

Governing Agricultural Data: Challenges and Recommendations



Medha Devare, Elizabeth Arnaud, Erick Antezana, and Brian King

Abstract The biomedical domain has shown that *in silico* analyses over vast data pools enhances the speed and scale of scientific innovation. This can hold true in agricultural research and guide similar multi-stakeholder action in service of global food security as well (Streich et al. *Curr Opin Biotechnol* 61:217–225. Retrieved from <https://doi.org/10.1016/j.copbio.2020.01.010>, 2020). However, entrenched research culture and data and standards governance issues to enable data interoperability and ease of reuse continue to be roadblocks in the agricultural research for development sector. Effective operationalization of the FAIR Data Principles towards Findable, Accessible, Interoperable, and Reusable data requires that agricultural researchers accept that their responsibilities in a digital age include the stewardship of data assets to assure long-term preservation, access and reuse. The development and adoption of common agricultural data standards are key to assuring good stewardship, but face several challenges, including limited awareness about standards compliance; lagging data science capacity; emphasis on data collection rather than reuse; and limited fund allocation for data and standards management. Community-based hurdles around the development and governance of standards and fostering their adoption also abound. This chapter discusses challenges and possible solutions to making FAIR agricultural data assets the norm rather than the exception to catalyze a much-needed revolution towards “translational agriculture”.

M. Devare (✉)

International Food Policy Research Institute (IFPRI), Washington, DC, USA

e-mail: m.devare@cgiar.org

E. Arnaud · B. King

Alliance of Bioversity International and CIAT, Rome, Italy

e-mail: e.arnaud@cgiar.org; b.king@cgiar.org

E. Antezana

Department of Biology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Bayer CropScience SA-NV, Diegem, Belgium

e-mail: erick.antezana@gmail.com

© The Author(s) 2023

H. F. Williamson, S. Leonelli (eds.), *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*,

https://doi.org/10.1007/978-3-031-13276-6_11

1 Background

The COVID-19 pandemic has caused great human loss and economic suffering worldwide, but it may prove to be a ground-breaking model for agile collaborative science. This is exemplified by rapid and powerful approaches to data sharing, including the COVID-19 Open Research Dataset (Semantic Scholar, 2020) and mature biomedical ontologies (Bodenreider, 2005; Robinson & Haendel, 2020). The COVID-19 Open Research Dataset by the Allen Institute for Artificial Intelligence in collaboration with several research institutes gave researchers free and open tools and data to develop new insights about the novel coronavirus. The shared standards and data coupled with the collaborative application of massive computing power enabled research efforts worldwide to model and identify over 70 promising compounds for treatment in just under 2 days—a result that would otherwise have likely taken years (Quitau, 2020). This is a shining example of *in silico* analyses over vast data pools enhancing the speed and scale of scientific innovation that may also be applied towards agricultural research and guide similar multi-stakeholder action in service of global food security (Streich et al., 2020).

Responding agilely and hyper-locally to challenges in the agricultural sector necessitates building on prior research. While much of the conversation in the agricultural research for development sector focuses on the need to appropriately scale promising solutions, these solutions must also be agile in responding to changing local conditions, be they weather, markets, or others. This, in turn, requires decision support tools that mine problem-relevant open pools of data, and data products that not only meet the Findability and Accessibility (“Open Access”) criteria of the FAIR Data Principles but are also interpretable and reusable by humans as well as machines (Thessen & Patterson, 2011; Wilkinson et al., 2016). The biomedical sector began coalescing around the need for open, interoperable and machine-readable data by the mid-1980s to early 1990s with the creation of powerful open databases, standards and toolkits under the aegis of the National Center for Biotechnology Information (Smith, 2013). NCBI paved the way for rapid data-driven, transparent development of therapies and medical innovation.

In comparison, the agricultural sector has lagged in making data assets open and interoperable, with the possible exception of precision agriculture and work involving genetic and “omics”, and technologies such as those related to developing plant germplasm or insect pest detection. Agriculture has moved in this direction only in the last few years (Smalley, 2018) partly because data assets still too often exist on individual laptops. Even when data is accessible on public repositories, it has traditionally been summary tables or metadata, rather than the raw and well-described data needed for analyses and further innovation. Further, where such data has gradually become available over the last 5–7 years, it tends to be opaquely annotated – if at all – and not interoperable or easily reusable as data variables are not described using standards, but typically by individual choice. Private sector has been

increasingly amassing and mining location-specific agricultural data since the early to mid-2000s through Internet of Things (IoT), Big Data, AI, Blockchain and allied technologies in the service of precision agriculture and smart farming solutions (Rijmenam, 2013; Noyes, 2014; Pham & Stack, 2018). However, much of this data remains proprietary, and responsive only to – at best – company-specific standards and bespoke tools, making governance (including ownership) and linking of relevant but disparate data difficult (Rosenbaum, 2010).

It is only recently that agricultural public sector entities and researchers – and more importantly, their funders – are beginning to acknowledge the importance of data standards, and to specify open licenses and FAIR requirements (European Commission Expert Group on FAIR Data, 2018; Bill and Melinda Gates Foundation, 2021). CGIAR (<https://www.cgiar.org/>), the world’s largest global agricultural innovation network launched the Gates Foundation-supported Open Access, Open Data Initiative in 2015 to facilitate culture change and technological support for open research outputs across the 15 globally-dispersed CGIAR agricultural research for development centers. The initiative built on the ratification of CGIAR’s Open Access and Data Management Policy (CGIAR, 2013), and the momentum of this effort continued with greater emphasis on FAIR data through the Platform for Big Data in Agriculture (<https://bigdata.cgiar.org/>) which began in 2017. The Platform’s work has resulted in a number of open tools and services, a revised Open and FAIR Data Assets Policy (CGIAR, 2021), and capacity enhancement to support FAIR research outputs.

There are several ongoing efforts to build knowledge bases and open data portals, including by CGIAR (the [GARDIAN data ecosystem](#)), the European Union ([European Data Portal](#)), the United States Department of Agriculture ([Ag Data Commons](#)), and similar databases of compilations maintained by a number of research, academic, and funding entities in the agricultural space. These three exemplars explicitly pursue FAIRness, through alignment with established metadata schemas and semantic standards such as controlled vocabularies and ontologies to describe data variables. Such approaches enable mining and linking of data (e.g., as Linked Open Data), but adherence to standards remains challenging and is still elusive for a variety of reasons.

With the exception of bioinformaticians in fields like crop breeding or germplasm diversity studies, researchers encounter several hurdles to the adoption of data standards, and these are particularly entrenched in “non-digital natives”. The challenges include limited awareness on how to mine and derive value from standards-compliant, interoperable data pools; limited data science capacity for *in silico* analyses, with a related emphasis on the collection, rather than reuse of existing data; and limited fund and time allocation towards data management and the collaborative development of standards. Other, more community-based issues relate to the collective development and governance of standards, and to coalescing “critical mass” around consistent adoption. Thus, while FAIR data assets are foundational needs for an evidence-driven, agile, and collaborative approach to

enhancing the impact of research and development in the agricultural domain, the discipline is in its infancy in realizing the potential of consistent application of the FAIR Principles. Throughout an institution or set of entities in a disciplinary domain, the consistent adoption of the FAIR Principles and associated data standards and approaches relies on good governance (Koers et al., 2020).

But what exactly does “good governance” mean? It may be useful to first frame how we view this idea in the context of data, in line with Stedman and Vaughan’s recent writing (2020), that defines data governance to be a cross-cutting concern to assure success across the data life cycle. Thus, the availability, usability, security, and trustworthiness of data are all dependent on its governance, which also includes development and oversight of data standards, policies, and compliance with these. This paper discusses governance challenges and possible solutions to enabling interoperability of agricultural data assets as a critical requirement in catalyzing a move from prescriptive, “one size fits all” recommendations, to more site-specific options that are agilely developed in response to local constraints and scenarios.

2 Challenges and Solutions

Effective operationalization of the FAIR Principles towards agricultural research data assets that support easier interpretation and linking requires as a foundational paradigm that researchers accept that their responsibilities do not end with data collection and manuscript publishing, as was the norm in a pre-digital age. As stated by Wilkinson et al. (2016), data-intensive science increasingly means “. . . assisting both humans and their computational agents in the discovery of, access to, and integration and analysis of task-appropriate scientific data and other scholarly digital objects.” The reach of research therefore extends not just to data collection for personal analysis and publishing, but to stewarding or resourcing the stewardship of data assets to ensure long-term preservation, wide access and reuse. The development and adoption of common standards embodied by metadata schemas, ontologies and controlled vocabularies are critical to good data stewardship and reuse but developing and maintaining these efforts in agricultural research has been difficult. Despite data sharing and reuse being more accepted in other domains including the environmental and biomedical, the consistent use of standards is spotty even in these domains. For instance, in a survey of 100 ecological and evolutionary research datasets over half of the databases had issues including missing metadata; 64% were archived in a way that rendered reuse partially or entirely impossible due to poor or missing metadata, and/or non-machine-readable formatting (Roche et al., 2015).

2.1 *Research Culture and How Researchers Understand Scientific Inquiry*

The more traditional view of science is a prediction-based, hypothesis-driven approach as articulated by Karl Popper in 1963 (Brockman, 2015). Although this view is no longer central to some scientific domains, it remains quite relevant in agriculture. A consequence is that data is considered as more a by-product than a driver of research, and its governance, defined by Leonelli (2019) to be “...the strategies and tools employed to identify, manage, and disseminate data...” is typically not sufficiently valued or resourced. Leonelli challenges the traditional view of data as fixed and context-independent, and the notions of data quality and reliability as universal rather than influenced by context and purpose. The author’s relational view of data (Fig. 1) argues instead that the presentation, selection, and use of data based on purpose and context is critical to knowledge creation. Thus, this relational view posits that data are often altered through production, dissemination,

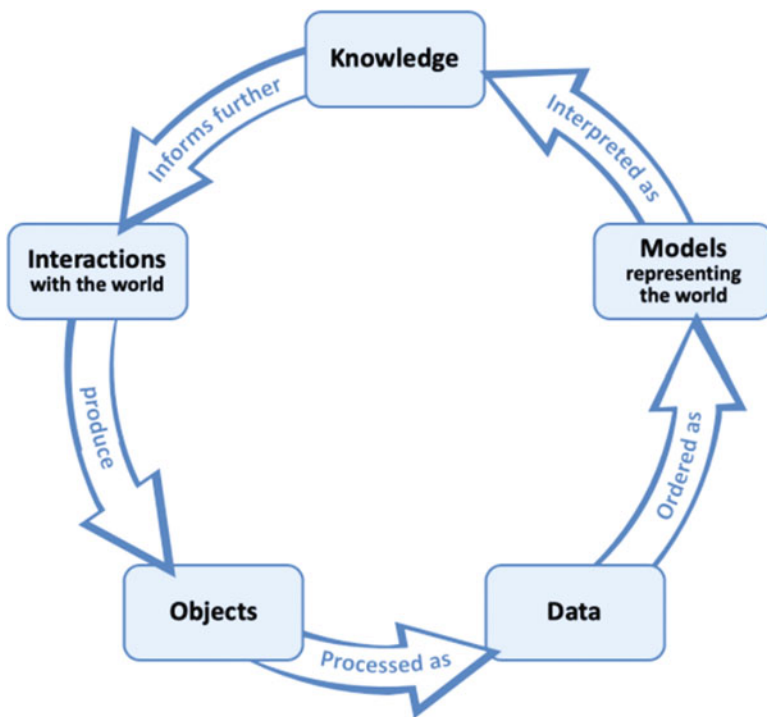


Fig. 1 Scientific inquiry according to the relational view of data (Leonelli, 2019). This mutually reinforcing view includes interactions of scientific subjects with the world, which produce objects that are documented as data. These data are managed and visualized to produce models that represent particular phenomena, leading to the creation of knowledge that can in turn inform future inquiry. (Reproduced without modification from Leonelli (2019), under CC-BY 4.0 licence)

and reuse for different purposes, imbuing their handling and management with more importance. Such a relational view is very relevant to the modern reality of digital technologies and capabilities, and particularly true for agricultural research – which necessitates context-based re-purposing of data.

Agricultural research culture is also influenced by the fact that it is traditionally field-based, involving time-consuming data gathering from experiments that typically run over several seasons/years and are generally conducted along the lines of Popper’s falsification-based hypothesis testing. Among the few exceptions to this, though relatively recent, are climate science, precision technologies, and disciplines like genomics in, say, germplasm development. Until recently, agricultural research rewarded those with strong field know-how and ability to employ the Popperian method over more quantitative or digital smarts, resulting in a culture of “my research, my data”. Our experience at CGIAR suggests that except for a few (e.g., geneticists, bioinformaticians, and the rare agronomist), the notion and use of *in silico* analysis involving secondary data is relatively new for agricultural scientists. In keeping with this, Denk (2017) suggests that researchers’ reluctance to use open data hinges on one or more of the following reasons: Insufficient knowledge to mine data effectively, a lack of awareness about the capabilities and power of big data analytics, and concern about data quality and reliability. Data is therefore seen as peripheral to research, and the notion of “data-centrism” espoused by Leonelli (2019) and other philosophers of science is the exception rather than the rule in agricultural research. Data governance, particularly around open and FAIR research data with its goal of widening access, mining, and reuse therefore remains relatively unimportant in the domain, with direct implications for the development and maintenance of widely accepted standards. Ongoing efforts towards data governance and linking through the International Treaty on Plant Genetic Resources for Food and Agriculture (PGRFA) represent an exception, as described in this volume by Manzella et al. Appropriate responses in the agricultural domain require that we acknowledge and address these challenges, learning from efforts such as the PGRFA.

Solutions to data governance issues are manifold and involve many actors and approaches. Some are highlighted here, based on experience across the CGIAR system:

- Data science is an active part of many life sciences areas but has come to the agricultural domain relatively late. Machine learning and big data analytics approaches that depend on FAIR agricultural data must be fostered through capacity building and continued institutional support and hiring/retention practices that make clear the link between standards-compliant data pools and the ability to derive insights from them. Fields such as bioinformatics that have been successfully deployed and accepted in key agricultural disciplines may be a model to follow, and indeed, the notion of “ag informatics” now exists.
- The adoption of best practices throughout the data life cycle including the use of standards that enable data aggregation should be an expected part of agricultural research, with high value assigned to contributions toward strong data outcomes.

Clarity around open and FAIR, and associated data schemas and standards must be part of contractual language for new hires. KPIs that explicitly acknowledge FAIR data and data-driven science and innovation should form part of researcher annual evaluations. As efforts around standards development, maintenance, and use require funding, allocation of budget towards best practices in data management that includes these aspects should be required, not recommended practice (10–25% is suggested by many project funders, including the EU). Together with these, data stewards must be valued and empowered for success.

- Data sharing requirements that specify repository, data, and allied standards must be implemented, ideally via data sharing templates and checklists that facilitate consistency across research units and institutions – and their partners – easing governance considerations. Addressing ownership issues by democratizing data authorship and upload to standards-compliant repositories is likely to be a foundational aspect of buy-in to these.
- Robust institutional data policy and strategy frameworks are crucial to prioritizing open and FAIR data, and formalizing many of the above points, yet several academic and research entities in agriculture lack these, thereby missing the opportunity to effectively prioritize and leverage a strengthened open and FAIR data culture. A case in point is the 76 Land Grant Universities (LGUs) in the United States, set up in 1862 to focus on curricula in practical agriculture, life sciences, and other disciplines. Most of the LGUs have no explicit policy governing open data sharing, with recommendations urging exploration of the relative advantages of selective commercialization vs. fully open access approaches to advance science and support for sustained investment in research and development (Barham et al., 2017). Uncertain or missing policy/strategy means that researchers are not held to expectations relating to data stewardship. It makes governance related to linking across multi-disciplinary agricultural data challenging within any institution, let alone across the LGUs and beyond. Where data policies do exist, few explicitly require the consistent use of data standards.
- Research funders and publishers play a key role in changing institutional data culture towards openness and FAIRness. Funders who require open and FAIR data to be shared in specified time frames along with publications, and who hold grantees accountable for this are crucial catalysts of culture change regarding data sharing and reuse. Although data journals are cropping up rapidly and the sharing of data underlying publications is increasingly expected by scientific publishers, this is still not the norm even in the biomedical realm. A study by Vasilevsky et al. (2017) indicated that just under 40 of 318 biomedical journals explicitly required data sharing as a prerequisite for publication.

It is important to note that a key reinforcer of open and FAIR data sharing is the “re-examination” of data, either for quality and/or reuse in new analyses. Without such benefits, the carrots and sticks outlined above may only result in partial success. This idea and several of those above are summarized in Fig. 2, from a 2020 manuscript by Sielemann et al.

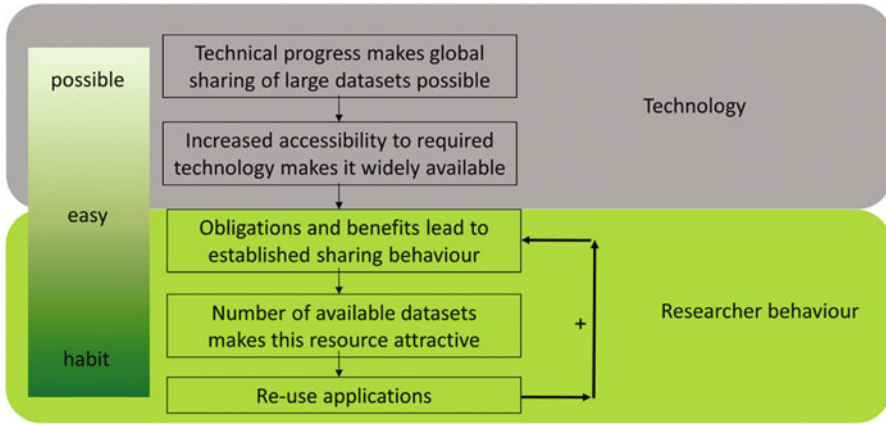


Fig. 2 The evolution of data sharing behavior. (Reproduced without modification from Sielemann et al. (2020), under CC-BY 4.0 licence. <https://doi.org/10.7717/peerj.9954/fig-1>)

2.2 Governance Issues and Repercussions Around Data and Data Standards

Technical challenges to governance towards greater openness and interoperability of data (which standards confer) are generally easier to address than those that are cultural, or subject to the legal frameworks of countries or rights of stakeholders (Sara & Devare, 2020). The latter may include intellectual property rights, confidentiality and/or privacy, farmers' rights, sensitivity (e.g., sensitive information relating to, say, harvesting forest species), farmers' rights and privacy (see also Leonelli and Williamson, this volume; Zampati, this volume).

Research data scenarios most likely to require robust governance frameworks include those that:

- Concern vulnerable peoples (including indigenous communities);
- Contain personally identifiable information that could be used to identify individuals or communities;
- Include anonymized data in which re-identification could result in significant harm;
- Concern genetic resources (including Digital Sequence Information) and any associated traditional knowledge;
- Include sensitive political data (including weather or health-related data, which in some countries is subject to formal or informal reporting restrictions).

Governance arrangements in the above scenarios require due diligence in how the data is described and managed, acknowledging and addressing restrictions that may arise due to the need for:

Prior Informed Consent Human subject data is typically subject to ethical standards requiring approval of an oversight body (such as an Internal Review Board), and prior informed consent from research participants which is purpose-specific. Prior informed consent also features prominently in the context of restricted use of data, privacy protection, and ABS compliance (see below).

Restricted Use of Data Including Commercialization and/or Commercial Use of Data Use of data in a manner inconsistent with the informed consent or contractual obligations under which it was obtained can have legal as well as reputational repercussions. Accordingly, it must be proactively handled subject to appropriate data protection measures.

Proprietary, Commercially Sensitive and or Confidential Data Public disclosure of data that is proprietary, commercially sensitive or confidential in nature can have legal as well as reputational repercussions, and must be subject to robust data protection measures.

IP and Contractual Rights Over the Data and Results or Innovations Generated Using the Data Access and use of data may be subject to intellectual property and contractual rights governing the use of data as well as derivatives of the data (e.g., CC-BY-SA and other licenses requiring share-alike terms) and downstream products developed using the data.

Privacy Protection and Human Subject Rights Personal data (i.e., directly identifying data) or data that could potentially be used to identify an individual (i.e., indirectly identifying data such as GPS coordinates on their own or in combination with other data) can be subject to requirements complicated by a fragmented regulatory landscape governing data protection, privacy and the rights of data subjects (e.g., the EU's 2018 General Data Protection Directive).

Access and Benefit Sharing (ABS) Compliance Accessing biological resources and associated information (such as genomic information and traditional knowledge) can be subject to best-practice or regulatory requirements concerning prior informed consent and mutually agreed terms governing access to, and the sharing of benefits (monetary and non-monetary) resulting from, research and development concerning the biological resources or associated information (e.g., as addressed by the Nagoya Protocol on Access and Benefit Sharing).

Agricultural Data Codes of Conduct These are tools to facilitate better data governance frameworks, particularly as agricultural data are increasingly collected through digital sensors often embedding Artificial Intelligence-based analytics. Few countries have a code of conduct for farm data; an exception is the European Union Code of Conduct for Data Sharing by Contractual Agreement developed in 2018 (Wiseman et al., 2019) Such codes encompass many of the above points, in attempting to provide principles about rights and responsibilities supporting a transparent data governance that engages farmers in decision-making, guaranteeing their full access to data collected from them. To address the lack of global guidelines, GODAN has recently published a generic toolkit (<https://www.godan.info/codes>) to

guide scientists or collectors of agricultural data to create a customized code that can be validated with the national authorities of the countries where data will be collected (Zampati, this volume). Implementation of such a code of conduct by public and private institutions may help plug the data governance gap particularly visible in the public sector.

As a case in point, agricultural research for development institutions in the CGIAR System have been attempting to tackle governance needs relating to the above while addressing cultural aspects in an *ad-hoc* way. With a new, more centralized CGIAR model envisioned, it is expected that governance frameworks will also be more uniformly applicable and backed by accountability. A nascent model is proposed for this new modality by a Data Assets Management Task Team operationalized in 2021 to address concerns around research data, with governance key among these (Fig. 3).

This data asset governance model recognizes that good governance goes beyond technical solutions, depending also on the ability of appropriately organized and empowered bodies with clear roles and responsibilities to create and assure a culture of best data practices, and compliance with legal structures. The proposed structure is briefly described here, and envisages 3 primary cascading areas of intervention, at the: (1) **strategic**, pan-CGIAR research portfolio plane; (2) **tactical**, research initiative level (with several initiatives forming the portfolio); and (3) **operational**, research team level within initiatives.

In this scenario, a strategic level Data Governance Committee (DGC) includes data scientists, domain experts (researchers), IT and legal personnel, and data stewards (with data asset management and standards expertise) and provides oversight across all three levels while primarily interfacing with tactical level teams. It is

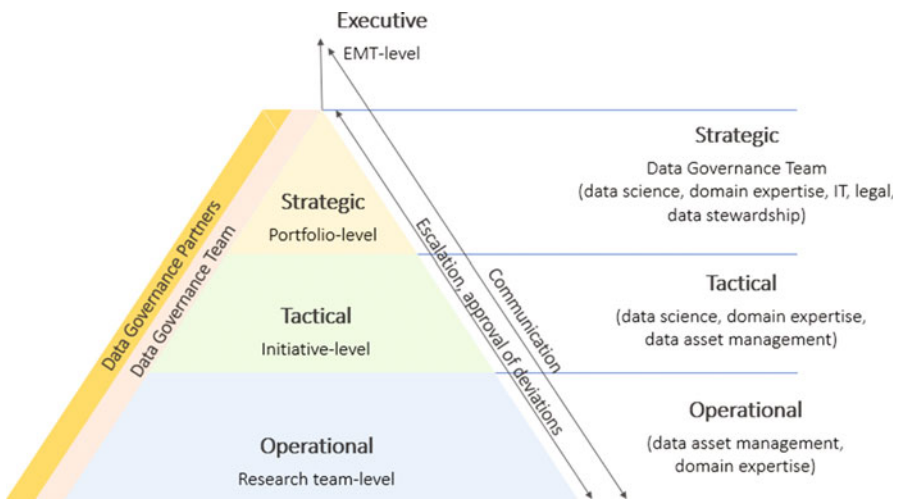


Fig. 3 Proposal for data governance under the One CGIAR model. (Modified from L. Mwanzia, pers. comm.)

solely responsible for strategic governance that determines organizational recommendations and decisions concerning all aspects of data governance, including policy implementation, repository management, standards governance and implementation, data asset management (e.g., concerning sensitive data, metadata), analytics needs etc.

Tactical level teams operate at the research initiative level and ensure that each initiative's data asset management and analysis approaches are aligned with the strategy and best practices suggested by the DGC. These teams are empowered to implement data governance principles, procedures and practices (including those around standards) as set out by the DGC, and involve data scientists, domain experts (researchers), and data asset managers (with standards expertise), with IT and legal expertise called on as needed. They provide oversight for operational level teams working on their research initiative's data asset management. Operational level teams include data asset managers and domain experts (researchers), working as part of or with research teams within an initiative to help them manage and share well-annotated, standardized data assets aligned with best practices as suggested by the DGT via tactical teams.

Considering effective use of standards in data management more specifically, governance relating to the creation, maintenance, and effective use of standards continues to be a hurdle in almost all scientific realms (McCourt et al., 2007; Zu & Wu, 2010), in no small part due to proliferation and overlap of the standards themselves. For example, there are a number of agricultural data standards, from metadata schemas aligned with industry standards such as Dublin Core (e.g. the CG Core Metadata Schema used by CGIAR Centers; <https://github.com/AgriculturalSemantics/cg-core>) to ontologies such as the Crop Ontology (CO; Shrestha et al., 2010; Cooper et al., 2018; Arnaud et al., 2020; <https://www.cropontology.org/>). The CGIAR Platform for Big Data in Agriculture also supports a beta version of the Agronomy Ontology (Devare et al., 2016; <https://bigdata.cgiar.org/resources/agronomy-ontology/>) and an early prototype socioeconomic ontology, with work just begun on small scale fisheries and aquaculture and livestock-related ontologies – with some overlap almost certain across these growing resources. As noted already, there are also several well-established standards used by researchers working in the crop genetics or genomics domains. The literature reflects the authors' experience across CGIAR and its stakeholders, that despite existing standards and growing awareness of their importance to enable linking across heterogeneous agricultural data, their development, maintenance, and adoption remains challenging (Wolfert et al., 2017; Bahlo et al., 2019; Drury et al., 2019).

Governance around standards in the public sector has been especially difficult, but efforts are ongoing even as new needs arise for interoperability standards around such concepts as Digital Sequence Information (DSI). As argued by Manzella et al. (this volume), interoperability across data systems is critical to enable legal solutions addressing access and benefit sharing (ABS) associated with the plant genetic material covered by the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA). The authors cite the need for an ontology to model the ways DSI is defined by scientists and policy makers, as it will enable mediation

across these two communities in arriving at common understanding of what DSI entails. While reaching consensus across these communities on DSI and a DOI-like standard applied to digital sequences (Digital Genomic Object identifier, or DGO) is not elementary, it will support traceability of data not only for scientific use, but also for ABS, by enabling provenance of genetic material. This work is likely to also spawn a new governance model for other use cases, that addresses interoperability for access and the needs of academic and non-academic communities. We envisage potentially similar ABS implications at the nexus of agricultural research (academic) and development (non-academic actors and beneficiaries) as machine learning, AI, and IoT applications are driven by multi-disciplinary and multi-instrument data streams.

A governance framework for the Crop Ontology project referenced above was solidified and could form a model for governance across the organization and the sector itself (Fig. 4). Governance around the ICASA variables developed by the Agricultural Model Intercomparison and Improvement Project (AgMIP) based at the University of Florida to help crop modelers harmonize crop simulation data is another noteworthy effort (White et al., 2013).

In sum, there are several reasons for such governance-related issues around data and data standards; here we have drawn on our experience across public and private sectors (i.e., CGIAR and Bayer CropScience) to illustrate a few, providing models and suggestions to overcome them where possible.

- “Invisible” data governance. While data governance may be embedded to various degrees in private sector organizations dealing with R&D data, it is not as common in the public sector, and in either case, is often not recognized as a critical function. Data governance efforts therefore tend to be ad-hoc, invisibly

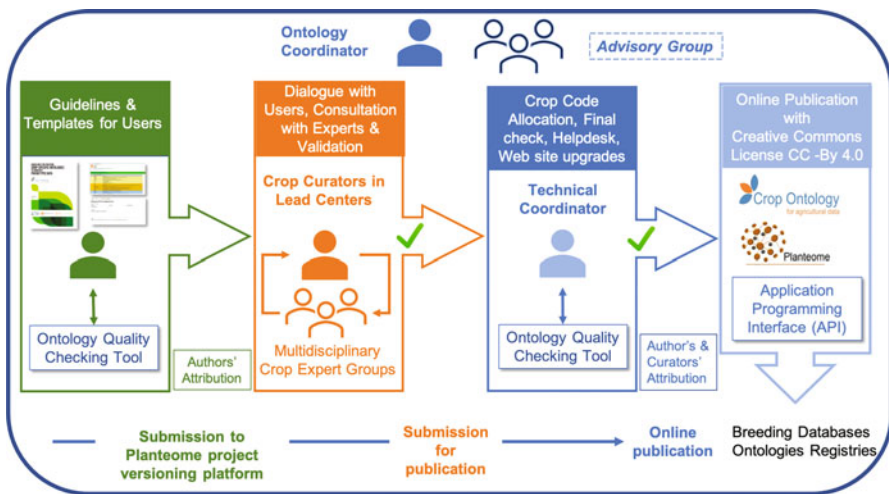


Fig. 4 Elements of a governance framework for the Crop Ontology. (Source: Crop Ontology Governance and Stewardship Framework (Arnaud et al., 2022))

keeping alive data systems, platforms and processes in spite of a lack of proper assignment and recognition of roles and responsibilities. Data stewardship, which usually ensures data accuracy, quality, and completion, is a role absorbed by data scientists or data enthusiasts. These persons must devote time to perform those data maintenance tasks on top of their core duties because a data governance strategy is absent. Another common problem when data governance strategies exist concerns their scoping of activities. Ideally all data should have a form of governance, ranging from a light setup with a few data stewards, for example, to a complex setup with several stewards, data owners and an overarching data council. In practical implementations only certain data areas are typically governed due to priorities, funding availability and staff resources. One model trialing at a couple of CGIAR Centers is the formation of institutional, multi-stakeholder governance teams, as suggested by Stedman and Vaughan (2020). Such a team might be composed of a leadership representative, research program leads and scientists, data scientists, IT professionals, data managers, and possibly, someone with IP or legal expertise. Data architects may be part of it, along with a Chief Data Officer or their equivalent. Growers are not part of these teams, but institutions including CGIAR are increasingly tweaking data consent statements towards dynamic consent models that empower farmers in voicing how data about their farms might be shared or used. This model, with some changes, is presented above (Fig. 3) and is gaining traction as one that could be widely implemented across the CGIAR System.

- **Strategy.** A data governance strategy should be driven by data practitioners' needs and not by IT tools or technical requirements (e.g., development of a data mapping tool) as is often the case. Governance bodies should devise a plan based on the relevant R&D data requirements, potentially resulting in IT tools or systems only if well-considered requirements dictate. In the agricultural research domain this is more often done the other way around, where platforms, tools, and systems dictate roles and responsibilities. An effective strategy must recognize that governance is primarily about people and not directly about tools or technologies. These latter are important but are far from the sole determinants of process and organizational efforts. Typically, R&D organizations and digitalization efforts start by implementing standards (e.g., controlled vocabularies, ontologies) across their data systems. They then move into the organizational aspects (that is, governance) as the data standards get used by more platforms and users, which demand better checkpoints and data maintenance for sustainable and reusable data and data products. Standardizing data is a very good initial step towards reusability and sustainability, but its success and the continuity of activities depends on a governance strategy and planning.
- **Leadership support, governance teams and valuing data management.** As already mentioned, the policy environment around the use and governance of data and standards is often poor. At CGIAR, the 2013 Open Access and Data Management Policy emphasized "open" but only tangentially referenced data standards and semantic interoperability, and accountability is missing. Recognizing the importance of data interoperability, CGIAR leadership supported a revision of this

policy in 2021 to address FAIR data standards and their governance – including implementation, oversight, and compliance – without which loopholes bloom and adoption can wane. As noted above, many agricultural research and education entities lack policy frameworks, high-level support, and governance teams with formalized roles. This needs to change for governance around open and FAIR data to become less challenging.

- Governance plan deployment. A grassroots data governance initiative without high-level management support is condemned to fade or fail altogether. Another cause of failure is the relatively long-term deployment of a governance plan, for which it is important to define clear and concrete deliverables (e.g., appointment of data stewards, definition of data shareability policies, integration of data standards). The benefit of a governance setup is typically not only invisible to high-level management but also to end users who could further influence the investment of resources towards governance activities. Different tactics could be employed to mitigate these situations: proof of concept implementation on small data sets and limited to a few systems, including key players as part of decision bodies; implementing adequate data stewardship recognition mechanisms; a non-disruptive governance model, ensuring a balanced distribution of data standardization efforts; avoiding over-engineered governance plans that slow processes; and partitioning the data asset ecosystem into manageable but relevant pieces (e.g. governing data on traits and related assets as events).
- Developing and maintaining standards to enable linking data is time and effort-intensive, and funder support elusive. Ontologies can help standardize the heterogeneous data that the agricultural sector deals with, thereby enabling humans and machines to more easily mine and link such cross-disciplinary data. Best practices are typically followed in developing these ontologies, including technical considerations and the involvement of domain experts working with ontologists to build and validate content (Rudnicki et al., 2016; Garijo & Poveda-Villalón, 2020). However, such consultative processes often present difficult governance issues, in that they involve compromise on preferred individual approaches in favor of standard terminology that works more generally. Some of these issues can be mitigated by inherent properties of ontologies (as compared with controlled vocabularies), one of which allows for the addition of synonyms with their contexts and definitions. However, the process of arriving at a consensus choice of concepts that accurately and sufficiently cover a particular domain can be fraught and involve huge amounts of time and discussion. Lastly, funders and institution leadership typically balk at supporting what is often seen as the tedious underpinnings of data management, making such efforts difficult to sustain. Some of these challenges were articulated by respondents to a survey conducted by Geller et al. (2018) to determine why ontologies typically tend to be sparsely updated (Table 1). These situations can be improved if (1) a more progressive data culture and explicit policy and accountability environment as referenced above is in place; (2) data is routinely re-examined and reused to generate new value, in turn demonstrating the value of standardization; and critically, (3) there is wide-ranging support to allocate budget for these efforts.

Table 1 Results from a survey to assess primary reasons for sparse updates of ontologies. Curators of 83 ontologies were contacted, with a response rate of 48/83, or 58%

Main categories of responses

Groups	Reason	Number of Ontologies	Ontologies
Group 1	Incorporated into other ontologies / systems	9	ICECI, DDO, PAE, TAO, CSEO, EHDA, MAT, AAO, EHDA
Group 2	Lack of funding, time and man-power, interruption in funding	15	BIOMODELS, RH-MESH, TGMA, ONSTR, HFO, SDO, IDOMAL, CCONT, DERMO, APAONTO, ATOL, GALEN, GMO, MCCL, NPO
Group 3	Updating in large time gaps / slow development	7	GEXO, RETO, REXO, ICNP, IDOBRU, HINO, CHMO
Group 4	Organization / Project ended	6	OBI_BCGO, NIGO, GENE-CDS, TRAK, ONTOPNEUMO, DINTO
Group 5	Paused for publication / redesigning	3	ESSO, ROO, GMM
Group 6	Concepts valid / Serves the purpose	3	EHDAA2, ONTOLURGENCE, IMGT-ONTOLOGY
Group 7	Not updating on BioPortal	3	OMIT, PMA, COGAT
Unassigned		2	HUGO, NIFSUBCELL
Total		48	

Reproduced with permission from Geller et al. (2018), under CC licence

- Collaborative development and maintenance of standards to link agricultural data. Who decides what a standard should encompass, what standards to use for particular types of data, how to build critical mass around adoption? While a governance team may have a critical role to play in these concerns, the development and maintenance of data standards are thorny issues that require broader collaboration. An example may be where the boundaries are drawn around a particular domain standard; for instance, ontology concepts to be added to the Agronomy Ontology vs. the Environmental Ontology. Successful governance and maintenance of these ontologies involves working not just with ontology and subject experts within an organization like the CGIAR system, or even within any given domain, but forging strong relationships across domain ontologies to suggest new terms in the right domain ontology and reduce concept proliferation. For a multi-center entity like CGIAR, the governance of repository-level metadata also requires agreement from data and information managers around a common, widely responsive but industry-aligned standard, in this case, the Dublin Core-based CGIAR Core Metadata Schema (<https://github.com/>

[AgriculturalSemantics/cg-core](#)) which is broadly applicable to wide-ranging agricultural use cases. One model for cross-institutional domain-based governance is provided by the CGIAR Platform for Big Data in Agriculture (<https://bigdata.cgiar.org/>), launched in 2017 with the objective of increasing the impact of agricultural research and development by turning open and FAIR data into a powerful tool for discovery, while integrating principles of responsible and ethical data use (see box).

The CGIAR Big Data Platform hinges on several Communities of Practice (CoPs) (Agronomy Data, Crop Modelling, Data and Information Management, Geospatial Data, Livestock Data, Ontologies, and Socio-Economic Data) which engage research domain experts and practitioners, data and information managers, and ethics and IP specialists from CGIAR and a variety of stakeholders. Such CoPs could play a key role in helping to provide data and standards-focused governance, in the form of interactions across entities and individuals developing standards and other data solutions and providing cross-learning opportunities and guidance (Arnaud et al., 2020). While this is not yet the case for all the Big Data Platform CoPs, some are deeply involved in such activities, including the Ontologies and Data, Geospatial, and Information and Data Management CoPs. All these CoPs are also instrumental in facilitating the use of data standards, helped by the Platform's GARDIAN data ecosystem (<https://gardian.bigdata.cgiar.org/>). GARDIAN enables the discovery of data assets produced by the CGIAR network and other key stakeholders in the public domain, and provides a data-to-analytics and visualization environment and model pipelines to realize the value of increasingly FAIRer data, bolstering the work of the CoPs.

- Standards make data easier to link and use – but what about data owners? Industry has already recognized that digital agriculture and an associated constellation of powerful technologies (e.g., IoT, remote sensing, AI) that can mine well-harmonized data offer huge potential for hyperlocal, tailored agricultural recommendations (e.g. for fertilizer). As addressed in more depth by Zampati et al. (this volume), such technologies raise critical legal and ethical questions. While farmers increasingly acknowledge the benefits of such standards-reliant technologies, they are also beginning to express concern about losing ownership over their data. Data ownership issues are often exacerbated by concerns about privacy as technology increasingly facilitates data triangulation to expose personally-identifiable information, yet the foregoing discussion has thus far omitted mention of the data owner. Data cooperatives are a recent model for governing farm data, with several examples in the US, such as the Ag Data Coalition (ADC) and the Grower Information Services Cooperative (GISC). Some data cooperatives (like ADC) offer secure data repository solutions that enable farmers to store their data and decide with which platforms, agencies or research entities to share

it. Others like GiSC, offer a repository and also perform analytics to give farmers greater insight into their production practices and can negotiate opportunities to monetize data on their behalf. There are subscription-based approaches like the Farmer Business Network (FBN) which offer data platforms and analytics over the data pool to offer the farmer farm-specific management and profitability insights such as yield by soil type or fertilizer, and input price comparisons. Similar approaches are also emerging in developing countries, for example, Digital Green is building FarmStack, a data exchange platform for farmers in India, with features similar to ADC (<https://farmstack.digitalgreen.org/>). Yara and IBM have also launched a joint effort to enable farmers to securely share data and retain determination around who uses it and how, benefiting monetarily in the process (Yara International, 2020). Central to these newer models is the placement and valuing of the data owner in the mix of stakeholders that determine how data gets managed and used.

3 Conclusions

Good data governance practices are the beating heart of innovation and impact, particularly through their impact on data interoperability and reusability by humans and machines. Data governance ensures that data assets remain widely available and interoperable, but are also secure, trustworthy and not misused (Stedman & Vaughan, 2020). As the digital landscape and data capabilities become more sophisticated, these latter concerns are especially important to assure that governance efforts address the gamut, from policy through standards, to ethics, to assure that sensitive data, rules for data use, data sharing agreements and allied efforts are considered in the light of managing both institutional and individual risk. We have attempted to address legal, technical and cultural challenges to data and standards governance, outlining some models for successful governance to enable data linking that cover a range of aspects, from administrative and financial enablers to the human and technical considerations. In doing so we have briefly addressed considerations such as policy environment and governance teams and data cooperatives – both within and across organizational structures; funding support; capacity and awareness of the human actors in the data ecosystem; and technical infrastructure that allow data owners to have a higher level of self-determination over their information.

All interventions aiming for impactful data governance must recognize the human experiences involved before improved practices can be recommended or required. Thus, we have touched upon the epistemology of scientific research in general, and agricultural research in particular as largely being hypothesis-driven rather than inductive and empirical as fields embracing data science and big data technologies tend to be. As might be expected, the business of how research is viewed and conducted is likely to be a key determinant of how the data it produces is handled, as appears to be borne out by our experiences with CGIAR researchers for whom the notion of *in silico* analysis involving secondary data is relatively new.

There are many unaddressed questions and gaps that remain regarding data governance as it relates to interoperability and enabling data linkages. Blockchain is being increasingly explored as a data provenance solution, a way to enable data security, traceability, and accountability (Liang et al., 2017; Ramachandran & Kantarcioglu, 2017; Devan, 2018; Shabani, 2019; Kochupillai, this volume). The Food Trust Blockchain has already been launched by IBM as a food traceability platform and adopted by large retailers, fruit and meat wholesalers, and multinationals in the food products sector (Stanley, 2018). Closer to the research world, Blockchain has been proposed as a solution to handling electronic medical records (EMRs) to give patients access to their medical records across providers and treatment sites via an immutable record. As envisioned by Azaria et al. (2016), the application of Blockchain to EMRs via a decentralized records management system called MedRec allows researchers and other medical stakeholders to mine aggregated, anonymized data. In return, these actors sustain and secure the network via a “Proof of Work” algorithm that is tamper-proof, involving individual nodes competing to solve computational “puzzles” before another block of content can be added to the chain. The work required of “miners” to append blocks assures that it is difficult to rewrite history on the Blockchain. Azaria et al. therefore propose empowering researchers through big data pools, while involving patients and care providers in choices around the release of their (meta)data.

Such models that include farm data and the farmer as the determinant of how her data is used, and by whom, are beginning to gain traction through the notion of data cooperatives. While Blockchain is still in its infancy as an enabler in these ecosystems, it is likely to gain prominence in the near-term, as data economies grow across sectors. How data standards mesh with and augment Blockchain capabilities is not clear but requires consideration in the near-term. What seems clearer, in fact, is the potential of Blockchain technology to provide accountability and traceability in the standards development process, even if privacy is not generally a concern. That data standards are critical for enabling interoperability is generally accepted; as this paper attempts to make clear, there remain some unexplored considerations around their governance, along with questions of what constellations of actors ought to be involved in standards development and maintenance.

References

- Arnaud, E., Hazekamp, T., Laporte, M.-A., Antezana, E., Andres Hernandez, L., Pot, D., Shrestha, R., Dreher, K., Castiblanco, V., Menda, N., Fabio Guerrero, A., Hualle, V., Salas, E., Mendes, T., Makunde, G., Chaves, I., Rathore, A., Das, R., Afolabi, A., Pietragalla, J., Pommier, C., Michotey, C., Detras, J., McNally, K., Borja, N., Winger, L., Cooper, L., Jaiswal, P., Mauleon, R., & Yu, J. (2022). *Crop ontology governance and stewardship framework*. Retrieved from <https://hdl.handle.net/10568/118001>
- Arnaud, E., Laporte, M.-A., Kim, S., Aubert, C., Leonelli, S., Miro, B., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P. L., Mungall, C. J., Pietragalla, J., Agbona, A., Muliro, J., Detras, J., Hualla, V., Rathore, A., Das, R. R., Dieng, I., Bauchet, G., Menda, N., Pommier,

- C., Shaw, F., Lyon, D., Mwanzia, L., Juarez, H., Bonaiuti, E., Chiputwa, B., Obileye, O., Auzoux, S., Dzalé Yeumo, E., Mueller, L. A., Silverstein, K., Lafargue, A., Antezana, E., Devare, M., & King, B. (2020). The ontologies community of practice: A CGIAR initiative for big data in agrifood systems. *Patterns*, 1(7). Retrieved from <https://doi.org/10.1016/j.patter.2020.100105>
- Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). *MedRec: Using Blockchain for medical data access and permission management*. 2nd International Conference on Open and Big Data. Retrieved from <http://www.pitt.edu/~babay/courses/cs3551/papers/MedRec.pdf>
- Bahlo, C., Dahlhaus, P., Thompson, H., & Trotter, M. (2019). The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. *Computers and Electronics in Agriculture*, 156, 459–466. Retrieved from <https://doi.org/10.1016/j.compag.2018.12.007>
- Barham, B., Goldman, I., van Rijn, J., Foltz, J., & Agnes, M. I. (2017). Land-Grant University faculty attitudes in and engagement with open source scholarship and commercialization. *Agricultural and Environmental Letters*, 2(1). Retrieved from <https://doi.org/10.2134/ael2017.03.0008>
- Bill and Melinda Gates Foundation. (2021). *Bill and Melinda Gates Foundation Open Access Policy*. Retrieved from: <https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>
- Bodenreider, O. J. (2005). Biomedical ontologies. *Pacific Symposium on Biocomputing*, 76–78. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4300097/>
- Brockman, J. (2015, July 5). *Popper vs. A Conversation with Peter Coveney*. Retrieved from Edge Conversations: https://www.edge.org/conversation/peter_coveney-popper-versus-bacon
- CGIAR. (2013, October 2). *CGIAR open access and data management policy*. Retrieved from <https://cgspace.cgiar.org/bitstream/handle/10947/4488/Open%20Access%20Data%20Management%20Policy.pdf?sequence=1&isAllowed=y>
- CGIAR. (2021, April 16). *CGIAR open and FAIR data assets policy*. Retrieved from https://cgspace.cgiar.org/bitstream/handle/10568/113623/CGIAR_OFDA_Policy_Approved_16April2021.pdf?sequence=1&isAllowed=y
- Cooper, L., Meier, A., Laporte, M.-A., Elser, J. L., Mungall, C., Sinn, B. T., Cavaliere, D., Carbon, S., Dunn, N. A., Smith, B., Qu, B., Preece, J., Zhang, E., Todorovic, S., Gkoutos, G., Doonan, J. H., Stevenson, D. W., Arnaud, E., & Jaiswal, P. (2018). The Planteome database: An integrated resource for reference ontologies, plant genomics and Phenomics. *Nucleic Acids Research*, 46(D1), D1168–D11804. Retrieved from <https://doi.org/10.1093/nar/gkx1152>
- Denk, F. (2017). Don't let useful data go to waste. *Nature*, 543(7643), 7. Retrieved from <https://www.nature.com/news/don-t-let-useful-data-go-to-waste-1.21555>
- Devan, G. (2018). *How Blockchain technology is revolutionizing data provenance*. Retrieved from medium.com: <https://medium.com/blockpool/how-blockchain-technology-is-revolutionizing-data-provenance-e47610019390>
- Devare, M., Aubert, C., Laporte, M.-A., Valette, L., Arnaud, E., & Buttigieg, P. L. (2016). Data-driven agricultural research for development – A need for data harmonization via semantics. In P. Jaiswal & R. Hoehndorf (Eds.), *7th international conference on biomedical ontologies, ICBO 2016* (Vol. 1747:2). CEUR Workshop Proceedings.
- Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLOS Biology*, 1–12. Retrieved from <https://doi.org/10.1371/journal.pbio.1002295>
- Drury, B., Fernandes, R., Moura, M.-F., & de Andrade Lopes, A. (2019). A survey of semantic web technology for agriculture. *Information Processing in Agriculture*, 6(4), 487–501. Retrieved from <https://doi.org/10.1016/j.inpa.2019.02.001>
- European Commission Expert Group on FAIR Data. (2018). *Final report and action plan: Turning FAIR into reality*. Retrieved from Publications Office of the EU: <https://op.europa.eu/s/oHHB>
- Garijo, D., & Poveda-Villalón, M. (2020). *Best practices for implementing FAIR vocabularies and ontologies on the web*. *arXiv.org*. Computer Science: Digital Libraries. Retrieved from <https://arxiv.org/abs/2003.13084v1>

- Geller, J., Keloth, V. K., & Musen, M. A. (2018). *How sustainable are biomedical ontologies?* Proceedings, AMIA Annual Symposium. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371329/>
- Koers, H., D. Bangert, E. Hermans, R. van Horik, M. de Jong, and M. Mokrane. (2020). *Recommendations for services in a FAIR data ecosystem.* <https://dx.doi.org/10.1016%2Fj.patter.2020.100058>
- Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K., & Njilla, L. (2017, May). ProvChain: A Blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability CCGRID '17: *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing* (pp 468–477). Retrieved from <https://doi.org/10.1109/CCGRID.2017.8>
- Leonelli, S. (2019). Data governance is key to interpretation: Reconceptualizing data in data science. *Harvard Data Science Review*. Retrieved from <https://doi.org/10.1162/99608f92.17405bb6>.
- McCourt, B., R. A. Harrington, K. Fox, C. D. Hamilton, K. Booher, W. E. Hammond, A. Walden, M. Nahm. (2007). *Data standards: At the intersection of sites, clinical research networks, and standards development initiatives*. Retrieved from Therapeutic Innovation and Regulatory Science: <https://doi.org/10.1177/009286150704100313>
- Noyes, K. (2014). *Cropping up on every farm: Big Data technology*. Retrieved from Fortune: <https://fortune.com/2014/05/30/cropping-up-on-every-farm-big-data-technology/>
- Pham, X., & Stack, M. (2018). How data analytics is transforming agriculture. *Business Horizons*, 61(1), 125–133. Retrieved from <https://doi.org/10.1016/j.bushor.2017.09.011>
- Quitau, A. (2020). *IBM Supercomputer Summit Attacks Coronavirus...* Retrieved from IBM Digital Nordic: <https://www.ibm.com/blogs/nordic-msp/ibm-supercomputer-summit-attacks-coronavirus/>
- Ramachandran, A., & Kantarcioglu, M. (2017). Using Blockchain and smart contracts for secure data provenance management. *arXiv.org. Computer Science: Cryptography and Security*. <https://arxiv.org/abs/1709.10000>
- Rijmenam, M. V. (2013). *John Deere is revolutionizing farming with big data*. Retrieved from Datafloq: <https://datafloq.com/read/john-deere-revolutionizing-farming-big-data/511>
- Robinson, P. N., & Haendel, M. A. (2020). *Ontologies, knowledge representation, and machine learning for translational research: Recent contributions*. Retrieved from <https://doi.org/10.1055/s-0040-1701991>
- Rosenbaum, S. (2010). *Data governance and stewardship: Designing data stewardship entities and advancing data access*. <https://doi.org/10.1111/2Fj.1475-6773.2010.01140.x>
- Rudnicki, R., Smith, B., Malyuta, T., & Mandrick, W. (2016). *White paper: Best practices of ontology development*. CUBRC Advantage Through Technology. Retrieved from https://www.nist.gov/system/files/documents/2019/05/30/nist-ai-rfi-cubrc_inc_002.pdf
- Sara, R., & Devare, M. (2020). *Excellence in agronomy (EiA) initiative: Best practice guidelines to support global access implementation* [Guidance note for CGIAR initiative].
- Semantic Scholar. (2020). *CORD-19*. Retrieved from COVID-19 Open Research Dataset: <https://www.semanticscholar.org/cord19>
- Shabani, M. (2019). Blockchain-based platforms for genomic data sharing: A decentralized approach in response to governance problems? *Journal of the American Medical Informatics Association*. Retrieved from <https://dx.doi.org/10.1093%2Fjamia%2Ffoc149>
- Shrestha, R., Arnaud, E., Mauleon, R., Senger, M., Davenport, G. F., Hancock, D., Morrison, N., Bruskiwich, R., & McLaren, G. (2010). Multifunctional crop trait ontology for breeders' data: Field book, annotation, data discovery and semantic enrichment of the literature. *AoB Plants*. Retrieved from <https://doi.org/10.1093/aobpla/plq008>
- Sieleman, K., A. Hafner, B. Pucker. (2020). The reuse of public datasets in the life sciences: Potential risks and rewards. *PeerJ* 8:e9954 <https://doi.org/10.7717/peerj.9954>
- Smalley, E. (2018). *In silico farming drives next wave in agriculture*. *Nature Biotechnology*, 36(9), 783–784.

- Smith, K. (2013). *A brief history of NCBI's formation and growth*. Retrieved from The NCBI Handbook [Internet]. 2nd Edition: <https://www.ncbi.nlm.nih.gov/books/NBK148949/>
- Stanley, A. (2018). Ready to rumble: IBM launches food trust Blockchain for commercial use. *Forbes*, October 8, 2018. Retrieved from <https://www.forbes.com/sites/astanley/2018/10/08/ready-to-rumble-ibm-launches-food-trust-blockchain-for-commercial-use/?sh=13736cb97439>
- Stedman, C., & Vaughan, J. (2020). *What is data governance and why does it matter?* Tech Target. Retrieved from Tech Accelerator. https://searchdatamanagement.techtarget.com/definition/data-governance?_ga=2.159940984.476600454.1612269086-1594224955.1612269086&_gl=1*14r7hxp*_ga*MTU5NDIyNDk1NS4xNjEyMjY5MDg2*_ga_RRBYR9CGB9*MTYxMjIOTA4OC4xLjEuMTYxMjI0MTAzMy4w
- Streich, J., Romero, J., Gazolla, J. G., Kainer, D., Cliff, A., Prates, E. T., Brown, J. B., Khoury, S., Tuskan, G. A., Garvin, M., Jacobson, D., & Harfouche, A. L. (2020). Can Exascale computing and explainable artificial intelligence applied to plant biology deliver on the United Nations sustainable development goals? *Current Opinion in Biotechnology*, 61, 217–225. Retrieved from <https://doi.org/10.1016/j.copbio.2020.01.010>
- Thessen, A. E., & Patterson, D. J. (2011). Data issues in the life sciences. *Zookeys*, 150, 15–51. Retrieved from <https://zookeys.pensoft.net/articles.php?id=3041>
- Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ*, 5, e3208. Retrieved from <https://doi.org/10.7717/peerj.3208>
- White, J. W., Hunt, L. A., Boote, K. J., Jones, J. W., Koo, J., Kim, S., Porter, C. H., Wilkens, P. W., & Hoogenboom, G. (2013). Integrated description of agricultural field experiments and production: The ICASA version 2.0 data standards. *Computers and Electronics in Agriculture*, 96, 1–12.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hoofst, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wiseman, L., Pesce, V., Zampati, F., Sullivan, S., Addison, C., & Drolet, J. (2019). *Review of codes of conduct, voluntary guidelines and principles relevant for farm data sharing* (CTA working paper 19/01). CTA. Retrieved from <https://hdl.handle.net/10568/106587>
- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big data in smart farming—a review. *Agricultural Systems*, 153, 69–80. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0308521X16303754?via%3Dihub>
- Yara International. (2020). *Yara and IBM launch an open collaboration for farm and field data to advance sustainable food production*. January 23, 2020. Retrieved from <https://www.yara.com/corporate-releases/yara-and-ibm-launch-an-open-collaboration-for-farm-and-field-data-to-advance-sustainable-food-production/>
- Zhu, H., & Wu, H. (2010). Assessing quality of data standards: Framework and illustration using XBRL GAAP Taxonomy. In S. Sánchez-Alonso & I. N. Athanasiadis (Eds.), *Metadata and Semantic Research. MTSR 2010. Communications in Computer and Information Science* (Vol. 108). Springer. https://doi.org/10.1007/978-3-642-16552-8_26

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

