Co-producing biodiversity data infrastructures: Mexico's Agrobiodiversity Information System as a case study

Irene Ramos[1,2,*], Francisca Acevedo Gasman[3], Vicente Arriaga[4], Caroline Burgeff[3], María-José Ibarrola-Rivas[5], Mahelet Lozada-Aranda[3], Oswaldo Oliveros-Galindo[3], María Andrea Orjuela Restrepo[3], Alicia Mastretta-Yanes[2,*]

1 Posgrado en Ciencias de la Sostenibilidad, Universidad Nacional Autónoma de México, Mexico
2 Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom
3 Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Mexico City, Mexico
4 Comisión Nacional Forestal, Mexico
5 Instituto de Geografía, Universidad Nacional Autónoma de México, Mexico

* corresponding authors (Irene Ramos: i.ramos@kew.org, Alicia Mastretta-Yanes: a.mastrettayanes@kew.org)

Summary

- Agrobiodiversity data, rooted in place and in community, expose tensions around sustaining the involvement of data collectors across the data lifecycle and embedding their context-specific expertise as data are mobilized beyond their sites of origin. These challenges extend to biodiversity data infrastructures that integrate diverse socio-ecological information. We examine how a co-production approach to infrastructure design and development can sustain data collectors' participation beyond data generation, drawing on our experience developing the Agrobiodiversity Information System (SIAgroBD) in Mexico.

- We analyze three participatory strategies implemented during SIAgroBD's development: data literacy workshops, collaborative assessment of data collection tools, and co-design of a data management plan. Our analysis draws on workshop materials, facilitation notes, surveys, focus groups, and project documentation.
- Data literacy workshops helped collectors situate their work within the wider infrastructure, thus creating conditions for participation beyond the data generation stage. Participatory tool assessment surfaced context-specific adaptations and tacit knowledge, which collectors translated into shared documentation. Through scaffolded co-design, context-specific expertise from data collectors informed access and reuse agreements for sensitive data.
- Involving data collectors did not, on its own, make the development of SIAgroBD a co-production process. Meaningful participation required spaces for knowledge exchange and mechanisms through which their expertise could inform infrastructure development beyond the data collection stage, with different levels of involvement across the data lifecycle. These activities required substantial time, coordination and resources. Capacity building and feedback processes should therefore be recognized as integral components of infrastructure development.

Keywords

Agrobiodiversity, biodiversity data, data collection, data management, infrastructure, participation

Introduction

Data collectors remain the backbone of biodiversity data infrastructures. Their situated expertise, such as the purposes for which data were collected and collection protocols, is essential for data interpretation and evidence-based decisions (Pasquetto et al., 2019). For example, data collected during a botanical expedition for a site inventory (e.g. Davies et al., 2024) are different in scope and granularity than seed data collection for *ex-situ* conservation (e.g. Kallow, 2014). Across these and other contexts, data collectors may be researchers, field technicians, citizen-scientists, or local community members bringing context-specific expertise, from taxonomic knowledge to the judgement needed to adapt standard sampling protocols to

field conditions. However, large-scale data aggregation, cross domain integration, and automated data generation pipelines increasingly stretch the distance between collection and reuse (Borgman & Groth, 2025); the farther data travels for reanalysis and reuse, the more challenging it is to retain data collectors' context-specific knowledge. This reflection is timely, given the growing body of research integrating data into biodiversity syntheses for both research (Callaghan et al., 2023; Weigelt et al., 2020) and environmental policy (e.g. IPBES, 2024). It is particularly important for plants: while plant data comprise only 19% of the records available through the Global Biodiversity Information Facility (GBIF), a flagship platform for open biodiversity data, they account for roughly 44% of published research reusing GBIF data (Heberling et al., 2021).

In this paper, we draw on our experience developing the Agrobiodiversity Information System (SIAgroBD) in Mexico as a case study of how a **co-production** approach (see Table 1 **Glossary**) to the design and development of **biodiversity data infrastructures** can sustain the participation of data collectors across the **data lifecycle**, shortening the distance from the field to research and policy. This distance between data collection and reuse is particularly consequential for agrobiodiversity data, which encompass cultivation practices, local uses, agronomic traits, and other socio-ecological information that can be deeply place-based, and are often collected in collaboration with **local knowledge holders** and local actors, including farmers and community leaders). Agrobiodiversity data is also often mobilized beyond its site of origin for analysis, policy and decision making that must remain relevant to a local context. Maintaining this connection to a local context becomes challenging as infrastructures aggregate and mobilize data.

If the initial growth and development of biodiversity data infrastructures was largely driven by digitizing existing collections, it has now expanded to a distributed system of data production (Heberling et al., 2021; Soberón, 2022). Today, biodiversity data, and especially plant data, are generated and aggregated across scales, ranging from individual research projects to coordinated institutional efforts (Soberón, 2022), government monitoring programs, and participatory or citizen science projects led by local communities (e.g. Mason et al., 2025; Reyes-Galindo et al., 2023). National and regional systems have become intermediaries by coordinating collection, curation, and mobilization at sub-national scales and linking these efforts to global aggregation platforms (Güntsch et al., 2025). Despite this increasing complexity, data collectors remain a key pillar of biodiversity data production. For example, 95% of plant data records available through GBIF derive from human observations or preserved

specimens, with only a small fraction (0.05%) from machine observations like photographs, videos or satellite images (see this dashboard for a breakdown of plant data records by basis of record).

At the same time, global environmental challenges such as climate change, biodiversity loss, and food security, require the integration of biodiversity information with diverse types of socio-ecological data (Heberling et al., 2021; IPBES, 2024). Biodiversity data platforms are thus expected to facilitate interoperability, mobilization, and reuse, while also supporting the communities of practice that produce, use, and steward data (Güntsch et al., 2025); their functions have shifted from repositories to **knowledge infrastructures**. As we explore here, this in turn suggests a shift in the role of data collectors serving not just as providers but also playing a key role in shaping the infrastructures.

Designing data infrastructures that foster the continued participation of the communities that produce and reuse data is a socio-technical challenge involving redesigned data pipelines, training, and governance considerations (Heaton, 2022; Ribes & Finholt, 2009). Participatory approaches to the development of biodiversity data projects have been mostly connected to citizen science initiatives (e.g. Grattarola et al., 2025) and focused on usability improvements, adoption of specific tools, or the development of data standards (e.g. Wieczorek et al., 2012), rather than on the participatory design and governance of infrastructures across the data lifecycle. While user-center design is a standard practice, and an important foundation, it can reinforce a distinction between users and developers by treating participation as consultation, and positioning communities as end users rather than co-designers or co-stewards of the infrastructures. This distinction matters in biodiversity infrastructures, where users may be both data collectors and reusers whose practice shapes data quality and interpretability. For instance, researchers working in a herbarium may collect data in the field, recording a species occurrence and other relevant information alongside a specimen; later, as curators, they may revise the species identification of the record based on their taxonomic expertise and knowledge of the environment where it was collected, a decision that affects how the record can be interpreted and reused. Infrastructures operating at national or regional scales, our scale of interest, are also often built on existing systems, constraining what can be redesigned and how.

In contrast with user-centered design, co-production has not widely informed biodiversity infrastructure development, despite its prominence in sustainability science as a way of producing knowledge that is scientifically robust, socially relevant, and legitimate for decision

making (Norström et al., 2020). Co-production, and closely related approaches like co-design or participatory research, emphasize shared problem framing, recognize diverse forms of expertise, and support iterative negotiation of priorities. Because data infrastructure development is typically framed as technical delivery, rather than a knowledge-making process, projects rarely include dedicated roles for embedded research; responsibilities are assigned to software engineering, data management, or other technical positions (Ribes & Finholt, 2009), and infrastructure design decisions and implementation are less often documented as scholarly outputs. We draw on co-production as a framework for this work because digital infrastructure design is, in effect, a knowledge-making intervention that shapes how information travels across contexts (Heaton, 2022), and more accurately describes our experience in this case study.

What, then, does it mean to meaningfully engage user communities in the design and development of biodiversity data infrastructures? For us, this question emerged as we embarked in the development of the Agrobiodiversity Information System (SIAgroBD). As part of a small team leading SIAgroBD's development, we encountered questions about how to engage data collectors beyond the data generation stage, how to maintain their contextual expertise legible as data were mobilized for policy, and how to accommodate design interventions while building within an existing public digital system.

SIAgroBD is part of the Mexican National Biodiversity Information System (SNIB) hosted by the National Commission for the Knowledge and Use of Biodiversity (CONABIO), a governmental organization with 34 years of experience in biodiversity conservation and the sustainable management of natural resources in Mexico (CONABIO, 2018; CONABIO et al., 2023). SIAgroBD provides access to open data on native Mexican crops and crop wild relatives, including maize, beans, cacao, agave, squash, chilli, among other crops of high cultural, economic and ecological importance. Beyond serving as a repository, the system facilitates data mobilization to inform conservation and food security policies, and it provides data collection and visualization tools, data management support, and upstream integration with national and global biodiversity platforms. SIAgroBD serves a **community of practice** working on agricultural biodiversity and sustainable food systems, including researchers, academic institutions, civil society organizations, and governmental agencies.

SIAgroBD offers an interesting case study to explore what co-producing biodiversity data infrastructures can mean in practice, under real-world constraints. It brings into focus tensions between contextual knowledge and data mobilization: maintaining interpretability and local

relevance even as data travel far from their site of origin, as they are repurposed, and are aggregated across contexts. SIAgroBD also reflects the complexities of co-production during ongoing infrastructure development. In this work, we use SIAgroBD as a study case to describe, analyze, and reflect on adaptations towards more participatory approaches to the design of data infrastructures, focusing on capacity building, participatory tool assessment, and data governance. Our objective is to examine how biodiversity data collectors can be meaningfully engaged during data infrastructure development beyond the data generation stage. We connect these insights to the broader development of national biodiversity infrastructures, where demands for aggregation and interoperability sit in tension with the need to safeguard local knowledge and community stewardship.

Methods

Case study: Mexico's Agrobiodiversity Information System (SIAgroBD)

Our work focuses on the participation of plant data collectors in the development of SIAgroBD beyond the data generation stage. In this case study, "data collectors" refer to members of fourteen research groups funded by CONABIO to conduct field data collection from 2020 to 2022. These research groups were selected through an open call for proposals supporting *in situ* data collection on agrobiodiversity taxonomic groups underrepresented in national biodiversity databases. The call prioritized teams who demonstrated established ties to the local communities where they proposed to conduct fieldwork. Selected proposals came from universities and civil society organizations across Mexico and combined field data collection with research aimed at improving taxonomic or ethnobotanical knowledge of crops and crop wild relatives. Teams were not expected to have prior experience in data management.

Research teams included project leads and principal investigators, students, technical staff and early career researchers. They typically came from a background in botany, taxonomy, ethnobotany or agronomy, and had varying levels of experience with biodiversity databases. Data collectors collaborated with local communities, local knowledge holders, seed and data donors to collect agrobiodiversity occurrences (observations, specimens, and accessions) and related socio-ecological information. CONABIO provided data management support, infrastructure, and training.

Field data were captured by data collectors through standardized digital surveys implemented in CONABIO's instance of KoboToolbox, an open source platform for digital data collection (Kobo, 2026). Data collected included biological, geographic, agronomic, social, and cultural variables, enabling records to retain contextual information (Figure 1, *data collection stage*). Upon data submission to KoboToolbox, data curators at CONABIO reviewed the information and shared feedback on data quality issues (from typos to taxonomy identification) that was then addressed by collectors (Figure 1, *data curation stage*).

We examine three participatory strategies aimed at steering SIAgroBD's development toward engaging data collectors beyond the data generation stage: data literacy workshops, collaborative assessment of data collection tools, and co-design of data management plans. These strategies were implemented iteratively in response to the challenges described below, which emerged at different stages of SIAgroBD's development and implementation. We draw on workshop materials, facilitation notes, feedback surveys, focus groups, and project documentation to describe, analyze, and reflect on how we designed these strategies and what forms of participation they enabled. We present each strategy by first describing the underlying challenge, followed by our response and its implementation.

Data literacy workshops

Challenge: Staff responsible for curation and data management (Figure 1, *data curation stage*) primarily communicated at specific project delivery checkpoints with data collection team leads, but there was limited engagement with other team members also working in the field. An initial survey about data skills revealed uneven adoption of data collection tools and uneven data literacy across teams, for example, collectors had little experience working with open data and open licenses. The limited visibility and inconsistent expectations about workflows, curation, and data quality, all central to data infrastructure literacy (Gray et al., 2018), restricted their involvement in later stages of the data lifecycle, including as reusers.

Response: To establish a shared understanding of data concepts and workflows, we organized a series of workshops on data literacy skills with mentorship from OLS's Open Seeds programme (Bernaldo et al., 2025). We began with an internal workshop for 22 CONABIO staff members in June 2021, bringing together curators, engineers, SIAgroBD leadership, and agrobiodiversity domain experts, individuals with distinct roles who often only interacted through specific tasks delegated across their different areas (e.g. repository administrators received a

spreadsheet from the data curation team). Similar to data collectors, technical staff had limited visibility on the redesigned workflows and practices introduced to CONABIO for SIAgroBD's development, like KoboToolbox. The session aligned staff members around open data concepts and FAIR principles, and provided an overview of workflows from data collection to publication. Participants from this workshop later helped facilitate workshops with data collectors.

A month later, in July 2021, we invited 27 data collectors to an online workshop delivered in three two-hour virtual sessions. These sessions introduced open data and FAIR principles, CONABIO's data portals, agrobiodiversity data lifecycles, and ethical data use. We prioritized hands-on activities, small group discussions, and structured reflection on data infrastructures, deliberately excluding data analysis and computer coding tools to focus on building a shared vocabulary around data. Participant surveys before and after the workshops documented self-reported familiarity with open data and tools, requested feedback on the activities, and asked about training needs. Workshop materials are archived in Zenodo (Ramos, 2022).

Participatory tool assessment

Challenge: Although data collector team leads had co-designed the standardized data collection surveys, differences in how teams used KoboToolbox became visible at the data cleaning and review stage, delaying quality control processes. For example, instead of collecting data directly on the app, some teams continued to collect data on paper, and then digitized their information after fieldwork. The usual communication channels between data collectors and CONABIO staff, which mostly relied on project delivery checkpoints, did not provide enough insight into these differences. Requests for further KoboToolbox training after the data literacy workshops also pointed to the need for additional support and alternative communication channels.

Response: We drew on the data journeys framework to implement a participatory assessment of KoboToolbox. Data journeys map how data move across stages, from field collection to cleaning, standardization and revision, highlighting the social dimensions of data production by making visible the roles individuals play at each step (Bates et al., 2016; Leonelli & Tempini, 2020).

We first distributed an anonymous survey to 255 data collectors and active KoboToolbox users. The survey included multiple-choice questions on participant's roles, perceptions of the tool's effectiveness, and open-ended questions about its main advantages and challenges.

Responses were collected over four weeks in November 2021, with 51 responses. We tagged open-ended responses thematically, refined iteratively across two rounds, to identify recurring challenges, perceived advantages, and insights that could support tool adoption during ongoing data collection.

Building on the themes surfaced in the survey, we then organized a four-hour virtual workshop over two days in January 2022, involving 25 representatives from data collection teams. On the first day, participants co-developed KoboToolbox documentation, in the form of a Frequently Asked Questions document, emphasizing adaptations to different field contexts and integration of the app into their team workflows. On the second day, participants were assigned to focus groups of four or five members, and collaboratively mapped their data workflows using a data journey template documenting key stages of collection, review, and dataset management. Facilitators guided discussions to ensure that participants captured detailed narratives of their workflows, focusing on challenges, strategies for team coordination, and adaptations to their own needs. The completed workshop templates were compiled and compared to identify commonalities and differences among the teams' approaches. Workshop materials are archived in Zenodo (Ramos, 2025).

Co-design of Data Management Plan

Challenge: Standard open data policies and donor consent had guided data collection from the onset, but during curation and preparation for public release, we identified data access and reuse considerations specific to agrobiodiversity records that could not be resolved through existing institutional practices. Socio-ecological information linked to data donors and households required protections for personal information not usually applicable to biodiversity occurrence records.

Response: We used a scaffolded process to develop a data management plan focused on access and reuse. An internal review first established aspects of data management that were not open to modification, including legal regulations and institutional practices for managing sensitive information, such as data donor's personal information (e.g. name, age, indigenous identification, household location), as well as guidance for publishing records of taxa listed under IUCN Red List categories (Chapman, 2020). We then drafted an initial framework and invited asynchronous feedback from data collection team leads, specifically on risks not captured by official classifications, such as species or landraces considered locally threatened.

In more nuanced cases, we also held follow-up discussions to negotiate recommendations alongside technical requirements.

Results and Discussion

As we show in this section, a series of small interventions can be strategically articulated to expand the participation of data collectors, even during ongoing infrastructure development and within established systems. Three major takeaways emerged (Table 1): capacity building enabled co-production, participatory tool assessment surfaced adaptations across contexts, and contextual knowledge shaped data governance through scaffolded co-design.

Capacity building enabled co-production

Staff who attended the internal data literacy workshops developed a shared understanding of how SIAgroBD adapted workflows for agrobiodiversity data management, and how those adaptations required coordination across different areas at CONABIO (Table 1). As one participant summarized in the post-workshop survey, the workshop helped them see that there were *more minds [in the institution] working on improving information management [...] and [that] it is possible to build a team across areas that can address areas for improvement*". This reflection was important because updating workflows had historically been a bottleneck at the institutional level. By inviting participants to lead sessions in the follow up workshops with data collectors, we created a pathway for staff to better understand field data collection constraints and opened communication channels beyond project checkpoints and team leads.

With respect to workshops with data collection teams, post-workshop surveys demonstrated increased awareness of open data as a collective effort. One participant noted that "*it was very gratifying to see an interest in engaging directly with the people involved in collecting information for CONABIO. This way, data quality will improve a lot, and it will be more valuable and trustworthy for everyone*". This response suggests that by making the data lifecycle visible, workshops helped data collectors locate their own work within a wider scale of infrastructure, reframing their role not only as data providers, but as integral contributors shaping data quality and reuse.

Participants valued the workshops as spaces for collective reflection and peer learning. A key takeaway for one participant was "*meeting collaborators from other data collection teams [...]*

willing to share very frankly the challenges and concerns that have arisen in their projects [...] meeting in these spaces helps very much with shared reflection and collaboration." Situating training within reflective processes that help connect practices to the contexts in which data are produced and reused has been done in other participatory processes involving biodiversity data tools. For example, Grattarola and collaborators (2025) describe how training in using citizen science tools for local biodiversity monitoring sparked broader reflections on access to information, and advocacy for local conservation priorities. In SIAgroBD, data literacy skills workshops similarly encouraged reflection on collaboration and the roles of staff and data collectors across workflows; opened a space for communication among data collectors, and between data collectors and CONABIO staff; and surfaced the need for further KoboToolbox support, which we describe next. In this sense, capacity building helped create the conditions for participation further along the data lifecycle.

Participatory tool assessment surfaced adaptations across contexts

Survey responses provided an overview of the perceived advantages and challenges of KoboToolbox. Data collectors valued its offline functionality and immediate data backup, while common difficulties included the time required to complete the digital forms during fieldwork, and editing data after submission. The teams mapped how team configurations and workflows interacted with these challenges in the data journeys workshop. Small teams, for example, struggled to balance interviewing and data entry, and sometimes preferred to record interviews and take written notes before uploading information to KoboToolbox when returning from the field. Larger teams could assign specific responsibilities to different people, but emphasized coordination issues. The narrative format of the workshop was fruitful for peer learning, as one participant noted: "*I really liked that the activities welcomed anecdotes so we could freely share our experience using Kobo [...] I was surprised at how each team organized their work in different ways, I learned a lot from them!*" A recurring theme across discussions was that KoboToolbox redistributed data management responsibilities that had been more centralized within CONABIO's staff. For instance, more of the data cleaning was carried out by data collectors themselves, creating an opportunity to strengthen their data management skills through hands-on practice. However, we didn't determine whether this made workflows more or less efficient.

Collaboratively developing documentation provided a way to translate tacit knowledge into shared learnings. Participants contributed recommendations for working with unstable internet

connections, collaborating with data donors in the field, and even recognizing when KoboToolbox was not the most appropriate tool, for example, some teams noticed that an external GPS provided more reliable location data than the app on their mobile devices. Heaton (2022) also recognizes the key role of this "intermediate knowledge" in allowing data to travel: in a case study of herbarium digitization, volunteers exchanged practical knowledge to improve workflows beyond what was documented in formal protocols; they aggregated this knowledge in a manual for newcomers, allowing the knowledge to persist. This echoes how adaptations that may have remained tacit knowledge within individual teams became tool documentation in SIAgroBD.

Overall, what both staff and data collectors had flagged as a gap in technical training, was better understood as a need to share how teams integrated KoboToolbox into their own workflows. Participants described this "*[knowledge] exchange across projects*" as "*an ideal way to find solutions to shared problems*". Another participant reflected "*on the way we organized ourselves within our team, if it was the best way or where we could improve*". These comments illustrate how, beyond mechanisms to report problems, feedback processes became a space for peer learning and broadened opportunities for co-production.

Contextual knowledge shaped data governance through scaffolded co-design

Co-designing the Data Management Plan brought contextual knowledge into data governance. Some data access agreements reflected in the data management plan derived directly from legal and institutional requirements. For instance, personal and personally identifiable information was restricted to internal use, and coordinates associated with households (e.g. a chili plant growing in a backyard) were obscured, meaning the precision of the published coordinates was rounded to a degree proportional to the sensitivity of the data. Other decisions fell outside standard institutional practice: based on their knowledge of locally threatened species or landraces not officially classified as such, data collectors recommended also obscuring location data; as a result, coordinate resolution was adjusted for these records.

Restricting location data may have implications for reuse, including misleading conservation strategies (Contreras-Díaz et al., 2023), so these decisions involved non-trivial tradeoffs between data accessibility and protecting sensitive information. Decisions regarding agave records were particularly contested because concerns about local conservation intersected with

commercial interests. In this case, a representative with expertise in agave conservation and knowledge of the agave industry was invited to contribute to the discussion. Formalizing these diverse, sometimes conflicting, concerns around who and how could access and reuse the data created a mechanism through which contextual knowledge could remain embedded in the metadata even through subsequent aggregation.

This was CONABIO's first experience collaboratively designing a data management plan for agrobiodiversity data. The consultation process was bounded by legal, technical and institutional constraints, involved data collection team leads rather than every collector or data donor, and left final decision-making authority with CONABIO. Within these constraints, meaningful participation depended on scaffolding the co-design process. This meant making constraints explicit, identifying where contextual expertise was most relevant and could shape governance decisions, and negotiating between standard institutional practices and the diverse needs of contributors.

Conclusion

Meaningfully engaging data collectors in infrastructure development meant creating spaces for knowledge exchange and reflection on shared challenges, as well as mechanisms through which their contextual knowledge could persist beyond the initial data collection stage. The involvement of many participants across institutions, roles, and areas of expertise did not, on its own, make the development of SIAgroBD a co-production process. The form and degree of participation varied across the data lifecycle depending on the expertise required and institutional constraints, illustrating how co-production spans a spectrum of involvement. Future research could explore what meaningful participation in infrastructure development could look like for a broader diversity of actors who also bring valuable agrobiodiversity knowledge, such as farmers, data donors, and local knowledge holders.

Our experience responding to challenges that emerged during the development of a biodiversity information system led us to reconsider our role within the development team. Facilitating spaces through which knowledge could be exchanged became an important part of our role, and it required significant time and coordination. We therefore recommend treating the maintenance of capacity building and feedback processes as part of infrastructure development itself, and resourcing them accordingly.

One limitation of this case study is that the responses were developed as challenges emerged. This suggests an opportunity for infrastructure projects to establish mechanisms for learning, reflection, and feedback with data collectors as part of their initial design. Our responses were also not embedded in a longitudinal monitoring and evaluation framework. While participant feedback gives us evidence of post-workshop learnings, we cannot determine which practices persisted, whether they transferred to other contexts, or how they influenced data reuse. Evaluations of biodiversity data infrastructures could therefore examine, alongside technical aspects, the capacities and forms of stewardship that involving data collectors enables over time. By documenting these strategies and their limitations, we hope to encourage other infrastructure teams to share and reflect on their own experiences.

Agrobiodiversity data, rooted in place and in community, expose tensions around how to mobilize data without losing the contextual knowledge needed to interpret them, and how to sustain the involvement of data collectors as data move farther from their sites of production. These challenges extend to biodiversity data infrastructures more broadly as they increasingly integrate heterogeneous socio-ecological information. National-scale systems, in particular, mediate access and reuse between local data providers and upstream systems; it is at this scale that contextual knowledge may be diluted through aggregation, but also where decisions about standards and governance can help such knowledge remain legible.

In Latin America, where biodiversity data infrastructures are maturing (e.g. Canhos et al., 2022; De Vega et al., 2020), these discussions are timely. Regional environmental policy frameworks, such as the Escazú Agreement, also emphasize access to environmental information and the capacities needed to collect and reuse it (ECLAC, 2023). As SIAgroBD illustrates, these capacities include the ability of infrastructure developers and data collectors to make contextual knowledge legible across the data lifecycle, learn across institutional and disciplinary boundaries, and participate in decisions about data governance and infrastructure design. Approaching infrastructure development through a co-production framework may help ensure that large-scale and automated aggregation does not overlook the expertise of the data collectors on whom meaningful reuse depends.

Acknowledgements

We are grateful to the data collector teams who participated in and contributed to the Mexican Agrobiodiversity GEF Project (project IDs RG001, RG002, RG016, RG019, RG023, RG025,

RG034, RG055, RG057, RG070, RG078, SG002, SG004, SG007, SG012, and SG019). We also thank CONABIO's Coordinación de Agrobiodiversidad y Recursos Biológicos, and Subdirección de Inventarios Bióticos for supporting the work described in this paper. We are particularly grateful to Vivian Bass Vega, Verónica Suaste, and Francisco Ramírez for their work as software developers for SIAgroBD, and to Alejandro Ponce for his support with workshop facilitation and insightful feedback. IR is grateful to Mariana Benítez for her valuable mentorship. This work was supported by GEF Project ID 9380 CONABIO-GEF-FAO. IR also acknowledges financial support from Coordinación General de Estudios de Posgrado (UNAM) and from SECIHTI under grant number 4014134, as well as support from Posgrado en Ciencias de la Sostenibilidad, UNAM.

Author Contribution

I.R. and A.M.Y. designed the research and developed the methodology. I.R., M.A.O.R., O.O.G., M.L.A. and A.M.Y. designed and facilitated workshops and participatory activities, and validated outputs. F.A.G., C.B., M.I.R. and A.M.Y. supervised the project and provided mentorship. I.R., M.I.R. and A.M.Y. wrote the manuscript. All authors edited and approved the final manuscript. F.A.G., C.B., and V.A. secured funding for the project.

Data Availability Statement

Selected workshop materials that support our analysis are openly available in Zenodo at https://doi.org/10.5281/zenodo.6640643 and https://doi.org/10.5281/zenodo.17657255. Full survey responses are not publicly available due to privacy restrictions. No other datasets were generated for this work.

Conflict of Interest Statement

The authors declare no conflicts of interest.

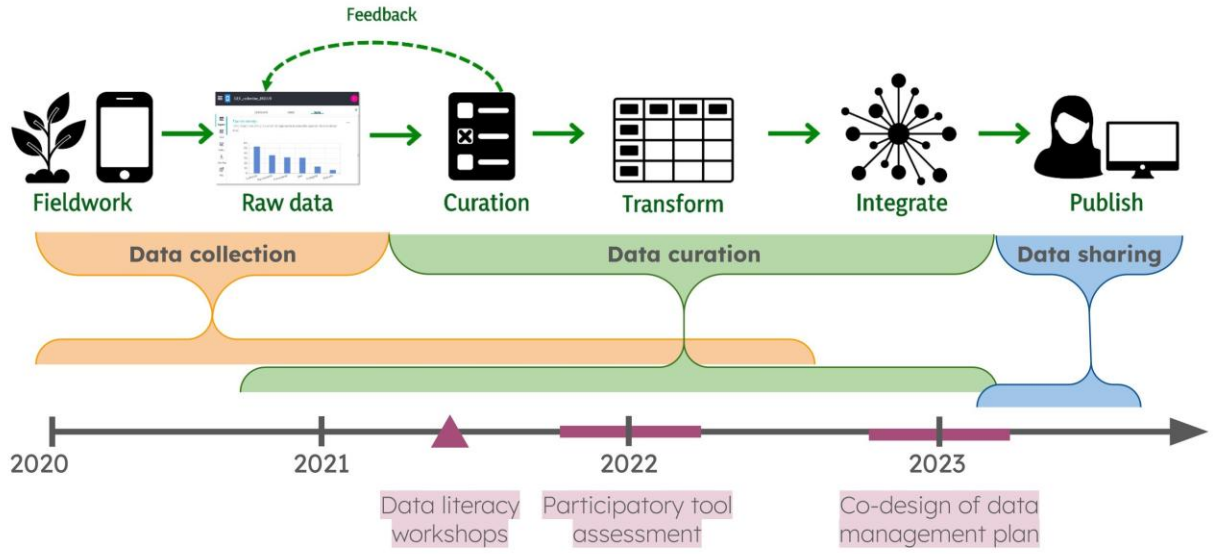


Figure 1. SIAgroBD's development timeline and data lifecycle. SIAgroBD's data lifecycle encompassed three stages: *data collection*, *data curation*, and *data sharing*; these stages overlapped throughout the development process of the system. During *fieldwork*, data collector groups collaborated with local donors to collect agrobiodiversity specimens, accessions, and observations, as well as associated socio-ecological data. Then, CONABIO's data curators checked data quality, including taxonomic validation, and gave feedback to collectors. When datasets were approved for publication, data managers transformed and integrated data in preparation for data sharing in SIAgroBD's data portal. Responses to challenges that emerged during the development process are highlighted in pink.

Table 1. Synthesis of challenges, responses and outcomes		
Challenge	**Response**	**Outcome**
Uneven data literacy limited engagement beyond data generation; fragmented awareness of the workflows connecting data collection to curation and publication.	Data literacy workshops with CONABIO staff and data collectors.	Capacity building created conditions for participation beyond the data generation stage, opening channels for communication, and helping collectors situate their work within the wider infrastructure.
Uneven adoption of digital data collection tools and limited knowledge exchange between data collection teams.	Participatory tool assessment through a survey, data-journey mapping, and collaborative documentation.	Participatory assessment surfaced tacit knowledge, enabling data collectors to compare adaptations and translate practical knowledge into shared documentation.
Standard institutional data governance practices only partially addressed access considerations specific to agrobiodiversity data.	Scaffolded co-design of a Data Management Plan within legal, technical, and institutional constraints.	Data collectors' contextual expertise informed decisions about sensitive information; access and reuse needs were negotiated within explicit boundaries, and embedded as metadata that could persist through data aggregation.

Table 2. Abridged glossary of key terms	
Term	**Definition**
Biodiversity data infrastructure	Biodiversity data infrastructures encompass databases, computational tools (e.g. online portals, APIs, dashboards), and communities of collectors, users, developers, and other stakeholders. These infrastructures span scales and functions, and at the national level, are better understood in terms of the services they offer as "providers, facilitators, mediators, and platforms for effective biodiversity data management, integration, and analysis" (Güntsch et al., 2025).
Community of practice	"Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly." (Wenger-Trayner et al., 2023)
Data lifecycle	"The flow of research data starting from planning, collecting, processing, analysing, preserving, sharing and finally reusing the data." (CODATA, 2026)
Knowledge co-production	"Iterative and collaborative processes involving diverse types of expertise, knowledge and actors to produce context-specific knowledge and pathways towards a sustainable future" (Norström et al., 2020).
Knowledge infrastructure	"Robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2013, p. 17).
Local data donor or	A person, often from a local community, who voluntarily provides

knowledge holders	knowledge, information, or data (e.g., observations, practices, local names, or uses) related to nature and culture to a research project, based on free, prior and informed consent.

References

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, *3*(2), 205395171665450. https://doi.org/10.1177/2053951716654502

Bernaldo, P., Pereyra Irujo, G., Yehudi, Y., van der Walt, A., Iley, B., Míguez, M. P., Ramos, I., Plomp, E., & Sharan, M. (2025). *Examining the impact and limitations of Open Science training: A case study of the "Open Seeds" programme*. Zenodo. https://doi.org/10.5281/ZENODO.15593896

Borgman, C. L., & Groth, P. (2025). From data creator to data reuser: Distance matters. *Harvard Data Science Review*, *7*(2). https://doi.org/10.1162/99608f92.35d32cfc

Callaghan, C. T., Borda-de-Água, L., Van Klink, R., Rozzi, R., & Pereira, H. M. (2023). Unveiling global species abundance distributions. *Nature Ecology & Evolution*, *7*(10), 1600–1609. https://doi.org/10.1038/s41559-023-02173-y

Canhos, D. A. L., Almeida, E. A. B., Assad, A. L., Cunha Bustamante, M. M. D., Canhos, V. P., Chapman, A. D., Giovanni, R. D., Imperatriz-Fonseca, V. L., Lohmann, L. G., Maia, L. C., Miller, J. T., Nelson, G., Peterson, A. T., Pirani, J. R., Souza, S. D., Stehmann, J. R., & Thiers, B. (2022). speciesLink: Rich data and novel tools for digital assessments of biodiversity. *Biota Neotropica*, *22*(spe), e20221394. https://doi.org/10.1590/1676-0611-bn-2022-1394

Chapman, A. (2020). *Current Best Practices for Generalizing Sensitive Species Occurrence Data*. GBIF Secretariat. https://doi.org/10.15468/DOC-5JP4-5G10

CODATA. (2026). *Research data lifecycle*. CODATA Research Data Management Terminology. https://terms.codata.org/rdmt/research-data-lifecycle

CONABIO. (2018). *CONABIO, 25 years of evolution*. CONABIO. https://bioteca.biodiversidad.gob.mx/janium-bin/janium_login_opac.pl?find&ficha_no=15149

CONABIO, FAO, & GEF. (2023). *Sistema de Información sobre Agrobiodiversidad (SIAgroBD)*. https://siagro.conabio.gob.mx

Contreras-Díaz, R. G., Nori, J., Chiappa-Carrara, X., Peterson, A. T., Soberón, J., & Osorio-Olvera, L. (2023). Well-intentioned initiatives hinder understanding biodiversity conservation: Cloaked iNaturalist information for threatened species. *Biological Conservation*, *282*, 110042. https://doi.org/10.1016/j.biocon.2023.110042

Davies, N. M. J., Drinkell, C., & Utteridge, T. M. A. (2024). *The Herbarium Handbook* (1st ed). Royal Botanic Gardens, Kew.

De Vega, J. J., Davey, R. P., Duitama, J., Escobar, D., Cristancho-Ardila, M. A., Etherington, G. J., Minotto, A., Arenas-Suarez, N. E., Pineda-Cardenas, J. D., Correa-Alvarez, J., Camargo Rodriguez, A. V., Haerty, W., Mallarino-Robayo, J. P., Barreto-Hernandez, E., Muñoz-Torres, M., Fernandez-Fuentes, N., Di Palma, F., & the Colombian Cyberinfrastructure Consortium for Biodiversity. (2020). Colombia's cyberinfrastructure for biodiversity: Building data infrastructure in emerging countries to foster socioeconomic growth. *Plants, People, Planet*, *2*(3), 229–236. https://doi.org/10.1002/ppp3.10086

ECLAC. (2023). *Regional Agreement on Access to Information, Public Participation and Justice in Environmental Matters in Latin America and the Caribbean*. https://hdl.handle.net/11362/43583

Edwards, P. N. (2013). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (1st ed.). The MIT Press.

Grattarola, F., Bergós, L., Carabio, M., Montiel, R., & González, S. (2025). NaturalistaUY in Uruguay: A case of community science in Latin America from a critical perspective. *Desenvolvimento e Meio Ambiente*, *66*, 142–161. https://doi.org/10.5380/dma.v66i.95896

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 205395171878631. https://doi.org/10.1177/2053951718786316

Güntsch, A., Overmann, J., Ebert, B., Bonn, A., Le Bras, Y., Engel, T., Hovstad, K. A., Lange Canhos, D. A., Newman, P., Van Ommen Kloeke, E., Ratcliffe, S., Le Roux, M., Smith, V. S., Triebel, D., Fichtmueller, D., & Luther, K. (2025). National biodiversity data infrastructures: Ten essential functions for science, policy, and practice. *BioScience*, *75*(2), 139–151. https://doi.org/10.1093/biosci/biae109

Heaton, L. (2022). Chains of participation in producing biodiversity infrastructures: Digital reconfigurations of scientific work. *Science as Culture*, *33*(3), 297–320. https://doi.org/10.1080/09505431.2022.2025774

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, *118*(6), e2018093118. https://doi.org/10.1073/pnas.2018093118

IPBES. (2024). *Summary for Policymakers of the Thematic Assessment Report on the Interlinkages among Biodiversity, Water, Food and Health of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (P. D. McElwee, P. A. Harrison, T. L. van Huysen, V. Alonso Roldán, E. Barrios, P. Dasgupta, F. DeClerck, Z. Harmáčková, D. T. S. Hayman, M. Herrero, R. Kumar, D. Ley, D. Mangalagiu, R. A. McFarlane, C. Paukert, W. A. Pengue, P. R. Prist, T. H. Ricketts, M. D. A. Rounsevell,

… D. Obura, Eds.). IPBES secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.13850289

Kallow, S. (2014). *UK National Tree Seed Project Seed Collecting Manual*. Royal Botanic Gardens Kew. https://doi.org/10.34885/1

Kobo. (2026). *KoboToolbox*. https://www.kobotoolbox.org/

Leonelli, S., & Tempini, N. (Eds.). (2020). *Data Journeys in the Sciences*. Springer Open. https://doi.org/10.1007/978-3-030-37177-7

Mason, B. M., Mesaglio, T., Barratt Heitmann, J., Chandler, M., Chowdhury, S., Gorta, S. B. Z., Grattarola, F., Groom, Q., Hitchcock, C., Hoskins, L., Lowe, S. K., Marquis, M., Pernat, N., Shirey, V., Baasanmunkh, S., & Callaghan, C. T. (2025). iNaturalist accelerates biodiversity research. *BioScience*, biaf104. https://doi.org/10.1093/biosci/biaf104

Norström, A. V., Cvitanovic, C., Löf, M. F., West, S., Wyborn, C., Balvanera, P., Bednarek, A. T., Bennett, E. M., Biggs, R., De Bremond, A., Campbell, B. M., Canadell, J. G., Carpenter, S. R., Folke, C., Fulton, E. A., Gaffney, O., Gelcich, S., Jouffray, J.-B., Leach, M., … Österblom, H. (2020). Principles for knowledge co-production in sustainability research. *Nature Sustainability*, *3*(3), 182–190. https://doi.org/10.1038/s41893-019-0448-2

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and reuses of scientific data: The data creators' advantage. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.fc14bf2d

Ramos, I. (2022). *Taller Datos abiertos de agrobiodiversidad*. Zenodo. https://doi.org/10.5281/ZENODO.6640643

Ramos, I. (with Mastretta-Yanes, A., Lozada, M., Oliveros, O., Ponce, A., & Ramírez, F.). (2025). *Taller Intercambio de experiencias sobre el uso de Kobo Conabio*. Zenodo. https://doi.org/10.5281/zenodo.17657254

Reyes-Galindo, V., Jaramillo-Correa, J. P., Carrasco Nava, K., De-la-Rosa-González, A. E., Flores Flores, D., Martínez, M., Monroy-De-la-Rosa, L. A., Morelos Zamora, M. Á., Ramírez Morales, B. E., Ramírez Morales, O. T., Rodríguez, M. D. P., Salazar Zamora, M., Zamora Callejas, C., Zamora Callejas, R., Zamora, C., Zamora, T., González-Camacho, V. A., Rebollo, E., Torres-Jardón, R., … Mastretta-Yanes, A. (2023). Evaluating pollution-related damage and restoration success in urban forests with participatory monitoring and digital tools. *Conservation Biology*, *37*(5), e14112. https://doi.org/10.1111/cobi.14112

Ribes, D., & Finholt, T. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, *10*(5). https://doi.org/10.17705/1jais.00199

Soberón, J. (2022). Biodiversity informatics for public policy. The case of CONABIO in Mexico. *Biodiversity Informatics*, *17*. https://doi.org/10.17161/bi.v17i.18270

Weigelt, P., König, C., & Kreft, H. (2020). GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography. *Journal of Biogeography*, *47*(1), 16–43. https://doi.org/10.1111/jbi.13623

Wenger-Trayner, E., Wenger-Trayner, B., Reid, P., & Bruderlein, C. (2023). *Communities of practice within and across organizations. A guidebook.* (2nd ed.). Social Learning Lab.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE*, *7*(1), e29715. https://doi.org/10.1371/journal.pone.0029715