DATOS+ Collaboration

Facilitating Data Flows through Data Collaboratives

A Practical Guide to Designing Valuable, Accessible, and Responsible Data Collaboratives

AUTHORS
Uma Kalkar
Natalia González Alarcón

EDITORS
Arturo Muente Kunigami
Stefaan Verhulst

THEGOVLAB IDB

Facilitating Data Flows through Data Collaboratives

A Practical Guide to Designing Valuable, Accessible, and Responsible Data Collaboratives

AUTHORS
Uma Kalkar
Natalia González Alarcón

EDITORS
Arturo Muente Kunigami
Stefaan Verhulst

THEGOVLAB

IDB

JEL Codes: O30, O32, O33

Keywords: data governance, data collaborative, digital transformation

Banco Interamericano de Desarrollo
1300 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org

The Institutions for Development Sector was responsible for the production of this publication.
EXTERNAL VENDORS:
Production Editor: Sarah Schineller (A&S Information Partners, LLC)
Editor: Amy Scott, Nomad Enterprises
Design: Cleiman

CONTENTS

CONTENTS

PROLOGUE

Interactions with our surroundings are becoming increasingly digitized, immediately generating data that is captured by a myriad of devices and stored in servers around the world. The consequences of this can impact our lives in both positive and negative ways.

In fact, data can provide value through better decision-making and design and delivery of goods and services. It can expand our knowledge of available economic opportunities. Data may also help by increasing awareness and understanding of simple as well as complex problems. Sometimes, it helps prevent harm from coming our way, while at others it may be used for unintended purposes or even to achieve harmful goals.

Data may be collected without one's consent for marketing purposes or sold in non-regulated markets, sometimes falling into the hands of people or organizations engaged in illegal activities. In many instances, data required by one agent has already been collected by someone else, usually for a different objective. Despite the potential value of this existing data, accessing it can be challenging: the non-rival nature of data (meaning it can be used multiple times by multiple parties simultaneously) can raise concerns around competition and privacy that impede the flow of data between agents.

Data has no location and, in principle, no expiration date. There is likely more information about us stored outside our homes than inside, and it may remain there in perpetuity. With this in mind, who will protect us if our data is misused in another jurisdiction? Can public institutions in our jurisdictions do something about it? Can firms using data do something about it?

So how do we deal with these challenges? How do we promote an environment in which the safe use of data can be expanded? This publication takes these questions head-on and provides guidance to address them through an innovative approach: data collaboratives.

Data collaboratives, also called data trusts or data spaces, are "trust spaces" that allow for the sharing of data among a defined number of participants under a defined set of rules and guidelines. These rules and guidelines are intended to enhance the proper use of data, identify opportunities for sharing data, and especially to highlight risks for misuse of data.

This publication builds on an IDB event organized with The GovLab and in partnership with the Bank's Knowledge, Innovation, and Communication Sector under their Cutting-Edge Program. It presents the opportunities that data collaboratives can help address, describes different models for implementation, showcases specific examples, and elaborates on the governance and regulatory challenges that need to be addressed to implement them.

The goal is to share the highest standards on data flows and operationalize a collaborative environment as well as offer specific recommendations for policymakers. We hope this contribution helps expand the best in data collection and dissemination. The road is long and winding, but it is better if we walk it together.

Roberto de Michele

Division Chief
Innovation for Citizen Services Division
Institutions for Development Sector
Inter-American Development Bank

EXECUTIVE SUMMARY

Data has emerged as an indispensable asset in today's society. Despite its pervasive presence, all countries and organizations are increasingly confronted with situations where one needs data that has already been collected by another but is not easily accessible. Moreover, the inherent nature of data poses significant challenges to traditional market mechanisms, hindering the awareness and accessibility of crucial information. As a result, innovators from governments, corporations, and social impact organizations are tirelessly exploring novel approaches to accessing and reusing data, aiming to unleash its full potential in designing effective services and promoting the greater good. The diverse array of data collaboration models that have emerged underscores the absence of a one-size-fits-all model for designing valuable, accessible, and ethical data collaboratives. Instead, what is needed is a methodology and set of common elements that may enable experimentation and the creation of environments and collaboratives that are fit-for-purpose and aligned with the context in which they emerge.

The Governance Lab (GovLab), along with the Inter-American Development Bank (IDB), sought to understand the key requirements and common steps to initiate data (re) use initiatives and operationalize data collaborative models.

Specifically, we asked: What are the shared parameters of success across data flows, and what conditions enable them through robust data collaboratives? On November 29, 2022, the GovLab and the IDB held **Datos+VAR**, an event featuring public sector officials and private sector experts from across Latin America and the Caribbean, the United States, and Europe to understand how these two sectors foster a data (re)use culture and facilitate data access. With these insights, the GovLab and the IDB present here the policy and regulatory considerations and the concrete steps to develop and enable data collaboration across data flow and data (re)use use cases.

Our findings break down how to operationalize a data collaborative environment and how to tackle the considerations needed to set up a data collaborative.

First, we provide seven recommendations for decision-makers at government and policy levels seeking to create an enabling environment for data collaboratives:

1. **Establish governance principles to build impactful collaborations:** It is imperative to define and articulate a set of core principles that will serve as guiding beacons for data access and (re)use across sectors. These principles should prioritize data protection while enabling responsible and ethical data utilization.

2 **Develop comprehensive governance and regulatory frameworks:** To strike the delicate balance between promoting innovation and safeguarding data rights, it is essential to establish governance, regulatory, or legal frameworks. These frameworks will provide a solid foundation for public and private actors to (re)use data while upholding the principles of data protection and fostering innovation.

3 **Cultivate data stewardship capabilities:** Invest in building the expertise of "data stewards" who can play a pivotal role in identifying opportunities for data reuse and generating public value. These data stewards should possess the necessary knowledge to navigate the social and technical complexities involved in steering responsible data collaborative initiatives.

4 **Invest in robust data infrastructure and standards capacities:** To effectively structure data (re)use initiatives, it is crucial to allocate resources toward developing reliable data infrastructure and standardized formats. This will ensure that data is presented in a manner that is conducive to collaboration and maximizes its value.

5 **Foster a widespread data culture through leadership:** Attracting and nurturing data talent is essential for cultivating a resilient data

collaboration culture. Encourage leaders who understand the transformative potential of data and advocate for data stewardship. This will create an environment that fosters innovation, collaboration, and responsible data practices.

6

Focus on building sustainable governance processes: Rather than focusing solely on individual solutions, prioritize the development of sustainable and replicable governance processes. These processes should address the operational aspects of the data collaborative, effectively manage challenges, and ensure long-term success.

7

Cultivate a social license for data (re)use projects: To establish enduring trust and garner multisectoral support for data collaboratives, it is crucial to cultivate a social license. This means actively seeking permissions from the public to (re)use data, ensuring transparency, privacy protection, and meaningful engagement with relevant stakeholders. Building and maintaining a strong social license will be instrumental in securing the legitimacy and longevity of data (re)use initiatives.

Second, we outline five steps for practitioners (i.e., data stewards) seeking to establish data collaboratives:

1. **Define the need to address a problem with a data collaborative:** Recognize the purpose of a data collaborative by defining the problem it seeks to solve so as to home in on the priority focus areas and required data.

2. **Define the supply of data needed:** Test the feasibility of the data collaborative idea by identifying where the needed data resides and who has the skills to analyze it in order to ensure that its responsible (re)use is possible.

3. **Define the value proposition behind data access and reuse:** Pinpoint the type of data flow that is best suited to address the problem and utilize the resources and skills available.

4. **Match demand and supply:** After scoping these prerequisite needs for a data collaborative, support data stewards to operationalize the data collaborative by considering the regulatory, operational, and technological structures and considerations that need to be at play. Use these specifications to identify the type of data collaborative model best suited to the problem.

5 **Design a monitoring and evaluation (M&E) framework to measure impact:** An M&E framework can evaluate the impact of data access and (re)use to iteratively improve data stewardship and data collaborative practices.

To help decision-makers, specifically in public institutions, realize the value proposition of data (re)use and guide them through the policy and implementation development stages of facilitating data flows through six forms of data collaboratives, we provide worksheets (Appendix A) to sketch out the enabling environment considerations and identify steps toward building data collaboratives.

INTRODUCTION

Data has emerged as an indispensable asset in today's society, playing a pivotal role in driving innovation, enhancing decision-making, and fueling economic growth.

However, despite its pervasive presence, countries and organizations worldwide are grappling with the complex reality of data fragmentation and limited accessibility. Frequently, situations arise where one entity requires data that has already been collected by another, yet obtaining access becomes a cumbersome and arduous process. The inherent characteristics of data—including its volume, variety, and velocity—present significant challenges to traditional market mechanisms, impeding the efficient dissemination and availability of critical information. This scarcity of accessible data not only hampers the progress of scientific research, public policy formulation, and business strategies but also impedes the development of innovative solutions to pressing societal challenges. Consequently, there is an urgent need for data collaborative efforts and innovative approaches to overcome these barriers, unlock

the untapped potential of data, and create a more inclusive and data-driven future.

There are myriad ways that data access and (re)use can be facilitated. Finding the best governance, operational, institutional and technical requirements for all parties involved is where the challenge lies. Coordination efforts between actors highlight different data flows, or the directions of data exchange. Different mechanisms are used to experiment with data collection, collation, processing, and sharing. These mechanisms, both at innovation and regulation levels, lead to a diversity of data flow systems, better known as data collaboratives. Data collaboratives take data initially collected for one use and unlock it for new uses. They bring together multiple data holders who may not typically collaborate to access and (re)use data to avoid duplicating efforts and generate innovative insights.

BOX 1.
Data Collaborative Examples across the Health Industry

There are many types of data collaboratives (detailed in the following sections) that engage industry players and encourage data and knowledge sharing rather than operating in silos.

Take the health field, for example: individuals' health data, private sector innovation, and public sector information,

when (re)used responsibly, can mutually benefit all parties involved. From a government perspective, unlocking health data can help researchers across industries to contribute to public health. For instance, the UK's National Healthcare System's (NHS) Research Passport is a mechanism for non-NHS-affiliated stakeholders to receive an Honorary Research Contract to access NHS data for their research. The research passport creates a streamlined process through which the researcher submits their information for verification by research institutes and NHS organizations to get access to data more efficiently (NHS Health Research Authority, 2019).

The research community can also benefit by sharing insights with other academics and government or private sector partners to improve and apply research. One example of this is Yale University Open Data Access (YODA),[1] which facilitates clinical data sharing and (re)use among researchers. YODA receives data from different users, who sign a Data Use Agreement to allow their data to be reshared for projects that follow YODA's policies. YODA has the right to grant or deny a data access request, based on the agreement that it made with the data holders. All data is provided free of charge and the data transfer mechanism depends on the data partner. Another example is the Moore-Sloan Data Science Environments project (MSDSE).[2] Over a five-year partnership between

1 https://yoda.yale.edu/

2 http://msdse.org/about/

New York University; the University of California, Berkeley; the University of Washington; the Gordon and Betty Moore Foundation; and the Alfred P. Sloan Foundation, this initiative created Data Science Environments to harness data scientists and data science practices in an interdisciplinary manner. These environments brought data science talent, methods and tools, and institutions together to better support data science–driven research and facilitate effective use of data scientists' skills within and among partner institutions.

Data collaboratives also provide an avenue for individuals to increase collective bargaining power. For example, Salus Coop[3] is a data cooperative led by citizens that hosts personal health data and facilitates its sharing with researchers aligned with the cooperative's values. This model leverages data straight from individuals to build a direct consumer-to-corporation data sharing channel while empowering individual data awareness, literacy, and agency.

Each of these data collaboratives have technical, operational, and governance differences and cater to different data flows. Yet they all are able to achieve the common goal of promoting responsible, sustainable, and systematic data use and reuse to improve understanding of medical and scientific research.

3 https://www.saluscoop.org/

In this report, we ask: What are the shared parameters of success across data flows, and what conditions enable them through robust data collaboratives? To answer this question, we look at three things. First, we consider *why* understanding data flows and data access for decision-making is important. Second, we explain *what* data flows and data collaboratives entail. Last, we demonstrate *how* to operationalize a data collaborative environment and tackle the considerations needed to set up a data collaborative.

To help decision-makers, specifically in public institutions, realize the value of data (re)use and guide them through the policy and implementation development stages of facilitating data flows through six forms of data collaboratives, we provide worksheets (Appendix A) to sketch out the enabling environment considerations and steps toward data collaboratives. We also assess these forms against the spectrum of data collaborative attributes (Appendix B) to provide a rough baseline for each model's traits.

Governments and organizations need to consider how to walk the line between fostering innovation and protecting data rights while promoting the (re)use of data by public and private actors.

REPORT METHODOLOGY

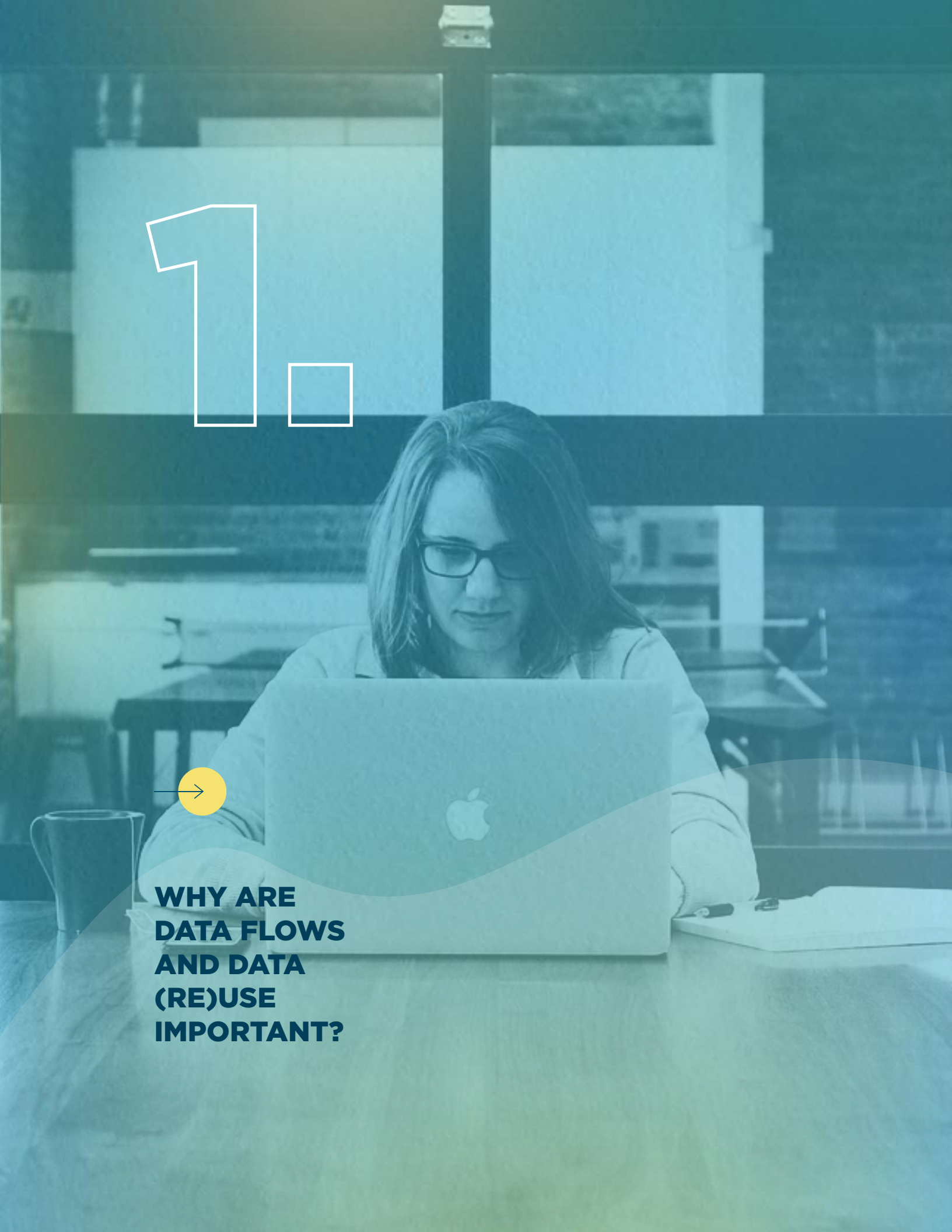
On November 29, 2022, the GovLab and the IDB held **Datos+VAR**, an event featuring public sector officials and private sector experts from across Latin America and the Caribbean, the United States, and Europe to understand how these two sectors foster a data (re)use culture and facilitate data access.

The experts discussed elements of successful national data strategies, such as training programs and multisectoral policy boards, and data collaboration models, including data platforms and data cooperatives, with a focus on how to accelerate access, sharing, and use of data for the public good. Following the panels, IDB department teams took part in a hands-on studio session to explore ways to introduce or improve data collaboration models for better

use of data in their specific sectors.

We sought to learn about how public institutions and private sector organizations design data collaboration efforts, given the numerous (and often overwhelming) avenues available for data management and operationalization. We looked at how different sectors can be motivated to (re)use data while maintaining competitive advantages. We believe that overcoming confusion around data access and (re)use pathways is essential because the potential of accessing the "treasure" of data resulting from digital transformation is immense.

To this end, this report serves as a guide for decision-makers to appreciate the value of identifying data flows and participating in data collaboratives. And, for those ready to dip their toes in the data (re)use pool and build these models, we offer practical tools to create the necessary enabling environment and follow the established steps to achieve data collaboratives.



1.

WHY ARE DATA FLOWS AND DATA (RE)USE IMPORTANT?

THE NEED FOR DECISION-MAKERS TO CONSIDER DATA FLOWS AND DATA (RE)USE

The datafication phenomenon, described by Mejias and Couldry (2019) as the "quantification" of human life, actions, and needs into "chunks" of data for public and private use, is in full swing. The volume of available data and the velocity by which it is generated by actors across the board hold enormous promise for data use and reuse by different actors (Arbesman, 2013). Yet this potential requires two things: first, the cultivation of a data access environment

that accommodates all stakeholders, from government officials to corporate decision-makers to civil society members, and second, purposeful and responsible "by design" principles and practices to drive data (re)use.

Creating an intersectoral data access environment allows decision-makers to unlock additional and perhaps unseen values of data. This creates opportunities for data (re)use, which entails responsibly repurposing data that was initially collected for one purpose and utilizing it for another initiative.

Many examples of data (re)use arose during the COVID-19 pandemic. For instance, the United Kingdom (UK) government rolled out a strategy for repurposing data from its National Healthcare System (NHS) to create the NHS COVID-19 Data Store database,[4] which harnessed this data to create data visualizations for pandemic policy responses. Similarly, telecommunications company Vodafone

4 https://www.england.nhs.uk/contact-us/privacy-notice/how-we-use-your-information/covid-19-response/nhs-covid-19-data-store

provided the European Commission's Joint Research Council and a number of international organizations—such as the World Bank, UNICEF, and the WorldPop project at the University of Southampton—with anonymized and aggregated data insights during the height of the pandemic to understand changes in mobility trends, track gender and social disparities, and inform epidemiological models to address infection rates (Lourenco et al., 2021). Similarly, the private global location data provider Veraset provided the IDB with geo-referenced data from mobile phones to measure the impact on human mobility of social distancing policies implemented in 18 Latin American and Caribbean countries (Aromi et al., 2020). Furthermore, the COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University[5] provided daily global administrative raw data on COVID-19 cases, vaccinations, and deaths via a data dashboard accessible by governments and citizens. These efforts, among others, helped decision-makers react to the pandemic in a more coordinated and informed manner to manage the spread of the virus.

Even before COVID-19, opening vertical and horizontal access to data has helped inform business and government decisions and foster a community of trust. For instance, InvestmentMap, an IDB initiative, provides individuals and businesses across Latin America and the Caribbean with

5 https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6

real-time capabilities through digital platforms to effectively monitor public spending and resource allocation and mitigate corruption risks.

Data flows help foster a culture of purposeful data sharing between data holders and data users that balances the ever-growing datafication phenomenon with the rights of data subjects. New, non-traditional data sources can introduce previously unexplored approaches to protracted problems that can improve quality of life. For example, during the COVID-19 pandemic, mobility data gathered from mobile devices, smart wearables, and GPS/Bluetooth devices helped understand people's movements (Khatib et al., 2021). Similarly, economic data shared by private financial institutions gave insight into macroeconomic health and consumer spending changes during the pandemic (Beasley, 2022). In short, when carried out purposefully and responsibly, data collaboration mutually benefits all parties involved in discovering insights, running operations, and providing value to target audiences. Incentivizing data collaboration falls under three main forms of return on investment: corporate social responsibility, brand equity, or knowledge and insights (see Figure 3 for more details) (Zahuranec, Young, and Verhulst, 2021).

CHALLENGES OF ASYMMETRIES AROUND DATA FLOW EXCHANGES

If intersectoral data access and (re)use help generate new insights and public value, why are these practices not more mainstream? At present, data (re)use is held back due to asymmetries that stem from how data is being collected and treated across its lifecycle.

The absence of an adequate data protection law, or the incompatibility of these legal frameworks between countries, can lead to significant challenges in the realm of cross-border data flows. These issues can create substantial asymmetries, generating problems not only for promoting data flows but also for potentially affecting the security and privacy of citizens' personal information. Despite the potential tensions that may arise between innovation and

regulation, both are necessary for reducing uncertainty regarding government oversight and delineating avenues for innovators. From a legal perspective, data protection laws are not only necessary but also crucial for ensuring an adequate treatment of sensitive personal data of citizens. Moreover, they act as enablers for innovation based on the responsible use of data.

Many data collection practices remain opaque about how and what data is being collected from data subjects, how these subjects are informed about what data is being collected and for what purposes, and how their consent is sought. As well, the algorithms and tools used to analyze data may operate in a closed "black-box" method, which leads to difficulties in explaining how conclusions are drawn from data. These (lack of standard) practices make data (re)use a difficult concept for the public to endorse for fear of potential data-based bias, targeting, and distrust of data users. Moreover, the ways in which data is collected and processed can be unequal across different divisions or teams of the same institution, creating information silos or data quality issues, and eventually leading to inefficient decision-making (Verhulst, 2022b). Data is also especially "hoarded" by private actors for fear of giving other firms and public institutions competitive business advantages (Verhulst, 2022b). These behaviors, among others, end up creating several asymmetries that jeopardize data flows (Verhulst, 2022a), including:

- **Data asymmetries**, which occur when those who can benefit from data are unable to access it. This can be due to a "divide or disparity in control of and access to data" which can change depending on the relationship between data owners and users (contributing to *agency asymmetries*) (Verhulst, 2022a). Data asymmetries enable a culture of opacity around data gathering and use and disenfranchise data subjects from exercising agency and control over their data (Kalkar, 2022).

- **Information asymmetries**, which result when stakeholders possess different abilities and capacities to translate data to relevant and actionable information, even when they have access to the same data (Verhulst, 2022a).

- **Intelligence asymmetries**, which are caused by the use of black-box algorithms and automated decision-making systems. Without understanding how these tools process and analyze data and what underlying biases they hold, decision-makers could be led toward outcomes that harm vulnerable populations and erode trust in data (re)use initiatives (Verhulst, 2022a).

- **Resource asymmetries**, which are a result of different stakeholders who can benefit from data being unaware of the data's existence and/or not having the tools to utilize data. This results from differences in actors'

"abilities to translate data into actionable information" due to technical, financial, or human resource limitations (Verhulst, 2022a). Resource asymmetries can lead to "too big to fail" data monopolies and/or perpetuate the "missed use" of data due to institutional inefficiencies (World Economic Forum, 2019).

Agency asymmetries, which refer to imbalances in data collection and (re)use relationships, with the pre-existing vulnerable party becoming further disenfranchised, skewing the ability and capacity of each stakeholder to equitably contribute to a data access and (re)use initiative (Verhulst, 2022b).

To combat these asymmetries, decision-makers should look to establish rigorous governance principles and processes across the data lifecycle to foster responsible and equitable ways of using data. For example, as part of the data collaborative design, the principle of "targeted transparency"[6] becomes an important approach to promote data flows and data sharing under scenarios where the traditional mechanism of transparency fails. Targeted transparency creates the right conditions for data collaboratives by building trust among stakeholders based on the commitment to using data for agreed purposes

6 Targeted transparency refers to publicly required disclosure of specific information in a standardized format to achieve a clear public policy purpose. It therefore goes beyond the disclosure of information as an end in itself. (Weil, Graham, and Fung, 2013).

while safeguarding against data misuse or abuse[7]. These practices help organizations enter the Third Wave of Open Data, which centers itself around purposeful and impactful data (re)use to benefit the public good (Verhulst, et al., 2020). The last sections of this document present ways that decision-makers can implement multiple practices to reduce these asymmetries.

A key pillar of the Third Wave of Open Data is **data collaboratives**, public-private partnerships that leverage data for public benefit (Verhulst et al., 2019). Strong data collaborative design and implementation help pave the way for **data stewardship**, or the advancement of data and data science experts who can liaise and instill trust between different sectors and the public for data (re)use initiatives, as well as uphold robust **data governance** to cement open data principles and data rights through legal and regulatory channels. These concepts will be further explored in later sections.

---

7 Targeted Transparency policies work best under the following scenarios (Fung, Graham, and Weil, 2007):

- A bridgeable information gap contributes substantially to risks or public service failures.
- The policy problem lends itself to consensus metrics.
- Communication is practical.
- Information users have the will, capacity, and cognitive tools to improve their choices.
- Information disclosures can help to reduce risk and improve performance.
- Variable results are acceptable.

2.0

WHAT CAN WE DO TO ENCOURAGE RESPONSIBLE DATA COLLABORATION?

MAIN TYPES OF DATA FLOWS

Data access and (re)use practices occur across a host of different sectors and actors, from small organizations to large corporations.

These include supranational, national, and subnational government bodies (G); small, medium, and large private businesses (B); and citizens, academic institutions, and community groups (C). Working effectively in a responsible and collaborative manner requires an understanding of data flows, which, as mentioned previously, refers to the different types of intersectoral interactions (see Figure 1 for an overview).

Government Business Citizen/Society

The following presents the main types of data flows, including:

G2G (government-to-government): Data shared between different government agencies and/or branches that breaks down cross-agency silos and promotes standardization and interoperability of government data.

- **Example:** In 2019, the City of Boston, Massachusetts, created a data warehouse for information coordination between public departments. The data warehouse holds data from 31 city departments and features a centralized data mapping platform (Johnston, 2019). Officials from the Citywide Analytics Team can filter the city-wide and department-wide data to isolate a specific data source to spot trends or streamline related data sources for a more holistic understanding of public service needs.

G2B (government-to-business): Data collected by the government, such as in the form of official statistics, that is shared with private firms to steer their business decisions.

- **Example:** In 2015, European Union (EU) government satellite data was shared with private shipping companies to help them adapt

to changing winter conditions in the Baltic Sea (Sawyer et al., 2015). This data sharing allowed for sea lanes and ports to function more effectively during the winter, benefiting economic gains and maintaining consistent supply chains.

G2C (government-to-citizen/society): Data presented to the public via open data databases or made publicly available by freedom of information acts.

- **Example:** 3D Data Hack Dublin[8] was a hackathon organized by Dublin City Council (via the Smart Docklands Programme) in 2019. The aim was to combine both existing and specially commissioned data from the City of Dublin, Ireland, to unlock exciting new opportunities and solve public challenges.

- **Example:** He Arotahi Tatauranga[9] presents official, high-quality statistics on New Zealand's Māori residents for research. The database and its framework for collecting and using data are available to everyone. This open data platform helps make indigenous people more visible in data-driven work, promoting inclusive policies.

8 https://data.gov.ie/organization/about/3d-model-datalyticon

9 https://www.stats.govt.nz/methods/he-arotahi-tatauranga/

B2G (business-to-government): Data collected by private companies by and for their business decisions that are shared with government actors to inform policymaking.

- **Example:** Created in 2014, the Accelerating Medicines Partnership[10] data pool allows genetic and molecular data gathered from 10 private pharmaceutical companies to be used by the U.S. National Institutes of Health and the Food and Drug Administration. This data sharing helps reduce the repetition of efforts and improve research and treatment for currently incurable diseases.

- **Example:** The Joint Research Centre of the European Commission used data from Facebook's advertising platform to estimate the number of expats in 17 European countries (Spyratos et al., 2018). This data helped model intra–European Union mobility trends in a low-cost and timely manner.

B2B (business-to-business): Data gathered by one firm that is shared with another one for mutual benefit.

- **Example:** Spanish private banking group BBVA collaborated with CARTO, a company that provides

10 https://www.nih.gov/research-training/accelerating-medicines-partnership-amp

spatial intelligence data and tools, to generate insights about city dynamics by analyzing credit card transaction data. Insights from the data shared between the firms were fed into the Urban Discovery platform, an open tool that can help urban planners or city officials recognize how a neighborhood organizes itself, regardless of the administrative boundaries set by the local government (BBVA, 2018). The Urban Discovery initiative provided visualization and insights for Madrid, Barcelona, and Mexico City.

- **Example:** Open Traffic, an initiative founded by the World Bank jointly with the transportation consulting and software company Conveyal and open-source mapping company Mapzen, facilitates the exchange and use of data from transportation operators and makes it accessible to select partners (World Bank, 2016). Data contributors can apply to participate in a larger community engaged in sharing anonymized traffic statistics while making use of open-source software tools.

B2C (business-to-citizen/society): Data made available to the public such as through open databases or APIs for general awareness and use.

- **Example:** Numina is a tech startup that offers insights about people, vehicles, and the street

environment using sensors (Ding, 2019). Its public API gives developers access to this data in order to enable smarter technology tools to improve city street planning. Numina's examination of construction scaffolding in downtown Brooklyn, New York, helped give insight into what times and days pedestrian safety is most likely to be compromised by blocked sidewalks.

- **Example:** Through its Data for Good Initiative, Cuebiq, a location intelligence company, shared its location data with researchers at MIT Media Lab. These researchers analyzed the data and visualized a map that shows how economic inequality segregates movement across public spaces, building an Atlas of Inequality.[11]

C2G (citizen/society-to-government): Data voluntarily collected and shared by citizens with governments to inform decision-making, such as through data altruism, citizen science, or citizen assemblies, or through coordinated crowdsourcing means.

- **Example:** Driver's Seat[12] offers performance analyses, such as hourly earnings and hourly activities, to rideshare and delivery drivers through

11 https://inequality.media.mit.edu/

12 https://driversseat.co/

a mobile application so that these service providers can optimize their driving times, strategies, and earnings with evidence-based data analysis. The application collects drivers' mobility/location intelligence data and sells it to city agencies in order to inform transportation planning decisions about congestion, pollution, and affordable transit.

- **Example:** The Colombia Longitudinal Survey (ELCA, for its acronym in Spanish) is a longitudinal survey developed by the Universidad de los Andes (Colombia) in 2007. It follows approximately the same 10,000 Colombian households in urban and rural areas of Colombia for 12 years, every three years. It aims to understand the social and economic changes at the individual and household levels, providing information on their behavior over time and their effect on welfare (Universidad de los Andes, n.d.). This survey data has been shared and used by government agencies at the local and national levels to inform policymaking processes.

C2B (citizen/society-to-business): Data provided by the public to a business, usually in exchange for access to a private service.

- **Example:** Open Humans[13] is a community-based

13 https://www.openhumans.org/

data sharing and (re)use platform that gives individuals the agency to access and share their personal data. This includes personal genetic data, mobility data from wearable devices, and/or health monitoring data for education, health, and research purposes.

- **Example:** Pecan Street[14] provides access to data on consumer energy and water consumption behavior. Their Dataport, the largest source of disaggregated customer energy and water data in the world, is used by numerous university researchers, entrepreneurs, and advocacy groups.

C2C (citizen/society-to-citizen/society): Data shared between citizens and community groups.

- **Example:** Launched in 2014, the Humanitarian Data Exchange (HDX)[15] aims to make humanitarian data easy to find and use for analysis undertaken by humanitarian workers. Managed by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), the site houses more than 9,800 datasets from 253 locations and 1,200 sources. Datasets containing personal information are only made available by request for approved users.

14 https://www.pecanstreet.org/

15 https://data.humdata.org/

- **Example:** OpenStreetMaps[16] originally started as an effort to map the UK in 2004, and it currently covers the entire planet, boasting over 10 million registered users who edit and contribute geo-referenced data daily. The information is presented in an open data format (attribution is required) and is widely utilized by numerous social and grassroots organizations.

16 https://www.openstreetmap.org/#map=6/4.632/-74.299

FIGURE 1.
Matrix of Data Flow Types

	Government	Business	Citizen/ Society
Government	**G2G** Data shared between different agencies (e.g., health and finance ministries) for new insights	**G2B** Official government statistics and data shared with private businesses	**G2C** Open data platforms and freedom of information acts to publicize government data
Business	**B2G** Private data shared with governments for policy action (e.g., traffic and rideshare data for smart city design)	**B2B** Data shared between two firms	**B2C** Data provided, usually free of charge, for public consumption
Citizen/ Society	**C2G** Data gathered from society by the government (e.g., census surveys)	**C2B** Data accrued from a product user (e.g., preference data gathered by social media firms)	**C2C** Crowd-sourced data actions (e.g., collaborative city mapping post-disaster)

In order to tap the flow of data between different sectors and disciplines, decision-makers across the board need to recognize that data can be a public good when applied in a thoughtful and responsible manner. Specifically, overcoming information silos and data sequestering, as well as enforcing good data governance policies for the collection, handling, and (re)use of data, can lead to stronger data flows and spur data collaboratives.

MAIN TYPES OF DATA COLLABORATIVES

After identifying data flow potentials, guiding the ways in which access to data is given and governed could be the answer to overcoming the aforementioned asymmetries and getting the most benefit from the data we hold.

Data collaboratives are how we operationalize data flows. The term *data collaborative* refers to a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors provide access to their data for (re)use in the public interest (GovLab, n.d.). They create processes to facilitate data access and (re)use across internal and external partners that can be retrofitted for various industries and objectives.

Data collaboratives are the means by which governments, businesses, and citizens/society capitalize on data (re)use

incentives for their various data flows. They seek to present these incentives through an optimal governance and operational structure for all parties and data types involved in a data access and (re)use project to realize a public good. Depending on their underlying goal and the stakeholders and data involved, data collaboratives can take on different shapes and structures to provide access to data without undercutting corporate advantage (see Figure 2).

FIGURE 2.
The Six Forms of Data Collaboratives

Public Interfaces

Open Access
Independent Use

(APIs, Data Platforms)

Trusted Intermediaries

Restricted Access
Independent Use

(Data Brokers, Third-Party Analytics Projects)

Data Cooperatives or Pooling

Open Access
Cooperative Use

(Public/Private Data Pools)

Research Partnerships

Restricted Access
Cooperative Use

(Data Transfers, Data Fellowships)

Prizes & Challenges

Open Access
Directed Use

(Open/Selective Innovation Challenges)

Intelligence Generation

Restricted Access
Directed Use

(Data Tanks, Internal Research)

The six most common types of data collaboratives[17] that provide a blueprint for managing data flows are outlined below.

Public Interfaces

A public interface involves a single entity providing open, public access to certain types of pre-processed data and data-driven tools, such as maps or dashboards, to enable independent use of the data by others. Although data assets are publicly accessible online, these interfaces are created with a particular audience or type of use case in mind. This model can appeal more to G2B, G2C, B2B, B2C, C2B, and C2C data flows.

Two of the current approaches to public interfaces are Application Programming Interfaces (APIs) and data platforms. APIs are software interfaces that publish data in an automated fashion on a near real-time basis. For example, Google Earth Outreach's[18] API provides nonprofits and the public sector with geospatial data and analytic tools to analyze that data in order to help developers generate new tools and insights. Data platforms make data assets and tools accessible to the general public through web or mobile applications, which require less data and

17 For a more comprehensive overview of data collaboratives, see Verhulst et al. (2019).

18 https://www.google.com/earth/outreach/

software development expertise than APIs. For example, the Uber Movement platform[19] provides mobility insights for urban planning and research using data collected from over 2 billion rides made with Uber globally.

Trusted Intermediaries

A trusted intermediary is a third-party actor who helps match supply and demand between data providers and data users. It does not necessarily require the trusted intermediary to hold the data; rather, the third party could act as a facilitator to establish data cooperation mechanisms, handle legal formalities, and bring in needed experts. This model supports the sharing of pre-processed data or insights between selected companies, governments, and the public in a directed or independent style and either a time-bound or open-ended manner. It facilitates collaboration and can also provide technical expertise and insights. This model can appeal more to G2G, G2B, B2G, and B2B data flows.

Two examples of trusted intermediaries are data brokerages and third-party analytics. A data brokerage engages third parties who facilitate the matching of the supply of data with its demand in a purpose-bound and time-bound fashion. For instance, in the UK, the Economic and Social

19 https://movement.uber.com/

Research Council established the Consumer Data Research Centre[20] to act as an intermediary to make data held by consumer-related businesses available to researchers from the University of Leeds, University College London, the University of Liverpool, and the University of Oxford for social and economic research. A third-party analytics project sees trusted partners access private sector data and conduct targeted analyses. The partners share insights from these analyses, but not the underlying data, with public or civil sector clients. For example, the nonprofit Flowminder analyzed private SIM card data to map population movement in Nepal after the 2015 earthquake, helping to coordinate on-the-ground humanitarian responses (Wilson et al., 2016).

Data Cooperatives or Pooling

A data cooperative or data pool brings data holders together to create a unified presentation of datasets that is accessible to multiple parties. These insights can be available to the general public or only accessible to select private groups and experts. This model can appeal more to B2G, B2B, B2C, C2G, C2B, and C2C data flows.

Two of the current approaches to public interfaces are public data pools and private data pools. A public data

20 https://www.cdrc.ac.uk/about/

pool mixes data assets from multiple data holders and shares these assets openly online. Though open to public access, contributions to the data pool are often limited to approved partners. For instance, the Accelerating Medicines Partnership[21] is a public data pool that brings together government, private sector, and nonprofit foundation partners to identify biological targets for therapeutics for currently incurable diseases, such as Alzheimer's disease or Type 2 diabetes. Data providers make their data and analyzes publicly accessible to the broad biomedical community through an online portal. In the case of private data pools, partners from different sectors pool data assets in a controlled-access environment. For instance, the U.S. Food and Drug Administration established the Sentinel Initiative[22] to actively monitor adverse reactions to medical products after they are on the market. The model securely shares and analyzes five sensitive data types from local health providers: administrative data, clinical data, registry data, inpatient data, and mother-infant linkage data.

Research Partnerships

A research and analysis partnership model engages companies directly with public sector partners to share certain private data assets to generate public value. This

21 https://www.nih.gov/research-training/accelerating-medicines-partnership-amp

22 https://www.sentinelinitiative.org/

A research and analysis partnership model engages companies directly with public sector partners to share certain private data assets to generate public value.

highly cooperative work allows corporate partners to augment existing business capabilities, incubate new product ideas, and analyze questions beyond the scope of internal business operations. This data collaborative model is usually purpose- and time-bound to address specific intractable problems. This model can appeal more to B2G and B2C data flows.

Examples of research and analysis partnerships include data transfers and data fellowships. For a data transfer, external actors are given access to corporate or government databases. The Valassis Address Data and Post-Katrina Repopulation project is an example of such an initiative; the marketing company shared its massive mailing address database with a nonprofit, which then used Google Maps and Street View to visualize and track the block-by-block repopulation rate after Hurricane Katrina (Neff, 2008). Similarly, the University of Minnesota's Integrated Public Use Microdata Series (IPUMS)[23] project receives census microdata from countries across the world, which is then harmonized and publicly released for researchers. In a data fellowship, actors temporarily join private sector organizations to work on a data-based project in-house, such as through Amazon Web Services (AWS) and Azavea's Open Source Fellowship.[24] Azavea, a social enterprise that uses geospatial technology to tackle civic and

23 https://www.ipums.org/

24 https://fellowship.azavea.com/about/

socioeconomic issues, recruits fellows to work in its office and provides them with access to AWS's earth data.

Prizes and Challenges

This type of data collaborative involves competitions where data holders make data available to participants to pioneer innovative uses of data for the public interest and to provide business value for prizes. Depending on the nature of the challenge, data access can be open or restricted and consist of pre-processed data or insights. The data is usually shared in a time-bound setting for a specific focus. This model can appeal more to G2C, B2B, and B2C data flows.

Two examples of prizes and challenges include open innovation challenges and selective innovation challenges. In an open innovation challenge, a company provides open access to datasets to attract participants to develop data-driven solutions to public challenges. DrivenData's challenge platform,[25] for instance, hosts data competitions for tools that create public value in collaboration with civil society or international organizations. Participants receive data from both public and private institutions to tackle issues like heart disease prediction, earthquake damage, and more. Selective innovation challenges, on the other hand, provide restricted data access to external parties

25 https://www.drivendata.org/competitions/

after thoroughly vetting candidates. In 2018, Turkey's state-owned telecommunications company, Türk Telekom, made anonymized mobile phone data available to a select group of researchers from nonprofit organizations and universities to develop ideas for improving living conditions for the 3.5 million Syrian refugees living in Turkey.[26]

Intelligence Generation

Intelligence generation is a fully restricted access model, with no data being shared with external parties. Entities internally develop data-driven analyses, tools, and other resources. This work often aims to aggregate and disseminate insights and analysis to inform policymaking and service delivery, before sharing those insights publicly. This model can appeal more to G2G and B2B data flows.

An example of intelligence generation is the JP Morgan Chase & Co. Institute's Insights Reports.[27] The institute uses its expertise and proprietary data from its parent multinational financial corporation to produce research reports on economic issues. These reports are intended for policymakers, businesses, and nonprofit leaders to address public challenges, such as how families manage tax refunds and how local commerce adapts in the digital age.

26 https://datacollaboratives.org/cases/turk-telekom-data-for-refugees-d4r-challenge.html

27 https://www.jpmorganchase.com/institute/research

Each of these data collaborative structures provide an operational, regulatory, and technical structure around which to center data innovation efforts in a conducive manner for all parties. As the data ecosystem grows and evolves, other forms of subsets of data collaborative models have emerged, including Data Commons, a form of Data Cooperatives or Pooling, wherein a data resource is shared and governed by multiple stakeholders, and Data Trusts, a form of Trusted Intermediaries that collect data from different sources via donation and serve as a middle man in maintaining and managing access to the data. Each model comes with its own set of advantages and disadvantages that decision-makers should pay close attention to when deciding on suitable models for their data initiatives. Appendix B assesses the six models of data collaboration across a spectrum of data access attributes.

3.

HOW CAN DATA COLLABORATIVES BE IMPLEMENTED?

Much of the data needed to address societal goals is being generated and collected by external parties, most of them as part of the private sector (National Academies of Sciences, Engineering, and Medicine, 2017). As such, the success of data collaboratives relies on the willingness and ability of these parties to provide access to their data—and, importantly, to do so in formats that are accessible and understandable to recipient organizations.

Thus, it is imperative that decision-makers seeking to tap into data flows and (re)use initiatives understand how to implement data collaboratives at both an ecosystem level and a systems level.

In doing so, it is critical to map the different roles that decision-makers can play according to each context when considering a data collaboration, most of which will depend on who owns or holds the data (see Box 2).

BOX 2.
Key Roles in Data Collaboration

Four roles can be identified when considering the design and implementation of a data collaborative:

- **Data holders:** These are organizations or individuals that possess and control access to data that is relevant for this particular project or collaboration. The data holder can be the owner of the data or have been given permission by the data owner to hold and manage the data. It can be governments, private sector, or a community.

- **Data subjects:** These are individuals or entities from whom the data is collected. The data subject is usually the focus of data protection laws and regulations, and has certain rights over the data, such as the right to access, correct, and delete it (IMY, 2021).

- **Data users:** These are organizations or individuals that access and (re)use data that has been made available by data holders through the data collaborative. Data

users may include governments, businesses, nonprofit organizations, or researchers (G, B, C).

- **Data stewards:** These are organizational leaders or teams with data and data science skills who are responsible for reusing their organization's data to create public value by identifying opportunities for cross-sector collaboration and responding proactively to external requests for access to data, insights, or expertise (Verhulst, Zahuranec, et al., 2020).

FOSTERING INSTITUTIONAL CAPACITIES: POLICYMAKER RECOMMENDATIONS FOR ENABLING A DATA COLLABORATIVE ENVIRONMENT

Effectively operationalizing data flows requires a thorough understanding of the ecosystem, stakeholders,

data, and purpose of data collaboration initiatives.

This enables decision-makers to address asymmetry challenges and correctly determine which data collaborative model is best suited for their project purpose. To answer the question "What does a government or a policymaker need to consider to cultivate a conducive environment for data collaboratives?" We present seven enabling recommendations for operational, organizational, and governance success. Worksheets to operationalize the concepts explored below can be found in Appendix A.

1. Establish governance principles to build impactful collaborations

Governments and organizations can steer data (re)use by prioritizing their guiding principles and mechanisms for getting access to and using data. It is imperative to define and articulate a set of core principles that will serve as guiding beacons for data access and (re)use across sectors. These principles provide a framework for how responsibility, accountability, transparency, and equity are embedded throughout the entire data lifecycle and should prioritize data protection while enabling responsible and ethical data utilization. Yet these principles can only go as far as the mechanisms and processes by which they are operationalized—namely, through the

data stewards that uphold and monitor them (see below Enabling Recommendation 3, Cultivate Data Stewardship Capabilities).

Thus, data protection can be demonstrated through collaboration by creating a credo of core principles, promoting processes for data governance, and enforcing oversight for each stakeholder—that is, the data holders and data users—involved in the effort. Further, assessing current practices against established ethical frameworks and soliciting input via participatory working groups can improve co-creation, understanding, and adherence to these principles.

Questions to ask:

→ What principles and values should the data collaboration follow and foster?

→ Is there any principle that stands out as a priority for the organization? What should not be happening with the data flows in the organization?

→ Through what metrics or indicators will the organization and its data stewards measure adherence to these principles and values?

→ What action will be taken to address infractions of these principles and values?

2. Develop comprehensive governance and regulatory frameworks

Regulatory frameworks can be an essential enabler for data innovation across data flows. There is a well-known and long-standing debate between innovation and regulation. To strike the delicate balance between promoting innovation and safeguarding data rights, it is essential to establish governance, regulatory, or legal frameworks for data collaboratives. A balanced regulation-innovation approach that bears in mind data protection measures while fostering data (re)use innovation can be achieved with clear and standardized legal frameworks for data collaboratives, along with well-governed, secure data governance structures. With these foundational frameworks, data collaboratives have a solid foundation on which to build out data governance structures. Indeed, data (re)use structures might be surrounded by multiple legal parameters, which can vary between and across governments, sectors, or use cases. Barriers to easily navigating this web of compliance requirements can disincentivize data holders from participating in data sharing, creating a missed opportunity to use potentially transformative data.

Governments and organizations need to consider how to walk the line between fostering innovation and protecting data rights while promoting the (re)use of data by public and private actors. Data stewardship helps ensure

adherence to regulation as well as promote internal data use when experimenting in regulatory gray zones responsibly. For instance, the Contractual Wheel of Data Collaboration (see Figure 5 on page 69) provides a starting point for organizations to assess the legal considerations of a data sharing agreement for a data collaborative.[28]

Questions to ask:

- What local and national regulations and laws need to be taken into account?
- Do existing regulations need to be adapted to allow data collaborations? Are new regulations required?
- Does the data collaboration share personal data? If yes, what is the associated regulatory framework to manage personal information?
- Who will ensure regulation is adhered to?
- Will compliance be audited? If so, by whom and when?
- How can existing processes and practices be improved to better embody regulatory aims?

3. Cultivate data stewardship capabilities

Data stewardship is integral to creating value through data governance arrangements and sparking data-focused

28 Learn more at https://contractsfordatacollaboration.org/our-framework/.

inspiration across an organization. It refers to the policies, functions, and competencies to enable access to and reuse of data for public benefit in a systematic, sustainable, and responsible way. Investing in data stewardship capabilities can play a pivotal role "to identify opportunities for data sharing and seek new ways of creating public value through cross-sector data collaboration." Data stewards are "bilingual" in data science and policy, which allows them to be an essential go-between for governments, businesses, and people to define a data flow initiative, persuade key stakeholders, and set up operational structures (i.e., data collaboratives).

Data stewards can play a crucial role in steering the process of using data and the insights it can generate by deciding who has access to data, for what purposes, and to whose benefit (Open Data Institute, 2022). They also play an active role in ensuring that data is being used in a responsible and ethical manner (Barrett, 2018). Data collaborative initiatives need data stewards to champion and sustain data (re)use to generate value that can withstand changing political or social tides and ensure sustained data responsibility. Attracting data stewardship talent and embedding a responsible data focus across stakeholders is essential for a successful data collaborative initiative.

Questions to ask:

- Is there a data governance framework established and operationalized within the organization? How is it upheld/enforced?
- Are responsibilities clearly assigned for managing all data assets in the organization?
- Does the organization have a data steward or data stewards?
- If there are no data stewards yet, what responsibilities and qualities should a data steward have?
- How can organizational and operational strategies be updated to enhance data stewardship?

4. Invest in robust data infrastructure and standards capacities

Data infrastructure and standards capacities are the enablers for setting up the structural and technical aspects of data collaboration. Data infrastructure comprises the ecosystem of technology, processes, and actors/organizations needed for the collection, storage, maintenance, distribution, and (re)use of data (Open Data Charter, n.d.). Specifically, it consists of the data assets (i.e., data centers, data exchange structures, cloud computing storage), the organizations that operate and maintain them, and the collaborative models available to handle data (Open Data Charter, n.d.).

Secure data infrastructure helps uphold the quality, accuracy, completeness, consistency, and reliability of the data under consideration. To make the most from managing these systems, decision-makers could also define data standards—standardized agreements on how data collection, handling, storage, and processing are to be conducted by participating organizations. Data standards describe the format, definition, structuring, tagging, transmission, manipulation, use, and management of data.

Data infrastructure and data standards needs depend on the purpose, type of data at hand, and data collaborative model in play. To this end, data stewards can outline the appropriate structures and formats for a data (re)use initiative to operate by and present data in a manner that is valuable and conducive to collaboration.

Questions to ask:

→ What current data infrastructures are in place? Are they secure, tested, and cohesive enough for establishing data collaborations?

→ What are the requirements (i.e., sourcing, format, tagging, granularity, etc.) for data to be collected, handled, and stored?

→ What oversight is in place to make sure these standards are adhered to? Who is responsible for this monitoring?

5. Foster a widespread data culture through leadership

Leadership has a critical role in triggering data innovation and unlocking boundaries among stakeholders. Encouraging leadership to realize the transformative potential of data and fostering an organization-wide data sharing and (re)use culture can embed widespread data expertise and ensure the sustainability of data collaboratives.

Indeed, human capital is an important success factor for long-lasting and informed data actions and agendas. Attracting data talent and promoting data stewardship are key to ensuring that both the systems and skill sets needed for data interoperability can withstand turnover in leadership or political administrations and foster a culture of data collaboration.

Questions to ask:

- Does the organization have a data-driven culture for decision-making?
- If not, what factors influence this lack of data culture?
- Is the leadership team (policymakers, managers, and executives) aware of opportunities through data collaboration?
- What steps can be taken in the short and long term to build an organic, fit-for-purpose data culture? Is there any existing organization to take inspiration from?

6. Focus on building sustainable governance processes

Defining the purpose of data collaboration helps guide initiatives as processes instead of technological solutions. This means that actors can focus on developing sustainable and replicable methodological frameworks for how the data collaborative operates and addresses challenges such as legal barriers, data ownership, asymmetries, and fears of misuse as well as privacy, ethical, and fairness issues (Ruijer, 2021). By concentrating on the structures through which challenges can be mitigated and addressed, the data collaborative builds resilient governance structures, processes, and practices before obstacles arise (Susha et al., 2018).

In this sense, a data collaborative might be driven around operational and organizational structures, meaning that people and processes that can manage the gap between an idea's potential impact and the necessary data considerations are paramount to address the question at hand.

Questions to ask:

- Are the processes to make decisions clearly defined? What are the decision-making inflection points?
- Who is responsible for executing and overseeing these actions? How are they supported?
- How will you ensure that the actions put in place will be scalable and sustainable?

7. Cultivate a social license for data (re)use projects

Last, to guarantee long-lasting and sustainable trust and multisectoral buy-in of a data collaborative, decision-makers need to cultivate the project's social license. A social license refers to the informal conversations between institutions such as governments or corporations with members of the public, to share and receive feedback on a particular set of activities to be carried out related to the use and reuse of their data (Shaw, Sethi, and Cassel, 2020).

In particular, reducing information and agency asymmetries through open communication, transparency proceedings, and auditable findings helps bridge the gap between data providers, data users, and data subjects. Practices such as citizen assemblies and mini-publics help stakeholders gauge public opinion over data (re)use and understand what data people are willing to contribute for the common good. Thus, not only is public awareness about data innovation important, but citizens' active engagement in creating and monitoring these processes can improve acceptance of data sharing and (re)use.

Questions to ask:

→ Does the general public know how their data is being managed?

→ Does the purpose of accessing and (re)using data resonate with the greater public? How are they consulted?

→ What concerns could stakeholders (e.g., watchdog groups, the public) have over the (re)use of data? How are these concerns addressed?

→ Are avenues for continual feedback set up? How frequent and involved are they?

Not only is public awareness about data innovation important, but citizens' active engagement in creating and monitoring these processes can improve acceptance of data sharing and (re)use.

STRENGTHENING THE OPERATIONAL APPROACH: PRACTITIONER RECOMMENDATIONS FOR ESTABLISHING A DATA COLLABORATIVE

The enabling environment recommendations help to cultivate high-level operational, organizational, and governance principles, policies, and processes for data collaboration.

Now, the following five steps will guide practitioners in answering the question "How can we identify the correct data collaborative to implement?" These steps will help practitioners select, design, and set up an appropriate data collaborative for their collaboration and data (re)use needs. A worksheet to operationalize the concepts explored below can be found in Appendix A.

1. Define the need to address a problem with a data collaborative

Setting up a data collaborative first requires identifying the need to address a problem and then identifying what data is required to solve it. Defining the purpose, goals, and objectives to be achieved by using and reusing the data are central to understanding the value brought by a data collaborative. Identifying the purpose helps design data initiatives that align with the ongoing Third Wave of Open Data and that consider the demand for a data initiative rather than focusing on the available supply of data to work with (Verhulst, Young, et al., 2020). To do this well, decision-makers—usually data holders/owners—need to define and understand the problem they hope to solve by scoping the data demand. This step requires scanning the subject matter around the problem, prioritizing areas of focus, and crafting specific questions that data can help to answer.

Specifically, defining the demand for data can be organized into four parts:

Problem Definition or Statement

When defining a good problem statement, it is important to break the problem down into smaller manageable pieces to understand its different elements and identify which one should be prioritized from a data reuse perspective. This helps determine how the data collaborative can help address the root causes of the problem.

- **Action:** Gather pre-existing statistics and relevant

studies or reports that shed light on the current situation, the gaps around the problem, what has and has not worked, and why the issue at hand still persists.

- **Action:** Craft a detailed problem statement that includes what the problem is, its underlying causes, and why it is important to address.

Topic Mapping

Topic mapping refers to the process of scoping out a particular issue or domain. A topic map locates different issues associated with a problem or topic within a larger context to help identify important subcomponents of an issue, the organizations already working in the space, notable research, and other information. Creating a topic map helps to structure and identify key areas of focus.

Following the GovLab's R-Search methodology (Young, 2015) to rapidly yet systematically scan a problem space helps develop comprehensive topic maps through two key actions:

- **Action:** Develop an understanding of the problem and solution area and undertake a rapid review of the issues and topics in the field for practitioners.

- **Action:** Identify the topics that organizations or experts in the field are already working on by conducting a rapid comparison of their online presence.

Topic maps have value because they can quickly get a data steward or other leader up to speed on a general problem they are facing. Topic maps can also help with actor mapping to identify who is already working on a topic or to identify potential "orphan issues" that no one seems to address.

Problem Prioritization

Equipped with a problem definition and topic map, a data collaborative initiative can engage with data holders, data subjects, data users, and data stewards to evaluate the potential impact, feasibility, and relative importance of each aspect of the problem and determine the most pressing issue areas.

- **Action:** Map the potential indirect effects of addressing the problem to identify potential outcomes of creating a data collaborative. This should include exploring the potential positive and negative effects of (re)using specific data and how to mitigate risks.

- **Action:** Identify any "capability traps" that may be present, including external factors like financing or

lack of expertise, that could lead to ineffective data strategies being sustained.

Question Formulation

To leverage the data, partners, and models of a data collaborative effectively, strong questions need to be asked to inform the topic or problem at hand. Different types of questions can be used to gain a better understanding of a problem, such as:

- **Ask: Situational awareness questions**, which leverage data to enable stakeholders to better understand trends and geographic distribution of the problem.

- **Ask: Cause-and-effect questions**, which help stakeholders better understand the key drivers and consequences of the problem, such as how the problem came to be and what effects it has.

- **Ask: Prediction questions**, which leverage data to allow stakeholders to assess future risks, needs, and opportunities.

- **Ask: Impact assessment questions**, which try to determine whether and how various interventions affect certain conditions and the obstacles that

hamper the achievement of certain objectives or the success of particular interventions.

At the end of this exercise, you should have a well-defined problem statement that is evidence and stakeholder based and framed in a data-actionable way.

- **Action:** Update the initial problem statement with the insights generated from this problem-focused approach.

2. Define the supply of data needed

Once the purpose is identified, the second step is identifying the data that would be needed to solve the problem, where the data supply resides, and who has the skills to analyze it (i.e., data stewards). At the end of this mapping of data needs and data supply, the data collaborative leaders and data stewards will be able to assess if the minimum viable data needed to answer the question is available and accessible. This process is called a pre-feasibility analysis—essentially, a litmus test on whether the data needed to address this issue can be effectively accessed and used. This analysis allows decision-makers to gauge the two central needs for a data collaborative project: expertise and data needs.

Data Needs

It is important to consider where the required data lives and how it can be accessed. For instance, privately held data may require a more stringent regulatory framework prior to (re)use, while information on open data platforms may be freely available for (re)use. After understanding where the data useful for the data collaborative is located, it is helpful to define the minimum viable dataset—the most minimal amount of data needed to make progress in answering the question. Determining the minimum viable dataset needed to answer a question and where it is available helps ensure that data (re)use is sufficient and proportional to provide meaningful insights and inform decision-making.

After crafting a problem statement, it is important to consider the following when defining the minimum viable dataset:

- **Action:** Identify the type of question being asked. Different types of questions will require different types and amounts of data. For example, a question about the current state of play may require a single data point while a question about cause and effect will require data on multiple variables.

- **Action:** Analyze if the data is accurate, complete,

and reliable, and what level of detail is required from the data. Depending on the question, different levels of detail may be required. For example, a question about broad trends may require data at a high level, while a question about specific causes may require data at a more granular level.

- **Action:** Establish when and where the data comes from. A question regarding geographic distribution will require data on location. In addition, the time frame of the data being collected is also an important consideration. For example, if the question is about recent trends, data from the last few weeks or months will be needed, while a question about historical trends will require data from a longer period of time.

- **Action:** Determine the sample size. The size of the sample also affects the data needs; depending on the question, a large sample size may be needed to be able to generalize the findings to the population.

Data Stewardship Needs

Do the data stewards involved have the necessary expertise in terms of data handling, analysis, context, and so on to run a data collaborative? Can the data steward(s) properly identify if this expertise exists

across the organization(s) and/or where to source it from? Data expertise can be internal from within the organization or from external sources; regardless, it is essential to identify where data stewardship is needed and if any gaps occur across the project duration. For more niche subjects, external expertise and capacity become particularly important to ensure responsible data use.

- **Action:** Map organizations' sources in terms of data analytics, engineering teams, and data librarians that would help handle and analyze the data, manage the data collaborative, and conduct quality assessments.

- **Action:** Map external suppliers by using expert networking platforms such as Vivo, which provides a network of scientists and scholars to make them more accessible to actors outside academia.

3. Define the value proposition behind data access and reuse

After mapping data, sources of data, and related stakeholders, there are several components of a data collaborative that its leaders—typically organizational leaders who are supported by the data stewards and data holders—should consider. This includes pinpointing the

type of data flow that might be more suitable to address the problem (e.g., is it a B2G, C2G, or G2G data flow?) and information about data accessibility and data attributes (see Appendix B). This information helps show the value proposition of using data to address the challenge at hand.

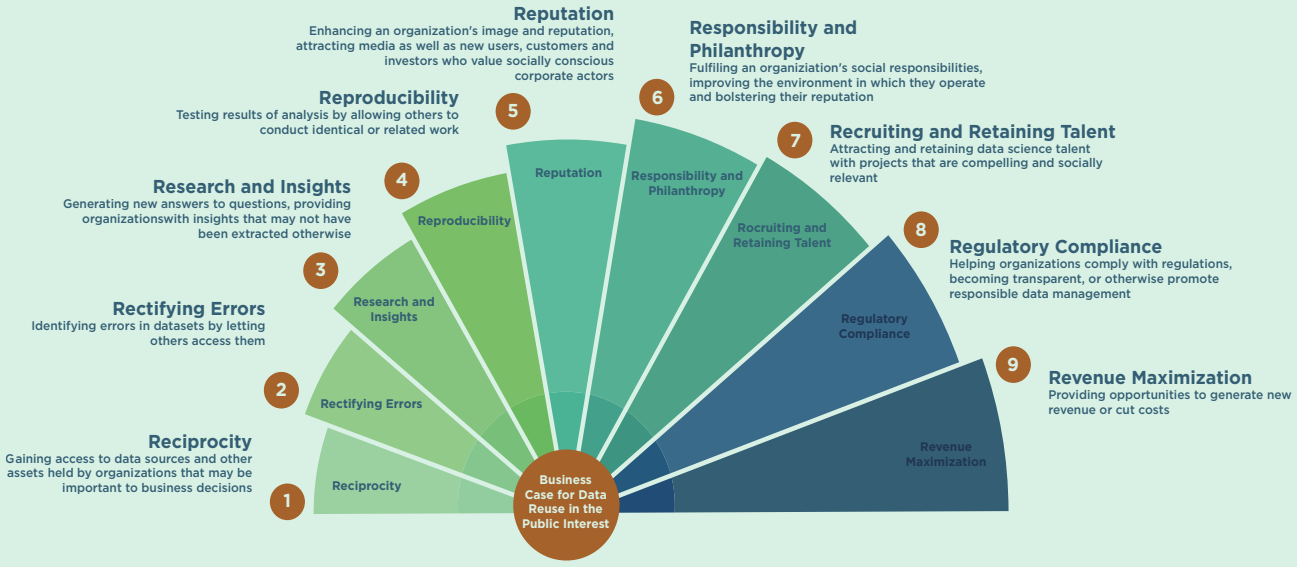
Value Proposition

The 9Rs Framework (see Figure 3) consists of nine motivations for establishing data collaboratives, categorized by different types of return on investment: license to operate, brand equity, or knowledge and insights. Considered together, these nine motivations compose a model to help organizations (public or private) understand the business value of making their data assets accessible.

- **Action:** Identify the main purpose and motivation for which the leading organization will pursue the data collaborative.

- **Action:** Evaluate the potential impact and value of data access and (re)use.

FIGURE 3.
The 9Rs to Open and Reuse Data in and for the Public Interest



Source: Zahuranec, Young, and Verhulst (2021).

1. **Reciprocity:** Sharing data with others can guide mutually beneficial business decisions.
2. **Rectifying Errors:** Sharing datasets with others can help identify and correct any errors.
3. **Research and Insights:** Sharing data can spark new and innovative approaches to issues.
4. **Reproducibility:** Sharing data allows for external validation of the results of analysis.
5. **Reputation:** Sharing data, especially to advance public issues, can bolster the image and reputability of a firm, attracting new socially conscious clients, talent, and followers.
6. **Responsibility and Philanthropy:** Data collaboratives

allow businesses to drive meaningful corporate social responsibility programs.

7. **Recruiting and Retaining Talent:** Data collaboratives attract data science talent with a compelling and socially relevant portfolio of data for good projects.
8. **Regulatory Compliance:** Data collaboratives can help corporations advance transparency and trust by establishing and following data sharing protocols.
9. **Revenue Maximization:** Corporate data can be sold to data collaboratives, generating novel revenue streams and business opportunities.

Risk Assessment

The collection, processing, sharing, analysis, and use of data introduce a number of risks and challenges for stakeholders involved in data collaboratives. In this step, stakeholders should seek to understand the risks at every stage of the data lifecycle (see Figure 4) in order to develop well-targeted strategies for mitigating them.

- **Action:** Map potential risks across the data lifecycle (with a particular focus on the sharing and use stages). Next, identify the procedures in place for conducting an end-to-end risk assessment across the data lifecycle—all the way from original collection to eventual (re)use. If there are none,

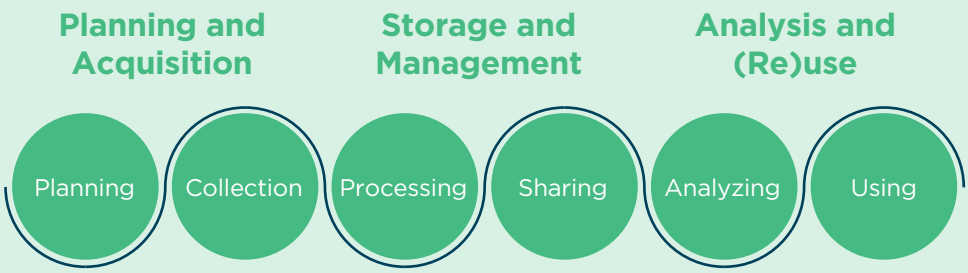
establish processes to pursue risk assessments across the data lifecycle.

- **Action:** Define mitigation actions to address the identified risks for each data lifecycle stage.

- **Action:** Determine which policy and legal frameworks at the local, national, regional, and global levels must be taken into account. This could include data protection laws, such as the EU's General Data Protection Regulation (GDPR), and/or sector-specific regulations relating to, for example, sensitive health or child protection data.

FIGURE 4.
The Data Lifecycle

Successful data sharing and (re)use need to consider all the phases of the data lifecycle. Although there are many approaches to define the data lifecycle and its stages, it can be summarized into at least three stages: Planning and Acquisition, Storage and Management, and Analysis and (Re)Use.



- **Planning and Acquisition:** During this initial stage, data-collecting organizations and agencies identify their current and future data needs and set out their objectives for using data. This includes an assessment of the existing data holdings, their maturity, and whether (or how) these will serve the public interest. Finally, the strategy defines and communicates standards regarding format, content, and quality for current data holdings and future data acquisitions.
- **Storage and Management:** Once the data is collected, the organizations need to define rules that determine how and by whom the data will be stored, shared, and accessed. These rules need to align with the legal and regulatory context, including data governance practices and technical requirements.
- **Analysis and (Re)Use:** After the data is collected and processed, it is essential to establish how organizations will generate insights and value to inform decision-making processes. During this stage, the organizations and agencies involved should promote data skills and foster partnerships to encourage the broader use of its data. This section can describe both primary use (by the collecting organization) and secondary use (by other organizations).

Once there is a clear problem definition and a mapping of the data demand and supply, data stewards can operationalize the data collaborative project and facilitate data sharing and (re)use.

4. Match demand and supply

Once there is a clear problem definition and a mapping of the data demand and supply, data stewards can operationalize the data collaborative project and facilitate data sharing and (re)use. At this step, it is also essential to consider different aspects of data governance within the participant entities—namely, the structures and mechanisms, operational models, and technological considerations necessary to match the data demand and supply effectively (see Box 3 for an in-depth definition of data governance). As previously mentioned, data collaboratives exist in several forms, each positioned to better address a specific problem or date type (see Appendix B). Below, the governance, operational, and technical considerations are outlined that can help select and set up a data collaborative model.

Governance Frameworks

When establishing a data collaborative, it is important to think about how the collaborative will be governed and ensure that the processes for making important decisions are clearly defined and understood by all parties—data holders/owners and data users (see Figure 5). Different actors approach the governance of data collaboratives in diverse ways, and each collaborative brings its own unique opportunities and challenges. The 4Ps of governance—purpose, principles, processes, and practices—is a good way to get started.

- **Action:** Remember the **purpose** for which the data collaborative is going to be developed. Revisit Enabling Recommendation 1 if needed.

- **Action:** State the prioritized **principles** that the design of the collaboration will be based on.

- **Action:** Identify the decision-making instances and define the **processes** that should be followed and by whom.

- **Action:** Explain and develop the mechanisms that will be implemented to apply the decision-making **practices**.

BOX 3.
Key Definition: Data Governance

Broadly speaking, "Data governance concerns the rules, processes, and behaviors related to the collection, management, analysis, use, sharing, and disposal of data" (Davies, 2022). These include but are not limited to: data security protocols, efficient processes for collecting and handling data, innovative ways of gleaning insights from data, designing accountable and effective feedback loops to improve data use and handling practices, training staff

to work with data responsibly, and adhering to local and national data use and protection regulation measures.

Data governance applies to both personal and non-personal data and seeks to "promote benefits and minimize harms" (Davies, 2022) across the planning and acquisition, storage and management, and analysis and (re)use stages of the data lifecycle.

With regard to data flows, data governance describes the governance and policy actions undertaken by data stewards to guide responsible data (re)use and data collaboratives.

Operational Structure

There is no one-size-fits-all approach when it comes to building fit-for-purpose collaboratives. Therefore, to determine which of the different data collaborative models is the best fit, these five variables can help: engagement, accessibility, scope, data, and collaboration. After defining the mentioned variables, it will be easier to decide between the six most common data collaborative models (described in Figure 2).

- **Action:** Identify and foster the **engagement** between stakeholders from the demand (data users) and supply (data holders) sides.

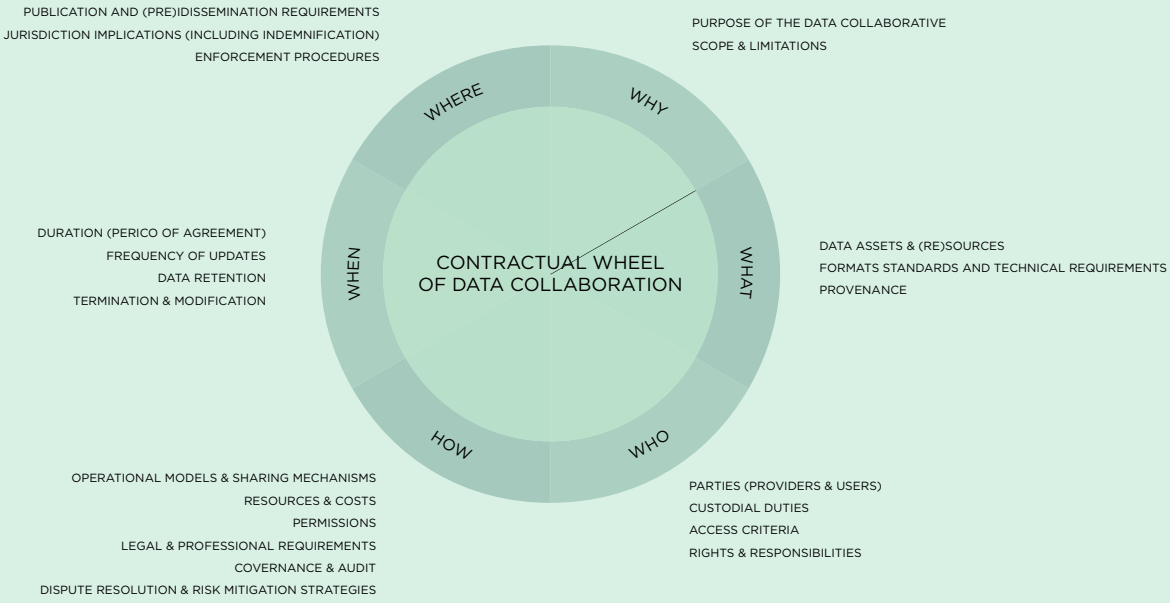
- **Action:** Analyze the **accessibility** of the data needed and evaluate it during the length of the data collaborative, with the support of the data stewards. For example, is the data open access or restricted? Who can access it and how?

- **Action:** Define the **scope** of the data collaborative—is it a temporary initiative, or will it become institutionalized? Pay specific attention to the time frame and define start and end dates while focusing on the project mission.

- **Action:** Identify and map the types of **data**, the data subjects who provide it, the data holders who own it, and their required characteristics for the data collaborative.

- **Action:** Identify the levels of **collaboration** across partners, the sectors involved, and the flow of data between partners.

FIGURE 5.
The Contractual Wheel of Data Collaboration



The Contractual Wheel of Data Collaboration walks data collaborative partners through discussions across six guiding aspects of a data sharing agreement:

- **WHY is data being shared?** What is the context and purpose?
- **WHAT kinds of data are being shared?** What are the sources, formats, and other technical requirements?
- **WHO is party to the agreement?** Who will be providing and using data resources? Are there any other third parties involved? Who has certain rights and duties?
- **HOW are data being shared?** How is the relationship

managed? How will issues such as security, privacy, and risk be handled?

- **WHEN will data actions take place?** At what point does the agreement start and end?
- **WHERE are data being shared to and from?** Are there jurisdictional issues to consider? Are there any international laws that apply?

Technical Infrastructure

For a data collaborative to be successful, it needs a solid yet flexible technical infrastructure. Both public and private institutions require and prioritize the development of e-infrastructures for data capturing, processing, analysis, and storage to enable correct and proper data (re)use (Jean-Quartier et al., 2022). Some key areas that should be considered are data standards and interoperability, as well as different data preparation, transfer, and access models.

- **Action:** Define and establish the required data standards across data sources and across the data lifecycle (see Figure 4).

- **Action:** Evaluate the level of data maturity among data providers to reveal and take into account technical disparities.

- **Action:** Define where the data is going to be stored and processed.

- **Action:** Explore interoperability between databases and, if necessary, how it will be achieved in an efficient manner if not currently in place. Note that interoperability can be broken down into three categories (Kruseman, 2022):
  - Data Formats, how data and metadata will be presented and queried;
  - Data Schemas, the structure for linking, sharing, and (re)using data; and
  - Ontology, or the meaning behind how data is described.

- **Action:** Consider the privacy features[29] needed to protect data and data subjects, based on the data sensitivity and the data collaborative model.

5. Design a monitoring and evaluation (M&E) framework to measure impact

By answering the previous questions and taking into account the categorizing of data collaborative components

29 One such tool of importance is Privacy Enhancing Technologies (PETs) that minimize the risks of data (re)use. PETs consist of a suite of tools including Trusted Execution Environments, Homomorphic Encryption, Secure Multi-Party Computing, Federated Learning, Differential Privacy, and Synthetic Data. See more at Royal Society (2023).

shown in Appendix B, data collaborative decision-makers will be able to identify the type of data collaborative that is the most convenient to answer the problem and accommodate data partners. Now last but not least relevant, it is vital to track and evaluate the impact of data access and reuse. A monitoring and evaluation (M&E) framework can help to improve data stewardship practices and achieve desired outcomes continuously. To measure progress throughout the lifespan of the data collaborative, iterative mechanisms need to be implemented to assess data (re)use, stewardship, and models against the baseline.

- **Action:** Define the most representative (and measurable) indicators that are critical to assess data handling, data stewardship, and the data collaborative model in use. This can include designing SMART goals and listing Key Performance Indicators (KPI) and Objective & Key Results (OKRs) of the initiative.

- **Action:** Define a common baseline of current practice to be able to measure the progress of the data collaborative's impact and reach.

- **Action:** Determine how the progress for the data collaborative across governance, operational, and technical angles is going to be measured.

- **Action:** Explore and identify the mechanisms to ensure the sustainability of the data collaborative.

CONCLUSIONS

By framing *why* data access and (re)use can spur social action and *why* specific actors should take part in sharing and using new data sources, data flows and data collaboratives can be leveraged effectively and purposefully. This report provides insight into *what* data flows and data collaboratives are, *why* they matter, and *how* to facilitate them at an ecosystem level and at an institutional and project level to identify useful flows and models for successful data (re)use.

REFERENCES

Arbesman, S. 2013. Five Myths about Big Data. *The Washington Post*. August 16. https://www.washingtonpost.com/opinions/five-myths-about-big-data/2013/08/15/64a0dd0a-e044-11e2-963a-72d740e88c12_story.html.

Aromi, D., et al. 2020. #StayatHome: Social distancing policies and mobility in Latin America and the Caribbean. IDB Working Paper Series No. IDB-WP-1147. Washington, DC: Inter-American Development Bank. Available at https://www.econstor.eu/bitstream/10419/237447/1/IDB-WP-1147.pdf.

Barrett, J. 2018. Up to 73 Percent of Company Data Goes Unused for Analytics. Here's How to Put It to Work. *Inc.* April 12. https://www.inc.com/jeff-barrett/misusing-data-could-be-costing-your-business-heres-how.html.

BBVA. 2018. BBVA uses big data to 'restructure' maps of Madrid, Barcelona and Mexico City. October 31. https://www.bbva.com/en/innovation/bbva-uses-big-data-restructure-maps-madrid-barcelona-mexico-city/.

Beasley, K. 2022. Covid-19 Altered the Supply Chain—Does Your Forecasting Strategy Reflect the New Reality? *Forbes*. July 14. https://www.forbes.com/sites/forbestechcouncil/2022/07/14/covid-19-altered-the-supply-chain-does-your-forecasting-strategy-reflect-the-new-reality/?sh=6aebefd55e24.

Davies, T. 2022. Data Governance and the Datasphere: Literature Review. Geneva, Switzerland: The Datasphere Initiative. Available at https://www.thedatasphere.org/wp-content/uploads/2023/05/Data-governance-and-the-Datasphere-Literature-Review-2022.-Datasphere-Initiative.pdf.

Ding, J. 2019. Downtown Brooklyn Case Study: Scaffolding and Safety. Numina.co. March 28. https://numina.co/downtown-brooklyn-case-study-scaffolding-and-safety/.

Fung, A., M. Graham, and D. Weil. 2007. Full Discoure: The Perils and Promise of Transparency. Cambridge, UK: Cambridge University Press.

GovLab. n.d. Data Collaboratives. https://datacollaboratives.org/.

IMY. 2021. The data subject's rights. Stockholm: Swedish Authority for Privacy Protection. Available at https://www.imy.se/en/organisations/data-protection/this-applies-accordning-to-gdpr/the-data-subjects-rights/.

Jean-Quartier, C., M. R. Mazón, M. Lovrić, and S. Stryeck. 2022. Collaborative Data Use between Private and Public Stakeholders—A Regional Case Study. *Data* 7(2): 20. Available at https://doi.org/10.3390/data7020020.

Johns Hopkins University. n.d. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6.

Johnston, R. 2019. Boston's new centralized data platform is the city's starting point for predictive analytics. *StateScoop.* March 29. https://statescoop.com/bostons-new-centralized-data-platform-is-its-starting-point-for-predictive-analytics/.

Kalkar, U. 2022. Digital Self-Determination as a Tool for Migrant Empowerment. *The GovLab—Data Stewards Network*. May 16. https://medium.com/data-stewards-network/digital-self-determination-as-a-tool-for-migrant-empowerment-f0dc386c8a41.

Khatib, E. J., et al. 2021. Mass Tracking in Cellular Networks for the COVID-19 Pandemic Monitoring. *Sensors* 21(10): 3424. Available at https://doi.org/10.3390/s21103424.

Kruseman, G. 2022. A Flexible, Extensible, Machine-Readable, Human-Intelligible, and Ontology-Agnostic Metadata Schema (OIMS). *Frontiers in Sustainable Food Systems* 6:767863. Available at https://doi.org/10.3389/fsufs.2022.767863.

Lourenco, P. R., G. Kaur, M. Allison, and T. Evetts. 2021. Data sharing and collaborations with Telco data during the COVID-19 pandemic: A Vodafone case study. *Data & Policy* 3: E33. Available at https://doi.org/10.1017/dap.2021.26.

Mejias, U. A., and N. Couldry. 2019. Datafication. *Internet Policy Review* 8(4). Available at https://doi.org/10.14763/2019.4.1428.

National Academies of Sciences, Engineering, and Medicine. 2017. Using Private-Sector Data for Federal Statistics. In B. A. Harris-Kojetin and R. M. Groves (eds), *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy.* Washington, DC: The National Academies Press. Available at https://www.ncbi.nlm.nih.gov/books/NBK425876/.

Neff, J. 2008. 'Junk Mail' to the Rescue in New Orleans: Valassis Database a Boon to Repopulation Efforts. *AdAge*. August 19. https://adage.com/article/news/junk-mail-rescue-orleans/130414.

NHS Health Research Authority. 2019. Research Passport. May 29. https://www.hra.nhs.uk/planning-and-improving-research/best-practice/research-passport/.

Open Data Charter. n.d. What is a data infrastructure? https://opendatacharter.net/agriculture-open-data-package/section-2-towards-open-data-infrastructure-agriculture/what-is-a-data-infrastructure/.

Open Data Institute. 2022. Bottom-up data institutions: Mechanisms for government support. London: Open Data Institute. Available at https://theodi.org/wp-content/uploads/2022/03/ODI-2022-Bottom-up-data-institutions_Mechanisms-for-government-support.pdf.

Royal Society. 2023. From Privacy to Partnership: The role of privacy enhancing technologies in data governance and collaborative analysis. January. London: The Royal Society. Available at https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/From-Privacy-to-Partnership.pdf?la=en-GB&hash=4769FEB5C984089FAB52FE7E22F379D6.

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

Sawyer, G., A. Dubost, M. de Vries, and I. van de Kerk. 2015. Copernicus Sentinels' Products Economic Value: A Case Study of Winter Navigation in the Baltic. Paris: European Space Agency. September. Available at https://esamultimedia.esa.int/docs/EarthObservation/case_report_winter_navigation_in_the_baltic_final.pdf.

Shaw, J. A., N. Sethi, and C. K. Cassel. 2020. Social license for the use of big data in the COVID-19 era. *NPJ Digital Medicine* 3(128). Available at https://doi.org/10.1038/s41746-020-00342-y.

Spyratos, S., et al. 2018. Migration data using social media: A European perspective. Joint Research Centre, European Commission.

Susha, I., et al. 2018. A Research Roadmap to Advance Data Collaboratives Practice as a Novel Research Direction. *International Journal of Electronic Government Research* 14(3): 11. Available at https://doi.org/10.4018/IJEGR.2018070101.

Thomas, L. 2022. Introducing the Charlotte Regional Data Trust. *UrbanCore.* University of Charlotte, North Carolina. July 13. Available at https://urbancore.charlotte.edu/introducing-charlotte-regional-data-trust.

Universidad de los Andes. n.d. ELCA Project. https://encuestalongitudinal.uniandes.edu.co/en/elca/elca-project.

Verhulst, S. G. 2022a. The ethical imperative to identify and address data and intelligence asymmetries. *AI & Society*. Available at https://doi.org/10.1007/s00146-022-01410-5.

Verhulst, S. G. 2022b. Operationalizing Digital Self Determination. October 25. Available at https://arxiv.org/pdf/2211.08539.pdf.

Verhulst, S. G., A. Young, M. Winowatan, and A. J. Zahuranec. 2019. Data Collaboratives: Leveraging Private Data for Public Good. October. Boston: The GovLab. Available at https://datacollaboratives.org/static/files/existing-practices-report.pdf.

Verhulst, S. G., A. Young, A. J. Zahuranec, S. A. Aaronson, A. Calderon, and M. Gee. 2020. The Emergence of a Third Wave of Open Data: How to Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing. Open Data Policy Lab. Available at https://opendatapolicylab.org/images/odpl/third-wave-of-opendata.pdf.

Verhulst, S. G., A. J. Zahuranec, A. Young, and M. Winowatan. 2020. Wanted: Data Stewards—Redefining the Roles and Responsibilities of Data Stewards for an Age of Data Collaboration. March. Boston: The GovLab. Available at https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf.

Weil, D., M. Graham, and A. Fung. 2013. Targeting Transparency. *Science* 340(6139): 1410–11.

Wilson, R., et al. 2016. Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake. *PLoS Currents* 8. Available at https://pubmed.ncbi.nlm.nih.gov/26981327/.

Wiseman, J. 2021. Two Models for Successful Intergovernmental Data Sharing. January 11. https://datasmart.hks.harvard.edu/news/article/two-models-successful-intergovernmental-data-sharing.

World Bank. 2016. Open Traffic Data to Revolutionize Transport. *World Bank News*. December 19. https://www.worldbank.org/en/news/feature/2016/12/19/open-traffic-data-to-revolutionize-transport.

World Economic Forum. 2019. Data Collaboration for the Common Good: Enabling Trust and Innovation Through Public-Private Partnerships. April. Cologny/Geneva, Switzerland: World Economic Forum. Available at https://www3.weforum.org/docs/WEF_Data_Collaboration_for_the_Common_Good.pdf.

Young, A. 2015. New GovLab Resource: R-Search – Rapid Re-Search Enabling the Design of Agile and Creative Responses to Problems. *The GovLab Blog.* January 8. https://blog.thegovlab.org/new-govlab-resource-r-search-rapid-re-search-enabling-the-design-of-agile-and-creative-responses-to-problems.

Zahuranec, A., A. Young, and S. Verhulst. 2021. The "9Rs Framework": Establishing the Business Case for Data Collaboration and Re-Using Data in the Public Interest. *Open Data Policy Lab Blog*. October 26. https://opendatapolicylab.org/articles/business-case-the-9rs-framework-establishing-the-business-case-for-data-collaboration-and-re-using-data-in-the-public-interest/index.html.

Zendle, D., and H. Wardle. 2022. Debate: We need data infrastructure as well as data sharing – conflicts of interest in video game research. *Child and Adolescent Mental Health* 28(1): 155–157. Available at https://doi.org/10.1111/camh.12629.

ACKNOWLEDGMENTS

We are grateful to the following individuals for lending their expertise on data flows to craft this report: Mara Balestrini, Data Lead Specialist at the IDB Lab; Dominique Duval-Diop, Deputy Chief Data Scientist for the U.S. Office of Science and Technology Policy; Peter Fatelnig, Minister Counsellor for Digital Economy Policy for the Delegation of the EU to the U.S.; Matt Gee, President of BrightHive; Alex Hutchison, Director of Data for Children Collaborative; Hyejeong Lim, Senior Manager at the National Information Society Agency of Korea; Luisa Medina, Deputy Director for Digital Public Capacities at the Ministry of ICT of Colombia; and Peter Rabley, Managing Partner of PLACE. We thank Jorge Alvarado, Founder and Chief Technology Officer of TrustRelay, and Julieth Santamaria, Economics Research Consultant in the Data and Digital Government Cluster at the IDB, for their invaluable feedback on an earlier version of this report.

ABOUT US

About the Governance Lab

The Governance Lab's mission is to improve people's lives by changing the way we govern. Our goal at the GovLab is to strengthen the ability of institutions—including but not limited to governments—and people to work more openly, collaboratively, effectively, and legitimately to make better decisions and solve public problems. We believe that increased availability and use of data, new ways to leverage the capacity, intelligence, and expertise of people in the problem-solving process, combined with new advances in technology and science, can transform governance. We approach each challenge and opportunity in an interdisciplinary, collaborative way, irrespective of the problem, sector, geography, and level of government. For more information, visit **thegovlab.org**.

About the Inter-American Development Bank

The IDB is the main source of multilateral financing and expertise for sustainable economic, social, and institutional development in Latin America and the Caribbean. The IDB Group is the leading source of development finance for Latin America and the Caribbean. The group comprises the **IDB**, which has worked with governments for 60 years; **IDB Invest**, which serves the private sector; and **IDB Lab**, which tests innovative ways to enable more inclusive growth.

APPENDIX A:
Worksheets for Decision-Makers to Identify Appropriate Data Collaborative Models

The following worksheets allow decision-makers to work through the responsible data requirements and considerations to design effective data flows and data collaboratives.

The first two worksheets pertain to the seven recommendations for decision-makers at government and policy levels seeking to create an enabling environment for data collaboratives, while the third worksheet covers the five recommendations for practitioners (i.e., data stewards) seeking to establish data collaboratives.

Checklist

Score your data flow initiative brainstorm by checking that you have accomplished the following requirements.

The principles and values are:

- [ ] Accountable
- [ ] Equitable
- [ ] Responsible
- [ ] Transparent

Regulatory and legal frameworks:

- [ ] Adhere to local and national rules and laws
- [ ] Keep current with emerging legislation/regulatory landscape
- [ ] Design responsible data protection and governance framework

Data stewards have been:

- [ ] Identified/hired for the initiative
- [ ] Translating data requirements into clear language across the organization
- [ ] Involved in defining the purpose, process, and strategic planning
- [ ] Monitoring practices to ensure adherence to principles and regulation

Data infrastructure and standards:

- [ ] Are appropriate for type and purpose of data initiative
- [ ] Uphold security and privacy of data
- [ ] Define data collection and format styles
- [ ] Have monitoring practices in place

Policymaker Recommendations for Enabling a Data Collaborative Environment

What do decision-makers looking to leverage data flows need to know?

ESTABLISH GOVERNANCE PRINCIPLES TO BUILD IMPACTFUL COLLABORATIONS

- What principles and values should the data collaboration follow and foster?
- Is there any principle that stands out as a priority for the organization? What should not be happening with the data flows in the organization?
- Through what metrics or indicators will the organization and its data stewards measure adherence to these principles and values?
- What action will be taken to address infractions of these principles and values?

DEVELOP COMPREHENSIVE GOVERNANCE FRAMEWORKS

- What local and national regulations and laws need to be taken into account?
- Do existing regulations need to be adapted to allow data collaborations? Are new regulations required?
- Does the data collaboration share personal data? If yes, what is the associated regulatory framework to manage personal information?
- Who will ensure regulation is adhered to?
- Will compliance be audited? If so, by whom and when?
- How can existing processes and practices be improved to better embody regulatory aims?

CULTIVATE DATA STEWARDSHIP CAPABILITIES

- Is there a data governance framework established and operationalized within the organization? How is it upheld/enforced?
- Are responsibilities clearly assigned for managing all data assets in the organization?
- Does the organization have a data steward or data stewards?
- If there are no data stewards yet, what responsibilities and qualities should a data steward have?
- How can organizational and operational strategies be updated to enhance data stewardship?

INVEST IN ROBUST DATA INFRASTRUCTURE AND STANDARDS CAPACITIES

- What current data infrastructures are in place? Are they secure, tested, and cohesive enough for establishing data collaborations?
- What are the requirements (i.e., sourcing, format, tagging, granularity, etc.) for data to be collected, handled, and stored?
- What oversight is in place to make sure these standards are adhered to? Who is responsible for this monitoring?

Main Challenge:

Data Flow Type:

Data Lifecycle Stage:

Data Collaboratives

Checklist

Score your data flow initiative brainstorm by checking that you have accomplished the following requirements.

The data culture is:

- [ ] Defined for the organization
- [ ] Communicated to the organization
- [ ] Practiced by senior management

Putting processes over solutions:

- [ ] Identified potential initiative challenges
- [ ] Identified levers to signal risk mitigation success
- [ ] Identified decision-makers throughout the process
- [ ] Identified decision-making points

Stakeholders for social license are:

- [ ] Informed and understand how data is being collected, handled, and used
- [ ] Able to communicate their concerns about data governance
- [ ] In touch with initiative to provide continual feedback

Policymaker Recommendations for Enabling a Data Collaborative Environment

What do decision-makers looking to leverage data flows need to know?

FOSTER A WIDESPREAD DATA CULTURE THROUGH LEADERSHIP

- Does the organization have a data-driven culture for decision-making?
- If not, what factors influence this lack of data culture?
- Is the leadership team (policymakers, managers, and executives) aware of opportunities through data collaboration?
- What steps can be taken in the short and long term to build an organic, fit-for-purpose data culture? Is there any existing organization to take inspiration from?

FOCUS ON BUILDING SUSTAINABLE GOVERNANCE PROCESSES

- Are the processes to make decisions clearly defined? What are the decision-making inflection points?
- Who is responsible for executing and overseeing these actions? How are they supported?
- How will you ensure that the actions put in place will be scalable and sustainable?

CULTIVATE A SOCIAL LICENSE FOR DATA (RE)USE PROJECTS

- Does the general public know how their data is being managed?
- Does the purpose of accessing and (re)using data resonate with the greater public? How are they consulted?
- What concerns could stakeholders (e.g., watchdog groups, the public) have over the (re)use of data? How are these concerns addressed?
- Are avenues for continual feedback set up? How frequent and involved are they?

Main Challenge:

Data Flow Type:

Data Lifecycle Stage:

Data Collaboratives

Practitioner Recommendations for Establishing a Data Collaborative

How to identify the correct data collaborative to implement?

1 DEFINE THE NEED TO ADDRESS A PROBLEM

Problem Definition or Statement

- [ ] Gather pre-existing statistics and relevant studies or reports that shed light on the current situation.
- [ ] Craft a detailed problem statement.

Topic Mapping

- [ ] Develop an understanding of the problem and solution area and undertake a rapid review of the issues and topics in the field.
- [ ] Identify the topics that organizations or experts are working on by conducting a rapid comparison of their online presence.

Problem Prioritization

- [ ] Map the potential indirect effects of addressing the problem to identify potential outcomes of creating a data collaborative.
- [ ] Identify any "capability traps" that can lead to ineffective data strategies being sustained.

Question Formulation

- [ ] Ask situational awareness questions
- [ ] Ask cause-and-effect questions
- [ ] Ask prediction questions
- [ ] Ask impact assessment questions
- [ ] Update the initial problem statement with the insights generated from this problem-focused approach.

2 DEFINE THE SUPPLY OF DATA NEEDED

Data Needs

- [ ] Identify the type of question being asked.
- [ ] Analyze if the data is accurate, complete, and reliable, and what level of detail is required from the data.
- [ ] Establish when and where the data comes from.
- [ ] Determine the sample size.

Data Stewardship Needs

- [ ] Map organizations' sources that would help handle and analyze the data, manage the data collaborative, and conduct quality assessments.
- [ ] Map external suppliers by using expert networking platforms.

3 DEFINE THE VALUE PROPOSITION BEHIND DATA ACCESS AND REUSE

Value Proposition

- [ ] Identify the main purpose and motivation for which the organization will pursue the data collaborative.
- [ ] Evaluate the potential impact and value of data access and reuse.

Risk Assessment

- [ ] Map potential risks across the data lifecycle (with a particular focus on the sharing and use stages).
- [ ] Define mitigation actions to address the identified risks for each data lifecycle stage.
- [ ] Determine which policy and legal frameworks at the local, national, regional, and global level must be taken into account.

4 MATCH DEMAND AND SUPPLY

Governance Frameworks

- [ ] Remember the purpose. Revisit Consideration 1 if needed.
- [ ] State the prioritized principles of the Enabling Recomendation will be based on.
- [ ] Identify the decision-making instances and define processes.
- [ ] Explain and develop the mechanisms that will be implemented.

Operational Structure

- [ ] Identify and foster the engagement between stakeholders.
- [ ] Analyze the accessibility of the data needed and evaluate.
- [ ] Define the scope of the data collaborative (time frame and focus).
- [ ] Identify and map the types of data, their providers, and their range.
- [ ] Identify the levels of collaboration and the flow of data.

Technical Infrastructure

- [ ] Define and establish the required data standards across data sources and across the data lifecycle.
- [ ] Define where the data is going to be stored and processed.
- [ ] Explore interoperability between databases.
- [ ] Consider the privacy features needed to protect data and data subjects,

5 DESIGN AN M&E FRAMEWORK TO MEASURE IMPACT

- [ ] Define the most representative (and measurable) indicators for the data collaborative.
- [ ] Define a common baseline of current practice to measure the progress.
- [ ] Determine how the progress is going to be measured.
- [ ] Explore and identify the mechanisms to ensure the sustainability.

Main Challenge:	Data Flow Type:	Data Lifecycle Stage:

Data Collaboratives

APPENDIX B:
Operationalizing Data Collaboratives

The following table considers the variables of Data Accessibility, Data Attributes, Collaboration Dynamics, and Scope that come into play when assessing different types of data collaborative models. It assesses where each of the six data collaborative models fall when considering these variables to help guide decision-makers in choosing an appropriate model for their purpose.

These variables consider who to give access to, how much access to give, where to share the data, which datasets to share, how involved a company can be, and what the scope of the collaboration is. The highlighted text indicates where

each data collaborative falls on the attributes' spectrum and which scenarios it fits best. This chart is an extension of the analytic and scoping framework presented in Data Collaboratives: Leveraging Private Data for Public Good (Verhulst et al., 2019).

Data Collaborative Model	Model Forms	Common Data Flow Facilitations	Data Accessibility	Data Attributes	Collaboration Dynamics	Scope	Use This Model For:
Public Interfaces: Companies provide open access to specific data assets, enabling independent uses of the data by external parties.	APIs and Data Platforms	G2B; G2C; B2B; B2C; C2B; C2C	Access: **Open Access**/ Restricted: • Unfettered public access to data Availability: On-Site/ **Online**: • Data published online	Data Temporality: On Demand/ **Ongoing**: • Ongoing public access to the data Types of Data: **Pre-Processed Data/Insights**: • Can exist as both granular data assets and compiled insights Data Providers: **Single Data Provider**/Multiple Data Provider: • One organization contributes data to the public interface Data Variety: **Single Dataset/ Multiple Datasets**: • Can exist as both; one or many datasets can contribute data to the public interface	Engagement: Directed/ **Independent**: • Independent use Flow: **Uni-Directional Data Flow**/ Multidirectional: • Data flows from the provider to the users Relationship: Bilateral/ **Multilateral**/ Directed: • Data flows from the provider to many users Sectoral Stakeholders: **Private/Public/ Civil Society**: • Data can come from all three stakeholders	Focus: Purpose-Bound/ **Flexible**: • Data can be interpreted by users in multiple ways Timeframe: Time-Bound/ **Open-Ended**: • Indefinite access to data	• Non-sensitive data that can be shared and (re)used by any and all organizations over a long time frame • Non-sensitive data originating from one or many sources and sectors • Non-sensitive data that can be interpreted in multiple ways without affecting the integrity of the data • Creation of knowledge banks and repositories

Data Collaborative Model	Model Forms	Common Data Flow Facilitations	Data Accessibility	Data Attributes	Collaboration Dynamics	Scope	Use This Model For:
Trusted Intermediaries: Third-party actors support collaboration between private sector data providers and data users from the public sector, civil society, or academia.	Data Brokers and Third-Party Analytics Projects	G2G; G2B; B2G; B2B	Access: Open Access/ **Restricted**: • Select actors allowed access to data Availability: **On-Site/ Online**: • Depending on data sensitivity, can be made available on-site or online via protected platform	Data Temporality: **On Demand/ Ongoing**: • Depending on data sensitivity, can be made available on-demand or ongoing to the data partners Types of Data: **Pre-Processed Data**/Insights: • Tends to exist as granular data assets Data Providers: **Single Data Provider/ Multiple Data Provider**: • Can exist as both; one or many organizations can contribute data to the trusted intermediary Data Variety: **Single Dataset/ Multiple Datasets**: • Can exist as both	Engagement: **Directed/ Independent**: • Can exist as both Flow: Uni-Directional Data Flow/ **Multidirectional**: • Data flows from the provider to the trusted intermediary, who then provides data to the user Relationship: Bilateral/ Multilateral/ **Directed**: • Data flows from the provider to the users via the trusted intermediary Stakeholders: **Private/Public/ Civil Society**: • Data can come from all three stakeholders	Focus: **Purpose-Bound**/ Flexible: • Data shared for a specific reason among the selected partners Timeframe: **Time-Bound/ Open-Ended**: • Can exist as both	- Sensitive data that could be beneficial for vetted users for a specific purpose • Data originating from one or many sources and sectors around a specific purpose • Data can be amplified if its gaps are filled by other information • Addressing specific intractable problems

Data Collaborative Model	Model Forms	Common Data Flow Facilitations	Data Accessibility	Data Attributes	Collaboration Dynamics	Scope	Use This Model For:
Data Cooperatives or Pooling: Companies and other data holders agree to create a unified presentation of datasets as a collection accessible by multiple parties.	Cross-Sector/ Cross-Institutional (Public or Private) Data Pools; Data Platforms and Portals; Data Commons (Data Trusts); and Data Cooperatives	B2G; B2B; B2C; C2G; C2B; C2C	Access: **Open Access/ Restricted**: • Can be both depending on data sensitivity Availability: On-Site/ **Online**: • Tends to be available online	Data Temporality: **On Demand/ Ongoing**: • Depending on data sensitivity, can be made available on-demand or ongoing to the data partners Types of Data: **Pre-Processed Data**/Insights: • Granular data assets joined together to create a meta-dataset Data Providers: Single Data Provider/**Multiple Data Provider**: • Many organizations can contribute data Data Variety: **Single Dataset**/ Multiple Datasets: • Presents many datasets as one	Engagement: **Directed/ Independent**: • Can exist as both Flow: Uni-Directional Data Flow/ **Multidirectional**: • Data flows from the providers to the cooperative/ pool, then to the user Relationship: Bilateral/ Multilateral/ **Directed**: • Data presented as insights from the provider to the user **Private/Public/ Civil Society**: • Data can come from all three stakeholders	Focus: **Purpose-Bound**/ Flexible: • Data shared for a specific reason to create the unified data presentation Timeframe: **Time-Bound/ Open-Ended**: - Can exist as both	• Data that could be beneficial for a specific purpose and can be used by both public and specific audiences • Data originating from one or many sources and sectors around a specific purpose • Data can be amplified if its gaps are filled by other information • Addressing specific intractable problems

Data Collaborative Model	Model Forms	Common Data Flow Facilitations	Data Accessibility	Data Attributes	Collaboration Dynamics	Scope	Use This Model For:
Research Partnerships: Companies engage directly with public sector partners and share certain proprietary data assets to generate new knowledge with public value.	Data Fellowships; Data Grants; Data Transfer; and Research Data Portals	B2G; B2C	Access: Open Access/ **Restricted**: • Select actors allowed access to data Availability: On-Site/ **Online**: • Tends to be available online	Data Temporality: **On Demand**/ Ongoing: • Data provided for the specific research only Types of Data: **Pre-Processed Data**/Insights: • Tends to exist as granular data assets so researchers can use and analyze it Data Providers: **Single Data Provider**/Multiple Data Provider: • Typically one organization donates data to research organizations Data Variety: **Single Dataset/ Multiple Datasets**: • Can exist as both	Engagement: **Directed**/ Independent: • Data providers brought together for a specific goal/ purpose; participants use data to innovatively address a set issue Flow: Uni-Directional Data Flow/ **Multidirectional**: • Data flows from the providers to the users; insights generated flow back to the providers Relationship: **Bilateral**/ Multilateral/ Directed: • Data flows from the provider to the users via the trusted intermediary **Private/Public/ Civil Society**: • Data can come from all three stakeholders	Focus: **Purpose-Bound**/ Flexible: • Data shared for a specific reason among the selected partners Timeframe: **Time-Bound**/ Open-Ended: • Data shared for a specific period of time (i.e., duration of the data grant)	• Sensitive data that could be beneficial for vetted users for a specific purpose • Data originating from one (usually private) source that can benefit public good if shared • Data can be amplified if its gaps are filled by other information • Addressing specific intractable problems

Data Collaborative Model	Model Forms	Common Data Flow Facilitations	Data Accessibility	Data Attributes	Collaboration Dynamics	Scope	Use This Model For:
Prizes & Challenges: Companies make data available to participants who compete to develop apps; answer problem statements; test hypotheses and premises; or pioneer innovative uses of data for the public interest and to provide business value.	Open Innovation Challenges and Selective Innovation Challenges and Hackathons	G2C; B2B; B2C	Access: **Open Access/ Restricted**: • Can be both depending on data sensitivity and nature of the challenge Availability: **On-Site/ Online**: • Can be both; tends to be available online	Data Temporality: **On Demand**/ Ongoing: • Data provided for the specific challenge only Types of Data: **Pre-Processed Data**/Insights: • Tends to exist as granular data assets so researchers can use and analyze it Data Providers: **Single Data Provider/ Multiple Data Provider**: • Can be both; typically one organization donates data to research organizations Data Variety: **Single Dataset/ Multiple Datasets**: • Can exist as both	Engagement: **Directed**/ Independent: • Data providers brought together for a specific goal/ purpose; participants use data to innovatively address a set issue Flow: Uni-Directional Data Flow/ **Multidirectional**: • Data flows from the provider to the user; users share insights with data providers Relationship: **Bilateral**/ Multilateral/ Directed: • Data flows from the provider to the users **Private/Public/ Civil Society**: • Data can come from all three stakeholders	Focus: **Purpose-Bound**/ Flexible: • Data shared for a specific reason among the selected partners Timeframe: **Time-Bound**/ Open-Ended: • Data shared for a specific period of time (i.e., duration of the data challenge)	• Sensitive data that could be beneficial for vetted users for a specific purpose • Data originating from one (usually private) source that can benefit public good if shared • Data can be amplified if its gaps are filled by other information • Addressing specific intractable problems

Data Collaborative Model	Model Forms	Common Data Flow Facilitations	Data Accessibility	Data Attributes	Collaboration Dynamics	Scope	Use This Model For:
Intelligence Generation: Companies internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public.	Data Tanks; Internal Research Operations; and Cross-Sector Data Analysis Tools and Applications	G2G; B2B	Access: Open Access/ **Restricted**: • Select actors allowed access to data Availability: **On-Site/ Online**: • Can be both; tends to be available online	Data Temporality: **On Demand**/ Ongoing: • Data provided for the specific research only Types of Data: **Pre-Processed Data**/Insights: • Tends to exist as granular data assets so researchers can use and analyze it Data Providers: **Single Data Provider/ Multiple Data Provider**: • Can be both depending on the scope and scale of the data (re)use Data Variety: **Single Dataset/ Multiple Datasets**: • Can exist as both	Engagement: **Directed**/ Independent: • Data providers brought together for a specific goal/ purpose Flow: Uni-Directional Data Flow/ **Multidirectional**: • Data flows from the provider to the user; users share insights with data providers Relationship: **Bilateral**/ Multilateral/ Directed: • Data flows from the provider to the users **Private/Public**/ Civil Society: • Data is shared within a sector between internal departments	Focus: Purpose-Bound/ **Flexible**: • Data can be interpreted by users in multiple ways Timeframe: **Time-Bound**/ Open-Ended: • Data shared for a specific period of time	• Sensitive data that could be beneficial for vetted users for a specific purpose • Data originating from one source that can benefit if shared; access and (re) use of data best suited for internal purposes • Data can be amplified if its gaps are filled by other information • Creating internal knowledge banks and repositories


GOVLAB


IDB