Open.Data.Policy.Lab

The Third Wave of Open Data Toolkit

Operational Guidance on Capturing the Institutional and Societal Value of Data Re-Use

March 2021

THEGOVLAB

Photo by Jeremy Bishop on Unsplash

About

This toolkit was developed by **Andrew Young**, **Andrew J. Zahuranec**, **Stefaan G. Verhulst**, and **Kateryna Gazaryan** of The GovLab at New York University Tandon School of Engineering. The toolkit seeks to operationalize the report The Emergence of a Third Wave of Open Data by Stefaan G. Verhulst, Andrew Young, Andrew J. Zahuranec, Susan Ariel Aaronson, Ania Calderon, and Matt Gee.

The authors of this toolkit would like to thank the expert participants in the Summer of Open Data initiative, which informed this analysis. They include: Arturo Franco, Arturo Muente Kunigami, Barbara Ubaldi, Christian Troncsco, Daniel Jarratt, Denise Linn Riedl, Felix Shapiro, Jaimie Boyd, Jean-Noé Landry, Jeni Tennison, John Wilbanks, Justine Hastings, Kara Selke, Malarvizhi Veerappan, Natalia Domagala, Nuria Oliver, Patrick McGarry, Paul Ko, Rhiannan Price, Rudi Borrmann, Stephen Chacha, Swee Leng Harris, Theo Blackwell, Tyler Kleykamp, Vanessa Brown, and Zachary Feder. We also appreciate our colleague Michelle Winowatan's design contributions. Finally thanks to Microsoft for supporting the work of the Open Data Policy Lab and its inquiry into the Third Wave of Open Data.

Table of Contents

Introduction

The Value of Open Data and Data Re-use

As Statistics Canada describes it, data refers to “observations that have been converted into a digital form that can be stored, transmitted or processed and from which knowledge can be drawn.” Every day, organizations collect all kinds of these observations to support their work. Public health agencies collect and analyze data to understand patterns of disease spread. Governments collect and analyze data to understand the impact of implemented programs and policies. Some research institutions collect and use data to study behavior.

Given the various ways data can be used, a rough consensus has emerged among data practitioners in support of open data. This movement supports expanding access to data already collected to enable new uses. Rooted in the notion of an information commons, it begins from the premise that data collected from the public, often using public funds or publicly funded infrastructure, should also belong to the public or, at the very least, be made broadly accessible to those pursuing public-interest goals. The types of data and methods through which it is shared have changed over time.

Open data practitioners have made significant progress in promoting universally and readily accessible platforms that allow data to be readily accessed, used, and redistributed free of charge. However, as The Open Data Policy Lab argued in its piece

on the third wave of open data, the movement is still evolving. New practices and institutions can allow the movement to be more impactful and to allow for more expansive forms of reuse. New practices can allow the conception of open data to expand to include data held by businesses, community-based organizations, academic institutions, and sub-national governments. It can allow open data to produce tangible benefits for an even larger group of stakeholders.

WHO IS THIS TOOLKIT FOR?

This toolkit is meant to be a resource to data stewards, responsible data leaders at public, private, and civil society organizations empowered to seek new ways to create public value through cross-sector data collaboration. Building on the findings of the Open Data Policy Lab's previous report, The Emergence of a Third Wave of Open Data, it provides them with ways they can foster a third wave built around equitable, impactful data re-use.

WHY DID WE DESIGN THIS TOOLKIT?

This toolkit is the result of several months of research and conversations with data practitioners from around the world during the Summer of Open Data. These discussions revealed significant barriers that prevented organizations from scaling open data and data collaboratives. Many of these barriers were organizational and contributed to a broader data ecosystem that replicated challenges writ large.

This toolkit aims to help organizations deal with these challenges so they can foster more data re-use within their organizations and encourage broader responsible access across the domains they work. It achieves this goal by offering a framework to think about data re-use, one that starts from central questions about how data is created before expanding outward. It also offers eight actions data stewards to foster re-use and specific ways they can change their day-to-day operations to make these actions possible.

This toolkit does not intend to provide specific recommendations as to data management practices. It does not stipulate specific ways that data can be collected, stored, or accessed. Recognizing the disparate conditions under which data stewards provide, it also does not aim to respond to any organization's specific needs. Instead, it aims to provide general principles that data leaders can implement at the organizations they oversee.

The Third Wave of Open Data Framework

Making Sense of the Present and Future Data Ecosystem

Open data is the result of many processes all aligned toward data re-use. These individual processes can be hard to understand when viewed together. However, by separating stages out—identifying the ways the data lifecycle plays into data collaboration, the way data collaboration plays into the production of insights, the way insights play into conditions that enable further collaboration, and so on—data stewards can promote better and more impactful data management.

By analyzing the data they collect or gain access to, organizations can produce insights that can help them take more informed actions. If done responsibly, organizations can also create societal value by making their data available for re-use by other parties. This analysis starts with:

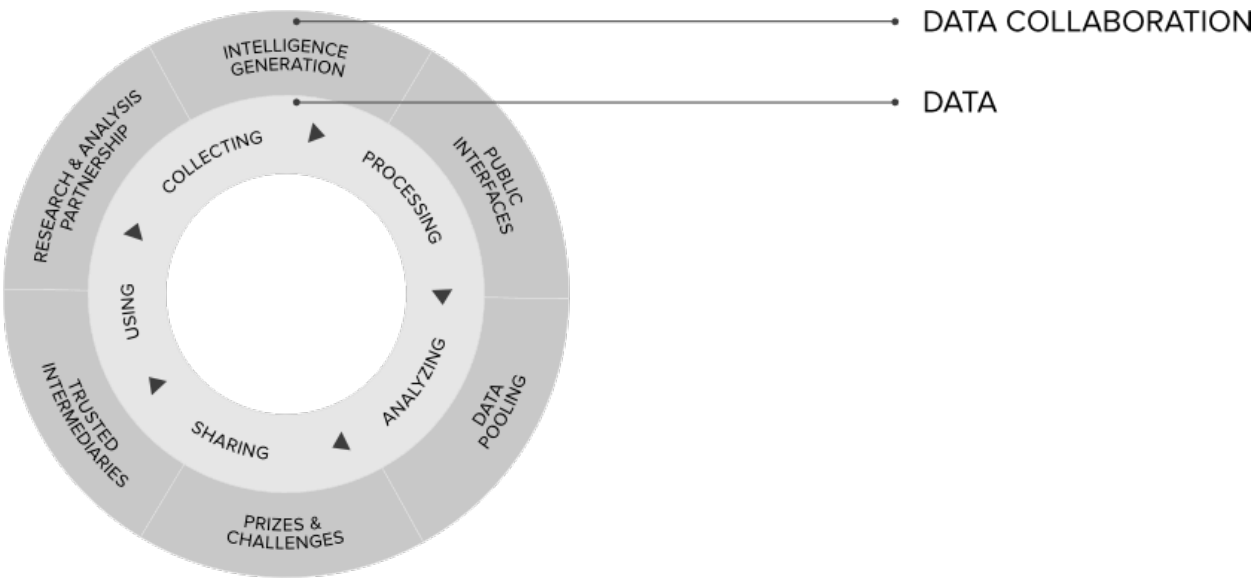
THE DATA LIFECYCLE

Data—which can be generated from things like smartphones, scientific studies, and financial transactions—has proliferated and is now an instrumental part of our modern world. Much of the increase in data is the result of datafication, the surge in data collection and storage that has resulted from exponential increases in networked computing and the rise of mobile phones and other digital devices with embedded processors. This phenomenon has created new opportunities to study complex problems but also has exposed people to new risks related to privacy and surveillance.

Though it can be used responsibly or poorly, data is not a simple asset itself. is the result of a process known as the data life cycle. This data life cycle includes:

- **Collecting:** The data practitioner gathers data from sources such as surveys, censuses, voting or health records, business operations, web-based collections, and other relevant, accessible sources. This data practitioner can be an individual, group, or organization.
- **Processing:** The data practitioner cleans the data to make it usable. Processing can entail removing irrelevant or inaccurate information, reformatting contents to be interpretable by an analytic software, and otherwise validating the data collection.

- **Sharing:** The data practitioner accesses the collected and processed data with relevant collaborators with the intent of deriving insights from it.
- **Analyzing:** The data practitioner assesses and interprets the data collection with a goal of extracting insights about the issue they are studying.
- **Using:** Finally, this data practitioner then acts on the insights derived. These actions can affect data collected for future operations.



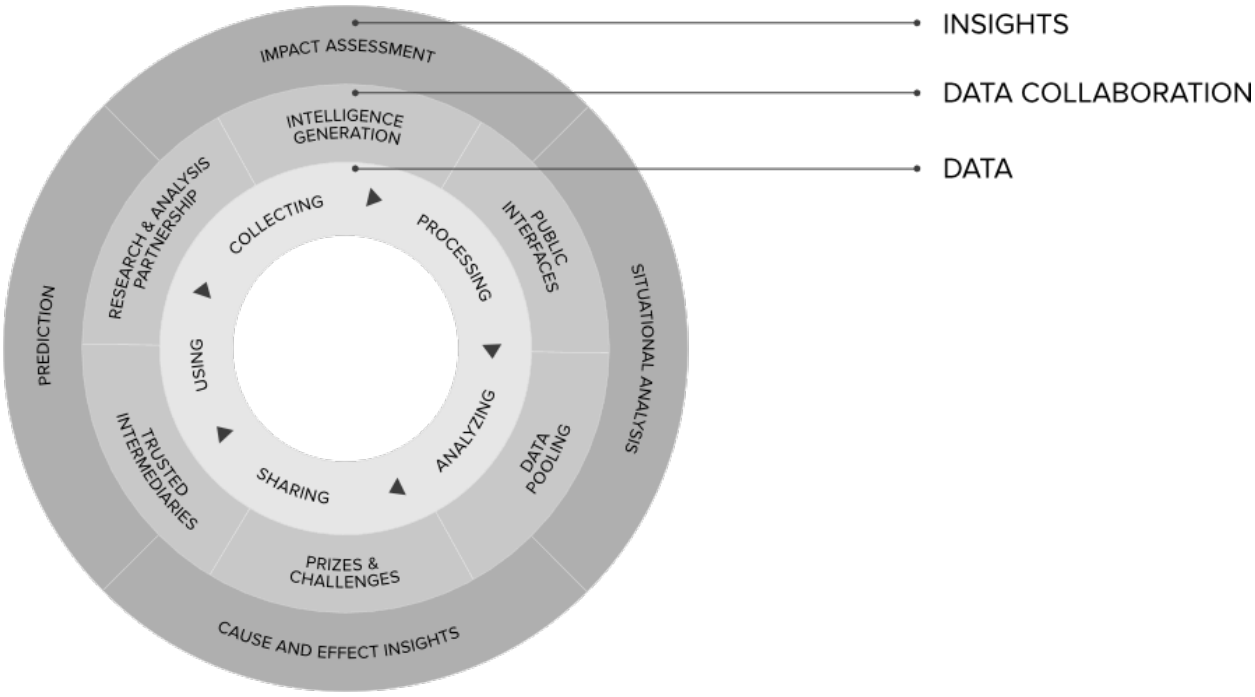
INCREASING ACCESS TO DATA THROUGH DATA COLLABORATION

Though we live in an era of data abundance, we also live in a period marked by tremendous asymmetries when it comes to data access. Much of the data generated today resides in silos controlled and often monetized by companies and others. New models for collaborating and accessing public and private-sector data, such as open data platforms or data collaboratives, can break these silos.

Data collaboratives are a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors exchange their data and data expertise to create public value. When done responsibly, data collaboratives allow for more complete, rigorous, or detailed analysis that supports public interests. As The GovLab discusses in its report, Leveraging Private Data for Public Good: A Descriptive Analysis and Typology of Existing Practices, data collaboratives can take the form of:

- ***Public Interfaces:*** Organizations provide open access to certain data assets, enabling independent uses of the data by external parties.

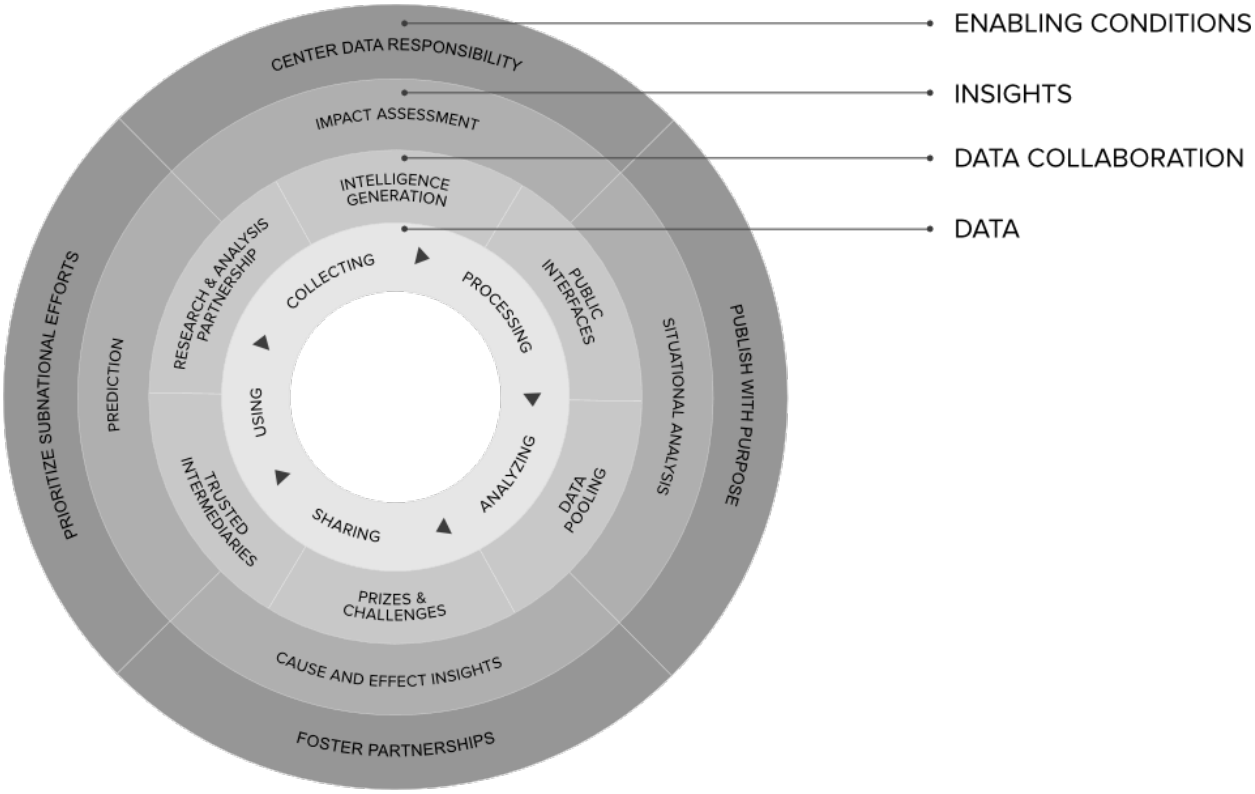
- ***Trusted Intermediary:*** Third-party actors support collaboration between data providers and data users from the public sector, civil society, or academia.
- ***Data Pooling:*** Data holders agree to create a unified presentation of datasets as a collection accessible by multiple parties.
- **Research and Analysis Partnership**: Organizations engage directly with public-sector partners and share certain proprietary data assets to generate new knowledge with public value.
- ***Prizes and Challenges:*** Organizations 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.
- ***Intelligence Generation:*** Organizations internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public.



INSIGHTS

Insights describe the findings that emerge from analysis of data. This analysis can be used to look forward or backward, providing insights about:

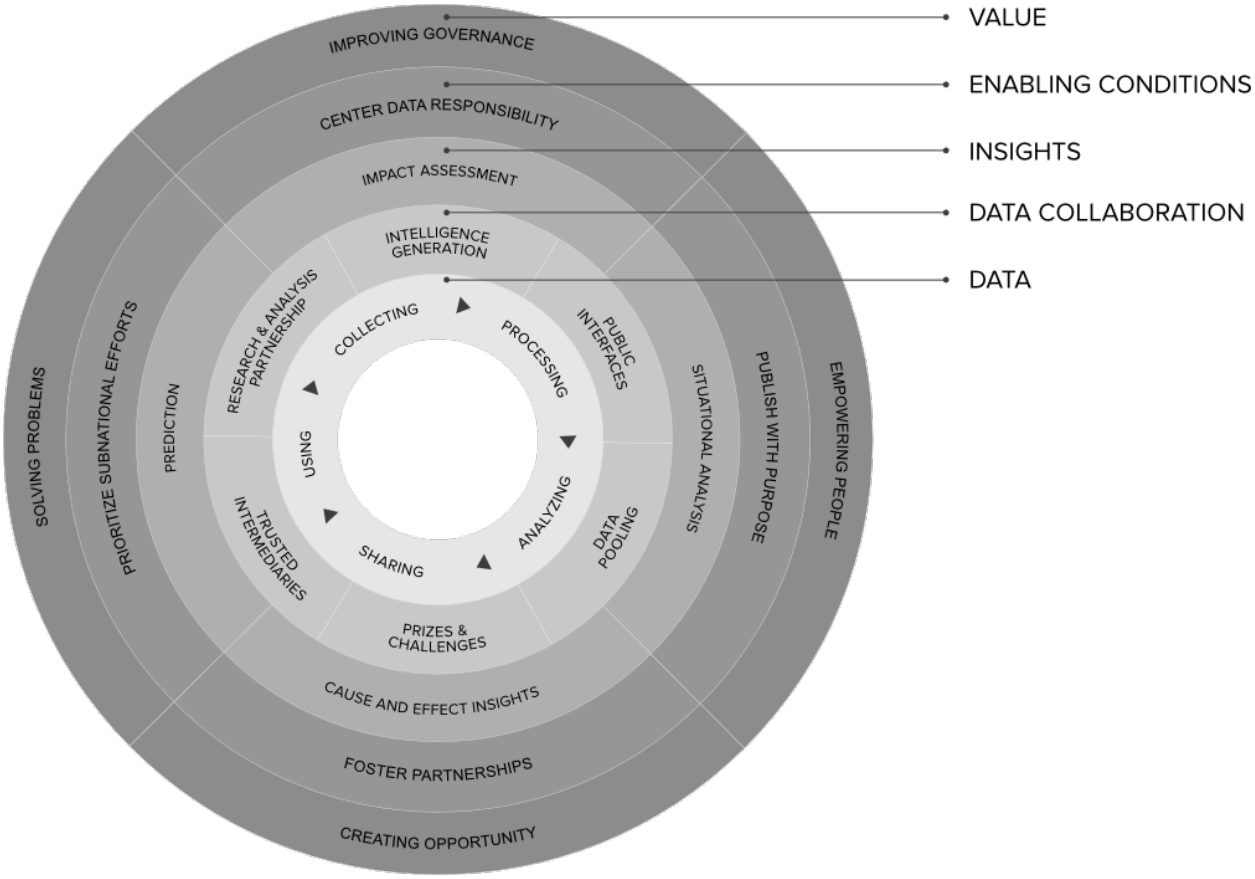
- ***Situational Awareness***: Answering what happened;
- ***Cause and Effect Insight***: Answering why it happened;
- ***Prediction***: Answering what will happen; and
- ***Impact Assessment***: Answering what should happen.



ENABLING CONDITIONS

While individual data projects can produce certain insights, broader changes to the data ecosystem to make data re-use and collaboration an effective tool to address public needs. These changes can be supported by organizations taking deliberate steps to be more open. As argued in the The Emergence of a Third Wave of Open Data, a more open ecosystem can be enabled by organizations that:

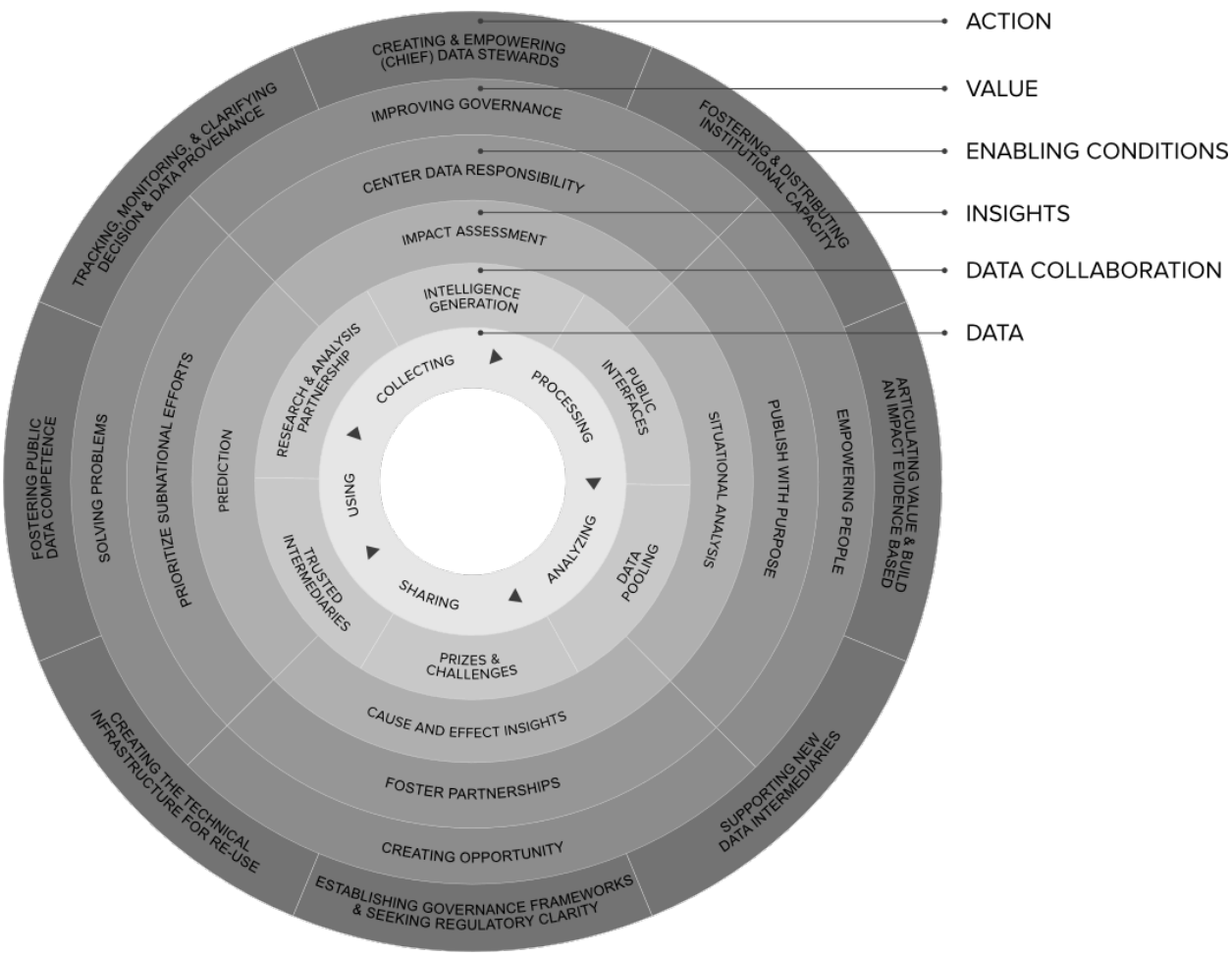
- **Publish with Purpose** by matching the supply of data with the demand for it, providing assets that match public interests.
- **Foster Partnerships** by forging relationships with non-professionals (e.g. small businesses and civic groups) who understand how data can inform meaningful real-world action.
- **Prioritize Subnational Efforts** by providing resources to cities, municipalities, states, and provinces to create new subnational data sources.
- **Center Data Responsibility** by promoting fairness, accountability, and transparency across all stages of the data lifecycle.



VALUE

When done effectively, responsibly, and in accordance with local expectations, data re-use can deliver real-world value across four broad categories:

- ***Improving Governance***: Insights from data can improve how organizations operate by making their processes more transparent to others, improving resource allocations, and enhancing their ability to deliver services.
- ***Empowering People***: Insights can empower people by communicating information they need to meaningfully act and make decisions about the challenges they face in their lives.
- ***Creating Opportunity***: Data-driven insights can inspire organizations to innovate in how they operate. For businesses, this innovation can be about identifying new business models while governments might use information to inform policy directed at economic well-being.
- ***Solving Public Problems***: Insights can optimize processes and services and better identify the needs of those who rely on those services. It can enable data-driven assessments of the environment and more targeted interventions.



RIDING THE THIRD WAVE OF OPEN DATA: PRIORITY ACTIONS

Lastly, organizations can capitalize on the emerging Third Wave of Open Data by taking eight key actions:

- **Creating and Empowering (Chief) Data Stewards**: Developing and nurturing responsible data leaders to support impactful data re-use.
- **Fostering and Distributing Institutional Capacity:** Taking steps to avoid consolidating and siloing of data skills and resources, and instead catalyzing such capacity to filter into daily institutional operations;
- **Articulating Value and Build an Impact Evidence Base**: Demonstrating the concrete, tangible value of increased access to and re-use of data;
- **Supporting New Data Intermediaries**: Engaging actors who can lower transaction costs in data collaborative relationships
- **Establishing Governance Frameworks and Seeking Regulatory Clarity:** Creating safeguards to mitigate risks of harmful outcomes;

- **Creating the Technical Infrastructure for Re-use:** Investing in innovative and sophisticated technologies to improve data use on data supply and demand sides;
- **Fostering Public Data Competence**: Engaging citizens to promote wider use of data informed by local contexts and priorities;
- **Tracking, Monitoring, and Clarifying Decision and Data Provenance**: Capturing data-handling and decision-making processes to ensure coordination;

In what follows, we provide brief primers on each of these priority actions to support users' efforts to maximize the institutional and societal impact of data re-use by riding the Third Wave of Open Data.

PRIMER #1:

CREATING AND EMPOWERING (CHIEF) DATA STEWARDS

Developing and nurturing responsible data leaders to support impactful data re-use

WHY

The Third Wave brings with it a focus on the institutional arrangements that need to be reimagined to achieve a more data-driven manner of working. This reimagination focuses particular attention on the role of the data steward, responsible data leaders that seek new ways to create value through cross-sector data collaboration. Data stewards may be either individuals or groups of individuals within an organization who are dedicated employees that initiate and contribute to sustainable data collaboration.

HOW

Designate a Common Authority: Often, data efforts are dispersed across an organization, making it difficult to ensure that all its components are working towards the same ends. A data steward, by contrast, serves to coordinate these disparate efforts and ensure that staff are not working cross purposes. By designating a singular team or individual to manage data collaboration, organizations can promote effective data practices.

Systemize Sharing: Data stewards are not only focused on technical aspects of data management but are empowered to engage in data sharing and data collaboration. They achieve this goal by filling three responsibilities. First, they collaborate with stakeholders and promote partnerships. Second, they facilitate the responsible management of data. Third, they ensure that relevant parties act upon the generated insights. Organizations

can ensure the success of stewards by granting them the institutional mandate and authority to complete these tasks.

Codify Data Stewardship in Policies and Procedures: Open data and data re-use pilot projects often fail to reach sustainability or scale due to several factors. These factors can be related to shifting institutional leadership, priorities, or funding. If data stewardship structures and practices remain ad hoc and are not codified in institutional policies and procedures, they risk being abandoned or scaled back if circumstances change.

MORE

The Data Stewards Network: A community of practice connecting responsible data leaders across sectors to facilitate cross-sector collaboration and generate new ways of creating public value, and a weekly newsletter sharing new insights, tools, and developments to support data stewards' work.

Wanted: Data Stewards: A position paper released by The GovLab which tackles questions regarding the profile and potential of Data Stewards.

Towards a European strategy on business-to-government data sharing for the public interest: The final report from the European Commission High-Level Expert Group on Business-to-Government Data Sharing, which highlights the need for data stewards across sectors to enable societally beneficial public-private data collaboration.

PRIMER #2:

FOSTERING AND DISTRIBUTING INSTITUTIONAL DATA CAPACITY

Taking steps to avoid consolidating and siloing of data skills and resources, and instead catalyzing such capacity to filter into daily institutional operations.

WHY

Organizations across sectors can increase the societal and organizational value created through data re-use by bolstering their personnel's data skills and ensuring those skills are distributed throughout the organization. When data skills and resources are relegated to small teams or units, organizations are unlikely to maximize the societal and organizational value of data re-use. Instead, capacity needs to be distributed evenly to ensure people in all parts of the organization understand the data they have, can use it to create value, and are willing to forge internal and external relationships around it. Focused efforts to invest in, foster, and distribute data skills can help an organization become more evidence-based and systematic across all its operations.

HOW

Building an Organizational Culture of Learning: The state of the art in data science and data stewardship is constantly evolving. New technologies and methodologies regularly emerge. Decision-makers can help their organization achieve the benefits of the Third Wave of Open Data by championing professional development and training programs. They can also enshrine policies, procedures, and support mechanisms to encourage personnel to bolster their data skills. These skills can be technical in nature, involving

training to collect, analyze, and derive insights from large datasets. These skills can also be related to issues of governance, related to defining problems, promoting data ethics and responsibility across an organization, or communicating and contextualizing data-driven insights.

Breaking Down Vertical and Horizontal Silos: Open data concerns itself with breaking down silos and increasing access to data for decision-making. Organizations can take steps to break down barriers between *vertical* domains, those silos preventing data being used across sectors toward a singular purpose. They can also take steps to break down barriers between *horizontal* silos, those barriers between disciplinary backgrounds that can stymie collaboration among actors working toward similar goals . Organizations stand to benefit from opening cross-departmental communication channels, creating interdisciplinary teams, and initiating peer-learning opportunities to ensure that data capacity feeds into normal business operations.

Nurturing Senior-Level Data Competency: Decision-makers who are responsible and accountable for setting organizational priorities can benefit from data science know-how. Even if these leaders are not likely to manipulate or analyze data, they will be better-positioned to make informed decisions in data investments if they are conversant in data science and data stewardship principles and practices.

MORE

Open Data Policy Lab Data Stewardship Training*:* Course offerings from The GovLab's Open Data Policy Lab initiative aimed at supporting decision-makers in the creation of a data re-use strategy.

Solving Public Problems with Data: The GovLab's series of online lectures from data experts on important data skills and competencies.

Open Data Institute Courses*:* Data competency training programs from the Open Data Institute.

Data Skills for Public Servants: A collection of openly accessible data competence training resources and course offerings curated by *apolitical*.

PRIMER #3:

ARTICULATING VALUE AND BUILDING AN IMPACT EVIDENCE BASE

Demonstrating the concrete, tangible value of increased access to and re-use of data

WHY

In the past, open data advocates have tended to argue for increased data re-use by relying on normative arguments. They noted how open data could enable greater transparency and provide accountability. While these arguments can be persuasive in some contexts, private-sector leaders, government officials, and the public sometimes need to understand how open data investments will tangibly benefit them. Otherwise, open data becomes another "nice-to-have" instead of an immediate need. In these circumstances, it is often better to appeal directly to personal or organizational interests, to provide simple explanations of how open data will support an actor's short, medium, and long-term goals.

HOW

Identify clear, specific use cases: A well-defined problem leads to targeted solutions where it is possible to understand who data work will help and how. After identifying a general issue it wants to improve (e.g. adolescent mental health or climate change), an organization can refine its focus toward an actionable problem and have in mind a clear, measurable outcome they intend their work to produce. Organizations might find it helpful to frame their concern as a question, one answerable through data science. They may find it useful to use participatory question formulation processes that allow stakeholders to develop questions they would likke answered.

Assess and segment demand: Data's value can often seem ephemeral to those with immediate needs, lacking direct impact. Organizations can avoid this sense by trying to identify which specific organizations data re-use work might benefit and engaging directly with them. By understanding which government actors, businesses, and nonprofits would gain, organizations can build a constituency determined to see a project to fruition.

Offer opportunities to use and contribute to datasets: Data re-use work can be a narrowly organized effort, conducted by only a handful of senior analysts. By limiting the user base, however, organizations also limit the effort's supporters, making it harder to launch and easier to end. When organizations allow data assets to be used by a large cohort, whether that be the public or another audience, they can build a constituency that can identify innovative new uses for data and are committed to seeing the project's success.

MORE

Open Data's Impact: A collection of resources from The GovLab examining the global impacts of open data, including tools, books, and 37 detailed case studies of illustrative projects.

Data Collaboratives Explorer: A repository of over 250 data collaboratives initiated around the world with the goal of creating new public value.

Data Collaboration for the Common Good: A report from the World Economic Forum outlining value propositions for public-private data partnerships and strategies for maximizing that value.

The 100 Questions Initiative: An exercise, relying on expert and public input, that seeks to identify important societal questions whose answers can be found through data science and collaboration.

Open Data Demand: Toward An Open Data Demand Assessment and Segmentation Methodology: A methodology that guides users through the process of identifying who would benefit from the insights open data can generate.

PRIMER #4:

SUPPORTING NEW DATA INTERMEDIARIES

Engaging actors who can lower transaction costs in data collaborative relationships

WHY

Matching the supply of data with those who demand it can be costly in terms of time, resources, and staff. Organizations need to identify relevant partners, develop the data infrastructure and capacity necessary to handle new information flows, and negotiate legal agreements. Any one of these actions can be difficult for an organization—especially a small one—and can dissuade data collaborative efforts. Consequently, third party organizations devoted to facilitating collaboration can be useful in addressing some or all of these issues. Data intermediaries can be important actors in the Third Wave by making the data value chain more fluid, working to facilitate data collaboration and lowering transaction costs between those supplying the data and those using it.

HOW

Assessing Challenges in Matching Supply with Demand: To realize the potential of data re-use and data collaboration, it is important to ensure that data supply is aligned with the user demand for it. Given that resources are often limited, it is important for organizations to identify priority areas where open data can provide the most benefit. This prioritization can involve engaging with relevant stakeholders to understand their needs. Open innovation challenges, public events, and data education programs are some of the mechanisms which organizations might use to engage with those involved in demanding data.

Determining What Types of Capacity Could Add Value: Organizations often exist in a world of constrained resources, lacking domain expertise, available funding or institutional connections. Understanding these gaps in capacity can be instrumental for data collaboratives as it can help organizations know what resources they need to seek out. Data intermediaries can empower organizations and support this work by providing necessary resources, whether that be secure platforms to collaborate and share data or support in drafting a data-sharing agreement.

Identifying Trusted Third Parties: Third parties who can create, mix, source and curate data from various sources can be useful in any data effort. However, these parties cannot support data collaboration if the participants in the data collaborative do not trust them to deliver requested work. Organizations might consider conducting due diligence on potential intermediary partners to ensure they possess the requisite experience, capacity, and ethical grounding.

MORE

What Are Data Institutions and Why Are They Important?: An overview from the Open Data Institute on benefits, models, and examples of intermediary institutions enabling the re-use of data in the public interest.

Guide to Starting a Local Data Intermediary: A step-by-step guide from the National Neighborhood Indicators Partnership on defining needs, opportunities, funding streams, and sustainability strategies for subnational data intermediaries.

Understanding MyData Operators: MyData's introductory paper which discusses the characteristics and roles of data operators, one type of data intermediary.

Open Data Intermediaries: Their Crucial Role: A piece from the Web Foundation explaining what open data intermediaries and the value they can provide.

PRIMER #5:

ESTABLISHING GOVERNANCE FRAMEWORKS

Creating safeguards to mitigate risks of harmful outcomes

WHY

Even 20 years into the digital revolution, rules and regulations dictating data usage remain sparse. While this failure offers additional flexibility to some, the broader effect is greater uncertainty about risk. Organizations do not understand how to respond coherently when data is abused or misused. Ad hoc responses to daily challenges have eroded public trust and damaged institutional credibility. Organizations have an obligation to create and foster safeguards to protect themselves and data subjects from harm. Institutionally, these safeguards can take the form of clear, public organizational policies about when and how an organization will share and re-use data. Societally, organizations can encourage governmental, inter-governmental, and international bodies to regulate data or create portals to facilitate data re-use.

HOW

Determining a Fit-For-Purpose Governance Model: Organizations have different needs depending on the issues they work on and the people they work with. Before starting a data project, organizations face the challenge of developing governance mechanisms that can reflect their specific circumstances. If working on a sensitive topic, an organization might seek out experts who can support ethical decision-making. If working on an issue that has implications for the general public, an organization might seek public comment and engagement.

Codifying End-to-End Data Responsibility: Data is not static but exists on a cycle. As part of a commitment to data responsibility, actors can assess and seek to prevent risks across the full data life cycle, including the collecting, storing and preparing, sharing, analyzing and using stages. This concept is called end-to-end data responsibility. It is essential for preventing harm and ensuring trust.

Identifying (or Creating) an Appropriate Data License: Data licensing regimes, as such, provide a way to secure and promote the re-use of data, either among a set of actors or among the public. They promote this re-use and sharing by requiring interoperability, articulating permissions and conditions around use, redistribution, modification, separation, compilation, non-discrimination, propagation, and/or application. Selecting a fit-for-purpose data license requires assessing different licensing regimes benefits and challenges, and could potentially involve the development of a new, customized data license to meet organizational needs.

MORE

The Responsible Data Re-Use Framework: The Responsible Data Re-Use Framework is a report seeking to identify appropriate data re-use standards amid COVID-19 in New York. It is the product of deliberations with data holders, civil rights advocates, and the public.

Contractual Wheel of Data Collaboration: A visual representation of the key elements involved in data collaboration developed by The GovLab.

Leveraging Private Data for Public Good: A report outlining the different structures through which data collaboration can take place.

Laying the Foundation for Effective Partnerships: A report describing good practices in creating fit-for-purpose Data Sharing Agreements to support responsible cross-sector data collaboration.

Report on Collected Model Contract Terms: A report from the EU Support Center for Data Sharing outlining common strategies and best practices for data licensing.

United Kingdom Data Ethics Framework: Guidance from the United Kingdom's government digitall service outlining how public sector organizations might appropriately and responsibly use data

PRIMER #6:

CREATING THE TECHNICAL INFRASTRUCTURE FOR RE-USE

Investing in innovative and sophisticated technologies to improve data use on data supply and demand sides

WHY

Open data portals have been key in enabling open data, combining various institutional datasets and allowing users to browse, filter, search, and download data to their machines. While the open data portal format will likely remain a common piece of technical infrastructure, new and sophisticated technological developments could facilitate greater collaboration and responsibility in data re-use. These developments could include improved computing capacity to analyze large datasets and new and secure ways of transmitting data. To facilitate this improved technological development, an intersectoral, multidisciplinary research and development effort will be useful.

HOW

Unlocking Funds and Resources: Technological innovation and infrastructure development are often cost-intensive exercises with extended time frames. Organizations can look to various internal and external sources of funding to develop the technical infrastructure necessary for systematizing impactful and responsible data re-use.

Prioritizing Purpose-Driven Innovation: With the support of governments and foundations, data scientists and researchers can co-design and co-develop technologies needed to

implement data collaboration at scale and in a responsible and sustainable way. This collaborative research could focus initially on core needs such as privacy-preserving technologies, security technologies, and access-control technologies.

Experimenting with New Innovations in Responsible Data: Several recent technical innovations are helping organizations to ensure safe and responsible re-use of data. New tools such as differential privacy, releasing information about a dataset without sharing personal or classified details it contains; and synthetic data, artificially created replica data; and safe sandboxes, tightly controlled data processing environments, can help support data re-use while mitigating risks.

MORE

State of Open Data: Data Infrastructure: A chapter from *The State of Open Data* explaining key elements of data infrastructure as it pertains to open data using an analogy to physical infrastructure.

Where Is Data Sharing Headed?: A piece from the Boston Consulting Group introducing several technical models to facilitate inter-organizational and cross-sector data sharing.

Differential Privacy for Privacy-Preserving Data Analysis: A blog series from the U.S. National Institute of Standards and Technology providing an introduction to differential privacy and practical implementation guidance.

Data Sharing and Interoperability Through APIs: Insights from European Regulatory Strategy: A paper by Oscar Borgogno and Giuseppe Colangelo which examines the importance of APIs within data sharing ecosystems.

PRIMER #7:

FOSTERING PUBLIC DATA COMPETENCE

Engaging citizens to promote wider use of data informed by local contexts and priorities

WHY

Facilitating greater data competence within the general public is an important step to ensuring that it could receive greater benefits from data re-use, as well as face fewer risks from it as data subjects. To advance full participation of the general public in data efforts there is a need to foster its data competence, going beyond the fundamental need for data literacy. This will help to bridge the gap between the public and the data ecosystem so that the public could both participate in and contribute to data efforts. This approach can provide the means necessary to address the persisting differences in power in the current data and digital era, as well as guarantee novel productive capacity while enabling creativity. Consequently, empowering the public to see itself as a producer of data, will put it in the 'position to negotiate' the ways in which data is re-used by different stakeholders.

HOW

Obtaining a Social License: Trust and legitimacy are key in the planning processes pertaining to data re-use. In order to ensure that data re-use initiatives create public good, they need to obtain a 'social license'. This means exercising the necessary due diligence and engaging with all relevant stakeholders to ensure that data re-use is aligned with public and stakeholder concerns and expectations. To make sure that data and technology are used responsibly, it is important that both the benefits and the risks associated with them are evaluated by local stakeholders.

Contributing to the Public Knowledge Base: Addressing modern challenges requires more than just basic digital literacy, but data competence. Organizations can support

efforts to foster public data competence by making not just data accessible, but also supporting documentation such as data dictionaries, glossaries and nutritional labels for data, and using plain language rather than jargon when communicating insights and lessons learned to the public.

Taking a People-Led Approach to Engagement: Taking a people-led approach means engaging in problem-solving processes that are centered around people and groups who organize and mobilize them. The approach consists of four phases: defining problems, definitions and priorities; ideating solutions by engaging with experts and leveraging data; experimenting and testing the solutions; and finally, expanding impact by sharing relevant lessons learned with others who may benefit from them. Some of the different types and groups of people which could be engaged under this approach include residents, domain experts, NGOs, community-based organizations, local business owners, anchor institutions, resource partners, and municipal government officials and civil servants.

MORE

How can stakeholder engagement and mini-publics better inform the use of data for pandemic response?: A piece written by The GovLab for the OECD's Participo blog explaining how The GovLab has used mini-publics to engage residents of New York on data re-use amid crisis events.

Data Infrastructure Literacy: A Big Data & Society paper which highlights the importance of digital literacy for public engagement with data infrastructures.

Data Literacy: In this State of Open Data chapter, Mariel Garcia Montes and Dirk Slater discuss the importance of data literacy in achieving organizational data maturity and realizing the full potential of open data.

People-Led Innovation: A series of tools, probing questions, and inspirational examples aimed at providing practitioners with a flexible guidebook for experimenting with new ways to solve public problems in an iterative, participatory manner.

Beyond Data Literacy: A Data-Pop Alliance white paper that outlines strategies for inclusive engagement and empowering individuals to effectively navigate their data and information ecosystems.

PRIMER #8:

TRACKING, MONITORING, AND CLARIFYING DECISION AND DATA PROVENANCE

Capturing data-handling and decision-making processes to ensure coordination

WHY

Data and decision provenance are key to reducing data risks while re-using data in the most impactful way. By identifying all the decisions that impact the collection, processing, analyzing, sharing, and re-use of data and the parties that impact these decision points, organizations can understand why systems exist as they do and react appropriately when systems do not produce intended results. Decision and data provenance bring to light gaps in data analysis strategies and can lead to improved awareness of the factors that influence decision-making.

HOW

Implement Processes to Identify Blindspots in Current Decision-making Processes: Implementing decision provenance does not require complex technical interventions—a common concern for organizations, especially smaller or less-resourced ones. Instead, process-oriented approaches, such as systems and constituency mapping, data responsibility journey mapping, decision trees, and other record keeping methods can provide frameworks to bring greater transparency to the decision-making process.

Determine Who Is Responsible vs. Who Is Accountable: Distinguishing between those working with the data and those who determine the strategies for its use is important. It allows organizations to implement mechanisms that can address harms where they occur

and ensure accountability for their actions throughout the process of collecting, transferring, storing, processing, analyzing, and using data.

Communicate Decision Provenance to Relevant Internal and External Stakeholders: Decision and data provenance processes can be useful in identifying the influences on data, but these systems are useless if the people who can use them are unaware of them. Organizations need to ensure that all those stakeholders involved in data processes are tracking their decisions and that all those stakeholders involved in assessing applications of data are aware that a paper trail of decision-making exists.

MORE

Decision Provenance: Harnessing Data Flow for Accountable Systems: A paper by Jatinder Singh, Jennifer Cobbe, and Chris Norval, which introduces the concept of decision provenance and explores its potential through a tech-legal perspective.

Decision Provenance Mapping: A methodology, developed in the context of the Responsible Data for Children Initiative and The GovLab, supporting users in capturing decision-making processes and assessing .

Towards a Sociology of Institutional Transparency: Openness, Deception and the Problem of Public Trust: An article by Sara Moore which examines open government initiatives and emphasizes the importance of trust in achieving government openness.

Data Lineage: A Wikipedia article describing the term, which includes “data origin, what happens to it, and where it moves over time.


Open.Data.Policy.Lab
opendatapolicylab.org
THEGOVLAB
thegovlab.org