

The Periodic Table of Open Data: A User's Guide

Policies and Practices for Enabling the
Third Wave of Open Data

August 2022



Open.Data.Policy.Lab

Periodic Table of Open Data Elements: Explanation and Rationale

Introduction:	3
Periodic Table of Open Data Elements:	4
Diagnostic Tool:	5
Problem and Demand	5
Capacity and Culture	7
Governance and Strategy	10
Personal and Partnerships	12
Risk Mitigation	13
Detailed Explanations:	16
Problem and Demand	16
Capacity and Culture	19
Governance and Strategy	24
Personal and Partnerships	27
Risk Mitigation	29



Introduction

Open data can be a valuable resource for society. When previously restricted data from the public, private, and academic sectors is made available for re-use by a broader range of actors, it fuels insights that would not have been possible otherwise, informing decisions on pressing social issues. This can empower citizens in making governments and businesses more accountable, enable disaster relief, and support people's livelihoods. In short, open data not only has normative value, but it can also lead to more innovative approaches to developing solutions to pressing public problems.

The question then is: *What makes open data "perform well"?*

At The Open Data Policy Lab (a collaboration between The GovLab and Microsoft), we not only seek to document the impact of open data initiatives around the world but we also try to identify the elements that lead to improved impact. This led to the exercise we undertook in 2016 when we sought to construct a "periodic table" that would list the enabling conditions and disabling factors that play an important role in determining the impact of open data initiatives. We were able to identify five central issue categories (as well as a number of sub-categories) that frequently play a key role in shaping the success or otherwise of open data projects across countries and sectors. Since then, we have validated the elements and complemented our research through numerous stakeholder interviews, workshops, and an on-going in-depth review of the literature.

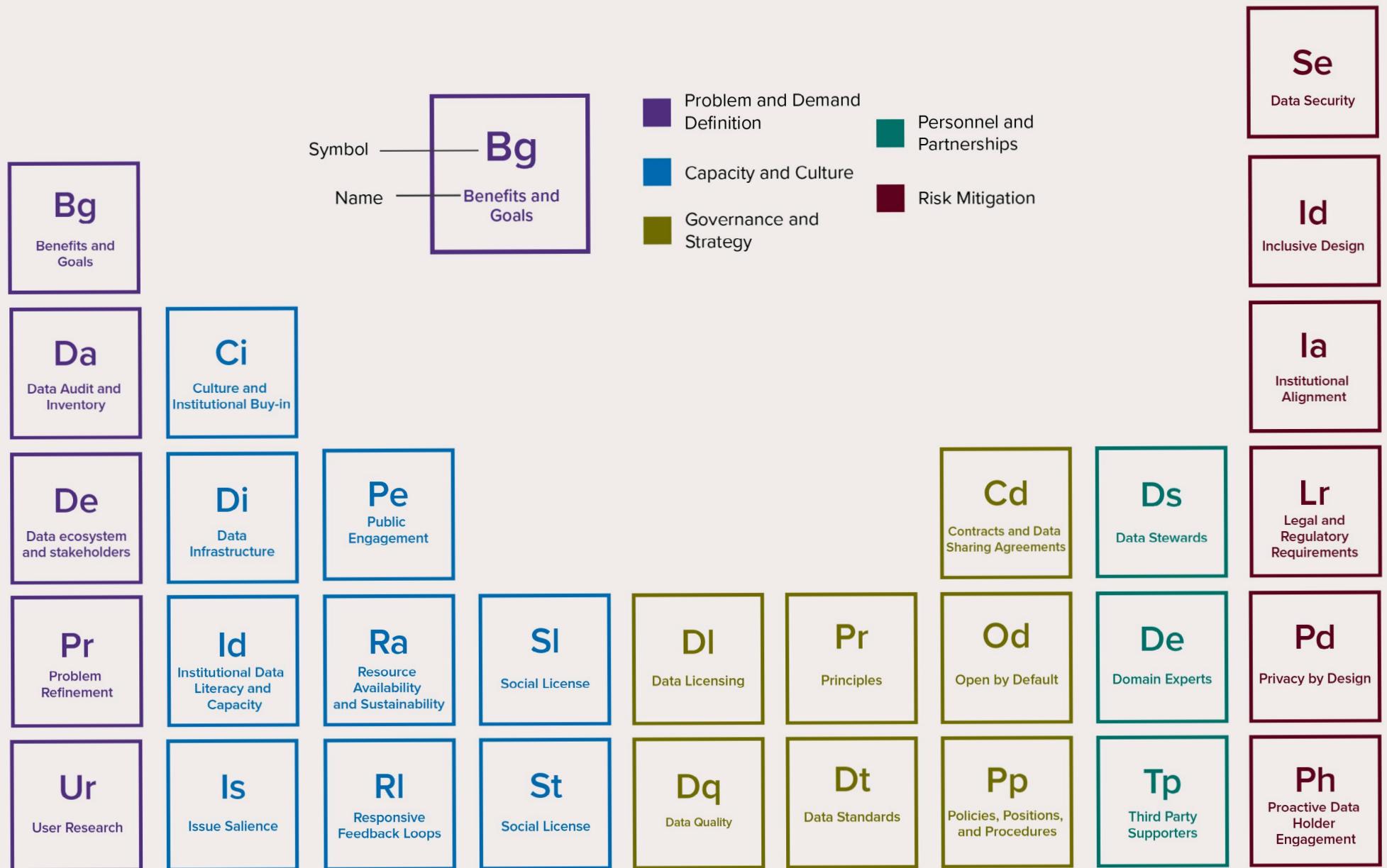
Taken together these elements provide a conceptual framework to diagnose when initiatives are successful, and when they are not. Such assessments are always valuable in their own right, but they can serve a further purpose too. By reverse architecting assessments of success (and failure) across initiatives, we can identify a number of conditions that can be built into the design of open data initiatives, thus improving their chances of achieving meaningful positive impact. This document seeks to provide such a practitioner's guide.

The Periodic Table of Open Data: A User's Guide summarizes some of the major issues that open data practitioners may want to consider as they develop strategies for releasing and re-using open data in the communities they serve. In the below you will find:

- **The Periodic Table of Open Data (Version 3):** depicting each of the elements of the Periodic Table of Open Data Elements and the categories in which they fit;
- **The Periodic Table of Open Data Diagnostic:** A worksheet that readers can use to assess if an element is present in their project and, if so, how much weight it should be given in terms of importance for the success of the open data initiative;
- **Detailed Explanations:** A longer, more detailed rationale behind each element emerging from evidence and research.

We invite practitioners to use this guide to identify ways they can promote the success of their open data initiatives or mitigate risks.

Periodic Table of Open Data Elements



Periodic Table of Open Data

Diagnostic

This section of the document is a diagnostic tool intended to help data practitioners understand which elements are present in their open data initiative and the extent to which they need to address each of these elements in their open data initiative. Practitioners can use the sheet as a checklist and tool to weigh the importance of each element to their overall effort. The importance given to each element is its “atomic weight” (which is always context specific) and suggests how much priority it should be given.

At the end of this document is a brief section where readers can indicate the total “weights” of each category. Sections with higher weights are those that practitioners can focus additional attention on while sections with lower weights can be deprioritized.

Problem and Demand Definition

Problem and demand definition include all those elements that can allow an open data project to be targeted and optimized. They are about increasing the return on investment and minimizing wasted effort.

Benefits and Goals

Have you articulated a tangible value that the project will produce for its sponsors and the public?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Data Audit and Inventory

Has a data audit been conducted to understand which data is collected and generated and which can still be made available?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Data Ecosystem and Stakeholder Assessment

Has the broader sector and data ecosystem been assessed to identify the data, expertise, and other resources that could be made available via partnership?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Problem Refinement

Have you refined the problem to identify a specific purpose or issue to solve, the data needed to produce insights, and how those insights can be acted upon to produce real public value?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

User Research

Have you used user research to aid in identifying, mapping, and understanding a problem, its components, or user needs?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Capacity and Culture

Capacity and culture refers to those elements that relate to an organization's ability to pursue open data. It involves intangibles—including how individuals inside and outside the organization view the project—and tangibles—such as specific hardware and funding sources.

Culture and Institutional Buy-in

Have organizational leaders bought into and come to actively support the open data initiative?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Data Infrastructure

Is your data infrastructure mature enough to facilitate the easy collection, storing, managing, and sharing of open data?

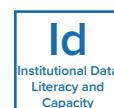


Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Institutional Data Literacy and Capacity

Have staff across the organization been trained on how they can use open data and data reuse to address the issues facing them?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Issue Salience

Has the open data project been connected to some issue of public or institutional interest that can be used to sustain active support?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Public Engagement

Have you promoted the open data initiative to the public in a way that conveys what it is and how it can meaningfully help them?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Resource Availability and Sustainability

Have you secured the internal and external sources of funding needed to systemize impactful and responsible open data?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Responsive Feedback Loops

Have you created mechanisms for data users and beneficiaries to provide input on an open data initiative?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Social License

Have you sought out the public's input and approval for the open data initiative which you are seeking to launch?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Strategic Leadership

Do you have an open data project leader who can enact policies, mobilize resources, and advocate for data openness?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

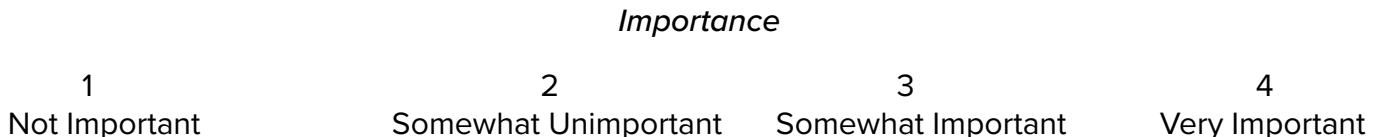
Governance and Standards

Capacity and culture refers to those elements that relate to an organization's ability to pursue open data. It involves intangibles—including how individuals inside and outside the organization view the project—and tangibles—such as specific hardware and funding sources.

Data Licensing

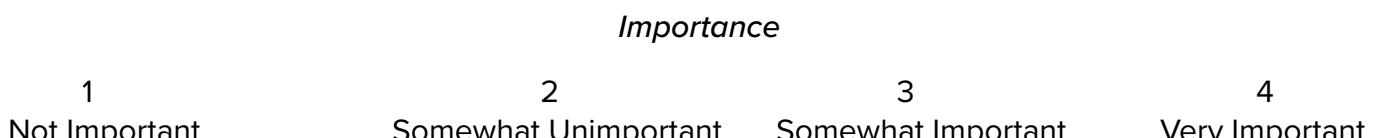


Is there a robust data licensing regime that can protect and promote the re-use of data by outlining the conditions under which practitioners can use and re-use data?



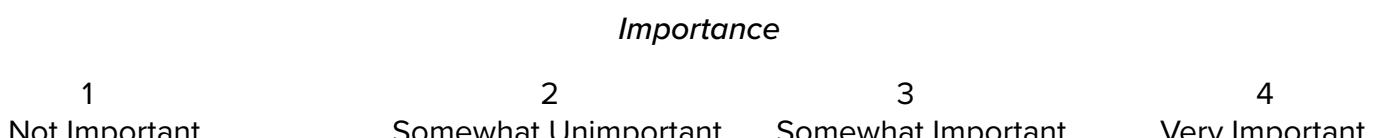
Data Quality

Has the data been assessed for quality to allow individuals to derive meaningful insights from it?



Principles

Do the organizations working together on the open data effort have principles that align with one another and the work they are pursuing?



Data Standards

Have technical specifications—such as machine readability and data portability—for the datasets been laid out to allow it to be easily shared, exchanged, and combined with other assets?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Contracts and Data Sharing Agreements

Is there a contract or agreement that, in clear and specific terms, establishes what the roles and responsibilities of each party in the open data initiative are and how data will be made open?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Open by Default

Has data been made “open by default” to reach as many people as possible?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Policies, Positions, and Procedures

Are there policies, professional positions, and procedures that guide how individuals can best contribute to the open data initiative?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Personnel and Partnerships

These elements relate to the specific actors—both individuals and institutions—whose support can enable an open data initiative’s success.

Data Stewards

Are there data stewards present who are empowered to seek new ways to create public value through cross-sector data collaboration?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Domain Experts

Are there individuals with knowledge of a sector or domain who can be called upon to better identify opportunities to (re)use data for the public’s benefit or otherwise help identify, design, analyze, and communicate insights about the open data?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Third Party Supporters

Are there other organizations, institutions and individuals invested in a project’s success who might be useful in building a broad base of support for an open data effort?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Risk Mitigation

Risk mitigation elements are all those other practices adopted by an organization to prevent or mitigate potential harms that may arise in an open data initiative. These harms can be direct and obvious or relate to the more subtle ways in which data can exacerbate existing inequities.

Data Security

Are there practices in place to protect data against attack and breaches and promote public trust in their work?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Inclusive Design

Is the open data effort designed in a way that is mindful of inequities and seeks to reduce asymmetries in how impact is felt?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Institutional Alignment

Do all institutions working on the open data project share a common end goal?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Legal and Regulatory Requirements

Have the open data project's sponsors verified they are in compliance with all legal and regulatory requirements?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Privacy by Design

Are there procedures that have been put in place at the start of the data effort to preserve the privacy of data subjects and promote public confidence in the open data effort?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Proactive Data Holder Engagement

Are open data sponsors actively engaging with data holders to avoid missing or ignoring opportunities to use siloed data?



Importance

1 Not Important	2 Somewhat Unimportant	3 Somewhat Important	4 Very Important
--------------------	---------------------------	-------------------------	---------------------

Scoring

After answering each of the questions above, please tally your scores according to section. The highest scoring categories include those elements that are the most important for the success of the open data initiative while the lowest scoring categories are those that can be deprioritized.

Total Scores	
Problem and Demand	
Capacity and Culture	
Governance and Standards	
Personnel and Partnerships	
Risk Mitigation	

Problem and Demand Definition

Given that resources for opening data are often in short supply, a clear, detailed understanding of the problem to be addressed by an initiative can help ensure that efforts are targeted and optimized for success. Some of the most effective open data projects we have examined are laser-focused on a specific user group or identified gap. Clearly defining a problem can also aid in the development of metrics of success and a strategy for monitoring progress against a well-defined baseline. Problem and demand definition include all those elements that can allow an open data project to be targeted and optimized. They are about increasing the return on investment and minimizing wasted effort. The following aspects contribute to better problem and demand definition.

User Research:

User research helps data practitioners identify, map, and understand a problem, its components, and users' needs. It pushes against the notion of "if you build it, they will come."



Data-driven initiatives tend to be more successful when they do not assume "["if you build it, they will come."](#)" The most successful projects are those clearly optimized for an intended audience of users from the start. They understand the broader context in which a problem exists, know how data could yield impact, and what the constituencies for the effort would be. This work requires researchers to study and understand the community in which they operate and to ensure their efforts match its technical, social, political, and economic context. It might involve [topic mapping](#) or other research techniques.

Problem Refinement:

Pr
Problem
Refinement

Effective data-driven initiatives are often those that identify a specific purpose or issue to solve, the data needed to produce insights, and how those insights can be acted upon to produce real public value.

It helps make the effort targeted and purpose-specific. Understanding a problem exists is one matter. Knowing how it can be best addressed is another, one best achieved by refining the problem at hand. As we note in our [Third Wave of Open Data Toolkit](#), 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, an organization can refine its area of concern toward an actionable problem with a clear, measurable outcome that will benefit a specific audience. Organizations might find it helpful to frame their concern as a research question, one answerable through data science. Drafting might take several iterations.

Benefits and Goals:

Bg
Benefits and
Goals

Open data projects can benefit from articulating the tangible value that they will produce for their project's sponsors and the public to maintain interest and engagement. Clear, quantifiable goals, coupled with metrics, can demonstrate progress.

In the past, open data advocates have tended to argue for increased data re-use by relying on [normative arguments](#). Proponents claimed open data enabled greater transparency or provided accountability. While these arguments can be persuasive in some contexts, private sector leaders, government officials, and the public often need to understand how it will specifically 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 or demonstrations (e.g. apps, platforms) of how open data will support their short, medium, and long-term goals. Importantly, this information must be understandable to the intended audience. An open data platform, for instance, that does not cohere to a logic understood by the public is unlikely to be used.

Data Audit and Inventory:

Da
Data Audit and
Inventory

An important aspect of a data-driven initiative is understanding what data is collected and generated and which can still be made available. A data audit allows data practitioners to understand what they have access to and who is accountable for the datasets.

Once the problem and value proposition are in place, practitioners are able to explore the availability of the datasets they hold.

A comprehensive [data audit](#), which allows practitioners to identify, locate, and assess their data assets, can uncover which internally held data sources could add value to an initiative. It could also inform strategies for collecting or accessing that siloed data. Assembling a catalog of assets following an audit might be useful in compiling information and indicating clearly which assets need to be prioritized when it comes to opening and disseminating.

Data Ecosystem and Stakeholder Assessment:

De

Data Ecosystem
and Stakeholder
Assessment

When data, expertise, or other resources are not immediately available, a robust understanding of the broader ecosystem and stakeholders can be useful in finding partner agencies, businesses, and nonprofits who can fill those gaps.

Just as no person is an island, no data project exists in a vacuum. Projects often engage a multitude of overlapping government agencies, businesses, and nonprofits who are interested in or have collected data about problems affecting them. Identifying who these stakeholders are and what data assets they have (particularly what data they have disaggregated demographically) can allow organizations to avoid unnecessary data collection and creation. Conversely, if a scan of the data environment reveals an absence (or inaccessibility) of a particular kind of data, this information can be used to motivate new collection.

Capacity and Culture Definition

A lack of available resources, insufficient human capital and immature technological capabilities can create major challenges to achieving meaningful impact with open data projects. These challenges can exist both within a country’s open data ecosystem—that is, the capacity of government, civil society, tech community, and the general public—as well as within actors on the demand side who use open data for certain objectives and the donor organizations funding them. Capacity and culture refer thus to those elements that relate to an organization’s ability to pursue open data. It involves intangibles—including how individuals inside and outside the organization view the project—and tangibles—such as specific hardware and funding sources.

Issue Salience:

When data practitioners can connect their project to an issue of substantial interest to their target audience, institution, or the public, they are more likely to receive support and funding than if their work had no resonance at all.

Is

Issue Salience

Issue salience directly affects how much and how quickly people and their leaders are willing to support data-driven initiatives. When open data can be related to an issue that is at the top of the agenda—such as due to some crisis or awareness of long-standing injustice—it is easier for data practitioners to launch projects related to them. Data practitioners might seek to connect their data-driven initiatives with some issue of special importance within the community they operate or otherwise understand how a local area’s priorities intersect with the issues they want to work on.

Data Infrastructure:

Mature data infrastructure is important for collecting, storing, managing, and sharing open data assets.

Di

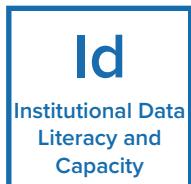
Data Infrastructure

As the researcher Tim Davies [notes](#), “data portals have been an integral part of the open data movement. Indeed, for many [...] the open data movement was crystallized with, or first discovered through, the launch of platforms like Data.gov and Data.gov.uk.” They have been key in enabling data openness, facilitating the combination of various institutional datasets and allowing users to browse, filter, search, and download data to their machines.

Data portals are one prominent example of data infrastructure. When they are available, data practitioners face fewer costs and barriers to launching their projects. New and sophisticated technological developments can [facilitate](#) greater collaboration and responsibility in data re-use. They can allow 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.

Institutional Data Literacy and Capacity:

Data practitioners can increase the societal and organizational value created through data reuse by bolstering their personnel’s skills and distributing those with skills throughout the organization.

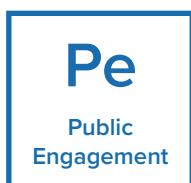


In the Data Foundation’s report “Data Literacy for the Public Sector: Lessons from Early Pioneers in the US,” the researchers Nick Hart, Adita Karkera and Valerie Logan [note](#): “Data and analytics are no longer ‘just’ for specialists [...]; rather, data literacy is now increasingly recognized as a core workforce competency.” This fact is true across regions, sectors, and disciplines. Knowing how data works is increasingly important for any institution’s success.

As such, organizations can increase the societal and organizational value created through their work by bolstering their personnel’s data skills and distributing those skills 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 allow people in all parts of the organization to 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—coordinated through efforts such as data governance programs—can help an organization become more evidence-based and systematic across all its operations.

Public Engagement:

People often will not engage with data efforts if they don’t know they exist and don’t know how data can help them. By promoting trust and understanding in data initiatives and processes, practitioners can increase the likelihood their work is adopted by the public.



Facilitating greater data competence within the general public is an important step toward verifying that it can receive greater benefits from data re-use and data openness, 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 can 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

This work falls into the larger category of public engagement. People will not engage with systems if they do not know they exist and do not know how they can help them. By working with residents and helping them understand and trust institutional systems and processes, organizations can increase the likelihood their efforts are adopted by the broader public. They can involve them in conversations about the kinds of data they would use in their regular lives so they are not merely data providers but data users as well.

Culture and Institutional Buy-in:

By gaining approval from the leaders of their organization and fellow associates, open data advocates can gather the resources needed to unlock institutional resources and overcome roadblocks.

Ci

Culture and
Institutional
Buy-in

As with all projects, data-driven initiatives need to have advocates within the organization sponsoring them. By gaining broad internal approval, especially from leadership, open data advocates can better acquire the resources needed to unlock institutional resources and overcome roadblocks. They can obtain this support through a clear articulation of the [potential business case](#) and purpose of the effort, recognition of institutional priorities (and pitfalls), an understanding of how the work would fit into the broader data economy and ecosystem, and a clear strategy for mobilizing existing tools and resources to operationalize the strategy.

Fostering a culture of data openness can be more difficult in some institutions than others. In risk-averse organizations, such as governments, there can be a tendency to obfuscate or avoid releasing data that could reflect poorly on them or require additional spending. In these cases, it can be important to emphasize how data openness fits into specific organizational objectives (e.g. how it adds value in the areas an institution cares about). It might also be useful to establish systems that guarantee accountability when data is not published.

Social License:

Using data for public good means involving the public in decisions that affect their lives. Projects that gain a social license to operate (after meaningfully informing subjects) are less likely to suffer backlash and may support important ideation.

SI

Social License

Trust and legitimacy are key in the planning processes pertaining to data re-use, particularly in an era defined by [declining trust in institutions](#) and [complex problems](#). To help make data re-use initiatives appropriate to operate and aligned with a public's understanding of what they need, they need to obtain a social license. This term means exercising the necessary due diligence and engaging with all relevant stakeholders so 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 risk associated with them are evaluated by local stakeholders.

Responsive Feedback Loops:

By creating mechanisms for data users and beneficiaries to provide input on a data initiative, practitioners can proactively address risks as they arise and promote public trust by demonstrating their commitment to act on concerns.

RI

Responsive
Feedback Loops

Open data initiatives tend to be less successful when they do not create meaningful mechanisms for users and beneficiaries to provide input to demand-side practitioners. By allowing experts, leaders, and the public to comment and addressing that feedback in a clear and direct manner, organizations can find new opportunities to create value and address harms they might have previously missed. As the World Bank's report "[Closing the feedback loop: Can technology bridge the accountability gap](#)" these methods can also promote public trust by demonstrating to individuals that their concerns have been heard, reducing the perceived gap in accountability.

Resource Availability and Sustainability:

Technological innovation and infrastructure development can be cost-intensive exercises with extended time frames. Organizations often benefit if they have internal and external sources of funding to systematize impactful and responsible data reuse.

Ra

Responsive
Feedback Loops

Technological innovation and data project development are often cost-intensive exercises that take place over extended time frames. Organizations need sources of funding before they can produce their intended results. These resources can be obtained by making use of internal assets, such as a company or organization's budget, or by seeking external sources through partnerships and collaborations.

Project and organizational leadership are central in the success of open data projects. Engaged leaders can enact policies, mobilize resources, and advocate for data openness.

While an organization might have bought into a data-driven culture, it requires a leader to enact the policies and bring to bear the resources needed to embed them in an institution. A senior leader can push for data openness. By strategically using the authority they hold within their institution, they can be a critical enabler for project success over the long-term. Acting on a sense of [public entrepreneurship](#) can act as an advocate for data projects and

Strategic Leadership:

Project and organizational leadership are central in the success of open data projects. Engaged leaders can enact policies, mobilize resources, and advocate for data openness.

SI

Strategic
Leadership

While an organization might have bought into a data-driven culture, it requires a leader to enact the policies and bring to bear the resources needed to embed them in an institution. A senior leader acting on a sense of public entrepreneurship can act as an advocate for data projects and push for data openness. By strategically using the authority they hold within their institution, they can be a critical enabler for project success over the long-term.

Governance and Standards Definition

A diversity of governing decisions affects the use and impact of open data efforts. Issues of governance exist at both the ecosystem level—especially related to standards and policies of data release—and on the demand side, with questions of risk mitigation and impact assessment leading the way.

Data Licensing:

Data licensing regimes protect and promote the re-use of data by outlining conditions under which practitioners can use and re-use data.

DI

Data Licensing

Data licensing regimes refer to the conditions under which an institution (such as a government, business, or nonprofit) is able to use and re-use data. These regimes provide a way to secure and promote the re-use of data, either among a set of actors or among the public. They can require interoperability, articulate 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.

Principles:

Principles help data practitioners and those they collaborate with to align toward the same goals. They can be useful for determining the type of data collaboration, as well as the practices and processes used.

Pr

Principles

To help their institution and all those they collaborate with are working toward the same goals, organizations might find it useful to identify normative principles that they hope to promote through their work. By stipulating, for instance, in clear, easily understood language their intention to be transparent, participatory, and people-centric, institutions can periodically check whether the practices they've adopted are well-aligned for those goals. They can also assess quickly which types of collaborations make sense and how to best engage with prospective partners.

Open by Default:

“Open by default” refers to the presumption that data should be published in absence of a legitimate justification to the contrary. This principle can allow insights to reach as many people as possible.



A central principle of the Open Data Charter is the notion of “[open by default](#),” the presumption that data should be published in absence of a legitimate justification to the contrary. Given the level of government resource allocation and time investment required to implement strong open data initiatives, high-level political buy-in and codified open data policies are needed to provide the incentives and flexibility to government officials to meaningfully advance open data goals.

Open by default can also be useful in protecting against “[open washing](#)”—superficial efforts to publish data without full integration with transparency commitments. By pushing data to be as open as possible, as closed as necessary, data practitioners can allow the impact of data to be as broadly felt as possible.

Policies, Positions, and Procedures:



Policies, professional positions, and procedures guide how to progress on a data initiative by outlining the necessary roles and guidelines for an open data initiative.

Clear policies pushing forward access to information and data—such as codifying data stewardship and data custodian positions, supporting mechanisms that encourage personnel to bolster their data skills, and “[Data for Good](#)” programs—can act as important drivers for open data initiatives. Without explicit policy backing, the sustainability of open data efforts can be called into question, and access to necessary data can dry up at any time. The existence of Freedom of Information policies can also provide means for accessing relevant information, though often at a much slower pace than open data. Taken together, all these elements can provide an actionable path for how to progress on a data initiative.

Contracts and Data Sharing Agreements:

Contracts and data sharing agreements establish the terms for how data is shared between parties. They can be useful for defining the structure of an open data or data collaborative initiative in a way that lets every party contribute what they are most capable of contributing.



Contracts and data sharing agreements are written agreements that establish the terms for how data is shared between parties. These documents are important, not merely for outlining roles and responsibilities but also for promoting accountability and trust between parties and avoiding misunderstandings. They can be used to [address](#) long-standing injustices by allowing ownership and access to data for certain marginalized groups and they can [ensure](#) data practitioners balance individual rights with public needs.

In short, these documents play a key role in governance between parties. They are often a prerequisite in any data reuse because they define not only what the purpose of data (re)use is but how it will be used.

Data Quality:



Good data quality allows data practitioners to reach insights that reflect the issue they are studying. If data quality is unaddressed, it can lead to inaccurate or harmful results.

A widely prevalent challenge to positive impact arises from poor data quality. As the quantitative researcher China Layne argues, quality issues can manifest in [a number of ways](#). Quality issues can relate to the unit of analysis, duplication of records, missing or invalid values, incorrect formats, outliers, or inconsistencies. These problems can cause inaccurate analysis that, if acted upon, can harm people.

When using datasets individually or together, it is important that practitioners try to promote data quality. By keeping data clean and accurate, practitioners can produce insights that are able to be acted upon to help others.

Data Standards:



Data standards are the technical specifications that make it possible to share, exchange and combine data. They also provide the framework necessary for data collaboration and enable machine readability and data portability.

As stated by [data.gov](#), data standards are “technical specification that describes how data should be stored or exchanged for the consistent collection and interoperability of that data across different systems, sources, and users.” These technical specifications are what make it possible to share, exchange, and combine data. They provide a framework for data collaboration. Standards can include topics such as machine readability, the ability of data to be read by a mechanical device, and involve matters such as [data portability](#), the right and ability of data subjects to move their information from one controller to another.

Personnel and Partnerships

Definition

In many high-impact open data projects, people and partnerships within and especially across sectors play a key role in enabling success. Whether creating touchpoints with citizens through partnerships with civil society, informing the public through media partnerships, or filling important data gaps through partnerships with private sector entities, open data suppliers and users often improve outcomes through collaboration.

The below elements relate to the specific actors—both individuals and institutions—whose support can enable an open data initiative’s success.

Data Stewards:

Data stewards are responsible data leaders empowered to seek new ways to create public value through cross-sector data collaboration. They can be useful for proactively collaborating, protecting customers, and acting on opportunities.

Ds

Data Stewards

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. Dedicated data stewards devoted to facilitating collaboration can be useful in addressing some or all of these issues. Data stewards are 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.

Domain Experts:

Domain experts are individuals with knowledge of a particular domain or sector whose knowledge can be used to better identify opportunities to (re)use data for the public benefit or otherwise identify, design, analyze, and communicate insights about open data.

De

Data Experts

In many cases, demand-side actors' expertise lies in technology or data science rather than the problem areas they seek to address through the use of open data. Tapping into the knowledge of stakeholders with relevant sector-specific expertise can improve efforts to optimize and target open data efforts based on a true understanding of needs, opportunities, and barriers.

Third Party Supporters:

Organizations, institutions, and individuals invested in a project's success can be critical for building a broad base of support for a data-driven effort. These supporters can be cultivated and engaged to build a project audience.

Tp

Third Party
Supporters

Open data practitioners can extend their capacity by collaborating with like-minded organizations, institutions, or individuals who are not directly related to the issue at hand but, nonetheless, invested in a project's success. These groups can include open data advocates and academics who see a data-driven initiative as a way to foster data openness and develop an evidence base. It can include journalists and members of the public who might raise awareness of the activities that an institution is undertaking. It can also include other governments, businesses, associations, and nonprofits who see an effort as a model for their own work and a way to identify new, promising practices.

Even for organizations that might be hesitant to do outreach because they are risk averse, these stakeholders can fill various gaps in human or technical capacity if engaged. They can also promote trust and legitimacy among other third party actors, such as data subjects in marginalized communities, who are often excluded from conversations about data reuse.

Risk Mitigation Definition

The release and use of open data comes with certain risks. An upfront mapping and consideration of risks associated with intended uses of open data can allow practitioners to design programs from the outset in ways that are well-positioned to overcome or mitigate those risks. The risks listed below, however, should not be considered arguments against opening or sharing data. Rather, they are reasons for taking a more fine-grained and considered approach, one that pays close attention to the empirical evidence, sifting out what works and what does not, and identifying conditions for scaling and replication. In addition, by adopting risk mitigation elements upfront organizations can prevent or mitigate potential harms that may arise in an open data initiative. These harms can be direct and obvious or relate to the more subtle ways in which data can exacerbate existing inequities.

Privacy by Design:

By instituting privacy in the design of open data efforts, organizations can minimize risk and boost public confidence in their work.

Pd

Privacy by
Design

Privacy concerns probably rank among the most commonly cited worries over opening up data. Especially on sensitive topics or in regions where there are high risks to data subjects, individuals' anonymity can be of life-or-death importance. Potential privacy harms can arise even from the release of ostensibly anonymized personally identifiable information.

Organizations can minimize these risks and promote broader trust in their work by adopting Privacy by Design, an approach to data responsibility in which data practitioners integrate privacy considerations into the design of their efforts. Already required by the European Union's [General Data Protection Regulation](#), it can allow data practitioners to think critically about the risks posed to others should data be inadvertently released or combined with other datasets to expose others.

Data Security:

Organizations and individuals can be reluctant to share data if they believe it will be exposed in ways that cause harm. By protecting against attacks and breaches, organizations can reduce distrust for their work.

Ds

Data Security

Because much government data contains sensitive information regarding individuals, industries, and national security, opening that data often leads to quite reasonable questions about data security. Cybersecurity remains a challenge across the world, and perhaps especially so in developing countries, which may lack the technical expertise to adequately protect information from sophisticated hackers and other intrusions. At the same time, though security concerns are very real and important, they must be balanced against the opportunity cost or risk of not sharing data; often, government decision makers can lean on tenuous security concerns to justify keeping data closed and restricting access, potentially limiting the solution space

Inclusive Design:

Data can often consolidate and reinforce privileges in ways that benefit elites. By remaining mindful of inequalities, data practitioners can work to reduce asymmetries and improve the impact of their work is broadly felt.

Id

Inclusive Design

Although data can be empowering, it can also consolidate or reinforce existing privileges and authority inherent in societies. This problem is closely linked (though not restricted) to digital divide challenges; when only the elite of a society have access to data and/or data science capabilities, releasing data is likely to disproportionately benefit that elite.

Open data projects often reach their best outcomes when they try to make the social and economic benefits widely, and evenly, distributed. Projects must be built with an understanding of how they interact with a society so as not to deepen inequities. By remaining mindful of these factors, data practitioners can try to make the impact of their work broadly felt.

Legal and Regulatory Requirements:

Legal and regulatory requirements can be an effective way of helping data practitioners understand their responsibilities and are accountable for misuse. These requirements can build trust and minimize harm.

Lr

Legal and Regulatory Requirements

Governance bodies across all levels are increasingly pursuing ways to regulate the use of data and algorithms in the places they operate. These standards—which include the [European Union's General Data Protection Regulation](#), [California's California Consumer Privacy Act](#), and [Children's Online Privacy Protection Act](#)—place real and specific requirements on data holders and users to minimize the potential harm to data subjects.

The costs for violating these provisions can be significant for both the institution found in violation (in terms of financial or legal penalties), data subjects (in terms of directly exposing them to harm and malicious actors), and the public (in terms of undermining trust in other data initiatives). Organizations must be cognizant of the legal and regulatory frameworks in the areas that they operate and have the resources needed to remain within them. These frameworks guarantee accountability and make clear the requirements facing stakeholders.

Institutional Alignment:

By encouraging all stakeholders to share a common end goal, data practitioners can increase the probability that all the work put into a project is complementary and effectively uses the resources available.

Ia

Institutional
Alignment

Institutional misalignment occurs when practitioners and stakeholders diverge in their goals and priorities. This can lead to inefficiencies in the project and can negatively impact the potential for public benefit.

To maximize public good, open data outputs should align to policy priorities. Organizational commitments and contracts between stakeholders, for instance, can be made to require that data holders share data related to outlined policy priorities. An example of such an initiative in practice could be global agreements to make specific data open to the public in the event of an emergency.

Proactive Data Holder Engagement:

Proactively engaging with data holders can minimize the chances of missing or ignoring opportunities of using siloed data. It can bolster projects by helping them make use of all the resources available.

Pd

Proactive Data
Holder
Engagement

While open data carries a number of risks, keeping data siloed or otherwise inaccessible also holds its own risks and ethical concerns. Whether it's public or private sector data, withholding access to data leads to "missed uses" or ways in which the data could be shared or re-used to create public benefit. Data is also an important resource for evidence-based policy interventions that have long term implications for public well-being.

Data practitioners can minimize the risk of "missed use" by proactively engaging with data holders and other stakeholders in the areas they operate. By working with partners, open data advocates can bolster their data projects and better make use of all the resources available to them.