

These are cornerstones of data sharing, as highlighted by Michal Gal and Daniel Rubinfeld (2019), who emphasize that "data **portability** (the ability to transfer data without affecting its content) and **interoperability** (the ability to integrate two or more datasets) significantly affect the efficient use of data and, resultantly, public and private welfare" (p. 739).

They add that it is through **standardization** that data portability and interoperability can be achieved. According to the authors, "Data standardization may be key to facilitating and improving the use of data, by increasing data portability and interoperability. Indeed, standardization is a precondition for the operation of industries in which cross-firm and cross-industry data exchanges are critical" (Gal and Rubinfeld, 2019, p. 740). In other words, portability and interoperability are achieved by adhering to processing standards that allow data to be easily transferred from one system to another and integrated.

There are a wide variety of standards—over one million national standards and 330,000 international standards according to Michel Girard (2018)—and digital data partnerships may be interested in adopting these pre-existing standards. In terms of digital management, these standards can affect nearly all fields of data application. For example, they allow to "bring clarity on definitions, systems architecture, data ownership, grading, pooling, storage, disposal and set the bar regarding privacy and aggregation requirements" (Girard, 2018, p. 1). Standards may govern the quality of data, establish specific computer formats, or define the kind of information contained in metadata.

These standards can sometimes be internal, i.e., specific to the partnership, but they are generally external technical standards established by independent standards bodies. Many of them have been developed by standards bodies such as the International Organization for Standardization (ISO) (for example, the international ISO/IEC 27001 standard for information security management), industrial consortiums (for example, HTML code), or individual companies (for example, the standardized *General Transit Feed Specification*, a format originally developed by Google). Some standards require users to purchase a licence, while others are open and available free

2 The term data portability refers to the ability to easily move, copy, or transfer data from one computer environment to another in a safe and secure manner without affecting its usefulness (Information Commissioner's Office, n.d.). The term has been popularized by the General Data Protection Regulation (GDPR), which states, in Article 20, the right to data portability.

of charge. Data standards also vary widely from one industry to another. For example, there are well-established standards for publishing open data (Guidoin et al., 2016).

The adoption and application of digital standards aims to improve the ability of parties to link and combine information, which enables the creation of increasingly complex datasets and thus the discovery of new perspectives through linked, pooled, and shared data, etc. (Girard, 2018). However, there are still no standards in some emerging fields. If a digital data partnership feels that no appropriate standards meet its needs, it can develop its own standardization practices. Over time, the partnership can examine whether there is a broader need for standardization in its sector or industry (for example, as organizations join the partnership) and whether time and resources must be invested to create new standards.

In short, identifying the standards and then implementing procedures and guidelines to ensure their application are significant aspects of data governance in digital data partnerships. This may require the creation of tools to support the adoption and use of standards; processes for evaluating, testing, and improving the selected standards; certified products that already meet security and interoperability standards, such as application programming interfaces (APIs); or policies and governance structures to ensure compliance.

BOX 18 : STANDARDS SEEM GREAT... BUT DO THEY REPRESENT EVERYONE?

While there are proven benefits to developing standards that facilitate the exchange of data between two or more systems, we acknowledge that data standards are criticized for prioritizing technical interoperability over human understandings of systems (Brandusescu, Canares, and Fumega, 2020). For example, reflecting on the adaptability of metadata standards to promote Indigenous data sovereignty, Montenegro (2019, p. 736) asserts that, "The most basic assumption regarding any standardization process is that everyone takes equivalent steps to adopt standards and that the standard that is successful for one group of people or institution works for all, or even more egregiously all-encompassing, that the adopted standard works better than any alternative method for documenting and managing information."

In her view, the standardization process ignores the many tensions and needs that must be considered, whether it be between the desire to organize and disseminate knowledge in a systematic way and the ideological inclination in favour of efficiency, or the existence of alternative ways of knowing and generating knowledge, as well as the need for local communities to maintain flexibility in how they document their own knowledge based on their beliefs. These tensions also manifest themselves through different power dynamics, as "different groups in society use knowledge and control of knowledge and its meanings in order to exercise power over other groups" (Battiste, 2008, p. 5).

Brandusescu, Canares, and Fumega (2020) make three recommendations for improving standards design to address these issues: include multiple perspectives in the standard design process; consider the contexts and needs of multiple users when defining standards; and be explicit about roles and relationships during implementation.

Managing data access and privileges

The primary reason for managing access in a digital data partnership is to protect data from unauthorized or abusive disclosure or modification, while ensuring its availability to legitimate users. This data protection requires that "every access to a system and its resources be controlled and that all and only authorized accesses can take place. This process goes under the name of access control" (Samarati and de Vimercati, 2001, p. 137).

Data access management refers to the system of permissions required to access data and carry out any activities related to storing, retrieving, or processing data that is stored in a database (Earley et al., 2017, p. 197). In this regard, the granting of minimal access rights is recognized as a basic security principle. Minimal access rights entails that "a user, process, or program should be allowed to access only the information allowed by its legitimate purpose" (Earley et al., 2017, p. 232).

In addition to controlling access, the permissions system may include various rules that determine and limit how the data can be manipulated. In other words, the system can control the data usage privileges associated with the access permissions.

Data access privileges can be determined based on various criteria, depending on the context and needs of the data partnership. For instance, access can vary in accordance with the **user's function**. For example, by applying *role-based access control*, some users will only be able to consult the database, while others will be able to extract or modify its data (Conrad et al., 2016, pp. 321–2). In this setting, it is common to find that only system administrators are able to permanently erase data. Next, there are other access categories based on **user type** (Samarati and Vimercati, 2001, pp. 139–40). For instance, in the public sector in particular, private companies, academic researchers, and government officials generally do not have access to the same data. Finally, another potential criteria which can determine data access is the **sensitivity** of the data (Kum and Ahalt, 2013).

In the United Kingdom, the Consumer Research Data Centre (CRDC) uses metadata to classify data at three levels of sensitivity: open, safeguarded, or controlled (CRDC, 2020). According to this classification, open data is freely accessible to all. Safeguarded data are non-identifying information that is accessible within the limits set by a licence and the law. Unsurprisingly, sensitive data is controlled, and as such must be kept in the most secure conditions possible, with highly restricted access privileges.

Decisions made around data access and privileges must be supported by various technical and security solutions. Many options and protocols can be considered, ranging from relatively simple processes, such as registration forms which require users to provide their name, to complex access systems, such as a user identity authentication process with role-based access control (for an overview of identity and access management techniques, see Conrad et al., 2016).

Overall, even when parties agree to and establish an access management framework and specific data security mechanisms, there is still a risk that data will be used in an unintended and harmful manner (whether intentionally or unintentionally) (OECD, 2019). This is why accountability mechanisms, which are covered in the following chapter, are essential.

BOX 19 : DATA SHARING AGREEMENTS

In a digital data partnership, where several individuals from different organizations have access to a centralized database, it is important to establish specific access rules, as well as reflect more broadly upon the terms and conditions for use of the data. For example, what happens if an employee violates the existing governance framework? What if data extracted using valid access permissions are nevertheless misused? These data sharing terms and conditions are generally set out in data sharing agreements between the various partners in the initiative.

While there is no prescribed format for data sharing agreements, the most important terms and conditions according to IMDA and PDPC (2019, p. 37) are the following:

1. the grant of the licence/permissions to use the data for the intended purpose;
2. restrictions to the permitted use of the data (if any), such as territorial or time limitations, exclusivity or commercialisation rights;
3. warranties or other assurances provided in relation to the Data Provider's rights in the data;
4. allocation of liability for contract breaches and other liabilities between the parties, as well as indemnification and other remedies when breaches occur;
5. confidentiality;
6. term/duration of the agreement; and
7. governing law and resolving disputes.

ACCOUNTABILITY: ASSESSING COMPLIANCE AND IMPACT ON AN ONGOING BASIS

The final key principle to be pursued by digital data partnerships concerned with protecting the public interest is accountability. Simply put, this means that the parties accept responsibility for their actions; are transparent about how they process, analyze, and use the data; and commit to assessing and responding to both the positive and negative impacts. As with the two previous principles, accountability can only be embodied by establishing concrete governance mechanisms.

The literature on data governance repeatedly highlights the importance of assigning responsibilities for data assets within the organization (Abraham, Schneider, and vom Brocke, 2019; Khatri and Brown, 2010; Otto, 2011). In digital data partnerships, where data crosses organizational boundaries, it is even more important to clearly define roles and responsibilities, knowing that the risks of inappropriate use or exposure (whether deliberate or accidental) increase significantly when multiple users can access the data or when the data are processed automatically (OECD, 2019).

Accountability can begin with internal review processes that aim to ensure data governance meets its intended objectives. Accountability also depends on clear decision-making structures and mechanisms that ensure compliance and pave the way for system auditability. However, accountability also has an external dimension, where the partnership can be called upon to be accountable to the public.

Despite its importance, we will not discuss the dimension of external accountability in greater detail in this section. Few authors in the data governance literature have studied citizen engagement and impact assessment in-depth, which in our view are essential to data sharing projects designed to serve the public interest (and which we have described in more detail in Chapter 1, in the section Foundations of a successful digital data partnership).

We find that these two dimensions – internal and external accountability – come together and mutually reinforce each other. Indeed, efforts to ensure transparency and public engagement must be first supported by clear data governance accountability structures, the establishment of internal processes to ensure compliance, and the documentation of decisions surrounding data throughout its lifecycle.

Governance mechanisms

Clear assignment of responsibilities

The introduction of many actors with diverse interests inevitably creates a need for more complex governance mechanisms (Abraham, Schneider, and vom Brocke, 2019). As The British Academy and The Royal Society (2017, p. 45) explain, "a key challenge for data governance is to find mechanisms for allocating responsibilities across this complex network, so that any fraudulent, unethical, abusive or discriminatory actions can be singled out, corrected and appropriately sanctioned." The existence of a clearly defined decision-making authority is often cited as a determinant of success in data governance, as is the case with digital data partnerships (Earley et al., 2017; Khatri and Brown, 2010; Otto, 2011).

However, there is no infallible model or structure to facilitate decision-making in digital data partnerships. The partnership may choose to depend on data sharing agreements that specify who is responsible for data governance, or to create an external entity, an intermediary, to which all data governance decision-making authority and responsibility is delegated. In both cases, however, it is essential that the responsibilities be clearly assigned.

In a digital data partnership **structured through agreements and contracts**, it may still be necessary to have a decision-making body with clear rules of representation and to deploy and oversee the implementation of a common data governance framework.

However, there is no infallible model or structure to facilitate decision-making in digital data partnerships. The partnership may choose to depend on data sharing agreements that specify who is responsible for data governance, or to create an external entity, an intermediary, to which all data governance decision-making authority and responsibility is delegated. In both cases, however, it is essential that the responsibilities be clearly assigned.

For example, partners may form a joint committee responsible for designing and selecting the protocols and procedures that define how they share data among themselves. Alternatively, they may have an ethics or audit committee responsible for resolving disputes or testing the security of the partnership members' systems.

It is worthwhile to note that when partners use data in the public domain or personal data, an arrangement based on a set of data sharing agreements may be insufficient to address public concerns about the ethical and responsible use of data. Moreover, questions may remain as to how the stakeholders should be held to account in the event of the inappropriate use of data. In this case, some digital data partnerships will prefer to entrust the authority and responsibility for data governance to a **trusted intermediary**, one that is external to all parties.

Such a trusted intermediary can take many forms. The most well known form is the data trust, a concept developed in England under common law (Open Data Institute, 2019). In the Québec legal context, this could be a non-profit organization, a cooperative, or a data protection trust, a new instrument of civil law currently being studied (Marchand, 2019).

Unfortunately, there are still few successful, mature, and well-documented case studies of data governance. As a result, the benefits of using a trusted intermediary, as well as the impacts that this legal form (or others) might have on data governance and the resulting outcomes, are little known to date (Coutts and Gagnon-Turcotte, 2019).

Nevertheless, we can imagine that using a trusted intermediary could help depoliticize and streamline negotiations between partners, particularly with respect to data ownership. If there is asymmetry between the partners in terms of their technical skills or available internal resources, bringing in a trusted intermediary can help level any imbalance in their

respective power of influence. On the other hand, using a trusted intermediary does not completely reduce the risk that the decision-making body will deviate from its members' interests or exercise poor judgment on behalf of its trustees (Porcaro, 2020).

Elements other than the governance structure can also influence if and how the partnership builds public trust. For one, the building of public trust can be determined by the business model used to maintain the financial sustainability of the intermediary. Other elements which can influence public trust include public perception or the data governance itself. In regards to data governance, the terms and conditions of membership, and the degree of individual control retained by the data holders or transferred to the intermediary, can impact the public's trust in the initiative. Therefore, all these elements will need to be carefully evaluated and decided upon by the digital data partnership's members.

Regardless of the governance structure favoured by the partnership, the fact remains that the assignment of decision-making authority over data governance must be clear and transparent in order to build the trust of both the partners and the public. Indeed, "If a nominally representative board is perceived to have no actual power, this could affect the perceived legitimacy of the governance model" (Coutts and Gagnon-Turcotte, 2019, p. 48).

Compliance monitoring

A digital data partnership must ensure that it uses its data in accordance with existing laws and regulations (such as those mentioned in Chapter 2), as well as the procedures and standards established by the partnership itself. Simply put, the partnership's decision-making bodies are required to engage in compliance monitoring.

Compliance can be broadly understood as referring to the parties' objectives, their ethical principles, their legal and contractual obligations, and their duty to the public. Compliance monitoring can be performed by an individual or a specific body, such as an ethics or audit committee.

The **role of data protection officer** is an increasingly common example of an individual performing the function of compliance monitoring. This role was first formalized under the General Data Protection Regulation (GDPR), which requires the designation of a data protection officer (DPO) for the bodies and institutions of the European Union. This officer is an independent, expert data protection authority and plays a central role in ensuring compliance with privacy legislation (European Parliament and Council Regulation [EU] 2016/679, 2016).

The DPO has a number of privacy responsibilities within the organization. According to the European Data Protection Supervisor (n.d.), these include:

- Ensuring that those in charge of processing as well as individuals are informed and aware of their data protection rights, obligations, and responsibilities;
- Providing the institution with advice and recommendations on the interpretation or application of data protection rules;
- Creating a record of data processing within the institution;
- Ensuring compliance with data protection within the institution and supporting the institution's accountability;
- Bringing to the attention of the institution any failure to comply with the applicable data protection rules.

Note that in Québec, private sector organizations may soon have to appoint data protection officers. Although still in its initial phase of passage, Bill 64 (2020, p. 3) proposes to create a data protection function within entreprises. In charge of the protection of personal information within the organization,

this individual would be tasked with ensuring that the parameters of the technological products or services that are used to collect personal information provide the highest level of confidentiality.

In addition to ensuring compliance according to laws and regulations, there may be other aspects to compliance monitoring that are relevant to data partnerships. First, the definition of compliance for all partners involves adopting codes of practice, codes of ethics, data use policies, etc., and then communicating them within the partnership. This often requires monitoring current legal developments and standards to ensure that partnership practices are always up to date. Control measures must then be implemented at specific points in the data lifecycle to assess day-to-day practices (or proposed practices and activities) and to determine if they align with existing policies and rules. Here, one can imagine putting in place periodic audits or security testing to verify compliance with access permissions. Moreover, capacity building, through training and the establishment of communities of practice, may increase the technical competence and digital literacy of partners, while reducing the risks of non-compliance.

A well-recognized practice in this area is to separate out the responsibility for compliance monitoring from the bodies that make policy decisions. This is especially important in public-private partnerships. For example, companies may be tempted to use 'high-value' data (such as human behavioural data), without adequately considering the ethical or legal implications. Creating a separation between decision-making thus ensures that there is a place for debate, reflection and oversight when dealing with contentious questions on how data should be used (Coutts and Gagnon-Turcotte, 2020).

In short, structural data governance mechanisms, such as compliance oversight bodies and DPOs, can play a critical role in digital data partnerships and their embodiment of the principle of accountability. However, in order to truly fulfil their functions, these bodies and individuals must be able to return to the source of the decisions and trace the path that data takes as it travels through systems and across organizations.

BOX 20 : *SAIL DATABANK*

The SAIL Databank is an anonymized repository of health data on the population of Wales, United Kingdom, that covers up to two decades of data and is accessible to researchers. Subject to safeguards and approvals, the data can be linked together to answer research questions.

The independent Information Governance Review Panel (IGRP) oversees access to the SAIL Databank. This committee is made up of representatives from various government agencies and sectors as well as the public. In addition to providing independent guidance and advice on policies, procedures, and processes, the IGRP reviews all SAIL Databank proposals to ensure they are appropriate and in the public interest (Jones et al., 2017b).

Auditability of decisions

The activity of documenting data collection, processing, and analysis decisions is often referred to as auditability (Zook et al., 2017). According to Zook et al. (2017, p. 7) "The goal of auditability is to clearly document when decisions are made and, if necessary, backtrack to an earlier dataset and address the issue at the root (e.g., if strategies for anonymizing data are compromised)."

Tracing decision-making in data systems is not an easy task, but it is nevertheless important for a digital data partnership that aims to respect the principle of accountability.

Ensuring the auditability of decision-making involves documenting not only the decisions made by individuals, but those made by automated systems as well. As more and more private and public organizations use algorithms in automated decision-making, auditability is a key principle being currently emphasized in literature on artificial intelligence and algorithmic transparency (Bertino et al., 2019; Abiteboul et al., 2016; Gasser and Almeida, 2017).

Auditability is closely linked to the notion of data traceability. The latter refers to documentation of the path taken by data through various systems, software and manipulations. In other words, auditability "provides understanding of how it was that the data came to be as it is" (Groth et al., 2008, p. 250). Identifying the source of the data helps create a data audit trail, determine data attribution and ownership, and improve data quality (de Lusignan et al., 2011). Good quality metadata is essential to ensuring data traceability and overall system auditability, especially when tracking data access and how the data has been modified from its original state. Adding this information to internal registers helps ensure that the data passes through the data lifecycle as expected (Allen and Cervo, 2015).

The creation of sticky policies is another approach described in the literature which serves to enhance data control and auditability throughout its lifecycle (Pearson and Casassa-Mont, 2011). Sticky policies are conditions and constraints associated with data that can be read by systems and software and limit how the data is processed. Sticky policies define permitted uses and obligations tied to data when the data is sent from one party to another. This offers users better control over the data. For example, a medical record sent from a hospital to a research institute and then to a research team may contain some variables which are coded (for example, medical results and personal information such as name, address, etc.). In this case, a sticky policy would explain how certain aspects of the medical record can be used (Pearson and Casassa-Mont, 2011).

BOX 21 : ALGORITHMIC REGISTERS

Algorithms are playing an increasingly important role in delivering public services in cities, making it increasingly important for citizens to access information about them and how they are used. On this front, Helsinki and Amsterdam are currently the first two cities in the world to have public registers in which the inner workings of their algorithms are carefully explained. Both registers provide an overview of their system, as well as additional information on the data they use, their operating logic, and application governance. The registers currently have just a handful of algorithms, but nevertheless represent a significant step toward greater transparency and accountability in the use of automated decision-making (Johnson, 2020).

Overall, we find that there are many ways to manage data auditability. Regardless of the method chosen, implementing measures to document how partners choose to collect, produce, process, or access data strengthens accountability overall.

Ultimately, compliance and auditability are but two aspects of accountability in data governance. In a time when initiatives for the common good are garnering greater attention, we hope that other aspects of accountability, such as impact assessment and citizen engagement, will be studied with greater focus in the data governance body of literature and practice.

As we saw in this chapter, a digital data partnership in the public interest should be based on three principles: responsibility, effectiveness, and accountability. A wide variety of mechanisms can be used to translate these principles into practice, whether they are structural, procedural, or relational in nature. In the next chapter we will look at the real experiences of organizations that are interested in data sharing or have participated in data sharing initiatives. We thus draw links between the topics discussed in this chapter and the real-world experiences of Montréal-based organizations in data governance.

CHAPTER 4

MONTRÉAL PERSPECTIVES

One of the main research objectives for this report was to identify key success factors, barriers, and risks associated with digital data partnerships. It was therefore essential for us to hear from people with real experience and interest in these topics.

As a result, during the summer of 2020, we interviewed eight representatives and experts from Montréal-based organizations that are involved in digital data partnerships or have an interest in them. The participants represent a wide range of backgrounds and experiences: some are from the arts and culture sector, while others are directly involved in *Montréal in Common* (Appendix A provides a complete list of the interview participants).

The objective of this research was not to obtain a representative sample — instead, we embarked on these discussions with an exploratory mind, equipped with a semi-structured interview guide. We thus discovered the participants' varying perspectives on data governance in the Montréal ecosystem. While the participants came from different fields and sectors, we found that a number of common themes emerged. These themes and their analyses are the primary focus of this chapter.

Readers will note that many of the topics covered here have been discussed in the preceding chapters. This allowed us to confirm that the orientations of our literature review reflect the real issues that organizations in Montréal are currently experiencing and searching to address.

Different conceptions of data governance

To start, most of the discussions held during the interviews focused on data governance: what it means in the day-to-day work of the interviewees, their organizations' projects, and their sectors.

The discussions with the participants illustrated that there is **no common or single definition of data governance** and that interpretations depend on the sector, the organizations, and the participants' roles within them. When asked what data governance means to them, many perspectives emerged, such as:

- Internal data management frameworks and processes that ensure data quality or control access to information;
- Common frameworks and processes for pooling or sharing data among multiple partners;
- Policies, procedures, and security measures to ensure legislative compliance and protect data confidentiality;
- The ability of a set of actors to adhere to a set of data use rules and make collective decisions.

We found that some participants (such as those occupying data management roles within their organizations) highlighted more traditional or corporate aspects of data governance: data management and compliance with legislation and internal policies. Others who had experience with digital data partnerships had a much broader view of data governance and described it in terms of frameworks and processes that enable multiple organizations to share data and make collective decisions about its use.

Data culture and organizational capacity

All the participants stressed that the organization's data culture is an important factor when participating in multi-stakeholder data sharing initiatives. This data use culture is influenced by many factors, such as the organization's industry, sector, size, technology choices, and the attitude of its employees toward data.

In particular, some participants saw data culture as a reflection of organizational culture. This was the case in our conversation with Jean-Sébastien Bélanger, Head of Membership and Customer Service of the Montréal Museum of Fine Arts, who emphasized the importance of reducing data silos within the organization to ensure adequate access to information across departments and to create more value from data.

> *"Here at the museum, we've decompartmentalized data access [...] Of course, there can be extremely closed off silo [...] I'll give you some examples: museums have management software for their foundations, they have ticket office management software, with data that are generated differently each time [...] In short, you're collecting all kinds of data, and it quickly becomes an issue. We don't have this issue, because we made the decision not to go down this path [...] Due to this known issue with museum data, we decided to use just one (software application), because we wanted data that could really be used by everyone, and have everyone understand the structure, and then know how to query it, and how to use it."*

Moreover, our discussions led us to conclude that organizational culture is an important factor in creating the conditions that enable successful digital data partnerships. Audray Fontaine, Knowledge Transfer Coordinator, Centre for Interdisciplinary Research on Montréal, explained that the *Montréal in Common* social data hub project aims to share and overlay data from multiple official and unofficial sources in order to "better inform decision-making by the City and their partners on various social issues." This will be difficult to achieve, as data culture varies across organizations, meaning that the organizations involved in the project may be less willing or less open to sharing their data because of the way they value and perceive data.

In addition, some participants made connections between data culture and the organization's understanding of its own data. For example, interviews with actors hailing from the arts and culture sector emphasized that before participating in a data pooling project, organizations must better understand their data assets (i.e. the characteristics of the data available to them, safeguards to be taken, etc.). Internal disagreements may occasionally arise regarding the very definition of what data represent for the organization. For this reason, developing a shared understanding of this definition across the organization was seen as an important first step in discovering the value of data in the context of a partnership.

> **We have a lot of questions. For example, if we share our data, even for the noblest cause, isn't there always a risk of error, of a data leak? What safeguards or remedies are there? And when a data breach occurs, say for usage data, what happens? These are very important issues being raised. Especially when you consider that usage data doesn't belong to us. Technically speaking, contact information belongs to the person who provided it. Given that, can we actually share this information?**
>
> **- Anastasia Vaillancourt,** Director of Development, Culture pour tous

Complex issues surrounding data sharing which require clarification

Given that personal information, privacy, and data ownership are matters of significant debate in the data governance literature, it is no surprise that these topics were the focus of discussions in the interviews. Due to their complexity, these topics were identified as barriers to participation in digital data partnerships.

In this regard, participants stressed the *dynamic nature of personal data*. For example, Sophie Tremblay, Lawyer and Chief Operating Officer, Novalex, who has provided legal advice on data pooling projects in the arts and culture sector, emphasized that while some data may not easily identify individuals, their context and how they relate to other data can create conditions for identifying individuals.

This observation has broad implications for digital data partnerships. For example, one critical step for organizations involved in data pooling projects is to identify the legislative frameworks that apply to their data and the safeguards that must be implemented to ensure data confidentiality. Organizations may need legal support to help them clarify these distinctions.

Interviewees also acknowledged that sharing data in a partnership raises questions and concerns for organizations that may go beyond concerns related to data breaches. For example, in discussing her organization's participation in a data pooling project, Anastasia Vaillancourt, Director of Development, Culture pour tous, said that there are many unanswered questions. In her words:

> *"We have a lot of questions. For example, if we share our data, even for the noblest cause, isn't there always a risk of error, of a data leak? What safeguards or remedies are there? And when a data breach occurs, say for usage data, what happens? These are very important issues being raised. Especially when you consider that usage data doesn't belong to us. Technically speaking, contact information belongs to the person who provided it. Given that, can we actually share this information?"*

She noted that the risks to data confidentiality and security become more complex when the ownership of data is unclear. This lack of clarity can be a barrier

to participating in data partnerships, as it makes it difficult to assign responsibility for the data and any outcomes derived from it.

Determining who owns the data is a complex task, requiring not only the interpretation of various laws, but also an understanding of individual attitudes towards data. Our discussion with Frédéric Julien, Director of Research and Development for the Canadian Association for the Performing Arts (CAPACOA) highlighted the following attitudinal aspects of data ownership:

> *"Even in the case of information that is otherwise freely available on the web, as soon as there is any coding effort, the individual who enters their data in a system feels that the data belongs to them, and we can debate whether the data belongs to them or not... Coding (information) leads to a sense of ownership over the data. And ownership tends to shut people down and block initiatives that would otherwise facilitate data reuse. So there's a non-negligible attitudinal element to this."*

Underestimation of the cost and value of data

In general, a number of data collection and clean-up steps may be required before data can be shared and pooled. The participants noted that these internal data management steps and processes, which aim to ensure high data quality, can be time-consuming and call for significant investments in human and technical resources.

For example, for Patrick Joly, Managing Director, Société de gestion de la Banque de titres de langue française (BTLF), the costs of collecting, cleaning, and maintaining good quality data and metadata tend to be underestimated. His organization has partially assumed these costs and the corresponding responsibility, as part of its role as a data aggregator in the book-publishing industry.

Even in the case of information that is otherwise freely available on the web, as soon as there is any coding effort, the individual who enters their data in a system feels that the data belongs to them, and we can debate whether the data belongs to them or not... Coding (information) leads to a sense of ownership over the data. And ownership tends to shut people down and block initiatives that would otherwise facilitate data reuse. So there's a non-negligible attitudinal element to this.

- Frédéric Julien, Director of Research and Development, Canadian Association for the Performing Arts (CAPACOA)

Recognizing inconsistencies in data formatting and silos in data and metadata production methods within the industry, BTLF has recently established a policy to guide organizations in structuring and formatting commercial book data. Ultimately, the organization hopes to improve the quality of data of all stakeholders in the value chain, in order to support better business intelligence.

Indeed, for interview participants like Frédéric Julien, the costs of creating and maintaining data in a competitive environment can become a barrier to participation in digital data partnerships. In other words, whenever an organization invests a lot of time and resources to maintain its data, it may be less willing to share it freely.

Linked data and semantic interoperability

The topics of linked data and semantic interoperability were a common thread in several interviews. As discussed in Chapter 3, semantic interoperability addresses "shared vocabularies and common language using common models, attributes and definitions, with outputs like: registers, taxonomies, vocabularies and ontologies" (Open Data Institute, 2018).

Linking data is a practical way for achieving semantic interoperability, as it enables data publishers to support data discovery and integration applications (Schmachtenberg, Bizer, and Paulheim, 2014). It entails linking data elements to a controlled and shared vocabulary that links human-readable Web content to machine-readable metadata.

In the arts and culture sector, CAPACOA is currently leading a series of linked open data initiatives to support digital discovery in its field. For instance, CAPACOA and the Conseil québécois du théâtre are currently leading an initiative to increase the presence of the performing arts through Wikidata—an online tool that can serve as a common source of linked data about people and places, historical events, socio-economic conditions, and culture (Marino and Neto Costa, 2020).

Linked open data can also be a useful tool for *Montréal in Common*. For example, FabMob QC, an organization leading efforts in the field of mobility data, has a long history of documenting and sharing its projects using semantic Web tools. It also manages a Wikidata site that aims to "capitalize all projects, feedback, and mistakes and create a common culture of innovation in action." Elsa Bruyère, Co-founder, FabMob QC, highlighted the potential of these semantic tools to help disseminate the activities and results of *Montréal in Common* while avoiding data silos:

> *"If we at least have the same semantics and same semantic format, for example by applying the Resource Description Framework (RDF), someone making a request could find all our results as part of the challenge, without having to worry about which platform they're going to search on, which website they're going to search on. From one of these websites, they could go back to other things. So that would allow us to have a much broader cross-referencing than what we have today. Because otherwise, we might end up in site silos."*

The importance of linked data is measured not only by its role in disseminating the results of *Montréal in Common*. Broader links may be established between semantic interoperability and smart cities. On this topic, Frédéric Julien noted that:

> *"The smart city is not just about having video cameras and sensors all over the city. For me, beyond the technical acquisition of the data, in order for a city to be truly intelligent, that data cannot exist in silos . [...] It requires both semantic and technical interoperability so that the user who needs the data can access them in a timely manner.*

> **"The smart city is not just about having video cameras and sensors all over the city. For me, beyond the technical acquisition of the data, in order for a city to be truly intelligent, that data cannot exist in silos . [...] It requires both semantic and technical interoperability so that the user who needs the data can access them in a timely manner. To be really smart, the smart city will have to be decentralized. It can't be done otherwise."**
>
> - **Frédéric Julien,** Director of Research and Development, Canadian Association for the Performing Arts (CAPACOA)

To be really smart, the smart city will have to be decentralized. It can't be done otherwise."

Growing interest in digital data partnerships

While digital data partnerships raise complex legal, ethical, and operational issues, the participants expressed an ongoing interest in exploring various forms of collaborative digital data partnerships that generate value and enable the pursuit of goals in the public interest.

In particular, data pooling and decentralized approaches to data sharing (such as linked open data) are recognized as promising, particularly for interviewees active in the arts and culture space. Indeed, the participants selected a number of different initiatives as sources of inspiration. These include pilot projects to pool data which have been spearheaded by Synapse C, or the notion of a "social utility trust" that is currently being studied by Territoires innovants en économie sociale et solidaire (Marchand, 2019).

Montréal in Common is also paving the way for the exploration of new digital data partnerships in the public interest. Various collaborative projects are emerging in the food system and mobility data spaces that will require adapted models of data governance. For example, Récolte is launching a project to create a shared food systems infrastructure that may eventually link data from various sources to track equipment and food products across the logistics chain. In addition, the social data hub is exploring the creation of a secure environment for data sharing between public agencies. While *Montréal in Common* partners are testing and experimenting new approaches, the participants expressed a strong interest in and concern about how they will ensure the longevity of

their initiatives beyond the initial funding provided by the program. The participants also stated an interest in learning more about the potential of data governance and open data business models, which can lead to benefits in the public interest.

Other success factors for digital data partnerships

In addition to confirming interest in digital data partnerships, the interviews with the participants identified potential success factors for these initiatives, including:

- It is essential to ensure that all stakeholders are aligned and have a **shared vision** for the digital data partnership and its objectives. Recognizing that the organizations involved in such partnerships may have different data cultures, the participants felt that the **adoption of a common vocabulary** by the parties was key to the initiative's success;
- The parties must also develop a **common understanding** of what it means to participate in the digital data partnership and to share the **benefits** that can be generated. The participants, especially those involved in data pooling projects led by Synapse C, mentioned that the potential benefits include producing industry-wide insights and knowledge that enable organizations to make better, evidence-based decisions;
- Organizations must develop the **internal capacity** to manage and understand their data before participating in digital data partnerships. The presence of a **data champion** within the organization can help strengthen governance and data culture, as well as foster engagement in data partnerships.
- Several participants noted that the public currently has very limited trust in our institutions' ability to properly manage and protect our data. This low level of public trust is likely exacerbated by recent events and ongoing public debates, including the Desjardins data leak (Benessaieh, 2020), and the current promotion of contact tracing applications to curb the spread of COVID-19, which raises concerns about government surveillance (Canadian Press, 2020).

It therefore appears that **public trust** will continue to be a critical factor in the success of digital data partnerships. To that end, citizen involvement in initiatives, public awareness and education, along with transparency measures (e.g. public reporting and impact assessment activities), will be the key to success of any future initiative.

- Finally, it is preferable that digital data partnerships be led and supported by a dedicated governance body (such as a working group or steering committee) and supported by expert technical and legal resources, especially when partnerships involve actors with limited internal capacity. We found that participants who had experience in more advanced data pooling projects succeeded in adopting reference policies and establishing agreements to define standards and the terms and conditions for data use.

The interviews ultimately covered a range of topics, from understanding data governance to semantic interoperability and linked data. While we were unable to use the results of the interviews to validate the individual findings of our literature review, our findings emphasized the complexity of issues at hand, as well as the risks and challenges associated with data governance, and how these play out in the daily work of organizations in Montréal.

CONCLUSION

Implementing data governance is not an easy task. It requires decision-making by partners involved in various areas, including privacy, access management, risk assessment, data quality, and more. Data governance within a digital data partnership is a complex and ongoing process that requires negotiation and compromise to align diverse objectives and foster collective decision-making and collaboration, while assigning responsibilities appropriately in areas with both human and technical components.

Our research shows that digital data partnerships are useful vehicles for putting technology and data to work for the public. They have emerged in a rapidly evolving context, which means that there is no single method of organizing their governance. The combination of structures, processes, and relational mechanisms necessary for responsible and effective data governance is, in fact, dependent on multiple factors and the existing conditions in which they exist. Therefore, the parties must make decisions on the most appropriate mechanisms and their implementation according to their context, needs, and objectives. Our research has shown that a partnership which values collaboration will use co-creation and agile approaches to achieve their vision.

Overall, we are seeing more and more initiatives that seek to generate societal benefits by sharing data, while recognizing data's public value and the opportunity to manage it on behalf of a group or collective. Data partnership initiatives in the public interest stand out among these, as partners commit to creating tangible benefits for the public, deploying citizen engagement strategies, and adhering to strong data governance principles.

In this regard, digital data partnerships ensure the preservation of the common good based on the three principles of responsibility, effectiveness and accountability. First, to ensure public trust and the legitimacy of their initiatives, digital data partnerships must implement all necessary measures to process data responsibly and ethically. Digital data partnerships must then establish governance mechanisms that promote effective and consistent data management. Finally, we characterize accountability as a variety of mechanisms through which there is clear and transparent decision-making in regards to stakeholders and the public.

Through a series of interviews, we validated how data governance fits into Montréal organizations' day-to-day workings and projects. The individuals we interviewed highlighted the barriers and challenges they face surrounding data, namely concerning data privacy, ownership, quality, and interoperability. They showed that to succeed, digital data partnerships need dedicated support and expertise, as well as strong leadership and a decision-making body.

Overall, our research shows that building digital data partnerships that aim to serve public policy objectives is of considerable interest, despite the aforementioned challenges. As governments, the private sector, universities, non-profit organizations, and charities continue to explore new digital data partnerships in Montréal and elsewhere, we conclude this report with some key findings from our research, which we hope will contribute to their future success.

1. Recognize that the public interest is defined and negotiated by citizens

The public interest is a living thing that exists in a state of constant deliberation and negotiation. As such, it can be challenging to identify. In some cases, a clear consensus on the nature of the public interest may emerge. However, when new social or technological issues are involved—such as the use of artificial intelligence—the public interest may be less clear.

A digital data partnership that integrates ongoing and sustained public participation in its governance is more likely to perceive public opinion changes and ensure that the partnership's objectives align with

citizens' needs. Also, a digital data partnership will be more credible and more likely to create clear benefits for the public if it develops its objectives in collaboration with a wide range of voices, including women, Indigenous communities, recent immigrants, low-income residents, among others.

2. Invest time in your collaboration and experimentation processes

Building a strong data partnership requires time and sustained effort from all parties. Partners must make an initial investment to build trust, demonstrate a willingness to work together, develop a common understanding of the issues they search to address, and rally around a set of common goals.

This cycle of collaboration also encourages a willingness to experiment, develop new pilot projects, and harness their impact to create new ways to realize value from data. The development of use cases is a useful way to start exploring the potential of the data to be shared, and to ultimately determine what benefits can be generated from a partnership.

3. Create data governance that is tailored to your needs

Various data governance 'models' such as data trusts or data collaboratives promise to be the solution for the ethical, effective, and responsible use of data. But in reality, there is no standard model for structuring data partnerships. Even if a partnership is interested in one of these models, it takes time and effort to implement the structures, procedures, and relationships that will best suit its context. Determining the appropriate mechanisms for a digital data partnership depends on understanding several factors such as the organizational scope, the data scope, and the domain scope. The good news is that the effort invested will be rewarded with a data governance framework that is tailored to the partners' needs.

Despite the absence of a turnkey model, there are promising legal frameworks in the context of digital data partnerships. For example, the social utility trust, a legal vehicle unique to Québec civil law, addresses a desire for common ownership and collective governance (Marchand, 2019). This legal framework is currently attracting the interest of many researchers. It may be the subject of future experimentation, despite the many questions it continues to raise, in particular, whether the data can constitute a form of property that could benefit from a social utility trust designation.

This example shows that digital data partnerships cannot be developed in isolation. Several data governance requirements are already defined in existing legislation, regulations, or standards. Among others, Québec and Canadian privacy laws establish a framework from which data governance cannot depart. For many, these laws are outdated and must be reviewed by legislators. This gap nevertheless offers ambitious organizations a space for innovation.

4. Document your impact and share your successes

It is beneficial to know the extent to which the digital data partnership has achieved its goals and vision. First, monitoring and communicating progress can help to ensure that local stakeholders and the public remain committed to the initiative. Second, documenting the starting conditions, decisions taken, challenges encountered, and lessons learned will help partners adapt and improve their initiative over time. Third, by telling the project's story, the partners will make an invaluable contribution to local, regional, and global communities of practice. Research on what works best for digital data partnerships is still in its infancy, which shows the importance of documenting the use cases, successes, failures, and impacts of these initiatives – and to ensure that we can continue to build on these successes.

APPENDIX

List of Interviewees

- Sophie Tremblay, Lawyer and Chief Operating Officer, Novalex
- Jean-Sébastien Bélanger, Head of Membership and Customer Service, Montréal Museum of Fine Arts
- Patrick Joly, Managing Director, Société de gestion de la Banque de titres de langue française
- Anastasia Vaillancourt, Director of Development, Culture pour tous
- Frédéric Julien, Director of Research and Development, Canadian Association for the Performing Arts
- Elsa Bruyère, Co-founder, FabMob QC
- Audray Fontaine, Knowledge Transfer Coordinator, Centre for Interdisciplinary Research on Montréal
- Lorenzo Daïeff, Project Manager, SALIM – Smart Cities Challenge, Récolte

REFERENCES

Abiteboul, S., Miklau, G., Stoyanovich, J., & Weikum, G. (2016). Data, Responsibly. *Dagstuhl Reports, 6*(7), 73.

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Abrams, M. (2014). The Origins of Personal Data and its Implications for Governance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2510927

Allen, M., & Cervo, D. (2015). *Multi-domain master data management: Advanced MDM and data governance in practice*. Morgan Kaufmann.

Ana Brandusescu, Michael Canares, & Silvana Fumega. (2020, August 21). Open data standards design behind closed doors? *ILDA*. https://idatosabiertos.org/diseno-de-estandares-de-datos-abiertos-a-puertas-cerradas/

Ansell, C., & Gash, A. (2007). Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory, 18*(4), 543–571. https://doi.org/10.1093/jopart/mum032

Artyushina, A. (2020). Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics, 55*, 101456. https://doi.org/10.1016/j.tele.2020.101456

Bass, T., & Old, R. (2020). *Common Knowledge: Citizen-led data governance for better cities* (DECODE). Nesta. https://www.nesta.org.uk/report/common-knowledge-citizen-led-data-governance-better-cities/

Bass, T., Sutherland, E., & Symons, T. (2018). *Reclaiming the Smart City: Personal data, trust and the new commons*. https://decodeproject.eu/publications/reclaiming-smart-city-personal-data-trust-and-new-commons

Battiste, M. (2008). Research Ethics for Protecting Indigenous Knowledge and Heritage: Institutional and Researcher Responsibilities. In *Handbook of Critical and Indigenous Methodologies*. SAGE Publications, Inc. https://doi.org/10.4135/9781483385686

Benessaieh, K. (2020, June 19). *Vol de données chez Desjardins: La catastrophe, un an plus tard*. La Presse. https://www.lapresse.ca/affaires/entreprises/2020-06-19/vol-de-donnees-chez-desjardins-la-catastrophe-un-an-plus-tard

Bennett, C. J., & Raab, C. D. (2018). Revisiting the governance of privacy: Contemporary policy instruments in global perspective. *Regulation & Governance*. https://doi.org/10.1111/rego.12222

Bertino, E., Merrill, S., Nesen, A., & Utz, C. (2019). Redefining Data Transparency: A Multidimensional Approach. *Computer, 52*(1), 16–26. https://doi.org/10.1109/MC.2018.2890190

Bhargava, R., & D'Ignazio, C. (2015, June 30). *Designing Tools and Activities for Data Literacy Learners*. Data Literacy Workshop at ACM Web Science Conference, Oxford, UK. http://www.dataliteracy.eita.org.br/wp-content/uploads/2015/02/Designing-Tools-and-Activities-for-Data-Literacy-Learners.pdf

Bill 64, An Act to modernize legislative provisions as regards the protection of personal information, 1st sess, 42th leg, Québec, 2020. http://www.assnat.qc.ca/Media/Process.aspx?MediaId=ANQ.Vigie.Bll.DocumentGenerique_159567&process=Default&token=ZyMoxNwUn8ikQ+TRKYwPCjWrKwg+vIv9rjij7p3xLGTZDmLVSmJLoqe/vG7/YWzz

Bolychevsky, I., Ruhaak, A., Bunting, M., McMillan, A., Cameron, S., Voznick, M., & Pasquarelli, W. (2019). *Exploring the potential of data trusts in reducing food waste*. Open Data Institute. https://docs.google.com/document/d/1v9O3exRdZFu6h-xqo11E-j3Ul4cyePfRCYYBDrpUoNBk/

Budin-Ljøsne, I., Teare, H. J. A., Kaye, J., Beck, S., Bentzen, H. B., Caenazzo, L., Collett, C., D'Abramo, F., Felzmann, H., Finlay, T., Javaid, M. K., Jones, E., Katić, V., Simpson, A., & Mascalzoni, D. (2017). Dynamic Consent: A potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics, 18*(1), 4. https://doi.org/10.1186/s12910-016-0162-9

Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X., & Greene, C. S. (2020). Responsible, practical genomic data sharing that accelerates research. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-020-0257-5

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri, 63*(2). https://doi.org/10.1515/libri-2013-0010

Cavoukian, A. (2009). *Privacy by Design: The 7 Foundational Principles*. https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

Cavoukian, A., & Dixon, M. (2013). *Privacy and Security by Design: An Enterprise Architecture Approach* (Open-File Report). Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-privacy-and-security-by-design-oracle.pdf

Centre for Information Policy Leadership. (2014). *A Risk-based Approach to Privacy: Improving Effectiveness in Practice*. https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_1-a_risk_based_approach_to_privacy_improving_effectiveness_in_practice.pdf

City of Montréal. (2020). *Digital Data Charter*. https://laburbain.montreal.ca/sites/villeintelligente.montreal.ca/files/25817-charte_donnees_numeriques_ang.pdf

Conrad, E., Misenar, S., & Feldman, J. (2016). *CISSP study guide* (Third edition). Elsevier, Syngress.

Consumer Data Research Centre. (2020). *Protecting Data*. https://data.cdrc.ac.uk/protecting-data

Coutts, S., & Gagnon-Turcotte, S. (2020). *Data Governance and Digital Infrastructure: Analysis and Key Considerations for the City of Toronto*. Open North. https://www.toronto.ca/wp-content/uploads/2020/08/95fb-2020-07-10-Open-North-Data-Governance-Report-Main-report-WEB.pdf

Curty, R. G., & Qin, J. (2014). Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1–4. https://doi.org/10.1002/meet.2014.14505101072

D'Addario, J., Dodds, L., Brown, W., & Maddison, J. (2020). *Sharing data to create value in the private sector*. Open Data Institute. https://theodi.org/article/report-sharing-data-to-create-value-in-the-private-sector/

de Lusignan, S., Liaw, S.-T., Krause, P., Curcin, V., Vicente, M. T., Michalakidis, G., Agreus, L., Leysen, P., Shaw, N., & Mendis, K. (2011). Key concepts to assess the readiness of data for international research: Data quality, lineage and provenance, extraction and processing errors, traceability, and curation. Contribution of the IMIA Primary Health Care Informatics Working Group. *Yearbook of Medical Informatics, 6*, 112–120.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports, 3*(1), 1376. https://doi.org/10.1038/srep01376

Du Perron, S. (2020a, June 17). Projet de loi 64: Une réforme à l'Européenne du droit à la protection des renseignements personnels. *Laboratoire de cyberjustice*. https://www.cyberjustice.ca/2020/06/17/projet-de-loi-64-une-reforme-a-leuropeenne-du-droit-a-la-protection-des-renseignements-personnels/

Du Perron, S. (2020b, November 24). *Le temps des réformes: Cinq comparaisons entre le projet de loi n° 64 et le projet de loi C-11*. Autonomisation des acteurs judiciaires par la cyberjustice. https://ajcact.openum.ca/2020/11/24/le-temps-des-reformes-cinq-comparaisons-entre-le-projet-de-loi-n-64-et-le-projet-de-loi-c-11/

Earley, S., Henderson, D., & Data Management Association (Eds.). (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd edition). Technics Publications.

Elliot, M., Mackey, E. et O'Hara, K. (2020). The Anonymisation Decision Making Framework 2nd Edition: European Practitioners' Guide. UKAN. https://msrbcel.wordpress.com/framework/

European Commission. (2020). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A Digital Agenda for Europe* https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52010DC0245

European Data Protection Supervisor (n.d.). *Data protection officer*. https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en

European Parliament and the Council of the European Union (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)* https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW), 19*(3–4), 355–375. https://doi.org/10.1007/s10606-010-9117-8

Faundeen, J. L., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., Hutchison, V. B., Martín, E., Montgomery, E. T., Ladino, C. C., Tessler, S., & Zolly, L. S. (2014). *The United States Geological Survey Science Data Lifecycle Model* (U.S. Geological Survey Open-File Report No. 2013–1265; Open-File Report). http://dx.doi.org/10.3133/ofr20131265

Gal, M. S., & Rubinfeld, D. L. (2019). Data Standardization. *New York University Law Review, 94*, 737–770. https://www.nyulawreview.org/issues/volume-94-number-4/data-standardization/

Gasser, U., & Almeida, V. A. F. (2017). A Layered Model for AI Governance. *IEEE Internet Computing, 21*(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835

Gellert, R. (2016). We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection. *European Data Protection Law Review, 2*(4), 481–492. https://doi.org/10.21552/EDPL/2016/4/7

Girard, M. (2018). *Canada Needs Standards to Support Big Data Analytics* (Policy Brief No. 145). Centre for International Governance Innovation. https://www.cigionline.org/sites/default/files/documents/PB%20no.145web.pdf

Groth, P., Munroe, S., Miles, S., & Moreau, L. (2008). Applying the Provenance Data Model to a Bioinformatics Case. In L. Grandinetti (Ed.), *High Performance Computing and Grids in Action* (pp. 250–264). IOS Press.

Guidoin, S., Marczak, P., Pane, J., & McKinney, J. (2016). *Identifying recommended standards and best practices for open data*. Open North & ILDA. http://geothink.ca/wp-content/uploads/2016/02/Identifying-Recommended-Standards-Open-Data-Open-North.pdf

Hardinges, J., Wells, P., Blandford, A., Tennison, J., & Scott, A. (2019). *Data trusts: Lessons from three pilots*. Open Data Institute. https://

docs.google.com/document/d/118RqyUAW-P3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit?usp=sharing

IEEE. (1990). Interoperability. In *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*. IEEE.

Infocomm Media Development Authority of Singapore & Personal Data Protection Commission. (2019). *Trusted Data Sharing Framework*. Government of Singapore. https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Information and Privacy Commissioner of Ontario. (2018, July). *Privacy Fact Sheet: General Data Protection Regulation*. https://www.ipc.on.ca/wp-content/uploads/2018/07/fs-privacy-gdpr.pdf

Involve, Understanding Patient Data and the Carnegie UK Trust, & Scott, K. (2018). *Data for Public Benefit: Balancing the Risks and Benefits of Data Sharing*. Involve, The Carnegie UK Trust, & Understanding Patient Data. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf

Jacobson, I., Spence, I., & Bittner, K. (2011). *USE-CASE 2.0: The Guide to Succeeding with Use Cases*. https://www.ivarjacobson.com/sites/default/files/field_iji_file/article/use-case_2_0_jan11.pdf

Janssen, M., Estevez, E., & Janowski, T. (2014). Interoperability in Big, Open, and Linked Data—Organizational Maturity, Capabilities, and Data Portfolios. *Computer, 47*(10), 44–49. https://doi.org/10.1109/MC.2014.290

Järvinen, T. L. N., Sihvonen, R., Bhandari, M., Sprague, S., Malmivaara, A., Paavola, M., Schünemann, H. J., & Guyatt, G. H. (2014). Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *Journal of Clinical Epidemiology, 67*(7), 769–772. https://doi.org/10.1016/j.jclinepi.2013.11.011

Johnson, K. (2020, September 28). Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI. *VentureBeat*. https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/

Jones, K. H., Ford, D. V., & Lyons, R. A. (2017b). *The SAIL Databank: 10 years of spearheading data privacy and research utility, 2007-2017*. Swansea University. https://saildatabank.com/wp-content/uploads/SAIL_10_year_anniversary_brochure.pdf

Jones, K. H., Laurie, G., Stevens, L., Dobbs, C., Ford, D. V., & Lea, N. (2017a). The other side of the coin: Harm due to the non-use of health-related data. *International Journal of Medical Informatics, 97*, 43–51. https://doi.org/10.1016/j.ijmedinf.2016.09.010

Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics, 23*(2), 141–146. https://doi.org/10.1038/ejhg.2014.71

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152. https://doi.org/10.1145/1629175.1629210

Klievink, B., van der Voort, H., & Veeneman, W. (2018). Creating value through data collaboratives: Balancing innovation and control. *Information Polity, 23*(4), 379–397. https://doi.org/10.3233/IP-180070

Kum, H.-C., & Ahalt, S. (2013). Privacy-by-Design: Understanding Data Access Models for Secondary Data. *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science, 2013*, 126–130.

La Presse canadienne. (2020, June 17). *COVID-19: Les applications de traçage, alliées ou ennemies?* Radio-Canada.ca; Radio-Canada.ca. https://

ici.radio-canada.ca/nouvelle/1712841/applications-tracage-coronavirus-debat-experts-canada-pour-contre-vie-privee-sante

Marchand, M.-A. (2019). *Les fiducies d'utilité sociale: Synthèse de connaissances*. Territoires innovants en économie sociale et solidaire (TIESS). https://bit.ly/FUS-synthese_pdf

Marino, V., & Neto Costa, J. (2020, June 22). *The Wikidata project for the performing arts is on! – LDFI*. https://linkeddigitalfuture.ca/2020/06/22/the-wikidata-project-for-the-performing-arts-is-on/

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society, 7*(2), 205395172094808. https://doi.org/10.1177/2053951720948087

Montenegro, M. (2019). Subverting the universality of metadata standards: The TK labels as a tool to promote Indigenous data sovereignty. *Journal of Documentation, 75*(4), 731–749. https://doi.org/10.1108/JD-08-2018-0124

Mozilla Insights. (2020). *Shifting Power Through Data Governance*. Mozilla Insights. https://drive.google.com/file/d/1XLlGWRbm2bu48GgTFjG2aU4DSCL0U1s9/

Office of the Privacy Commissioner of Canada. (2017). *Summary of privacy laws in Canada*. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/

Office of the Privacy Commissioner of Canada. (2018, May 24). *Guidelines for obtaining meaningful consent*. https://www.priv.gc.ca/en/privacy-topics/collecting-personal-information/consent/gl_omc_201805/

Office of the Privacy Commissioner of Canada. (2019a). *PIPEDA fair information principles*. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/

Office of the Privacy Commissioner of Canada. (2019b). *PIPEDA in brief*. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/

Office québécois de la langue française. (1999). Protection de la confidentialité. In *Le grand dictionnaire terminologique (GDT)*. http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=2074344

Office québécois de la langue française. (2000). Propriétaire de fichier. In *Le grand dictionnaire terminologique (GDT)*. http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=2074764

Office québécois de la langue française. (2004). Donnée. In *Le grand dictionnaire terminologique (GDT)*. http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8358482

Open Data Institute. (2018). *Types of open standards for data*. Open Standards for Data Guidebook. https://standards.theodi.org/introduction/types-of-open-standards-for-data/

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

Open Data Institute. (n.d.). *The Data Spectrum*. Retrieved September 9, 2020, from https://theodi.org/about-the-odi/the-data-spectrum/

Opendatasoft. (2020). *Choisir et décrire vos métadonnées: Nos conseils pour rendre vos données découvrables, réutilisables et interopérables*. https://www.opendatasoft.com/fr/livre-blanc-metadonnees

Organisation for Economic Co-operation and Development (OECD). (2019). *Enhancing Access to and*

Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. OECD Publishing. https://doi.org/10.1787/276aaca8-en

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Otto, B. (2011). A Morphology of the Organisation of Data Governance. *ECIS 2011 Proceedings*., 272. https://aisel.aisnet.org/ecis2011/272

Pearson, S., & Casassa-Mont, M. (2011). Sticky Policies: An Approach for Managing Privacy across Multiple Parties. *Computer, 44*(9), 60–68. https://doi.org/10.1109/MC.2011.225

Peer, L., Green, A., & Stephenson, E. (2014). Committing to Data Quality Review. *International Journal of Digital Curation, 9*(1), 263–291. https://doi.org/10.2218/ijdc.v9i1.317

Peña Gangadharan, S., & Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society, 22*(7), 882–899. https://doi.org/10.1080/1369118X.2019.1593484

Policy and Research Group of the Office of the Privacy Commissioner of Canada. (2016). *A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act*. Office of the Privacy Commissioner of Canada. https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/consent_201605/

Porcaro, K. (2020). *Failure Modes for Data Stewardship*. Mozilla Insights. https://mzl.la/2Zu34tH

Québec Access to Information Commissioner (n.d.). *Un renseignement personnel, c'est quoi ?* Retrieved October 16, 2020, from https://www.cai.gouv.qc.ca/entreprises/un-renseignement-personnel-cest-quoi/

R v Jarvis, 2019 SCC 10, [2019] 1 SCR 488 https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/17515/index.do

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. (2015). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*. SSHRC. http://rgdoi.net/10.13140/RG.2.1.1922.5044

Riley, J. & National Information Standards Organization (U.S.). (2017). *Understanding Metadata: What Is Metadata, and What Is It For?* National Information Standards Organization (U.S.). http://www.niso.org/publications/understanding-metadata-riley

Ritchie, F. (2017). *The 'Five Safes': A framework for planning, designing and evaluating data access solutions*. Data for Policy 2017, London. http://dx.doi.org/10.5281/zenodo.897821

Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*(1), 3069. https://doi.org/10.1038/s41467-019-10933-3

Royal Academy of Engineering. (2019). *Towards trusted data sharing: Guidance and case studies—Data sharing checklist.* https://www.raeng.org.uk/policy/publications-(1)/interactives/data-sharing

Samarati, P., & de Vimercati, S. C. (2001). Access Control: Policies, Models, and Mechanisms. In R. Focardi & R. Gorrieri (Eds.), *Foundations of Security Analysis and Design* (pp. 137–196). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45608-2_3

Scassa, T. & A. (2018a). *Data Ownership* (No. 187; CIGI Papers). Centre for International Governance Innovation. https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf

Scassa, T. & B. (2018b, June 7). *Enforcement powers key to PIPEDA reform*. Policy Options. https://

policyoptions.irpp.org/magazines/june-2018/enforcement-powers-key-pipeda-reform/

Schmachtenberg, M., Bizer, C., & Paulheim, H. (2014). Adoption of the Linked Data Best Practices in Different Topical Domains. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, & C. Goble (Eds.), *The Semantic Web – ISWC 2014* (Vol. 8796, pp. 245–260). Springer International Publishing. https://doi.org/10.1007/978-3-319-11964-9_16

Sebastian-Coleman, L. (2018). *Navigating the labyrinth: An executive guide to data management* (1st edition). Technics Publications.

Shamsi, J. A., & Khojaye, M. A. (2018). Understanding Privacy Violations in Big Data Systems. *IT Professional, 20*(3), 73–81. https://doi.org/10.1109/MITP.2018.032501750

Sheehan, M., Thompson, R., Fistein, J., Davies, J., Dunn, M., Parker, M., Savulescu, J., & Woods, K. (2019). Authority and the Future of Consent in Population-Level Biomedical Research. *Public Health Ethics*, phz015. https://doi.org/10.1093/phe/phz015

Smart Dubai & Nesta. (2020). *Data sharing toolkit: Approaches, guidance and resources to unlock the value of data*. https://www.nesta.org.uk/documents/1832/Data_Sharing_Toolkit_1.pdf

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review, 154*(3), 477. https://doi.org/10.2307/40041279

Sposito, F. A. (2017). *What do data curators care about? Data quality, user trust, and the data reuse plan*. 7. http://library.ifla.org/1797/1/S06-2017-sposito-en.pdf

The British Academy & The Royal Society. (2017). *Data management and use: Governance in the 21st century*. https://royalsociety.org/-/media/policy/projects/data-governance/data-management-governance.pdf

The British Academy, techUK, & The Royal Society. (2018). *Data ownership, rights and controls: Reaching a common understanding* [Discussions at a British Academy, Royal Society and techUK seminar on 3 October 2018]. https://royalsociety.org/~/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf

The Engine Room. (2016). *The Hand-Book of the Modern Development Specialist: Being a Complete Illustrated Guide to Responsible Data Usage, Manners & General Deportment*. https://the-engine-room.github.io/responsible-data-handbook/assets/pdf/responsible-data-handbook.pdf

Thuermer, G., Walker, J., & Simperl, E. (2019). *Data sharing toolkit: Lessons learned, resources and recommendations for sharing data*. Data Pitch. www.datapitch.eu

UK. Information Commissioner's Office. (2019). *Data sharing code of practice*. https://ico.org.uk/media/about-the-ico/consultations/2615361/data-sharing-code-for-public-consultation.pdf

UK. Information Commissioner's Office. (n.d.). *Right to data portability*. ICO. Retrieved November 2, 2020, from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/

Vangen, S., & Huxham, C. (2012). The Tangled Web: Unraveling the Principle of Common Goals in Collaborations. *Journal of Public Administration Research and Theory, 22*(4), 731–760. https://doi.org/10.1093/jopart/mur065

Verhulst, S. G., & Sangokoya, D. (2015, April 22). *Data Collaboratives: Exchanging Data to Improve People's Lives*. Medium. https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a

Verhulst, S. G., Young, A., Winowatan, M., & Zahuranec, A. J. (2019). *Leveraging Private Data for Public Good* (p. 57). https://datacollaboratives.org/static/files/existing-practices-report.pdf

Visceral Visions. (2020). *Culturebrew.art*. https://www.visceralvisions.com/culturebrewart

Wade, M., & Hulland, J. (2004). Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *MIS Quarterly, 28*(1), 107–142. https://doi.org/10.2307/25148626

Weill, P. (2004). Don't Just Lead, Govern: How Top-Performing Firms Govern IT. *MIS Quarterly Executive, 3*(1), 1–17.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). *Creating an Understanding of Data Literacy for a Data-driven Society*. 18.

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology, 68*(4), 946–956. https://doi.org/10.1002/asi.23730

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology, 13*(3), e1005399. https://doi.org/10.1371/journal.pcbi.1005399