DEEP DIVE

Data Collaboratives

Data Collaboratives

At-a-Glance

- Cross-sector, public-private collaborations between multiple organizations aim to unlock and share data for societal benefit.
- Many data collaboratives in LMICs focus on improving the data infrastructure, improving data access, and informing decision making.
- Drive innovation and create an enabling data sharing space for agriculture.
- Limited participation with and contribution by farmers compared to other models.

Data collaboratives are cross-sector, public-private collaborations aimed at data collection, sharing, and processing for the purpose of a societal benefit, emphasizing the collaboration between parties and going beyond data sharing. The term was first coined by the NYU Govlab.[1] There are an increasing and varying number of data collaborative initiatives. Data shared by companies for public good is of particular interest, since much of the data that is critical for addressing societal challenges lies in private hands.[2] The premise of data collaboratives is to create new value, foster innovation, deliver greater insights, and improve public governance, beyond the direct goals and capabilities of each individual actor.[3] The model provides partners with the option to choose the rules of data exchange, with the help of a data steward, and create greater control, trust, and confidence between partners.

Many in the development sector advocate for open data frameworks for public data and commercial incentives for private-sector data providers. There are ample examples where data collaboratives have gained access to data from social media platforms to track opinion and behavior in low- and middle-income countries (LMICs).[4] Many of these initiatives are made possible through "data donation," where anonymous company data sets are shared for nonprofit purposes. Donor organizations, nonprofits, and governments are also encouraged to provide data to the private sector. For example, the Grameen Foundation shared farm data from its Ugandan Community Knowledge Worker Initiative with data scientists and analysts from private firms.[5]

Data collaboratives in agriculture

This study examined over 20 projects that focus on public-private partnerships (PPP), such as the Sustainable Technology Adaptation for Mali's Pastoralists project and AfricaFertiliser, a PPP between input providers, to gather and operate data on markets and prices. There are also instances of private-academic partnerships, for instance, Intel's Big Data for Precision Farming on optimizing irrigation. Some use cases in agriculture are highlighted below.

1. DataCollaboratives.org is a resource on creating public value by exchanging data run by the NYU GovLab.
2. Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development.
3. Klievink, B., Van Der Voort, H., & Veeneman, W. (2018). Creating value through data collaboratives. *Information Polity*, 23(4), 379–397.
4. For example, in Indonesia, United Nations Global Pulse has used mobile data to understand food price fluctuations.
5. Krepp, S. (2013). Fostering entrepreneurship among Ugandan Smallholders, In *International Trade Forum* (No. 2). International Trade Centre.

Drive innovation and an enabling ecosystem for agriculture. Data collaboratives serve as incubation hubs to bring different technologies, companies, and data-sharing models together. The FarmStack Case Study includes an open-source protocol for transfer of data and provides reliable and regulated data exchange to advise solutions for farmers, limit costs, and increase the reach of tools. The platform developed a free and open-source solution to allow peer-to-peer sharing with selective and enforceable data protections via a local interface.

Data collaboratives can ensure a safe data sharing atmosphere. This was one of the main goals of WAGRI, where the terms of use between users and operators and an agreement linking WAGRI users and farmers were formed based on the guideline set by the Japanese Ministry of Agriculture Forestry and Fishery (MAFF). The platform demonstrates how a data collaborative can break silos. Special attention is given to public engagement and the need to understand perspectives of all stakeholders. The use of data is dependent on organizational arrangements and the aspects around the problems and expected outcomes.

Some data collaboratives have sought to leverage data from diverse sources, along with open public data to provide insights to farmers and others in the value chain. Clima y Sector Agropecuario Colombiano (CSAC), a partnership between the International Centre for Tropical Agriculture (CIAT), the Colombian Ministry of Agriculture, and farmers, combined data on weather patterns with farm data to measure conditions for rice varieties. CSAC provides predictive data and informs decision-making among farmers. Studies suggest that this predictive modeling could have a huge economic impact. The data collaborative predicted the drought and crop loss worth $3.6 million.[6] However, there is limited information on who owned and controlled the data. The role of the intermediary was important in providing information and tools to farmers and in representing their concerns at the legislative level to put forth their demands and requests.

Why data collaboratives matter

Data collaboratives provide for a number of benefits, including:
- Insights on or benefit to public and/or private sector goals in agriculture (e.g., channeling research and development, identification of risks and opportunities, signaling social responsibility, reciprocity, and revenue generation).
- Added value by creating data exchanges and coordination among data providers and users.
- More synergies by unlocking highly demanded standards and frameworks to make farm data interoperable and useful across organizations and sectors.
- The central hosting is usually done by one of the members or partners in a collaboration, acting as a steward or intermediary.
- Opportunities for farmer organizations to bargain on a more level playing field and advocate for fairer and more equitable treatment of farmer data.

Key components for success
- **Opening up siloed data.** Data collaboratives mobilize and incentivize stakeholders to effectively release data from silos in a purpose-driven way. A key challenge in the data economy is accessing often privately held datasets—possibly of significant value in solving public challenges.

---

6. Young, A. & Verhulst, S., (2017) Acclimate Colombia, Open Data to Improve Agricultural Resiliency Case Study

- **Enabling the secondary use of data.** The options for reuse can be discovered with greater ease by combining or analyzing different data sets. Data collaboratives offer this opportunity as stakeholders can pool multiple types of data from a range of sources.
- **Partnerships and trust.** Given the nature of data sharing, a space that enables trust and cooperation between partners is a crucial component. The fear of loss of control over data and competitive advantage often acts as a disincentive to data sharing, as was observed in the case study A Data Collaboration with Fertilizer Companies, Steps towards Farmer-Centricity. Research suggests that building relationships, consistent communication, and articulation of common goals are vital in building data collaboratives and fostering a conducive space.
- **Regulatory frameworks.** Provisions governing data collection and sharing determine the options available, for instance, looking at the use, nature, and location of data. In some jurisdictions, certain types of data, such as call detail records, are restricted. Some countries, like India and Indonesia, require the data collected in their country to be stored in their country.
- **Stakeholder engagement and participation.** Ensuring internal structures and processes that all parties are comfortable with is crucial to any collaboration.[7] For a collaborative to be successful, processes and rules must be decided upon with active engagement and participation from all data-sharing entities. The model can provide for a modified "delegated" arrangement to allow farmer organizations to negotiate the terms of use. However, there is limited information on the representative nature of these groups and the power imbalance they bring or uphold. Without adequate investment and genuine stewardship, there is danger that collaboratives can become "tokenistic" at best.
- **Multidisciplinary know-how.** Data collaboratives are in a unique position to generate value, and subsequent innovation, by employing a range of expertise, data, and perspectives. Cross-sectoral experience and technical or other forms of experience, along with inter-organizational data exchange, are crucial to solve complex societal problems.
- **Champions and sponsors.** Buy-in from leadership within parties is vital. It requires ongoing sponsorship from within those who have formal authority and can champion the initiative.
- **Data systems and standardization.** Sharing data across partners raises questions of data quality, standardization, and interoperability. Data collaboratives must define standards for data quality and formats for sharing and ensure compatibility of data sets. This requires interaction between data teams and can be time-consuming.
- **Assist institutions to build capacity.** Many organizations—especially public agencies, civil society organizations (CSOs), and farmer groups—lack technical infrastructure, capacity, and resources. Collaboratives can bring in expertise by setting up a data steward or intermediary that can manage, host, and govern data across stakeholders.

The challenges or pitfalls of data collaboratives

- **While data collaboratives allow organizations to pool data and resources, their focus is still on an organization as the primary data owner.** Our research suggests that these models are less useful in placing individual farmers at the center of data governance. They are better suited when farmers work as a collective. While there are instances of data collaboratives working to promote societal benefit, data-sharing norms and rules are discussed at the organization level, not individual level. **Most data collaboratives do not explicitly deal with farm data ownership and governance,** even initiatives that have engaged in extensive feedback to generate alignment on

7. Emerson, K., Nabatchi, T., & Balogh, S. (2012). Integrative framework for collaborative governance. Journal of public administration research and theory, 22(1), 1–29.

data sharing. Within this model, farmers have limited control in creating the rules of engagement with their shared data.

- **Data collaboratives require engagement and participation, which might be constrained by costs and farmers' data literacy.** An effective collaborative model requires active and informed participants to advocate for interests with other members. Such expectations might not be compatible with the existing data literacy level, especially of smallholder farmers. It is also imperative to form a common vocabulary for multi-stakeholder engagements, since it is observed that they will be ineffective due to lack of a frequent vocabulary.[8] Partners may have different agendas and values, and this may influence their attitude with regard to data sharing and access, making the process sometimes time-consuming.

- **There is limited experience of sustainable governance models for data collaboratives.** While they aren't new, common governance frameworks do not exist. The process of setting up these models usually need to be carried out from scratch and require significant effort and consensus. Most initiatives, especially in LMICs, are mediated by the government or donor organizations and are in an early stage of functioning. Lessons are therefore drawn from a limited base of knowledge.

Important considerations for implementing data collaboratives in LMICs

- **Data collaboratives can harness the power of farmer collectives in LMICs to advocate more equitable data governance mechanisms.** Collective mobilization for advocacy in policy, contract negotiation, and collaborative action are common among farmer networks and have been successful. They allow individual farmers to overcome scale and size problems, present greater negotiation power, and attract more funding. Farmer groups have been active participants in policy making in LMICs.[9] In their essence, data collaboratives align with the culture of farmer associations. However, there is limited experience and research done in translating farmer associations to data collaboratives.

- **Many data collaboratives are set to create a data infrastructure, improve access to data, and inform better policy and farm decisions.** Africa Soil Information Service (AfSIS), for example, is creating continent-wide digital soil maps via spatial-temporal analysis with combined datasets. AfSIS provides crucial market intelligence for governments, farmers, and emerging industry players. Data collaboratives with CSOs and private- and public-sector actors can help farmers and service providers harness the power of data insights to improve farm management.[10]

- **Governments can play an important role in building and mediating data collaboratives.** The CSACexample mentioned earlier highlights the central role governments in LMICs play in facilitating data partnerships for a clear cause, enabling access to data. Within WAGRI, the government hosts the data platform and brings stakeholders together to build trust.

- **Improved access to data, such as satellite data, can inform decisions.** An example is pooling climate data for better problem solving and development efforts. STAMP seeks to provide access to geospatial data to improve decision making for farmers in LMICs and provide reliable insights on availability of surface water. Intel's big-data initiative provides insights on precision farming, optimizes water use and irrigation, and forecasts droughts. The Africa Regional Data Cube layers

8. The Govlab. (2021). Designing Data Collaboratives to Better Understand Human Mobility and Migration in West Africa, Big Data for Migration Alliance. (BD4M).
9. IFAD, (2016) Asia-Pacific Farmers' Forum IFAD's Medium-term Cooperation Programme with Farmers' Organizations Phase Two (MTCP2).
10. See the VIFAA and MTIC Case Studies.

satellite imagery and earth observation data for Kenya, Senegal, Sierra Leone, Tanzania, and Ghana. It works with country partners to generate policy insights. In Senegal, this data has been used to help large pastoralist communities better deal with climate change and informed changes in their practices.

Financial viability and sustainability

Data collaboratives in agriculture are primarily funded by governments and donors. The funding of collaboratives is often provided for only a specific time, with limited avenues for extension. There are some initiatives with a revenue or subscription model, where members are charged to access the data and insights. WAGRI requires data users and data generators to make monthly payments of $465 and $280, respectively. Users may be required to pay charges depending on the nature of the data they want to access. Addressing resourcing and revenue sharing can determine the sustainability and participation in a data collaborative. It is unclear if these costs are enough to meet the operational cost of the collaborative.

How to create an enabling environment for data collaboratives

Role for the government
- Provide a clear regulatory framework that promotes trusted data sharing and clear roles and responsibilities to help overcome the trust and hesitation deficit. Clarity in data-sharing standards engenders trust among actors.
- Provide policy guidance on harnessing the collective power of farmer collaboratives to advocate for equitable data terms with service providers.
- Improve data literacy and engagement with farmers. If inclusive literacy programs do not receive sufficient investment, power asymmetries, attribution, and unequal economic benefit will continue. This could include data literacy modules in farm extension services, in courses and in farmer organizations.

Role for the private sector
- Define clear data standards when using farm data.
- Set up a formal data stewardship role or organization to sustainably manage the initiative.
- Engage in stakeholder consultations and continuous communication while discussing the terms of the data governance approach.
- Provide clarity on revenue models and potential economic benefits of data sharing.

Role for the development sector
- Promote research on the use of data collaboratives as a model of farmer-centric data governance.
- Where farmer collaboratives exist, provide technical support to improve their data understanding and operational management.
- Provide guidance on engagement of the private sector with farmer collaboratives.

Case study: **A Data Collaboration with Fertilizer Companies, Steps towards Farmer-Centricity.**

The goal is to underpin data supply and enhanced policies and investments to expand affordability, availability, and quality of data. The program works with partners to create dashboards and tools that help improve the understanding and use of fertilizer data in Kenya, Ghana, and Nigeria. VIFAA developed a consultative data-sharing approach that garners feedback from data generators and helps overcome the trust deficit. Issues of trust between stakeholders are largely solved, through the use of NDAs and extensive relationship building. The country governments provide unique identification codes, especially for farmers' data, in order to avoid price discrimination. VIFAA has elaborated on the elimination of unique identification codes (personalized information regarding the farmer) of farmers and has developed a protocol for them to engage in feedback. The program conducted multi-stakeholder meetings to ensure all providers are aligned with data sharing.

West Africa Cashew Project (CashewIN). The project aims to provide enhanced market outcomes for cashew farmers via a data management system with improved access and use of data and to enhance policy decisions, increase and reinforce capacities, productivity and efficiency of farmers, boost trade, and support supply chain linkages with agro-food companies. CashewIN recognizes gaps in the processes of data collection, storage, use, and dissemination in the cashew sector in Côte d'Ivoire, Benin, Burkina Faso, Ghana, and Nigeria. The platform is an application showing data (e.g., plots, associations, and processing areas) and indicators on international benchmarking. At present, an official method for aggregating cashew sector data is non-existent, even though multiple datasets are present in public and private sector organizations and institutions. Data is siloed, but to inform decision making, the data needs to be collected regularly, updated, aggregated, and presented in a way that gives value to the different users of the data. The project seeks to present this disparate data in a harmonized way. Data ownership has become an increasingly important issue since it is primarily owned by the implementing country, which places restrictions on use across borders. The data is connected through various access rights and profiles to smooth data collection, entry, and confidentiality policies at the respective country.