



BILL & MELINDA  
GATES *foundation*



# DEEP DIVE

---

## Data Commons

This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). This guide was produced under DAI's Digital Frontiers Project (Cooperative Agreement AID-OAA-A-17-00033) at the request of USAID. The contents are the responsibility of Development Gateway and do not necessarily reflect the views of USAID or the United States Government.





## At-a-Glance

- Pools and co-locates **data as a common resource** for shared gain or private benefit.
- **Prioritizes interoperability** and access to data.
- Can be **resource intensive** for community engagement, participation, and data infrastructure.
- Limited examples of true data commons in agriculture and LMICs.

**Data commons are a specific type of commons, for a field or discipline, to accelerate access and discoverability to increase the impact and benefits of the data.**<sup>1</sup> Data commons have a strong interoperability element, focusing not just on data access, but also on experimentation and interaction. Where a data trust pools data and sets terms of access, a data commons pools data and offers relatively **unrestricted access**. Degradation through restriction of access or malpractice is discouraged, although protective measures of the commons may not exist. Value is derived from services that use the data—, this is viewed as the legitimate claim of those who create it.

**Data commons co-locate data as a digital resource to store, manage, share, access, and interact with collectively owned data with and by a community.** The idea of collective knowledge commons dates to historical practices, as well as the “tragedy of the commons”<sup>2</sup> and Ostrom’s Design Principles.<sup>3</sup> Throughout history, there have been different types of “property regimes” that have focused on or are owned by individuals. Proprietary rights within customary law in low- and middle-income countries (LMICs) more commonly allow for a number of specific interests to be vested in different holders.<sup>4</sup> **Communal ownership continues to seize the imagination of people and disciplines**, with distinct types of knowledge, epistemologies, and approaches.<sup>5</sup> The legal framework can hold agreements between peers and should address governance, sharing, intellectual property, and access to resources via a regime of **collective ownership**. The governance needs to reflect community values and goals.

**Data subjects decide if and how data is shared.** The model is ideally structured democratically, reflecting members’ interests. Data commons provide a platform for equitable data rights for vulnerable groups and individuals. A well-governed data commons can unlock the power of shared data while respecting the data subjects. Some see it as the model to rebalance democratic control using data as a collective resource.<sup>6</sup> Others suggest that the model should include ownership (or control) of platforms.<sup>7</sup> Centralized and decentralized data commons can help prevent the formation of data monopolies that disenfranchise data subjects. Here, no one contributor is the data owner.

1. Grossman, R. (2022). [Ten Lessons for Data Sharing With a Data Commons](#). arXiv preprint arXiv:2207.11167.

2. Hardin, G. (1968). [The Tragedy of the Commons](#). *Science*, 162(3859), 1243–1248. <http://www.jstor.org/stable/1724745>.

3. Ostrom, E. (1998). [A Behavioral Approach to the Rational Choice Theory of Collective Action](#): Presidential Address. *American Political Science Review*, 92(1), 1–22.

4. Bennett, R. (2007). [Property rights, restrictions and responsibilities: their nature, design and management](#). University of Melbourne, Department of Geomatics, p. 654.

5. Hicks, J. (2022). [The future of data ownership: An uncommon research agenda](#). *The Sociological Review*, 00380261221088120.

6. Morozov, E., in DECODE presents: Data Commons and the City, watch [here](#), & Tarnoff, B. in Taking on the Tech Titans, Reclaiming our data commons, [see webinar](#).

7. LSE Lit Fest 2017: Platform Capitalism by Nick Srnicek.

**Data is stored on a cloud-computing infrastructure, usually through software, tools, and apps for managing, harmonizing and analyzing data.** Some see it as an evolution from earlier databases and data clouds. Data commons are not just a data repository—they are a unified database produced by normalizing and aligning schemes and entities across various datasets. The AgroDataCube, for example, aims to use common agro-semantic standards, stimulate the use of open source data, and exchange open knowledge across the agri-food chain.

The Data for Development (D4D) community has focused on securing a space for the free flow of public data. This is clearest in the data commons model, where data are held in one central and open repository, available to journalists, activists, and statisticians, as well as firms. Many sectors have expressed their need to find business models that incentivize private-sector participation. For corporate actors, the framework means releasing data either via data philanthropy or in return for business opportunities. Actors such as UN Global Pulse and WEF have framed efforts as a win-win for everyone, as more generation, personalization, and centralization reinforce strategies seeking control over LMICs' data infrastructures. However, this push to reinforce corporate—and often foreign—control over data systems can deepen existing dependencies and inequalities in power and expertise, as some have indicated.

## Data commons in agriculture

**Data commons improve access to data and develop a shared responsibility.** There are a number of interesting initiatives that share these goals. Often these projects are research communities focused on the FAIR principles (findable, accessible, interoperable, and reusable), like Agri-Food Data Canada. The Ag Data Commons, funded by the U.S. government, seeks to improve access and discoverability of agricultural data for research. The initiative aims to encourage innovation, data use, integration, and visualization for better policy and decision making and provides a data repository for long-term archiving. The Ag Data Commons uses an open license with few restrictions to access and use data. It authorizes the U.S. Department of Agriculture's compliance with open access and data requirements, as well as FAIR.



Citizen Science for the Amazon, created in 2017, protects fishing in the Amazon by collecting, monitoring, and sharing Indigenous, local, and international knowledge. The network of stakeholders and citizens share observations via an app and platform into the first database available for researchers, practitioners, and decision makers. It is designed to guide the data governance decisions by connecting local efforts, using interoperable standards and aggregating data and making the information open, safe, and accessible. Partners agreed on elements of the data governance, including guiding principles, variables, protocols, terms of use, credit, and privacy protection. The public engagement fosters the sustainable use and management of resources, improves livelihoods, and empowers local communities to negotiate with government agencies and others on issues such as securing fishing permits and selling sustainably caught fish in niche markets with higher prices, using collected evidence.<sup>8</sup>

---

8. World Bank, [Citizen science protects fishing in the Amazon](#), World Development Report 2021.

**The infrastructure can engender trust and transparency across the value chain.** Collaborating with stakeholders to design and develop an approach that engenders trust (via transparency and informed consent) is crucial to increase data sharing and adaptation of tools. The governance activities need to be transparent and enable farmers' meaningful control over data, which should indicate how and with whom data is shared. Successful examples engage the members in the design and innovation processes.

**A licensing model and certifications can help mediate relationships and promote participation.**

Some have suggested that a successful agricultural data commons requires a back-to-front license that moderates the rights and responsibilities of data collectors and users and allows for maximum use. A certification mark should be added to indicate that the data was sourced equitably and act as an incentive for others to participate. Efforts like the Fair Trade Movement provide a template for how data commons could evolve; this was made possible via a network of value-driven member organizations that worked together to devise and enforce the standards.<sup>9</sup>

### Why data commons matter

**Data commons provide for a number of benefits, including:**

- An approach to pool data as a common resource for shared gain or private benefit.
- Supported standardization and resource sharing via collaborative governance.
- Prioritized interoperability and access to data.
- The potential to democratize data through easy access and availability. Decentralized, citizen-owned platforms play an important role for a public utility.
- A bottom-up initiative to address challenges of consent and anonymization.

### Key components for success

- **A wide spectrum of community-driven governance.** Considerations revolve around data hosted and the licensing contract that governs data: data contribution, use, access, security, and conflict resolution. Data commons need to be able to deal with unique challenges from cross-institutional and multidisciplinary areas, resource integration, data sharing, and data harmonization for greater interoperability.
- **Building participation.** Its success and sustainability rely heavily on fostering a community capable of using an open and connected environment and motivating engagement. Incentives like certification could help explain the benefits of equitable data sharing.
- **Selecting a platform.** A key feature is the software infrastructure, suitable for each type of data, controlling access to storage, structure and analysis, or visualization. It should also support interoperability with other platforms, authorizing participation in a wider ecosystem.
- **Procuring the data.** Data commons assist and harmonize data and can produce many types of products for the community. It is a gradual, labor-intensive process to curate and harmonize data. The value addition would be through centralizing this effort.

---

9. Baarbé, J., Blom, M., & De Beer, J. (2019). [A proposed "agricultural data commons" in support of food security](#). *The African Journal of Information and Communication (AJIC)*, 23, 1-33.

## Challenges/pitfalls of the data commons approach

- **Developing a data commons can be resource-intensive** from a community engagement, participation, and data infrastructure lens. Successful commons need to invest accordingly. Sustainance can require specialized computing infrastructures and operational costs, especially in the initial stages of software customization (see the Ag Data Commons). Many programs struggle to find funding for the infrastructure. Challenges are often of attitudinal nature, with major barriers being a risk of compromise due to privacy concerns.<sup>10</sup>
- There is a dearth of consensus within the research community for a set of standards to support data commons. Currently, many standards are not accepted for indexing, nor are Application programming interfaces (APIs) for data access, and for authentication and authorization protocols to access controlled data. Developing these protocols will need significant external technical expertise and support.
- **There are only a few examples of genuine data commons in LMICs.** While this study highlights some initiatives that have data commons features, they are often still evolving. It is difficult to paint an accurate picture of the functioning of data commons in agriculture in LMICs.

## Crucial applications in LMICs

- **The inclusion of legal traditions as sources of knowledge about communal ownership rather than dominating market-oriented, individualistic approaches to property.** More research is needed on non-Western traditions as sources of knowledge about communal ownership.<sup>11</sup>
- **The occurrence of “polycentric systems of governance,”**<sup>12</sup> defined by multiple governing authorities at various levels, and the occurrence of “legal pluralisms,” as demonstrated in the context of the PescaData and Abalobi case study. Initiatives will have to deal with the fact that often no single institutional configuration is deployed.
- **Data commons have been used to better understand and access large scale government data sets.** The India Data Commons showcases datasets that are issued by ministries and governmental organizations, and offers better visualization via a knowledge graph. Data commons could assist in limiting dataset access inequities experienced by smallholders.<sup>13</sup>
- **Share knowledge on independent digital infrastructure through workshops and training.** Initiatives like Tierra Comun provide capacity-building workshops and training in the fields on privacy, security, and data protection. Their goal is to encourage a crucial digital culture for communities and collectives to technological re-appropriation. In Brazil, the feminist hacker collective MariaLab aims to make tech spaces varied, including autonomous networks and guides on building community networks.

---

10. Mills, S. (2019). [Who owns the future? Data trusts, data commons, and the future of data ownership. Data Trusts, Data Commons, and the Future of Data Ownership.](#)

11. Hicks, J. (2022). [The future of data ownership: An uncommon research agenda. The Sociological Review](#), 00380261221088120.

12. Carlisle, K., & Gruby, R. (2019). [Polycentric systems of governance: A theoretical model for the commons. Policy Studies Journal](#), 47(4), 927–952.

13. Baarbé, J., Blom, M., & De Beer, J. (2019). [A proposed “agricultural data commons” in support of food security. The African Journal of Information and Communication](#), 23, 1–33.

## Business viability and sustainability

A data commons approach provides for principles for revenue models that distribute benefits widely, without commodifying or privatizing the “goods.”<sup>14</sup> The commons constitutes a system to share wealth so that people can control, manage, and distribute data similar to common resources. Data monetization can work if members are actively involved in collective decision making, similar to union models. Unions are financially supported by their members, which allows them to work democratically and on behalf of themselves. Fees are paid on a regular basis to a committee, which acts as a data steward, and is elected by the members to work on behalf of them, within set data governance. Members pool their resources to achieve these goals set by the committee. Any profits can be used to sustain the data commons or support members.

It is important to communicate the scope and value addition of a data commons for users who want to contribute data and those who use it. This will be essential in case there are fees linked to accessing data. The organization could include incentives for sharing data, like providing credit for data citation, enhancing research performance, or providing access to latest data. There are three different business models that can be used to maintain data commons: (1) users will pay to access data, (2) users will pay to store their data within the repository, or (3) a central organization supports the repository.

The funding of a centralized data commons remains unclear. It is feasible that this is paid for by a national or international authority. For it to be efficient and effective, a data commons must be strengthened by stable funding, especially in its early stages. Data services may be compelled to pay a small fee for access, similar to how some authorities are funded in the financial sector. However, it is likely data services would assume the role in validating data held within the data commons, as it is in their interests to ensure they gain access to valid data.

An interesting example of government-led data commons infrastructure is presented in the DECODE project. The city of Barcelona attempted to create data infrastructure as a new meta-utility under the DECODE project. The project included a data sovereignty clause, which required suppliers to share the data using open APIs and open standards; explored cryptographic tools that allowed citizens greater autonomy over data sharing; and provided data commons to local institutions, cooperatives, and social-sector organizations that build public value via data-based innovations. While the project itself has now closed, it highlights the importance of governments in creating a data infrastructure.<sup>15</sup>

## How can stakeholders create an enabling ecosystem for data commons?

### Role for the government

- Identify opportunities for creation of government-led data commons architecture.
- Invest in creation and management of data commons infrastructure and set up regulatory bodies.

### Role for the private sector

- Adopt transparency principles that foster an open data approach.

### Role for the development sector

- Work with public institutions and farmer organizations to build data capacities for a data commons approach.
- Promote technical research and guidance on the feasibility of agricultural data commons.
- Invest and develop resources for knowledge sharing and training facilities.

---

14. Bollier, D. (2014). [Think like a commoner: A short introduction to the life of the commons](#). New Society Publishers.

15. Gurumurthy, A., & Chami, N. (2022). [Working Paper 23 Governing the Resource of Data: To What End and for Whom? Conceptual Building Blocks of a Semi-Commons Approach](#), Data Governance Network.



### Case study: **Farmer Cooperatives in Uganda United in a Data Collaborative.**

A consortium of public and private organizations carried out a pilot in 2019–2022, aiming to build a coordinated system with effective data governance requirements, with farmer organizations and the Ministry of Trade, Industry, and Cooperatives (MTIC) in Uganda. The governance structure resembles and aspires to be that of a data commons approach. For now, the National Cooperative Database is more structured as a data collaborative data governance approach, aiming to help actors understand the value of harnessing the power of data, create a sense of empowerment, and establish negotiation power and agency for farmer organizations. The case study provides insights into data commons, the challenges of data collection, participation with farmer cooperatives, and data ownership and control.



**Biobanks and Data Safe Havens (DSHs)** are platforms used to store research data, for access exclusively by approved stakeholders, with strict safeguards controlling access to medical and personal data. In the health sector, vast amounts of data remain in silos, managed and governed by individuals in myriad ways. A DSH allows access to data that is otherwise not available. The benefits of pooling health data have driven data commons platforms like BC OVCARE Biobank or CommonHealth, where individuals deposit personal data and decide how to share them. CommonHealth makes it possible for users to download their records to their device, store them securely for themselves, or share them consensually with their apps and services on their phone. The research team of BC Biobank built a governance model and platform to implement protocols and standard operating procedures for management and governance of interoperable data and to streamline data requests and inquiries, making data available to the research community according to research and ethics requirements. They developed the data holdings into a data commons. They found that no single solution could meet all the different data needs and opted instead to integrate multiple solutions to achieve the desired outcome. To make the data commons more participatory and patient-centric, the team is launching an online consent process and adding patient-reported outcomes. DSHs provide for a data governance approach that brings researchers together, providing support and guidance from various sources for different research projects. The stakeholders can facilitate the navigation for requirements and cater to a mechanism of data sharing from each project. A DSH may also be seen as a trusted custodian and supplier of data and data services, representing the interests of data providers.