Effective and Ethical Data Sharing at Scale

A cookbook for data producers, donors, policymakers, and other development practitioners

Global Partnership for Sustainable Development Data

With support from Google.org

About this cookbook

What is effective multiparty data sharing?

Multiparty data sharing in the development sector refers to situations in which two or more organizations collaborate to collect, share, and/or analyze data to address societal challenges.

Multiparty data sharing often happens among stakeholders from different sectors, including governmental agencies, private companies, and nongovernmental organizations, who come together to establish collaborative partnerships. This document focuses on data sharing initiatives that involve partners from across sectors, *excluding* business-to-business (B2B) and government-to-government (G2G) data sharing.

There is no formal or broadly accepted consensus on what makes a data sharing partnership effective in general, much less in the development sector. In this cookbook, we draw on input from development practitioners to characterize **effective data sharing** as a reduction of friction in data sharing over the long term that addresses a societal challenge or seeks to improve public wellbeing without producing negative externalities, and which is accountable and transparent toward key stakeholders and the broader public.

For effective data sharing you'll need:

A **reduction of friction** in data sharing over the long term...

that **addresses a societal challenge** or seeks to improve public wellbeing without producing negative externalities...

and which is **accountable and transparent** toward key stakeholders and the broader public.

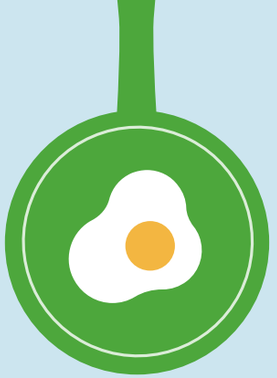
Why do we need a cookbook on multiparty data sharing?

Multiparty data sharing initiatives in the development sector have proliferated in the past decade as efforts to better leverage data for achieving the Sustainable Development Goals have intensified.

However, **few spaces exist to exchange knowledge** on best practices or for partners engaged in (more or less effective) initiatives to consolidate learning and draw out common insights. This deficit persists in spite of demands from members of the data for development community for recommendations and actionable insights on what enables effective data sharing.

The objective of this cookbook is to provide development practitioners and organizations with **actionable, evidence-based insights, recommendations, and examples to create successful data sharing initiatives**. This cookbook aspires to be a user-friendly, pragmatic, concrete, and concise tool for our community.

Success in data sharing within the development sector



While the concept of data sharing is straightforward, measuring the success of data sharing initiatives is more complicated. In conversations with development practitioners, three general concepts emerged as indicators of success:

Hitting key performance indicator targets

These definitions of success focus on the outputs of data sharing agreements such as the number of data users, applications or services developed, or reports produced.

Solving societal challenges

For some experts, the success of data sharing initiatives is represented by effectiveness in addressing real-world problems without creating negative externalities or harms (such as excluding marginalized groups). These measures look at longer-term outcomes rather than immediate outputs.

Becoming self-sustaining

Some experts consider success to be measured by the capacity of a data sharing initiative to become self-sustaining over time. This encompasses aspects of financial stability as well as relationships between key stakeholders in addition to maintenance of operations and data quality.

How was this cookbook created?

This cookbook is the outcome of a year-long project undertaken by the Global Partnership for Sustainable Development Data (Global Partnership) and sponsored by Google.org, aimed at identifying drivers of effective data sharing in the development sector.

It builds on insights collected through a landscape analysis of data sharing initiatives and feedback from a diverse working group of more than 20 experts who met four times between January and September 2022 (see Annex B). It also draws on the Global Partnership's experience in helping countries and organizations to fill their data needs and gaps by working with partners. The Global Partnership has facilitated over 120 data sharing collaborations since 2015.

If sharing data can be compared to diet, then ***effective* data sharing is like healthy eating**. To maintain a healthy diet requires knowing what foods are good for you. Cooking healthy recipes requires knowing which ingredients are important and how to combine them for each dish. Just like a healthy and balanced diet, effective data sharing in the development sector is essential to **ensuring that data activities benefit people** on the ground.

The similarities between a healthy diet and effective data sharing don't end there. Of course, none of us eats exactly as our ancestors did a few centuries ago, as food and tastes are not immutable. Not only do ingredients, recipes, and methods vary significantly across the world, with each country and region developing its specific "cuisine," but they are also constantly evolving. Likewise, this cookbook is merely a snapshot of current knowledge and experiences of data sharing from "chefs" all over the world. As with any cookbook, new and updated recipes should be added regularly to keep this tool relevant and to increase the diversity of sectors, regions, and initiatives represented here.

Finally, it must be acknowledged that, despite the best efforts to seek and consult a diverse range of data sharing experts and initiatives as well as to leverage the Global Partnership's experience working with organizations from the Global South, this document draws substantially from theoretical knowledge and recipes that have been developed in the Global North. Further efforts will be needed to rebalance our culinary knowledge to add more ingredients and tips from other contexts.

Within the recipes, this icon indicates that more research, experimentation and/or knowledge exchange is needed.

Understanding basic cooking

The data sharing food pyramid

The content in this cookbook is organized around five groups of key ingredients for effective data sharing initiatives.

5 — No diet is balanced without a healthy dose of fats or—in this case—flexibility and adaptability, which facilitate storing energy for when it is most needed.

4 — Dairy helps develop and strengthen human bones. Similarly, knowledge and skills support data sharing functioning and operations.

3 — Just like healthy proteins, data and infrastructure occupy the center of the data sharing plate and must always be accompanied by shared value and benefits.

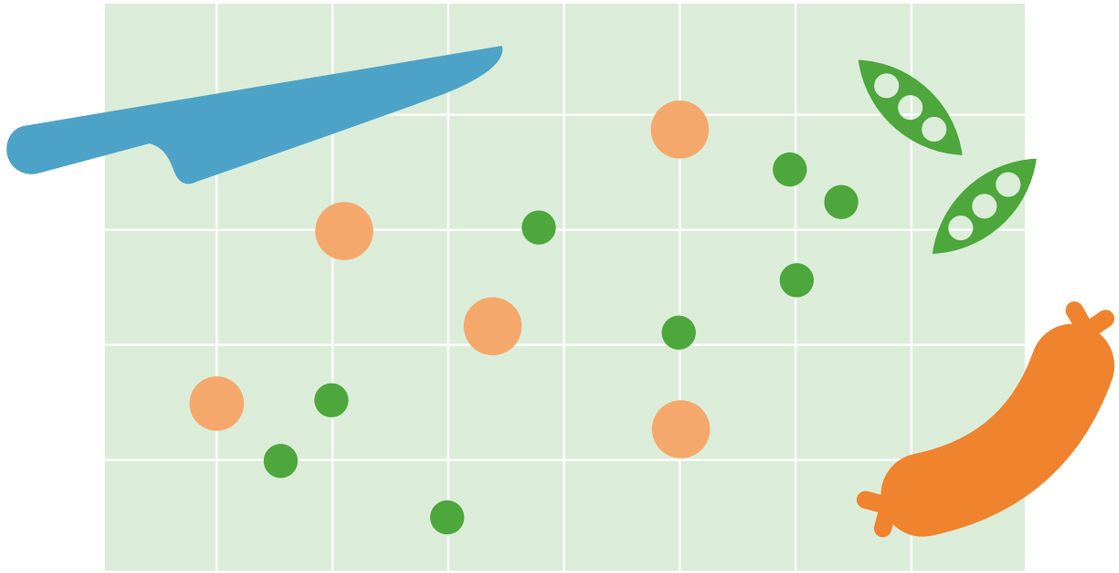
2 — When all stakeholders share in the value and benefits of data sharing, they get the necessary vitamins and minerals (like you'd find in fruits and vegetables) to reduce the risks of initiative-threatening "diseases."

1 — Any healthy diet starts with a foundation of trust among data sharing partners, data users, and other stakeholders.

These five key food groups are found across all successful data sharing initiatives. They form the basis of a varied and balanced data sharing diet. The quantity (or emphasis on) each of these varies according to individual data sharing initiatives, as do the recipes needed to create them.

Each of the five sections of this cookbook contains "recipes," i.e., tested methods and tools that can be replicated by others.

Data sharing in the development sector is a relatively new domain, and there are some clear gaps in knowledge and experience. For this reason, the cookbook points out where recipes or information on ingredients are missing. Highlighting these gaps helps aspiring chefs to understand what still needs to be done and paves the way for more thinking and experimentation.



Factors influencing choice and quantity of ingredients

There is no one-size-fits-all approach to effective multiparty data sharing. Instead, this cookbook offers suggestions for key ingredients and recipes to inspire creative approaches.

The importance (or quantity) of each food group, as well as the recipes to use, depends on two broad categories of factors related to:

- **The characteristics of the specific data sharing initiative**, such as stakeholders or sectors involved, types of data (i.e., personal, nonpersonal) being used, the stage the initiative is in, the objectives of the partnership, the number of stakeholders involved, and the openness of the data or initiative.
- **The context in which the data sharing initiative operates**, including, for instance, the regulatory and policy environment, other stakeholders within a particular data ecosystem, and other contextual factors such as whether data sharing happens during an emergency situation (e.g., a pandemic or natural disaster) or during "business as usual."

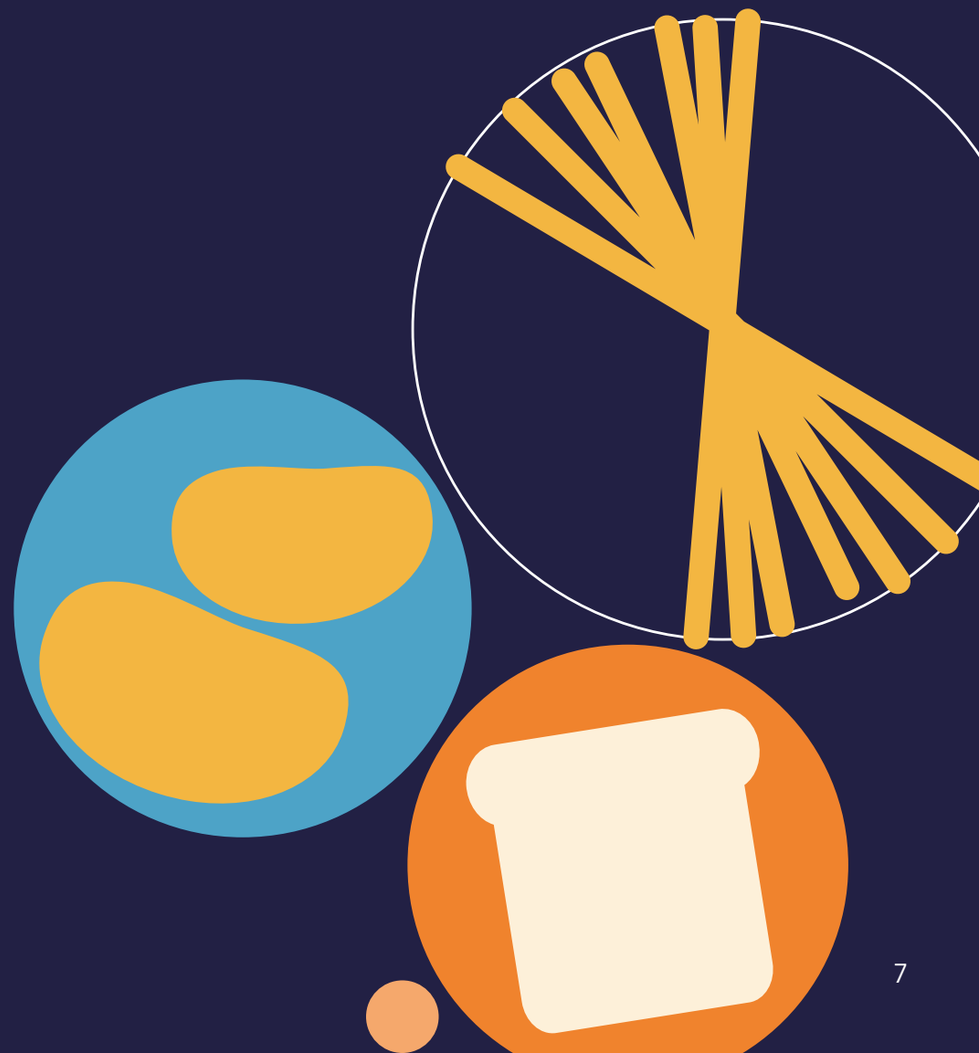
1

Mechanisms for building and sustaining trustworthiness

Carbohydrates

Recipes in this section

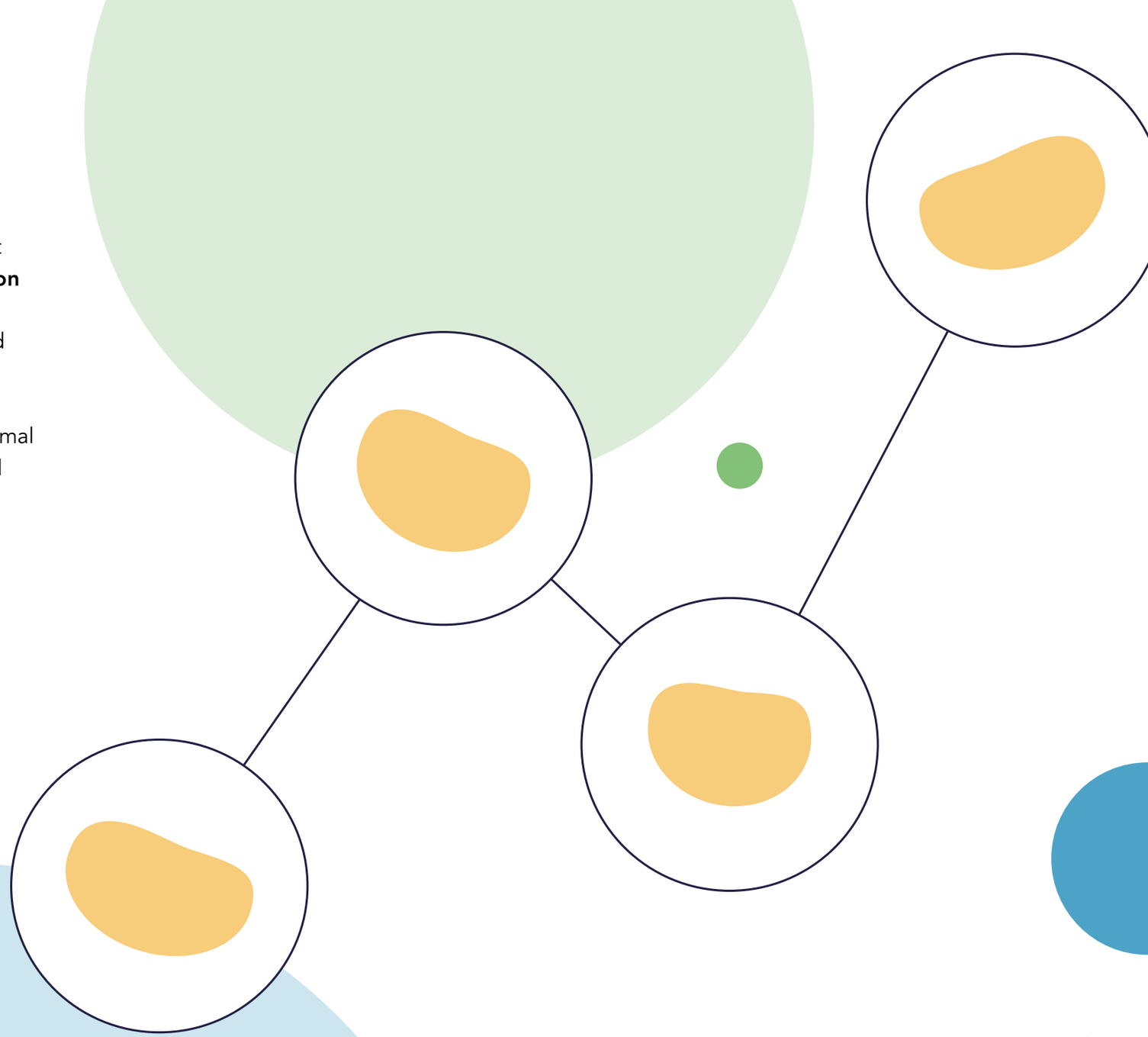
- Fit-for-purpose governance models
- Communicating with the public and stakeholders
- Accountability



Introduction

Trust is the most important ingredient for effective multiparty data sharing and, precisely like carbohydrates, it **provides initiatives with the energy they need to function and achieve their objectives.** Recent research highlights robust quantitative evidence that greater trust is associated with increased data sharing and that the impact of trust is particularly significant where initial levels of trust are low. However, the same research also reveals that, to reach optimal levels of data sharing, increasing trust needs to be coupled with other strategies or actions, which is why trust must be combined with other ingredients.

Simply saying that trust is important does not help organizations or individuals working on data sharing initiatives to clearly lay out how or what they need to do. For this reason, and in order to foster more effective data sharing, this section focuses on understanding effective mechanisms for building and sustaining trustworthiness in data sharing initiatives.



What are we making?

For data sharing initiatives to succeed, data partners need to build and maintain the trustworthiness of the initiative and embed core values and ethical considerations in its functioning.

The trustworthiness of a data sharing initiative is the result of three key elements:

- Trust among partners,
- Trust from the general community, and
- Accountability measures.

It follows that recipes for building and maintaining trust and for embedding values and ethical guidelines in data sharing initiatives can be grouped into three categories, which are described in the next section:

- Fit-for-purpose governance models
- Effective communication approaches toward the public and key stakeholders; and
- Audits, frameworks, and other mechanisms for accountability.

Recipes for fit-for-purpose governance models

Building and maintaining trust is an ongoing and iterative process that remains a focus throughout the life cycle of a data sharing partnership.

Of course, it's easier to build trust when data partners have previous experience working together. Prior successful collaboration improves baseline levels of trust and comfort. With prior positive experiences of working together, established communication channels and relationships facilitate working together on new initiatives.

But prior experiences aren't necessary precursors to building trust in new partnerships. Setting up relevant governance mechanisms can help build and maintain trust, even when partners have never worked together before. **Participatory governance models** such as **data trusts or collaboratives**, detailed and standardized **data sharing frameworks** or memorandums of understanding between partners, and **conflict resolution mechanisms** are among the most widespread and effective mechanisms for doing so.

Participatory governance models — Hong Kong Data Trust 1.0

Hong Kong's public transportation system consists of taxis, Mass Transit Railway (MTR), buses, minibuses, and tramways. Each transport sector is privately owned, and data sharing among providers has historically been limited. The critical barrier to data sharing has been lack of trust among stakeholders and the absence of a framework to govern data sharing.

A Data Trust at the University of Hong Kong (HKU) aimed to overcome this barrier through data sharing in real time. This Data Trust was a proof of concept for a trusted third-party model that encouraged data controllers (public transport operators) to securely share their data with the Data Trust. The Trust itself was an entity hosted by HKU (a neutral third party) with fiduciary responsibility to the data controllers and the technical capacity to analyze and process the data.

Relations between the public transport operators and HKU were formally established through memorandums of understanding. Data sharing via the Data Trust was also facilitated by a clear legal framework, constant communication with stakeholders, and open channels for feedback.

Data collaboratives — an example from California

The California Data Collaborative (CaDC) is an independent nonprofit working as a coalition of water supply agencies in California to facilitate data-driven water policy and operations. According to the GovLab, "Data Collaboratives are a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors — in particular companies — exchange their data to create public value."

CaDC seeks to make water data analytics central to improved water management. It maintains a secure, cleaned, and standardized water use database for each member agency. Using the data, CaDC creates data analysis tools and develops directly actionable insights.

The CaDC is governed by a steering committee composed of members from all participating water supply agencies. This Committee sets the research and policy priorities of the CaDC and identifies pilot projects for the collaborative. In its initial workshops, the CaDC developed a "trust framework" that included standardized data sharing and data transfer agreements that were available to all. Organizers hoped the "open" agreement would promote transparency and ensure everyone had access to the same data sharing rules. This equitable, transparent mode of governance has contributed to the smooth functioning of the CaDC.

Detailed and standardized data sharing frameworks

Development Data Partnership is an initiative that seeks to meet the needs of international development organizations. It created a platform to connect them with technology companies or data partners. Since building trust between partners, negotiating terms, and entering into a data sharing agreement requires time, Development Data Partnership created a highly detailed master licensing agreement that lays out procedures for crucial aspects of data management, such as ownership and access. All partners sign the same agreement, and this provides the framework under which the Development Data Partnership must gather, harmonize, process, and store data. The agreement makes the procedures predictable and trustworthy, lowering barriers and providing an easy system for a partner to share data.

Standardized agreements, however, are not always the best option. For example, Global Fishing Watch works directly with national governments to track illegal fishing and enforce action against rogue actors. National governments are both data providers (alongside other public and private providers of automatic identification systems and vessel monitoring systems) and users within the initiative. Given the direct engagement with governments and the specific regulatory requests of each partner, creating standard terms of engagement is not always possible.

Conflict resolution

Even though conflict is an inevitable aspect of cross-organizational and cross-sectoral partnerships, this review of data sharing partnerships found limited mention of formal conflict resolution mechanisms. The experts consulted suggested that disagreement tends to cluster around the establishment of the initiative rather than its implementation or functioning.

Practices such as meeting regularly with partners and establishing clear decision-making processes and governance bodies help build the tools for conflict management and ensure that issues that arise within the initiative are sorted out regularly. However, more research and experimentation are needed to understand which conflict resolution mechanisms are optimal for data sharing initiatives in the development sector.

Recipes for communicating with the public and stakeholders

Communication can make or break a data sharing initiative. For data sharing to be trusted, successful communication between data partners and with the public is paramount. Lessons from past initiatives, such as Sidewalk Toronto, suggest that failure to communicate clearly and effectively with the public can lead to the end of data sharing efforts, due to insufficient trust and high resistance from citizens.

For this reason, new data sharing initiatives need to invest considerably in establishing the right approaches to communicating with stakeholders.

Similarly to building and maintaining trust, communication among partners is facilitated by instances of previous successful collaboration. When new communication channels need to be established, mechanisms such as stakeholder consultations or workshops can be beneficial at the kick-off stage to start dialogue. Investments in communication among partners, however, cannot be relegated to the inception phase and must continue throughout the initiative.

Approaches to ensure effective communication include relying on **organizational champions** or data stewards, investing in **communication** throughout the lifespan of the initiative, and establishing **dialogue with civil society** and citizens right from the start and on a recurring basis.

Organizational champions — data stewards

Sharing data often requires a shift in the cultural mindset of top leadership. It can be hard to secure buy-in from organizations given the sensitivity with which organizations guard their data, along with the cost of data sharing in terms of investment in staff and funding. Having a champion among the key partners of the initiative or within the sector who is outspoken about the potential benefits of data sharing can help establish credibility and support.

For example, the California Data Collaborative, discussed in the preceding section, found its champion in the General Manager of a water district in California who endorsed the project and advocated for the value of data collaboratives from the beginning. His outreach to his peers helped the initiative gain buy-in from other organizations across the sector.

The role of communicating the benefits of data sharing can also be given to data stewards, individuals or teams within data-holding organizations who are empowered to proactively initiate, facilitate, and coordinate data collaboratives in the public's interest. The existence of data stewards is considered a key success factor in setting up sustainable and responsible business-to-government partnerships in the European Union. Such findings are transferable to data sharing within the development sector.

Investments in communication throughout the life span of the initiative

The Hong Kong Data Trust 1.0 discussed previously was established to promote data sharing among transport providers to inform policy. To address reluctance from transport providers to share commercially valuable data, the Data Trust engaged in communication with the stakeholders for 18 months prior to launching the initiative through consultations, workshops, roundtables, knowledge cafes, and regular newsletters.

These consultations allowed players to engage with each other and to understand the permissible use cases of data sharing and the resulting value exchange. They also provided stakeholders with a forum to exchange and discuss their concerns, and clearly identify the permissible use cases for the initiative.

Most of these communication channels have persisted after the inception phase, ensuring that partners keep up formal and informal exchanges.

Establishing dialogue with civil society and citizens

Analysis of existing data sharing initiatives within the development sector showed that very few invest in communication activities targeted toward concerned communities or the broader public. In many instances, citizens are not informed about their data being used (even if in anonymized or aggregated form) in the context of public-private data sharing partnerships. This has resulted in public backlash once partnerships have been established, as the scandal over a data sharing agreement between the United Kingdom's National Health Services and U.S.-based company Palantir shows.

Effective, clear, and ongoing communication with the public is often costly and difficult to implement. The stakes of not communicating, however, are high, and data sharing initiatives need to undertake the effort, both because it is part of trust-building processes and because informing the public is a first and necessary step to enabling accountability and public scrutiny.

There are few examples of initiatives investing time and effort in communicating with wider audiences, although some have begun to incorporate this approach into their activities. For instance, during the kick-off phase of a data sharing project in The Gambia, Spatial Data Commons together with The Gambian Bureau of Statistics and the Public Utilities Regulatory Authority organized a multi-stakeholder workshop for the private sector, government ministries and agencies, academia, civil society, and development partners. The workshop was intended to inform stakeholders about the prospective project, to present the use case for accessing mobile network operators' data, and to air potential concerns at the beginning stages of the project.

Recipes for accountability

Effective and trustworthy data sharing must be clearly rooted in societal values and held accountable to the highest ethical and operational standards.

In the last two years, more than 350 individuals and organizations from over 60 countries participated in a public consultation process that led to the launch of a #DataValues Manifesto calling on governments, companies, civil society organizations, donors, and others to make positive changes in how data is funded, designed, managed, and used. Its five calls to action and the evidence collected for its underpinning whitepaper, "Reimagining Data and Power," show that development practitioners truly want to establish value-driven and ethical data sharing.

Values and ethics can be embedded in the governance of initiatives and their operations in many different ways, such as through **ethical frameworks** or specific **governance structures**.

Ethical frameworks and guidelines

Frameworks and guidelines can help data sharing initiatives to interrogate the ethics of their governance and operations. Some are generally applicable, while others are tailored to specific types of data or contexts.

The Open Data Institute, for instance, has developed the Data Ethics Canvas, a tool for identifying and managing ethical issues in data projects. This tool provides a framework to develop ethical guidance to suit data sharing initiatives, regardless of their scope, size, or sector.

In 2021, the Data for Children Collaborative released its Ethical Assessment Framework. This document recognizes that children have specific vulnerabilities when it comes to data and that, in this context, "delivering an ethical Data for Children project requires more than just legal compliance." The framework provides a roadmap, including a series of questions to be answered at the inception stage, during the project, and at the final stage, with the objective of ethically navigating all activities.

The United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data recently developed Data Responsibility Guidelines, a set of principles, processes, and tools to support data responsibility in OCHA's humanitarian work. This document offers the minimum standard to follow for data responsibility across OCHA and recognizes the great risks that irresponsible data management poses to already vulnerable communities.

Boards and committees for ethical assessment

Since 2020, the Data-Pop Alliance has included a Council for the Orientation of Development and Ethics, or CODE, in most of its projects. A CODE is a group of independent stakeholders who voluntarily share their expertise in areas of direct relevance to a project. As an advisory group, it provides oversight to ensure a project abides by key ethical principles, including fair and safe use of data and local, context-specific concerns. CODEs are typically comprised of around 10 local experts from academia, civil society, government, and private enterprise. The advisory role of CODE members is not legally binding. Nonetheless, project implementation teams give considerable weight to their feedback and integrate CODE recommendations whenever possible.

When Dalberg Data Insights started work with the Belgian Government on the use of anonymized telecommunications data and localized incidence data to monitor the impact of lockdown measures during the COVID-19 pandemic, it set up an Ethics Committee composed of experts from different disciplines and backgrounds. This group of experts monitored the project throughout its life cycle and was tasked with raising any ethical concerns. The committee's role was particularly important considering both the urgency of the situation and the constraints that COVID-19 restrictions posed on collecting feedback from the public.

2

Shared value and benefits

Fruits and vegetables

Recipes in this section

- Lowering costs
- Delivering on innovation and data products
- Provision of tailored services

Introduction

Fruits and vegetables contain essential nutrients for people to thrive. Likewise, **a sense that all partners are benefiting** proportionately is essential to creating effective partnerships and reducing the risk of threats to the initiative. While humans can live for some time without fruits and vegetables, their dietary absence in the long run results in significant health problems. Unless organizations see the value in partnering for data sharing, they won't stay at the table for long. **Data sharing fails when the perceived value and benefits are unequally distributed for extended periods of time.** When benefits consistently go to a few instead of many, incentives to participate decline.

Benefits encompass **more than mere economic value**. Data sharing partners can profit from an initiative in terms of cost reduction, reputation, knowledge, skills, or even access to data generated by others. Therefore, the emphasis should be not only on the monetary value of data sharing but on understanding the broader scope of added value of the initiative for data partners. Tools such as the Data Ecosystem Mapping Tool and Guide developed by the Open Data Institute in the context of the Microsoft Open Data Campaign can help analyze potential value that can be exchanged between actors within a data ecosystem.

Sharing value in a fair way is already complicated in the context of bilateral data sharing between the public and private sector, as the literature suggests, because reconciling diverging interests and expectations can be hard, irrespective of the number of stakeholders involved. In the context of multiparty data sharing, it can become even more complex to balance competing interests.

Understanding value in data sharing

Value in economic terms is generally used to refer to "added value," meaning the difference between inputs and outputs for a certain product. In the context of multiparty data sharing, this often translates into a focus on financial benefits that are relatively easy to quantify. At the macroeconomic level, the Organization for Economic Development and Cooperation (OECD), for instance, suggests that improved public and private "data access and sharing can help generate social and economic benefits worth between 1 and 2.5 percent of Gross Domestic Product - GDP (in few studies up to 4 percent of GDP)". At the microeconomic level, a well known study by Deloitte on the impacts of Transport of London's (TfL) practice of openly sharing non-personal data showed that companies using TfL data generated a gross value added between GBP 12 million and GBP 15 million per year, including directly supporting around 500 jobs.

Many benefits from sharing data are not so easily measured. For the public sector, diverse and numerous social benefits may be attained through data sharing partnerships. The OECD highlights in particular positive impacts on transparency and accountability and increased user empowerment as associated with greater data sharing. For companies, reputational or knowledge benefits, as discussed above, may also serve as motivation to engage in data partnerships. The OECD also points at the opportunity for private sector data providers to crowdsource new insights and exploit user-driven innovation linked to the emergence of a community that creates additional value that an organization on its own would not be able to create.

What are we making?

Success in data sharing depends on establishing and maintaining shared value for all partners. It means that all partners benefit from data sharing to some extent and no partner benefits disproportionately more than others.

That said, value distribution is not set in stone and can change over time, either because the needs and expectations of data partners change or because the data sharing initiative evolves in terms of focus, activities, and level of input needed from the different partners. Just as adjusting the value distribution might be required, it is also important for initiatives to be clear about their initial approach and be able to monitor whether the promised benefits materialize for the various partners and to what extent. Doing so allows them to correct imbalances and change value propositions as needs arise.

Like other aspects of data sharing, there is no one-size-fits-all approach to creating and distributing value creation. Recipes to ensure that benefits and value are distributed as fairly as possible can take many forms, often customized to the needs of specific partners or to differing contexts and ecosystems. Three, in particular, emerged from the landscape analysis in preparation for this cookbook: **lowering costs, innovation and the provision of data products, and the delivery of tailored services.**

Recipes for lowering costs

Reducing the frictions hampering data sharing is a key objective for many multiparty initiatives. Data sharing between multiple partners is characterized by significant transaction costs that relate particularly to the time and energy required to reach and formalize an agreement and to translate it into appropriate infrastructure and well-functioning cooperation mechanisms.

Initiatives can deliver value to their members by lowering these costs, thus also reducing the time it takes to effectively share data. This most often requires developing reusable assets or tools that can be leveraged to speed up reaching an agreement and setting up data pipelines. For instance, the Development Data Partnership reduces time to agreements and transaction costs for both data providers and users by supplying a common legal agreement template (which does not need to be negotiated on a case-by-case basis) and readily available data sharing and storage infrastructure. These assets are ready-made and can be adopted by new potential partners without lengthy negotiations.

Reduction of costs as a strategy for the delivery of value to partners can also relate to stages of the data life cycle beyond the sharing phase. As an example, the California Data Collaborative spares its members the costs associated with the recruitment of large teams of data scientists and researchers, as these are hired and pooled together by the Collaborative at a lower fee for all the water management agencies. Cost reduction is a visible and quantifiable strategy for data sharing initiatives to deliver value to members.

Recipes for delivering on innovation and data products

Effective data sharing requires that each partner has clarity on the objectives, use cases, input required, and expected output of the initiative. This is paramount to define value distribution. However, many successful initiatives continuously innovate in order not only to achieve but also to evolve and strengthen their value proposition over time. The experience from these initiatives suggests that (a) data sharing can be a catalyst for innovation but also that (b) focusing on innovation is often an incentive that keeps partners engaged and encourages new partners to join.

Data innovation and research efforts in the context of data sharing can lead to the development of **data products** (or data goods), understood as methods, solutions, and tools produced by one organization that can be successfully used by others. Initiatives often create data products that are accessible to different extents, i.e., they can give free public access to some data goods while restricting access to others so that only data partners can benefit from them. In some cases, in fact, monetization of the data goods is considered a strategy for financial sustainability, so it is in the interest of the initiative to put the data goods behind a paywall. This is an approach taken by the California Data Collaborative — only members have access to data products. In other cases, where other financial sustainability mechanisms are in place, data products are considered public goods and made broadly available. Generally, data products are part of the value distribution among partners when their access is restricted but not necessarily when they are open access.

Production of and access to data goods

The Development Data Partnership created and maintains a marketplace for different types of data goods. The marketplace includes accessible code repositories for derived data products and algorithms. For example, temporal maps (instead of household surveys) were used by the World Bank in 2020 within a project to understand the digital divide in Ukraine. This approach was shared as a data good on the marketplace and later replicated by other partners of the initiative in projects in South America and Africa. In the case of the Development Data Partnership, data goods are shared with all partners but not made available more broadly to the development community.

Other initiatives, in accordance with their primary aim, grant universal access to their data goods. For instance, the Global Forest Watch, an online platform that provides data and tools for monitoring forests, provides open access to various tools and data goods, such as the new Radar for Detecting Deforestation, which detects forest disturbances using satellite-based radar data. Global Forest Watch needs maximum dissemination of data to achieve its goal of stopping illegal deforestation and unsustainable activities in forests by empowering policymakers and affected communities to monitor the state of their forests and develop appropriate policy responses for threats. Restricting access to the data goods would be counterproductive to achieving this ambitious objective.

Recipes for the provision of tailored services

While data products or goods described in the section above are useful resources for interested stakeholders and partners, tailored services respond to the specific needs and requirements of organizations.

Tailored services include delivering customized training, responding to specific analytical or research needs of partners, or developing tailored software. These tend to be resource intensive to produce, and few data sharing initiatives within the development sector offer them to partners. Research suggests that provision of customized services is more typical in situations in which the private sector is heavily involved and/or in contexts where the data partnerships take the form of data collaboratives, as in the case of the California Data Collaborative (see below).

Within the development sector, **tailored services** constitute a strategy to provide value to data sharing organizations that would not otherwise benefit from the provision of data goods or other mechanisms to distribute, such as partners who are the biggest contributors in a data sharing partnership. Tailored services might incentivize these partners to remain engaged for the long term. For instance, Global Fishing Watch, conscious of the high level of resources required for governments to share their data with the platform, provides them with customized advice and addresses their specific policy questions as a way to ensure data sharing partners receive a fair return.

Provision of tailored research services

The California Data Collaborative (CaDC) provides a clear value proposition to all its partners, lowering the transaction costs of sharing data and reducing the cost that individual water supply agencies incur by hiring independent data talent to respond to the regulatory requirements that the State of California created in 2015. In addition to the standard data sharing and analytics services it provides, the CaDC has worked to develop software that can facilitate water data management and analysis.

CaDC is also characterized by a focus on customized research services for its partners. Partners can come to the CaDC with their own data research questions, and the CaDC team will assist them in finding the answer. For instance, one of the partners of the collaborative was debating whether to invest in storage facilities. The CaDC team looked at the water supply agency's current water use and ran time series analyses to predict future water use before concluding that additional storage wasn't needed. That decision saved the agency $20 million.

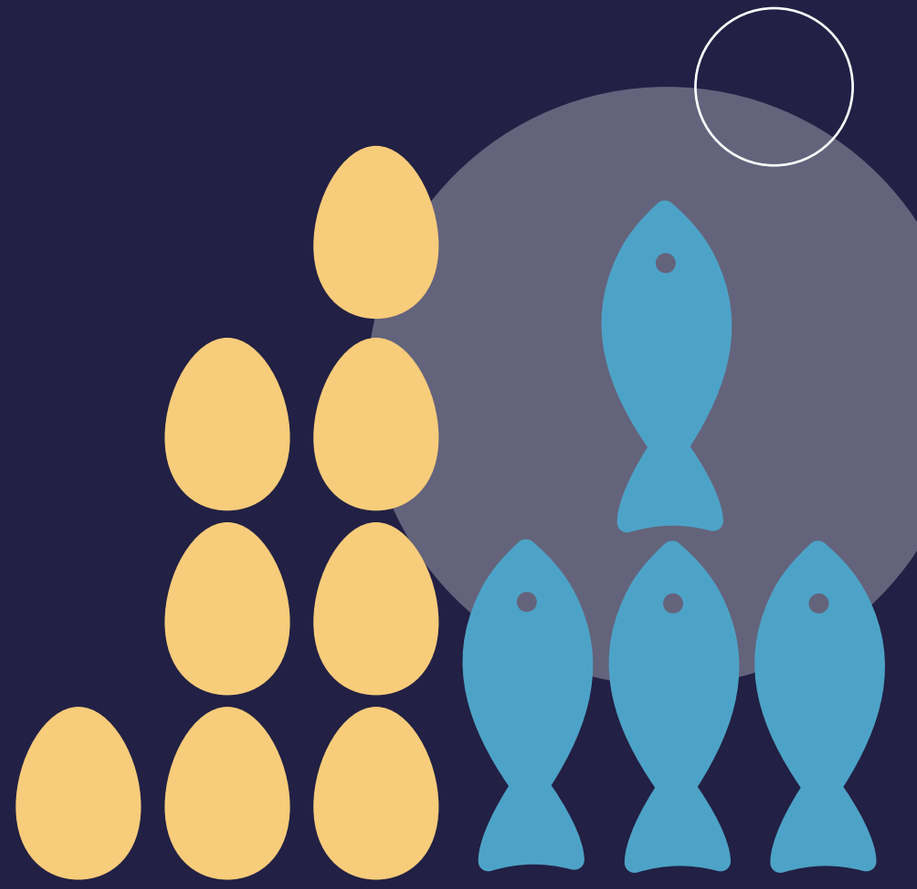
3

Dependable data and infrastructure

Protein

Recipes in this section

- Ensuring technical safeguards are in place
- Embedding quality assurance in data collection and analysis
- Interoperability and data format



Introduction

High-quality data and infrastructure (defined as the software and hardware underpinning data exchange) lie at the heart of successful data sharing. They can be compared to proteins, as they are generally put at the center of the plate and seen as the main course of a meal.

What is less obvious is that it is not the quantity of data shared nor the technological features of the infrastructure that ultimately lead to the success of data sharing initiatives, but rather their dependability.

Dependability encompasses two aspects of data and infrastructure in data sharing: the trustworthiness of these systems and the quality and appropriateness of the information created in terms of user needs.

- **Trustworthy** data and infrastructure are safe, unbiased, and regularly monitored.
- **Quality and appropriateness** are linked to the security and interoperability of datasets.

These two concepts reinforce each other and ensure that data partnerships result in dependable products that organizations can use to develop new applications, services, policies, and operations.

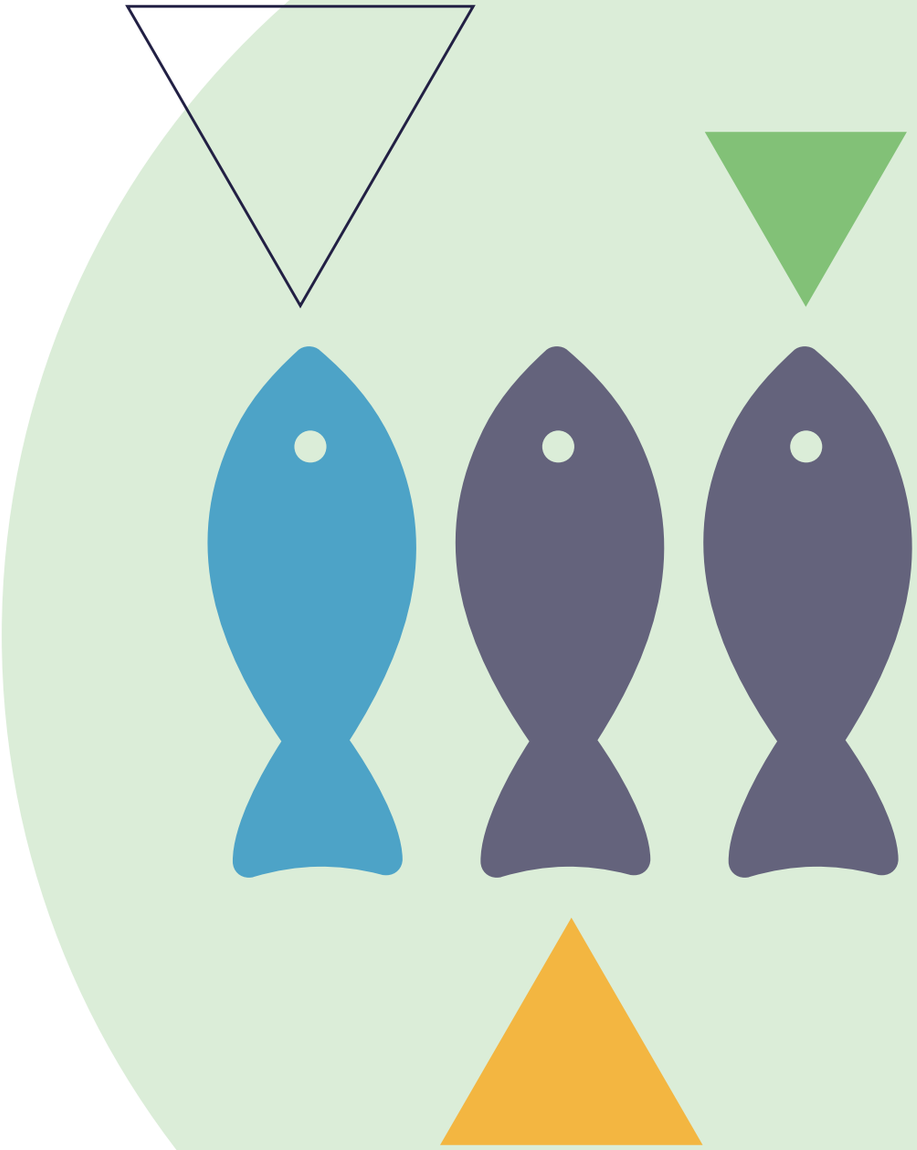
What are we making?

Data sharing initiatives may take different approaches to ensure their data and infrastructure are dependable, but what they generally have in common is their emphasis on:

- Ensuring technical safeguards are in place.
- Embedding quality assurance mechanisms in data collection and data analysis.
- Adopting appropriate interoperability and data format approaches.

A stronger focus on one or another of these elements depends mainly on the sensitivity of the data at hand, the characteristics of the data themselves (i.e., in terms of variety of sources or granularity), the targeted use cases and objectives of the initiatives, and the resources available. **Levels of trust between partners also influence the attention paid to technical safeguards and quality assurance in particular.**

For example, data aggregators such as the Humanitarian Data Exchange, which is an open-source and open-access web platform enabling humanitarian organizations to share data, spend more time and resources on ensuring interoperability of datasets, whereas initiatives such as Global Fishing Watch, which seek to inform government response, focus their efforts on ensuring data quality. This is because the former collects a wide variety of data from different sources whose interoperability is key to making the data usable, while the latter needs very precise data to create accurate maps of illegal fishing activities.



Recipes for ensuring technical safeguards are in place

Data sharing partnerships are characterized by opening up access to and sharing of information. However, doing so raises security concerns, which can be particularly serious in the context of personal data, defined as **any information that relates to an identified or identifiable living individual.** Even though most successful data sharing initiatives in the development sector use only anonymized, aggregated, or non-personal data, concerns among data providers about the fallibility of anonymization methods and the risk of re-identification remain.

Security and storage policies aim to address these concerns and are the primary tools to ensure technical safeguards are in place.

Data sharing initiatives can focus on ensuring data security by having in place the people, processes, and tools necessary for protecting data confidentiality and integrity against malicious attacks or unintended accidents throughout the data life cycle.

Approaches to **securely storing data** can be centralized, federated, or distributed, with each utilizing different ways to ensure the safety and security of the data concerned. While most initiatives analyzed rely on centralized storage systems, there is no one-size-fits-all approach and, in the development sector and beyond, a growing trend toward decentralization of storage is linked to increasing concerns over power imbalances and data hoarding.

Creating controlled data environments to secure data privacy

Since 2014, the LinkedIn Economic Graph challenge has invited research and analytic partnerships in which LinkedIn data can be leveraged to identify macroeconomic labor and economic trends.

To ensure the security of its members' data, the challenge provides extensive security training to each participating team and mandates that work be performed only on LinkedIn-issued laptops on the LinkedIn network within a monitored sandbox environment. Data downloaded outside of this network is heavily restricted, and a LinkedIn employee collaborator supervises all access to and use of data. Data use is restricted to the specific goal identified in the research partnership. In addition, an internal review board evaluates all research products created through the partnership.

Recent research carried out on LinkedIn data, for instance, shows that green skills are increasingly demanded on the job market, with at least 10% of job postings from the last year requiring them. It also suggests that more workers are green-skilling and transitioning into green and greening jobs, driving positive net transitions into these jobs.

Shifting from centralized to decentralized data storage

The Implementation Network for Sharing Population Information from Research Entities (INSPIRE) is a partnership to share data from Health and Demographic Surveillance Sites in five countries in East Africa. Initially, INSPIRE was established to create a large health data repository, and the data were stored centrally in a cloud-based facility, with the initiative, led by the African Population and Health Research Center, as the custodian of the data.

The INSPIRE team, however, in line with their strategy of increasing data providers' capacity and strengthening the skills of the research institutes involved in the initiative, intend to transition from being a repository of data to serving as a platform for services. This entails a shift toward a federated storage system in which the data providers remain custodians of their own data. INSPIRE's platform would then mine the data remotely for particular use cases.

As one of the key initiative stakeholders put it, the objective of this transition is to explore an approach in which no single partner holds all the data.

Recipes for embedding quality assurance in data collection and analysis

Data must adhere to agreed standards of quality to be usable. Data quality standards refer to the level of accuracy, currency, precision, and reliability of performance. Each data sharing initiative sets its own quality standards and defines what is acceptable and what is not based on the objectives it sets out to achieve. Clearly, some use cases (i.e., in the context of healthcare or humanitarian relief) require higher data quality standards than others. As such, no universal definition of "good enough" data can be established.

Investing in ensuring adherence to agreed-upon data quality standards can consume time and resources. However, setting up a system to accomplish this saves resources at later stages of projects, as errors and bias can be spotted before the data is put to use, saving costly efforts to correct errors once the initiative or platform is established.

Adopting appropriate **quality frameworks** at the data collection phase and establishing **transparent approaches to limit and mitigate bias** at the data analysis phase are useful steps for increasing data quality.

Defining appropriate data quality approaches

Approaches to data quality and veracity vary by initiative. For example, data exchange platforms like the Humanitarian Data Exchange (HDX) do not check the quality of the data they receive from partners. HDX's data architecture is not geared toward cleaning submitted data. This initiative adopts a "buyers beware" approach, where the veracity of the data is evaluated by the user.

Other initiatives work extensively with data partners to ensure the quality of the data that is shared. Organizers of the Global Fishing Watch (GFW), for instance, can spend months conducting quality checks of the data received from governments because each country reports its data differently. The GFW team standardizes the data format and checks for errors. Discrepancies such as missing data fields or wrong time zones are common, and the GFW team works with governments to fix them. Only once the GFW team is convinced of the quality of the data does it proceed to the analysis stage.

Another approach is to put the onus of data cleaning and quality control on the data suppliers. This is generally discussed at the beginning of the initiative, and the data partners agree to the initiative's data format requirements. For example, INSPIRE requires partners to complete the necessary data cleaning, quality checks, and quality assurance measures before sharing.

Transparency for bias mitigation and limitation

Haiti witnessed widespread violence in April 2022 due to fights between two gangs. The conflict led to the displacement of approximately 35,000 people from the affected area. Flowminder is a nonprofit foundation that specializes in analyzing Big Data such as call detail records, satellite imagery, and household surveys to solve development problems. To provide more evidence and details about the displacement, it formed a data sharing partnership with a telecom operator representing 74% of the national market share.

The objective of this partnership was to generate evidence to better understand the large-scale movements of the displaced population and support the provision of appropriate policy response. In its final report, Flowminder provided an extensive disclaimer about the limitations of data from mobile network operators, which are not statistically representative because access to phones isn't universal. The report cautions readers to consider the limitations of the data in drawing conclusions from the report.

Recipes for interoperability and data format

Data sharing by different partners and aggregating different data sources on common platforms requires significant investments in terms of **interoperability** and data format. Data are interoperable when they follow commonly agreed upon **technical standards** and have sufficient **metadata** for an analyst to identify provenance, concepts, methods, and usage guidance.

Both standards and metadata are necessary for interoperability. Without standards, even well-described data (i.e., with metadata) are not user ready, and, without metadata, it is extremely time consuming to understand data or put them to use.

Even within single initiatives, different types of data can require different approaches to standardization and interoperability. For example, the California Data Collaborative receives two kinds of data — billing and program participation data. The former requires little standardization, as the billing format is uniform across agencies. However, the formats of the latter can vary. The CaDC has therefore developed a software that partner agencies can use to standardize participation data before sharing them.

While there is often a strong emphasis on standards, it is important to acknowledge that successful initiatives invest heavily in producing high-quality metadata. Good metadata can smooth the process of transferring and using data. To guide data providers, the Humanitarian Data Exchange has developed and made available clear instructions on resource and metadata fields that are required and provides support to new partners to apply them.

Ensuring interoperability among disparate datasets

Any humanitarian organization can upload its data to the Humanitarian Data Exchange. To ensure data sharing on the platform is seamless, HDX supports contributions in any standard data format. However, from the beginning, HDX has recognized that some standardization was necessary to reduce the friction of sharing data.

Its original approach to data standardization required time and technical expertise from contributing humanitarian organizations. But, in 2014, HDX rebooted this approach to develop the Humanitarian Exchange Language (HXL). As most data on HDX was stored in spreadsheets, the new approach merely required the data to include an identifier (a symbol using a hashtag) in the column title, making it possible to merge the same fields across multiple datasets. While there are still challenges in the adoption of HXL, studies carried out by the data sharing initiative show that it has greatly improved data use and saved data processing time.

4

Mechanisms for supporting knowledge and skills strengthening

Dairy

Recipes in this section

Development of formal capacity building and training

Informal skills development and knowledge transfer

Introduction

A common characteristic of successful data sharing initiatives is their **emphasis on knowledge, skills and capacity building.** Sharing knowledge and skills strengthens data sharing agreements, just as the calcium in dairy products helps build and maintain strong bones.

It is widely recognized within the development community that it is not sufficient to make data available and expect organizations to access or use it. Instead, there is a consensus on the need to build data use skills among communities of users and on the importance of increasing data literacy to achieve sustained data use.

Often, however, users are not alone in lacking data skills. The landscape analysis, for instance, pointed out that partners of successful data sharing initiatives are concerned about their own organization's lack of data skills and the difficulty of finding and hiring the right people. Furthermore, the Global Partnership's work has highlighted that data governance skills, for instance, are among the major obstacles to establishing public-private partnerships.

Finally, it emerged from this work that limited attention is paid to building capacity at the community level. The landscape study indicates that almost none of the data sharing partnerships reviewed engage directly with the communities from which the data originate. Some models with limited engagement mechanisms included participation from civil society organizations and national representatives. However, these did not aim explicitly at building the data capacity of these communities.

What are we making?

Successful data sharing requires **building the skills of the user community alongside the skills of key stakeholders and partners.**

Identifying the set of skills needed for a data sharing partnership to achieve its objectives and staffing it with individuals who possess these skills is crucial to achieving success.

Furthermore, data sharing partnerships need to know their users very well and to understand their barriers to and requirements for accessing the data. This requires partnerships to focus on the capacity of external stakeholders who are expected to use the data.

Finally, to make data sharing fairer and more sustainable requires strategies to increase the data confidence of the communities whose data are being used. However, examples of how this can be done in practice are limited, in part due to the high costs of large data literacy programs that are perceived as falling outside the scope of specific data sharing initiatives.

To increase the capacity of external and internal stakeholders, approaches can range from producing formal training tools to establishing more informal mechanisms for transferring knowledge. A few successful recipes for **building institutional capacity** are provided below.

Recipes for the development of formal capacity building and training

The development of institutionalized capacity building and training represents a core activity for many successful data sharing platforms within the development sector. Capacity-building and training programs can take different forms, from virtual training, as in the case of the Humanitarian Data Exchange, to data fellowships, such as the Global Partnership's Data Science Fellowship Program, and frameworks and guidance documents like the Open Data for Social Impact Framework developed by the Microsoft Open Data Campaign.

While the focus on capacity building for data sharing is a clear driver of success, its importance varies depending on the type of stakeholders involved in the initiatives.

For instance, data sharing initiatives that work directly with national and local public sector organizations to inform decision-making are more likely to invest in capacity-building activities for these partners. This is because the initiative's objective requires that the ability to use data is permanently embedded in the public sector organization's functioning.

Furthermore, the landscape analysis suggests actors based in the Global South are more likely to focus on building local capacity. Within the landscape analysis, initiatives led by actors in the Global North were less likely to consider the direct, long-term benefits to actors within the Global South. Interviews indicate that actors based in the Global South are more concerned about addressing this skills imbalance through building internal capacity.

Capacity building for governmental organizations

In the context of the Global Fishing Watch platform, governments are both data providers and users of information.

However, not all governments are at the same level in terms of being able to collect and share data or use the data available on the platform to inform decision-making. Therefore, GFW has adopted a conscious and structured approach to capacity building.

GFW works with national governments to ensure their baseline capacity to collect and use the data concerning vessel-based human activity. GFW's team collaborates with the relevant data officers to clean and standardize the datasets. Further, once the data has been analyzed, the team works with the government department to translate data trends to identify bad actors and inform enforcement action.

Capacity building in the context of South-South data partnerships

The Implementation Network for Sharing Population Information from Research Entities (INSPIRE) aims to build capacity among its partners and give credit to them for their work. To achieve this goal, it provides tools and services to the data producers that enhance data sharing, reproducibility, and ancillary data production processes. These tools also assist data producers in discovering, accessing, and appropriately using the shared data and metadata. In the coming years, INSPIRE aims to build in-house technical skills of data producers, which will enable them to participate in the extraction, transformation, and loading of data into common data models.

Recipes for informal skills development and knowledge transfer

Knowledge transfer and skills development does not necessarily need to be embedded in formal training or capacity-building programs. Very often, peer-to-peer learning or informal mechanisms can increase the capacity of data partners or data users in the context of data sharing initiatives. Additionally, adopting an institutionalized approach to training doesn't exclude other informal mechanisms of transferring knowledge. On the contrary, these two approaches complement each other.

It emerged from the landscape analysis that, even in the context of successful data sharing initiatives, every partner did not possess the entire range of skills needed to establish and maintain effective data sharing at the beginning of the initiative. Rather, partners often had complementary skills that, taken together, led to sustained and successful data sharing.

While a variety of skills across partners is a good starting point for data sharing, in the long term, an initiative benefits by increasing the capacity of partners through sharing their unique knowledge and strengths (i.e., technical skills, legal understanding) with each other.

Informal skills development and knowledge transfer mechanisms can take a plethora of different forms, ranging from **peer learning networks** to **knowledge hubs** and collaborative workspaces like GitHub and resource-sharing platforms.

Knowledge cafes

Since its inception, the Hong Kong Data Trust 1.0 has organized periodic cafes to facilitate knowledge exchange between data partners and explore topics of interest in an informal setting. These forums are open to the data sharing initiative stakeholder community (made up of more than 500 organizations from the local transport sector) and to the public. The knowledge cafes often include presentations from and conversations with external experts.

For instance, in 2021, Data Trust 1.0 hosted a knowledge cafe on the topic of "data on walking" and invited experts from the working group on International Walking Data Standards and from the Walk21 Network to share insights about international developments to increase collection of, access to, and use of walking data.

These events benefit the stakeholder community and provide an opportunity to exchange knowledge and information between partners in a more casual way and outside the context of formal governance structures. This contributes to capacity building and enhances trust between partners.

Knowledge repositories

The Microsoft Open Data Campaign's primary aims were to facilitate data collaborations and create common resources to support data sharing across organizations. From the outset, the initiative was keen to develop a knowledge repository with tools and resources to benefit data sharing more generally.

The initiative joined with the Open Data Institute and The GovLab at New York University to develop resources, courses, and networks to help organizations and leaders navigate the transition to more open and collaborative data approaches. The initiative has worked with its knowledge partners to develop resources to provide data stewardship guidance for public sector and private sector actors. This effort includes the creation of a new Data Stewardship Academy; an Open Cities project to build community and share insights among cities using data to drive change; and a Peer Learning Network, which will act as a community of practice and learning.

5

Flexibility and adaptability

Healthy fats

Recipes in this section

Embedding flexibility in governance, operations, and architecture

Sustainable financial models

Introduction

Data ecosystems in which the data sharing initiatives operate are never static, but rather continuously evolve, particularly from technological, stakeholder, and cultural perspectives. It follows that governance, architectural, and operational choices cannot be set in stone, and **the possibility of adapting or changing must be envisaged at the outset.**

Financial and funding approaches also cannot be immutable. Most data sharing initiatives start with specific project funding but need to transition to something else at a later stage when the funding ends. Changing the funding model can mean diversifying sources of cash or adopting subscription or fee-based models. Whichever approach is chosen, this transition from one funding approach to another remains a delicate moment that can break even very successful (from a user perspective) initiatives.

Healthy doses of flexibility and adaptability are therefore as important in data sharing as healthy fats are for human diets. Reasonable doses of flexibility and adaptability are critical for data sharing to succeed in the long term and for partners to remain involved.

What are we making?

Like trust, flexibility and adaptability are generic terms that do not help organizations to figure out what exactly they should be doing to ensure success in data sharing. More concretely, what data sharing initiatives should be paying attention to is **how to embed flexibility in their governance, operations, and data architecture** and **how to identify suitable sustainable financial models** that allow them to evolve over time.

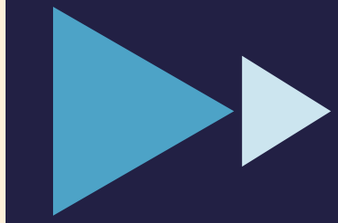
Unfortunately, the development sector does not provide many examples of flexibility embedded in operations, governance, architecture, and financing. On the contrary, this area remains under-explored and represents a topic for which successful recipes are missing.

This is not to say that no initiative has successfully included flexibility and adaptability, but rather that there is limited knowledge and research on how it can be done. Even successful initiatives have not explicitly theorized how to ensure the implementation of an evolutive approach. The sections below highlight a few findings and examples from the landscape analysis, but gaps remain concerning concrete steps for data sharing partners in this area.

Recipes for embedding flexibility in governance, operations, and architecture

Flexibility in governance and operations is difficult to achieve due to tension inherent to formalizing ways of functioning to ensure stakeholders' trust and building in margins of maneuverability for implementing changes. The way in which organizations have traditionally addressed this challenge is through adopting change management processes. These refer to a series of tasks for a seamless transition from a current state of affairs to a new one without affecting the data sharing operations or decreasing trust among stakeholders. Change management processes can be negotiated at the beginning of a partnership in conversations around governance. While no examples of change management models in data sharing partnerships within the development sector surfaced in this study, discussions with experts suggest that this is a relevant option for data sharing initiatives that are built to last.

From an architectural perspective, experts agree some of the biggest challenges are the scalability of existing infrastructure and how to finance information technology (IT) updates down the road. In this respect, there is a consensus on the importance of developing scalable IT plans at the very beginning of an initiative, anticipating how IT needs will evolve based on, for instance, an increased number of data users or the emergence of new technologies and standards. As an example, the uptake of application programming interfaces included the need to update much of the existing open data infrastructure to accommodate a different way of consuming the data by users. In the last few years, application programming interfaces have facilitated access and use of data by developers and programmers, reducing the need to download files or push paper.



No person or data sharing initiative can predict the future. **But thinking ahead to plan for necessary updates to the infrastructure** helps clarify future funding needs and ensures the initiative will remain relevant for stakeholders and users.

Repurposing data sharing initiatives to face new challenges

The Microsoft Open Data Campaign experience shows how important it is to build flexibility and change into data sharing collaborations.

One of its facilitated partnerships was aimed at monitoring air quality in the United Kingdom and involved the Alan Turing Institute – the U.K.'s national center for data science and artificial intelligence. After the emergence of the COVID-19 pandemic, this project was rapidly repurposed to help London authorities measure the impact of lockdown restrictions on the city. The Alan Turing Institute did this by deploying the infrastructure of the air quality study to measure the city's "busyness" and the public response to government interventions.

While this was not the original intent of the data collaboration, the data streams had already been put in place, and the data infrastructure was able to generate insights relevant to the COVID crisis. High levels of trust between data partners allow for such flexibility.

Recipes for sustainable financial models

A data sharing initiative's ability to fundraise and/or generate revenue impacts the sustainability of its operations and investments in continued growth. Ensuring financial sustainability and exploring a diverse set of funding models was a common concern across most data sharing initiatives included in the landscape analysis.

While a majority of the initiatives remain donor funded, all of them recognized the importance of diversifying funding sources and/or building revenue models into their operations. Most initiatives expressed interest in moving from donor-based to revenue-generating (or partially revenue-generating) approaches, but there were few examples of successful transitions. Similarly, there are not many examples of initiatives that managed to diversify the origin of their funding or to ensure sustainability based on this approach.

Some **funding models** currently being explored include subscription fees, fees for data services offered, and fees for data storage/processing. However, such models are still in their early stages, and their scalability and sustainability are not yet clear.

Adopting appropriate financial models

The California Data Collaborative is an example of a self-funding initiative based on a revenue-generating model. Most of CaDC's operational costs are met by the annual fee paid by members. While this makes the CaDC more expensive than a trade association, it is cheaper than the cost each agency would incur to hire a data manager or software developer. CaDC also receives supplementary grant funding to underwrite its research projects.

The Humanitarian Data Exchange, on the other hand, relies on core funding from key partners of the initiative. Initially, in 2014, HDX received project funding from the United Kingdom's Department for International Development's Humanitarian Innovation and Evidence Programme; the Government of Sweden; and the Humanitarian Innovation Fund. Today, most of HDX funding is ensured by the United Nations Office for the Coordination of Humanitarian Affairs, and this enables the continuity of the operations of the platform. HDX also seeks additional funding for specific projects (IT updates) from other donors, but the availability of these funds does not affect the initiative's core functioning. Applying for funding to various donors has allowed HDX to refine its vision and get constructive feedback. Diversifying funding sources for core and additional activities helps mitigate financial risks and allows for tapping into additional resources when needed.

Finally, changing the funding approach while the initiative is ongoing is always possible. The Data for Children Collaborative, for instance, is currently funded by the Scottish Funding Council and the Data Driven Innovation Programme and will receive this financial support for the first three years of existence. However, the Collaborative aims to become a consultancy that builds data collaboratives as a service. This will entail switching from donor-based funding to revenue-generating approaches in the near future.

Conclusions

The burgeoning number of multiparty data sharing initiatives within the development sector testify to the importance of data collaboration for the achievement of the Sustainable Development Goals. While increased data sharing is good news, new data platforms on their own won't ensure we meet the 2030 agenda. Data sharing initiatives like the ones described in this cookbook will not support the objectives of the development community unless they are effective. Effective data sharing within this sector requires more knowledge sharing to identify best practices that can be replicated across the world.

The recipes in this cookbook aim to do just that. This cookbook lays the groundwork for discussions about the key ingredients for effective data sharing and provides a number of recipes for inspiration. This is only the beginning. Much remains to be done to close gaps in the existing knowledge, due in part to the limits of this work and the relative novelty of many data sharing initiatives, especially in the development sector.

Contribute to the cookbook

Are you involved in implementing data sharing initiatives in this sector? There are lessons to be learned from both successful and less successful ventures. We welcome your contributions and invite you to share ingredients and recipes for effective data sharing with our community.

If you have a recipe or suggestion to share, do not hesitate to leave your input through this Google Form, and we will get back to you about how to share your knowledge with the community.

Annex A

This CookBook draws on the landscape analysis commissioned by the Global Partnership and carried out by Athena Infonomics between March and September 2022. The final report of the analysis is available here and its executive summary here. Interested readers can also find the summaries of the case studies here.

The Global Partnership is grateful to the Athena Infonomics and Atlas AI team (Shruti Viswanathan, Deepa Karthykeyan, Vivek Sakhrani, and Saiyed Kamil) for their relevant and insightful work.

Annex B

Effective Data Sharing Reference Group members:

Alewaert, Caroline
Data Privacy Specialist
UN Global Pulse

Arai, Ayumi
Project Assistant Professor &
Co-lead of the Spatial Data Commons
Center for Spatial Information
Science, University of Tokyo

Barrett, Alyssa
Senior Product Strategist,
Land & Carbon Lab
World Resources Institute

Calderon Machicado, Claudia
Partnerships Lead,
Development Data Partnership
World Bank

Cecconi, Gianfranco
Lead for Collaborative Data Ecosystems
Capgemini Invent

Clause, Victoria
Digital Climate Smart Agriculture
Manager
Mercy Corps AgriFin

Cooper, Sonia
Assistant General Counsel,
Open Innovation Team
Microsoft

Díaz, Alexander
Senior Manager, AI for Social
Good & Crisis Response
Google.org

Hodes, Cyrus
Co-founder & Chair, AI Initiative
The Future Society

Hoffman, William
Former Project Lead, Data for
Common Purpose Initiative
World Economic Forum

Jansen, Ronald
Assistant Director & Chief of Data
Innovation and Capacity Branch
United Nations Statistics Division

Kiragga, Agnes
Research Scientist Data Science
African Population and Health
Research Center

Krambeck, Holly
Program Manager and Founder,
Development Data Partnership
World Bank

Maloney, Christopher
Program Officer, Gender Equity
and Governance Program
William and Flora Hewlett Foundation

Orrell, Tom
Director
DataReady

Powell, Joshua
Chief Executive Officer
Development Gateway:
An IREX Venture

Scrollini, Fabrizio
Executive Director
Open Data Latin American Initiative

Schüür, Friederike
Former Fellow, AI Ethics
and Digital Governance
UN Global Pulse, now UNICEF

Sellitto, Michael
Affiliate Fellow
Stanford Institute for Human-Centered
Artificial Intelligence

Teran, Javier
Statistician
United Nations Office for the
Coordination of Humanitarian Affairs

Ure, John
Director of Fair Tech Institute
AccessPartnership

Villarreal, Pablo
Head of the National Digital Platform
Executive Secretariat of the National
Anticorruption System, Mexican
Government

Vollenweider, Xavier
Director of Mobile Data Partnerships
Flowminder

Global Partnership for Sustainable Development Data

This document was drafted by Martina Barbero with support from Janet McLaren and reviewed by Jenna Slotin, all of the Global Partnership for Sustainable Development Data.

This project was made possible by a grant from Google.org. The authors sincerely thank the research team from Athena Infonomics and Atlas AI (Shruti Viswanathan, Deepa Karthykeyan, Vivek Sakhrani, and Saiyed Kamil), who produced the landscape analysis on which this document draws (see Annex A), as well as the experts from the Reference Group for Effective Data Sharing: Beyond Platforms project, who contributed their time and feedback to elaborate on key findings. The members of this group are listed in Annex B.

With support from
Google.org