Data Management Challenges for Translational Agricultural Research in West Africa: The Case of the Crops Research Institute, Ghana

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

COLLECTION:
DATA MANAGEMENT PLANNING ACROSS DISCIPLINES AND INFRASTRUCTURES

RESEARCH PAPERS

]u[ubiquity press

JOYCE KORANTENG-ACQUAH
SABINA LEONELLI
*Author affiliations can be found in the back matter of this article

ABSTRACT

The establishment of robust, long-term data management strategies remains especially challenging in low-resourced and applied research contexts. Drawing on qualitative methods including ethnographic fieldwork and survey data collected from August to December 2023, this paper discusses ongoing concerns around data practices within Ghana's Crops Research Institute (CRI), a premier West African facility for translational agricultural research. This case exemplifies the concerns emerging from a lively and productive research environment, effectively contributing to basic and applied research for a variety of publics, where, however, the infrastructure is fragile and institutional memory resides mainly in people rather than in computational systems. CRI researchers recognise the key components of effective data stewardship but encounter significant material and administrative obstacles that prevent these ideals from being fully realised. Using social scientific understandings of institutional memory and epistemic cultures, the paper highlights three key reasons for such difficulties: (1) frustrated epistemic cultures, which entrench the community's difficulties in implementing known necessary practices; (2) data sedimentation, with successive layers of processed and summarised knowledge accumulating on top of earlier research records; and (3) epistemic precarity, that is, a widely perceived gap between understanding what is needed and having the capacity to meet those needs. The picture that emerges is not about insufficient commitment to data sharing and effective long-term management, but rather about structural conditions that make continuity in data practices precarious. If open science aims to effectively engage resource-constrained settings, it must address the lack of relevant infrastructure and administration rather than presuming inadequacies and lack of commitment in research cultures.

CORRESPONDING AUTHOR:
Joyce Koranteng-Acquah
Technical University of Munich, DE
joyacquah9@gmail.com

KEYWORDS:
data infrastructure; institutional memory; research cultures; data precarity; open science

TO CITE THIS ARTICLE:
Koranteng-Acquah, J. and Leonelli, S. 2026 Data Management Challenges for Translational Agricultural Research in West Africa: The Case of the Crops Research Institute, Ghana. *Data Science Journal*, 25: 35 pp. 1–27. DOI: https://doi.org/10.5334/dsj-2026-035

INTRODUCTION

In contemporary scientific research, robust data management and sharing practices are critical for several reasons. Data are foundational to knowledge production and play a key role in the reproducibility and validation of scientific results (Borgman, 2015; Leonelli, 2016).[1] Particularly in interdisciplinary research responding to global challenges, integrating diverse datasets can yield insights beyond those of single-discipline studies. Fields such as health, environmental, and agricultural sciences, which are often interconnected on both local and global scales, depend on such data-sharing practices to conduct cross-contextual analyses and tackle complex, multi-dimensional challenges. Effective data sharing facilitates independent verification, enhancing the reliability and credibility of findings (Mwelva *et al.,* 2020; Peng, 2011; Piwowar and Vision, 2013; Soundararajan and Mishra, 2023; Tenopir *et al.,* 2011). It may also facilitate the reproducibility of outcomes, which is both a methodological concern and a matter of scientific integrity, particularly when taking diverse forms of reproducibility into account (Leonelli, 2018). These practices ensure that discoveries extend beyond the originating lab to a broader research community, fostering resilient knowledge-making and cumulative knowledge-building (Baker, 2016; Diaba-Nuhoho and Amponsah-Offeh, 2021; Errington, 2024).

Alongside these foundational practices, technological advances have enabled real-time data collection, digital archiving, and intelligent sharing across platforms (Wu *et al.,* 2026). These advancements are expected to enhance access to data and the insights derived from it, potentially facilitating swift responses to evolving research needs. Moreover, the technologies associated with promoting transparency in scientific processes may also bolster public trust and enhance the perceived legitimacy of scientific institutions (Mittal *et al.,* 2023). Trust in applied sciences such as agriculture is particularly significant because research outcomes are widely perceived to influence public policy on critical issues such as food security and rural development (Williamson and Leonelli, 2022). Transparency and accessibility in data practices may foster public confidence in the scientific process, which in turn may inform decision-making and the implementation of effective policies. The notion that well-managed data ecosystems support scientific progress and uphold the social contract[2] between scientific institutions and the communities they serve is frequently underscored (e.g. Borgman and Groth, 2025), although the dynamics of such relationships warrant closer examination (Leonelli, 2025; Wyatt, 2025)—especially when it comes to the incorporation of indigenous knowledge, where foundational principles such as CARE need to be implemented in ways that are sustainable and responsive to local communities (Belarde-Lewis *et al.,* 2024).

The Crops Research Institute (CRI) exemplifies the complex interplay between scientific research, data management, and policy relevance in a resource-constrained environment. As a key agricultural research institution in Ghana, CRI operates at the intersection of knowledge production and practical application, translating scientific findings into actionable agrarian innovations. CRI research includes basic and applied elements and focuses on translation into local agricultural systems, thereby prioritising innovations that benefit local communities and markets. This makes the ties between CRI and its social and geographical surroundings particularly strong, as reflected in the everyday challenges and opportunities encountered by CRI researchers in fulfilling their goals. Data stewardship, including the production, storage, curation and dissemination of research data, presents specific requirements in such a context, which we detail in what follows.

Advancements in digitalisation offer new opportunities for efficient data handling. However, infrastructural constraints, diverse stakeholder needs, and varying levels of technological adoption among researchers complicate CRI's data management landscape. By investigating how CRI navigates data management, knowledge exchange, and technological integration within its sociotechnical landscape, the study examines the influence of these practices on

1 According to Leonelli (2016), data is not inherently meaningful or valuable on its own; instead, its role as the 'bedrock' of knowledge production is contingent on how it is curated, contextualised, and used within specific research practices. Borgman (2015) also emphasises data's foundational role in research, focusing on its organisation, documentation, and preparation for sharing and reuse, as well as the need for meticulous data management throughout the research life cycle, arguing that clear, structured, and accessible data practices are essential for enhancing reproducibility, validation, and interoperability.

2 Douglas and Branch (2024) examine the concept of a 'social contract for science', questioning the ideal of value-free science by analysing how societal values and public trust intersect with scientific research practices.

research continuity, institutional memory, and collaboration within the institute's extensive networks. Building on survey and ethnographic data gathered at CRI between August and December 2023 (details below), the paper aims to illuminate the challenges and opportunities in establishing resilient data ecosystems that support both scientific advancement and policy development in translational agricultural research in the West African context.

THE ROLE OF CRI IN GHANA'S AGRICULTURAL SECTOR

As part of the Council for Scientific and Industrial Research (CSIR), CRI's core mandate is to conduct research and develop technologies that improve crop productivity[3] and sustainability. Nevertheless, the process of creating these technologies is shaped as much by institutional priorities and resource constraints as by scientific ideals. CRI is a research institute that generates technical knowledge. It is a space where scientific inquiry, bureaucratic structures, and political mandates collide. Unlike global agricultural research powerhouses like the Consultative Groups for International Agricultural Research (CGIAR), comprising, for instance, International Maize and Wheat Improvement Centre (CIMMYT) in Mexico and the International Institute of Tropical Agriculture (IITA) in Nigeria, CRI is a state-sponsored organisation and operates within a framework of national accountability. It is tethered to the Ministry of Food and Agriculture (MOFA), whose policy agenda steers research priorities. While also affiliated with the Ministry of Environment, Science, Technology and Innovation (MESTI) through CSIR, CRI's research focus is influenced by MOFA's strategic objectives.

CIMMYT, as part of the CGIAR network, sets its research agenda in response to global challenges identified by CGIAR, such as climate change, food security, and sustainable agriculture. Its funding comes mainly from international donors, foundations, and multilateral organisations, enabling it to pursue long-term, strategic research programs without being tightly bound to any single government's immediate policy goals. Similarly, IITA, though focused on sub-Saharan Africa, operates within a broader regional framework that allows it to align with major global development initiatives. Researchers at these institutions are in a prime position to engage in international collaborations and explore cutting-edge agricultural innovations without immediate political constraints. CRI also participates in global partnerships through donor-funded projects and in cooperation with CGIAR institutes. In fact, such collaborations are highly valued as a crucial source of funding and inspiration. At the same time, its institutional structure and strong dependence on state institutions make CRI more sensitive to national agricultural research needs, making its programme more tightly tethered and responsive to national priorities than its more international counterparts, CIMMYT or IITA.

The necessity to align research outputs with government priorities means that scientific agendas can shift with political cycles, occasionally restricting the pursuit of long-term agricultural innovations that may not immediately correspond with national policy directives. This dual accountability—to both scientific communities and state bureaucracies—creates an environment where research agendas are constantly negotiated. This is particularly visible at CRI, where scientists need to produce academically rigorous knowledge while also fulfilling the practical need to deliver actionable data to policymakers and farming communities across the country. What distinguishes CRI from comparable institutions worldwide is not its geographic and institutional focus but its integration within Ghana's agricultural governance structure. CRI acts both as a generator of knowledge and as a conduit through which agricultural policies are developed and legitimised. Its scientific outputs serve dual purposes: as empirical evidence for policy decisions and as indicators of institutional credibility in securing research funding, particularly from international donors. CRI's dual role complicates the seemingly straightforward task of 'doing science', with significant implications for its data management strategy, as we shall discuss.

The institute's departments are organised around specific crops, including maize, rice, cassava, yams, vegetables, and legumes, which coexist alongside cross-cutting units dedicated to biotechnology, plant health, and socioeconomic research. Recent restructuring of the institute

3 CRI has contributed to the development of improved maize and cassava varieties, which have significantly bolstered food security and rural livelihoods (Pinamang Acheampong *et al*., 2021). For instance, in 2020, through the Modernising Agriculture in Ghana (MAG) project, 187 farmers in four districts of the Volta Region doubled their revenue and increased productivity by 61% from the cultivation of improved cassava in three years. [source]

has introduced a commercialisation division, reflecting an increasing emphasis on generating revenue through technology transfer, seed production, and consultancy services. This shift highlights the evolving expectations placed on public research institutions to balance scientific inquiry with financial sustainability. For CRI, this aspect is vital, as the institute does not receive government funding for research and must secure funding through external grants, partnerships, and commercial activities. This organisational hybridity shapes the types of research conducted and the epistemic value assigned to different forms of knowledge as CRI navigates its role between knowledge production and commercial viability.

Given the complexities of data management and its critical role in both the scientific and policy dimensions of CRI's work, we are interested in exploring how data management practices shape the institute's ability to facilitate translational agricultural research and policy development in Ghana. Specifically, we seek to understand how CRI's approaches to data management, knowledge exchange, and technological integration influence research continuity, institutional memory, and collaboration within its networks. At the heart of this inquiry lies a fundamental question: what are the challenges and opportunities in establishing resilient data ecosystems that support both scientific advancement and policy development in the context of translational agricultural research in West Africa?

By examining these dynamics, we aim to uncover the broader implications of CRI's data practices on agricultural policy formulation in Ghana, the mechanisms of institutional knowledge transfer, and the evolving role of research institutes in addressing societal challenges through scientific inquiry. Through this study, we seek to shed light on the essential yet often underexplored role of data infrastructures in agricultural research and their significance at the intersection of science and policy. Understanding how CRI navigates these challenges and opportunities provides deeper insights into ways research institutions can strengthen their impact on agricultural development and food security in Ghana and beyond.

LITERATURE REVIEW

DATA SHARING: FROM IDEAL TO REALITY

There is a puzzle at the heart of open science: everyone agrees data should be shared, yet sharing rarely happens as straightforwardly as policy aspirations suggest. Despite widespread endorsement of open science principles, empirical studies consistently identify a gap between researchers' stated support for data sharing and their everyday practices. Longitudinal work by Tenopir *et al.* (2011, 2015, 2020) shows that although researchers increasingly recognise the value of making data available, they worry about misuse, inadequate attribution, and whether others possess sufficient contextual knowledge to interpret their data appropriately. As Wallis, Rolando and Borgman (2013) put it, researchers frequently ask: 'If we share data, will anyone use them?'—particularly when datasets require the kind of tacit knowledge that does not easily travel beyond its original setting. Even within established collaborations, successful data reuse depends less on technical accessibility than on confidence in a dataset's provenance, quality, and relevance to particular analytical goals (Faniel and Jacobsen, 2010; Rolland and Lee, 2013). Consequently, data sharing is often driven as much by long-standing collaborative relationships and shared epistemic commitments as by formal mandates and technical infrastructures (Birnholtz and Bietz, 2003; Borgman, 2015; Leonelli, 2016).

But here is the catch: much of this literature is grounded in research environments where digital infrastructure and research funding are relatively robust (Borgman, 2015). The conditions under which data are expected to circulate are therefore taken for granted. Yet these assumptions become far less tenable in resource-constrained settings. Bezuidenhout *et al.* (2017a), drawing on experiences from African research institutions, argue that infrastructural limitations fundamentally shape what forms of data management are possible, capturing this disparity in the memorable observation that '$100 is not much to you'. In such contexts, intermittent connectivity, unreliable electricity, limited storage infrastructure, unstable project funding, and weak institutional support mean that data management frequently depends on constant improvisation and resilience strategies (Bezuidenhout *et al.*, 2017b). These constraints affect not only whether data can be shared but also whether they can be securely preserved and located over time. Given these challenges, West African agricultural research institutions remain largely absent from these debates on open science and research data management, notwithstanding their central role in food security, agricultural innovation, and development policy. Translational research in these contexts must bridge laboratory science, field trials, farmer knowledge, and policy implementation, each operating according to different evidential standards and practical needs (Leonelli *et al.*, 2018). This raises broader questions about how organisational memory, data continuity, and research cultures are sustained when institutional precarity is not the exception but the norm.

These questions draw attention to the infrastructures through which research data are expected to endure beyond individual projects and researchers. If long-term preservation and reuse are central aspirations of open science, then the distinction between formal data repositories and more improvised forms of digital storage becomes analytically important. It is therefore necessary to clarify what we mean by a data repository, since participants in this study often use the term in ways that differ from its technical meaning within the research data management literature. Throughout this paper, we use the term data repository to refer to a managed digital infrastructure designed for the long-term preservation, curation, discovery, and reuse of research data. Unlike simple data storage systems, repositories typically incorporate persistent identifiers, metadata standards, documentation requirements, preservation strategies, and governance arrangements that support data integrity, discoverability, accessibility, and long-term stewardship (Borgman, 2015; Briney, 2015; Lord and Macdonald, 2003; Lynch, 2003; Whyte and Tedds, 2011; Wilkinson *et al.*, 2016). Examples include repositories such as Dryad and Dataverse, which are designed to store files and facilitate their interpretation, citation, verification, and reuse through structured curation and documentation (Crosas, 2011; Borgman, 2015).

This understanding differs fundamentally from ad hoc storage arrangements, such as personal or commercial cloud storage, external hard drives, laboratory computers, or shared network folders, which may provide redundancy or enable collaborative access but generally lack the metadata, preservation planning, governance structures, and curatorial oversight that repositories are intended to provide (Borgman, 2015; Briney, 2015). We flag this distinction because, as the findings demonstrate, participants at CRI do not always differentiate between these forms of infrastructure. References to 'a repository' often express a desire for a stable, shared location where research outputs can be retained and accessed, rather than for a formally curated repository as conceptualised within the data stewardship literature (Whyte and Tedds, 2011; Borgman, 2015). This divergence between participants' functional expectations and the technical definition of a repository is itself analytically significant. Rather than reflecting conceptual confusion, it illustrates how experiences of infrastructural scarcity shape researchers' expectations of what digital infrastructures are expected to accomplish. We return to this distinction in the Results and Discussion.

INSTITUTIONAL MEMORY AND RESEARCH CONTINUITY

So, what exactly gets remembered and how? Organisational memory is not stored in one place—it is distributed, spanning infrastructures, routines, people, and archival practices. What an organisation remembers emerges from the interplay of multiple repositories, retrieval pathways, and decisions about what is recorded, classified, and made retrievable (Bowker, 2005; Walsh and Ungson, 1991). What looks like solid, stable data infrastructure depends on continuous, often invisible work to keep different materials compatible with one another. Studies in environmental and earth sciences show that bringing datasets together requires alignment of measurement techniques, metadata standards, and ontological categories—what Edwards *et al.* (2011) aptly call 'science friction'. Each act of harmonisation embeds assumptions about comparability, differences that can be ignored, and what must be preserved.

But keeping research going over time involves much more than just storage. As Star (1999) argues, infrastructures are typically invisible when they function well, receding into the background of everyday practice. Yet that apparent stability depends on repetitive work that rarely gets noticed: migrating file formats, updating software, repairing systems, and constantly renegotiating protocols as personnel and technologies change (Karasti, Baker and Millerand, 2010; Ribes and Jackson, 2013). When these systems fail, the infrastructure and the labour that sustains it suddenly become visible (Star, 1999), revealing that stability is an ongoing achievement of maintenance, repair, and improvisation (Jackson *et al.,* 2007). In smaller research settings, where data management is just part of what researchers do rather than someone's dedicated job, personnel turnover can make entire bodies of work unusable (Cragin *et al.,* 2010). In fact, Reichmann (2023) connects high staff turnover rates to broader increases in academic precarity, suggesting that whether research data is managed and shared properly fundamentally depends on the material conditions and job security of the people doing the work. When someone leaves, the loss of knowledge is not just technical—yes, the files might still exist, but the expertise needed to interpret them or reconstruct how decisions were made often walks out the door with the person who created them.

EPISTEMIC CULTURES IN TRANSLATIONAL RESEARCH

Collaboration provides the social scaffolding for research, yet the form it takes—and what it ultimately produces—depends fundamentally on epistemic culture. Think of epistemic cultures as the historically built-up combination of material setups, social relationships, symbolic practices, and shared assumptions that shape how different communities produce knowledge (Knorr-Cetina, 1999). Epistemic cultures encapsulate the characteristics and ways of knowing of specific research communities, spanning the objects of inquiry, methods, evidence, standards, and norms. For example, Knorr-Cetina has famously compared high-energy physics communities that rely on distributed consortia and standardised pipelines with molecular biology groups whose work is anchored in embodied skill and local know-how (Knorr-Cetina, 1999). These are not just stylistic differences; they shape what knowledge becomes possible, how reproducibility is established, and how uncertainty is handled (Leonelli, 2016).

Translational agricultural research occupies a particularly tricky position within this landscape. Field trials try to maintain experimental control while being embedded in environments that vary ecologically and socially. Hence, station-based data might meet all statistical criteria yet fail to predict on-farm performance, whereas farmer evaluations carry real practical authority but tend to lack formal publication properties. Thus, evidence moves across standards that do not quite fit together: there are ongoing tensions between demands for reproducibility and local relevance, or generalisability and situated performance. Agricultural research institutions end up functioning as what Knorr-Cetina (1982, 1995) called trans-epistemic arenas—places where scientific work intertwines with regulatory issues, economics, and politics. Research outputs must satisfy multiple audiences simultaneously—journal reviewers, policymakers, extension officers, and farmers—each using different evaluative logics. These boundary crossings shape what counts as data, how it gets formatted, and how it circulates.

DATA MOBILITY: IMMUTABLE MOBILES AND DATA JOURNEYS

If data are going to move beyond where they were created, how exactly do they travel? Here is something easy to overlook: scientific data are not naturally mobile. They become what Latour and Woolgar (1986) call 'immutable mobiles' through carefully built networks of instruments, calibration standards, protocols, and institutions. Making data mobile takes real work: standardising formats, adding metadata annotations, fitting data into ontologies, and curating everything to make it accessible (Leonelli, 2016; Leonelli, 2020). But even then, portability does not automatically mean usability. Data must be actively reinterpreted for each new setting, aligned with local questions, and integrated with site-specific knowledge. Strip away too much context, and the data become meaningless; keep too much local embedding, and they cannot generalise. Making data circulate successfully requires what you might call strategic contextualisation—finding that careful balance between abstraction and staying grounded in specifics.

And infrastructure matters enormously here. Well-supported 'big' data benefit from systemic calibration and harmonisation, while 'little' data rely on informal networks, and 'no' data risk disappearing entirely (Borgman, 2015; Edwards, 2010). At CRI, field trials produce heterogeneous data—from replicated plots to farmer observations—that must be rendered commensurable without erasing local significance. Figuring out which measures can travel intact, which metadata are indispensable, and how to keep links to tacit knowledge involves negotiations, revealing how technical, epistemic, social, and political dimensions all intersect in data work.

EPISTEMIC INJUSTICE AND RECOGNITION ECONOMIES

Beneath these technical and epistemic challenges lie questions about recognition and fairness. Data production and management are embedded in political economies of recognition, where the allocation of credit shapes whose contributions endure and whose disappear (Merton, 1968; Ross-Hellauer *et al.*, 2022). Inequalities in infrastructure, funding, and technical capacity create what philosophers call epistemic injustices (Dotson, 2014; Fricker, 2007)—situations in which certain kinds of knowledge and certain knowers are systematically marginalised. These inequalities extend beyond the visibility of published research to the often overlooked labour that makes research possible. Activities such as documenting experiments, organising datasets, maintaining records, preserving contextual information, and ensuring continuity across projects are essential to the production of reliable knowledge, yet they rarely carry the same institutional recognition as publications and scientific discoveries. When this work remains undervalued, responsibilities for data stewardship become concentrated in individual researchers while the institutional support needed to sustain them receives little attention. Decisions about curation, preservation, and visibility, therefore, shape more than the accessibility of research outputs; they also influence whose labour is recognised, what forms of expertise become durable, and which knowledge can persist within knowledge institutions.

METHOD

This paper draws on findings from CSIR-CRI, part of a broader multi-site ethnographic study of Ghana's agricultural research sector. Between August 2023 and May 2024, the first author conducted sequential ethnographic fieldwork at four institutions: the CRI, the Food Research Institute (FRI), the Science, Technology and Policy Research Institute (STEPRI), and the Centre for Agriculture and Biosciences International (CABI) West Africa, while the second author conducted fieldwork at these institutions in August 2023 and subsequently followed the study remotely. The wider project examined how diverse research environments articulate conditions under which open science may promote good research practice, yet this paper focuses exclusively on CRI. CRI is the only site where both ethnographic and survey methods were employed, enabling an integrated analysis of everyday data practices and their institutional conditions. Specifically, this strand of the project investigates how knowledge travels across the science–policy nexus through researchers' everyday work, attending to the infrastructures, people, routines, and social relations that shape data production and use. Ethical clearance was obtained from the [ANONYMISED FOR REVIEW] Ethics Committee as part of the [ANONYMISED FOR REVIEW] project.

FIELDWORK AT CRI

Fieldwork at CRI ran from late August to December 2023, approximately four days per week. Sustained engagement enabled close observation of daily scientific work and of how institutional realities shape research trajectories and practices.

ETHNOGRAPHIC OBSERVATION

Observations spanned laboratories, crop trial sites, departmental meetings, seminars, workshops, and informal spaces. Laboratory observation focused on how researchers interacted with instruments, managed protocols, recorded data, organised samples, and navigated equipment failures and resource limitations. Field visits documented how research designs engaged with farmer needs and how data travelled from paper-based field records into laboratory and digital environments, revealing points of transformation, translation, and occasional loss.

Meetings and organisational events illuminated how priorities, resources, and policies were negotiated. Conversations with library and database staff provided insight into infrastructure histories and gaps between formal systems and everyday use. Informal interactions during lunch breaks or field visits were especially important for accessing tacit norms, the affective dimensions of research, and trust-based data-sharing practices. Field notes were written during or immediately after observations and captured practices, physical conditions, social dynamics, and affects such as frustration, pride, and anxiety. Ethnography proved particularly valuable for understanding trust, reciprocity, and the social organisation of collaborative practices.

ETHNOGRAPHIC INTERVIEWS

Twenty-four semi-structured interviews were conducted with 13 research scientists, 5 technologists, 2 technicians, and 4 administrators, capturing diverse perspectives across career stages, roles, and disciplines. The interview protocol (Appendix A) explored research trajectories, methods and infrastructure, institutional constraints, ideal research environments, collaborative networks, and orientations towards open science practices. Interviews lasted between 45 and 90 minutes, were audio-recorded with consent, and transcribed verbatim. Analysis was iterative and inductive, guided by an analytic sensibility drawn from constructivist grounded theory (Charmaz, 2014). Coding proceeded alongside ongoing fieldwork and in close dialogue with field notes, enabling emerging analytic insights to shape subsequent observations and interviews through constant comparison and reflexive interpretation.

POSITIONALITY AND ETHICAL CONSIDERATION

As an embedded researcher, the first author navigated shifting insider-outsider dynamics, which facilitated access while also shaping what participants felt comfortable expressing. Reflexivity, transparency about research aims, and ongoing attention to power relations were essential throughout the research process.

DATA MANAGEMENT SURVEY

The survey[4] was created as a brief, institutionally feasible questionnaire to complement, not replace, extended ethnographic research at CRI. While 13 questions (Appendix B) may seem few, this concise format was a purposeful methodological decision influenced by practical and institutional factors. Furthermore, the survey aligned with CRI management's goal of establishing a formal data management policy. It was developed collaboratively with management and incorporated both closed and open-ended questions focused on data recording and storage practices, access to historical data, perceived difficulties, concerns about centralised infrastructure, interest in training, and basic demographic and professional details.

For analytical purposes, the questions were divided into two main sections: (1) demographic and professional positioning (Q10–Q13) and (2) data practices, infrastructures, and orientations (Q1–Q9). Some closed questions—especially Q1 and Q2—offer predefined response options. As shown in the questionnaire, respondents selected from a set of given choices, which limited the range of possible answers. The more detailed variety of data types and collection practices is examined through ethnographic observation and interviews. The survey was sent via Google Forms to 127[5] researchers between 10 October and 15 December 2023, achieving a 43% (55/127) response rate, which surpassed typical benchmarks for online surveys in organisational research (Baruch and Holtom, 2008; Nulty, 2008; Wu *et al.*, 2022).

ANALYSIS

Quantitative data were analysed descriptively, and qualitative responses were thematically coded manually to identify recurring patterns. Data visualisations were generated programmatically using Python to produce bar charts and pie charts that display both raw counts and percentages for each survey question, available in both PNG and PDF formats. Percentages do not always sum to 100% because respondents could select multiple options for some questions and provide multiple responses to open-ended items. Survey results were interpreted alongside ethnographic observations and interviews, with divergences between reported and observed practices treated as analytically meaningful rather than as error.

INTEGRATED ANALYSIS

Survey and ethnographic data were analysed in dialogue. The survey mapped the prevalence of selected practices, while ethnography revealed the mechanisms, meanings, and material conditions underlying them. For example, low reported use of institutional servers became intelligible through observations of unreliable or absent infrastructure, while widespread

4 Survey results are published on Zenodo [ANONYMYSED FOR REVIEW].

5 This was obtained from the institute's administrator.

informal data sharing was shown to operate through trust-based personal networks. This integrative approach aligns with the 'data journeys' perspective (Leonelli, 2020), which emphasises the significance of tracking the conditions under which data are transformed as they move across contexts, infrastructures, and communities of practice.

CONTEXT, SCOPE, AND LIMITATION

Although fieldwork took place at four institutions, CRI is the focus of this study because we can combine both survey and ethnographic data; it represents our longest period of immersion in the institutional culture, and it displays an actively evolving data management policy during the fieldwork. The findings reflect practices from late 2023 and may not account for future infrastructural or policy updates. Ethnographic observations were necessarily partial, and survey data may be influenced by self-selection and social desirability biases. The goal of this study is therefore not statistical generalisation but rather contextual analysis to explore mechanisms of data practice applicable to similar research contexts.

RESULTS

This section presents survey findings alongside ethnographic insights to provide a contextualised account of practices at CRI. The survey offers a broad profile of respondents and maps the prevalence of selected practices, while interviews with researchers (R), technologists/technicians (T), administrators (A), and principal investigators (PI) reveal how these practices are enacted and experienced in daily work. The section proceeds in two parts. First, descriptive statistics are presented to characterise the survey sample. Second, findings related to data recording, storage, access, challenges, and orientations toward institutional data infrastructures are discussed.

SAMPLE CHARACTERISTICS

Survey respondents represent a highly experienced and educated cohort: 63.6% (35/55) have worked at CRI for over 10 years, 58.2% (32/55) hold PhD degrees, and 69% occupy senior or principal researcher positions. This profile makes their responses particularly significant for understanding institutional memory dynamics, as these researchers are the ones whose eventual retirements pose the greatest risk to data continuity.

Senior research scientists constitute the largest group (17; 30.9%), followed by principal technologists (16; 29.1%) and research scientists (12; 21.8%). Smaller proportions include principal research scientists (5; 9.1%), assistant research scientists (3; 5.5%), and other technical staff (2; 3.6%). This distribution indicates that the sample is weighted toward mid- to senior-level personnel—those most directly involved in securing research funding, leading projects, designing experiments, overseeing trials, and producing reports for donors and policy actors, as illustrated in Figure 1.

Length of service at CRI further underscores this institutional embeddedness. Nearly two thirds of respondents (35; 63.6%) reported having worked at the institute for over ten years, while 12.7% (7) indicated between five and ten years, 20% (11) between one and five years, and only 3.6% (2) less than a year. Figure 2 visualises this distribution, suggesting that the survey largely captures the perspectives of actors with extensive experience of CRI's evolving research and data infrastructures.

Gender composition shows a predominance of male respondents (37; 67%), with women (18; 33%) forming a smaller but clearly present proportion of the sample (Figure 3). This pattern mirrors broader gendered and hierarchical structures within Ghana's public research institutions and provides important context for interpreting subsequent findings.

Educational background is similarly skewed toward advanced qualifications. Most respondents hold PhD degrees (32; 58%), followed by those with master's degrees (21; 38%). Only a very small minority reported undergraduate-level qualifications or other credentials (2; 4%). As shown in Figure 4, this high level of formal training reflects CRI's position as a national research institute and shapes expectations surrounding data production, documentation, and reporting.

Data Science Journal
DOI: 10.5334/dsj-2026-035

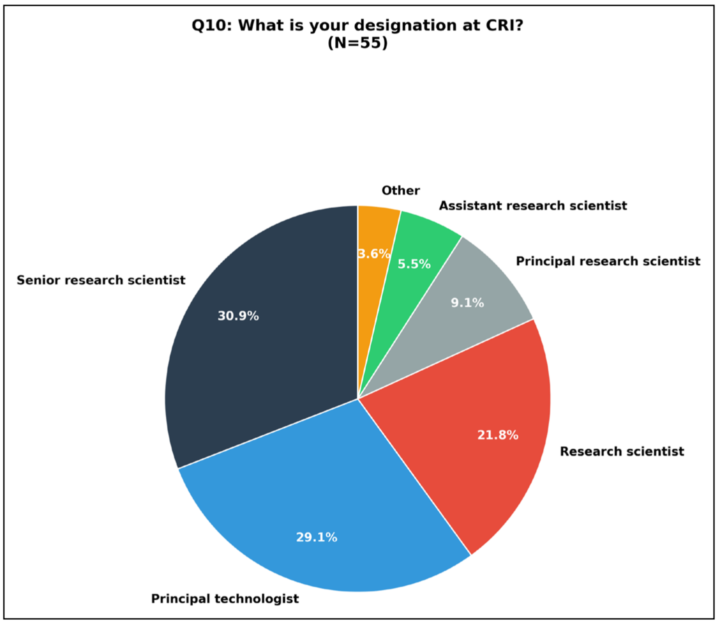

Figure 1 Participants' designation at CRI. Abbreviation: CRI, Crops Research Institute.

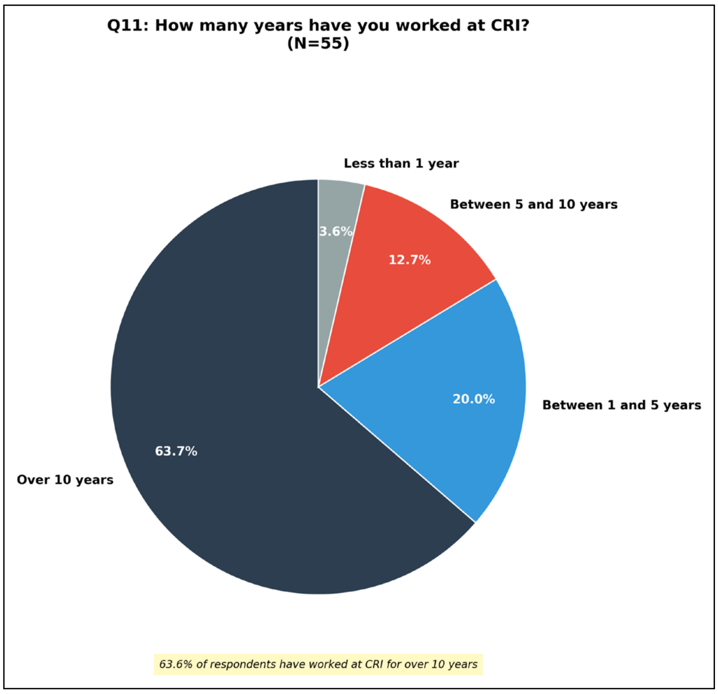

Figure 2 Participants' experience level at CRI.

Abbreviation: CRI, Crops Research Institute.

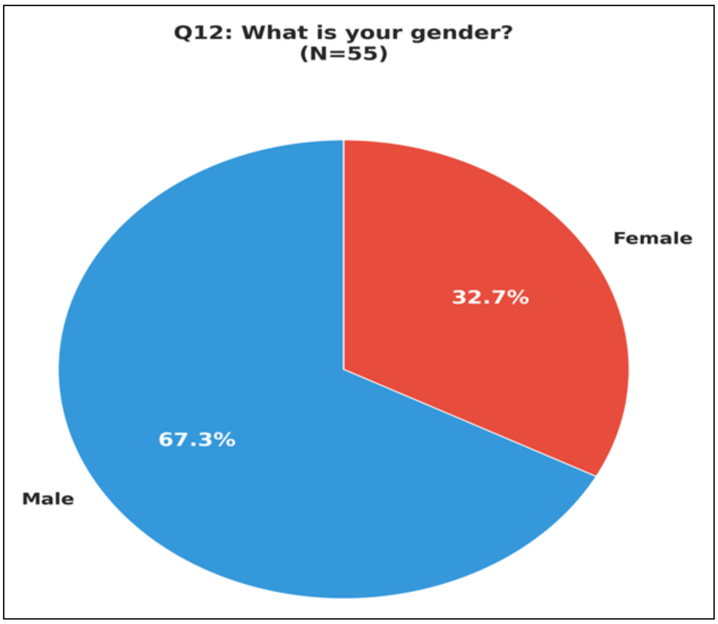

Figure 3 Gender distribution of survey participants at CRI.

Abbreviation: CRI, Crops Research Institute.

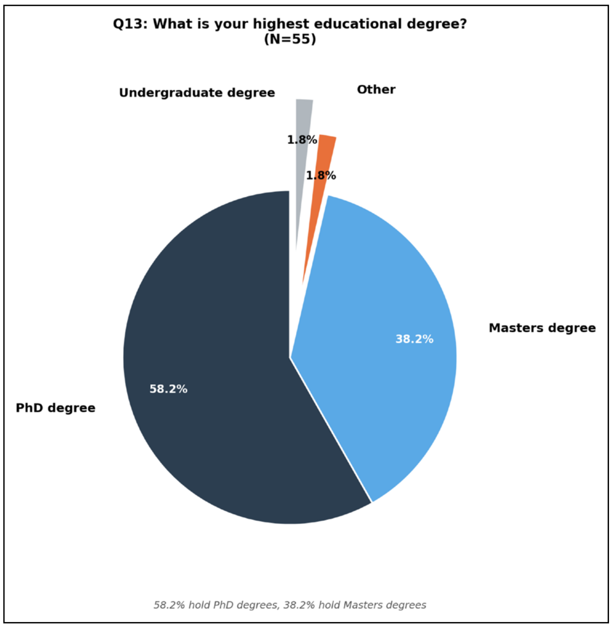

Figure 4 Educational levels of survey participants at CRI.

Abbreviation: CRI, Crops Research Institute.

SURVEY ANALYSIS

Data collection methods

Survey respondents employ multiple methods for data collection, with most using combinations of analogue and digital approaches (Figure 5). Paper notes remain the dominant method, used by 72.7% (40/55) of researchers, followed closely by Excel spreadsheets at 65.5% (36/55). Handhelds (phones or tablets) are used by 56.4% (31/55), while only 10.9% (6/55) use online data management platforms such as breeding databases.

Interview data reveals the diversity of data types collected across CRI's research programs. One researcher (R_10) described comprehensive documentation practices in participatory research.

> *As researchers, we don't take anything for granted. ... We take video footages, we take field diaries, record everything. Personally, I would say it is something I do all the time. Take pictures, evidence that this is what is being done. Record this is what I'm being told. So, even if I forget, I have my data, I can go back, play for myself ... Remember it is not going to be you sitting and doing it somewhere ... That didn't help us years back, so it has come to the point where you need to document or record everything that they [farmers] share with you because whatever you tell them, go 20 years and come back, they will have every fine detail and tell you.* (R_10)

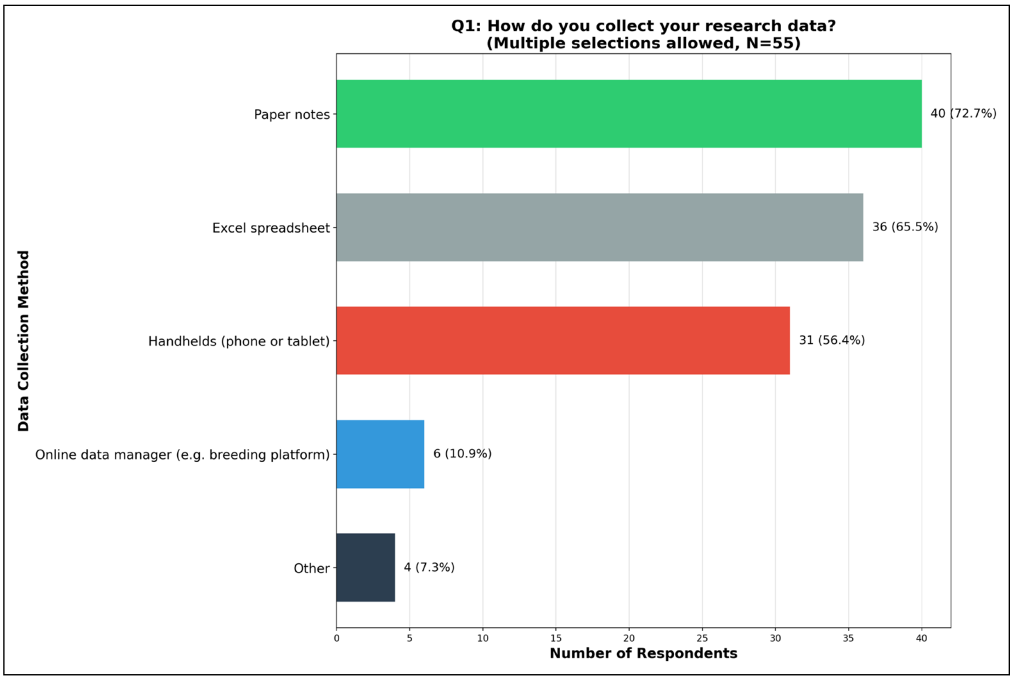

Figure 5 Data collection methods.

This suggests an institutional sensibility in which memory is understood as something made, not possessed: assembled through numbers, images, observations, and farmers' embodied knowledge. It acknowledges the fallibility of recollection—'even if I forget'—the necessity of inscription, and anticipates a future gaze, in which today's records become tomorrow's evidence—'go 20 years and come back, they will have every fine detail'.

Despite decades of digital technology availability, the continued use of paper-based data collection is not due to technological conservatism but results from adaptation to working conditions. Field trials across various agroecological zones require methods that are reliable without electricity or internet access, and the environment often renders digital tools impractical.[6] During participatory varietal selection (PVS) trials, the first author noticed that researchers were recording measurements in field notebooks rather than on tablets. Likewise, in the biotechnology laboratory, technologists and technicians kept personal notebooks for data recording.

> *[For] documentation, the lab has a desktop, aside [from] that we have laptops, those are ways that we document, we also have lab notebooks that we use and our own personnel books too.* (T_4)

This creates what one respondent described as 'doubling the workload'—data must be recorded on paper in the field (or lab), then manually entered into digital formats later, introducing opportunities for transcription errors and consuming time that could be spent on analysis. Data collection, however, is not limited to paper, as shown by survey responses. Some programmes, especially donor-supported breeding initiatives, have researchers using tablets and digital platforms to design trials, record measurements, store data, and analyse information.

> *For data collection, we have a digital system [for] collecting our data. EIB [Excellence in Breeding] has provided us with some tablets that we use for data collection, and then we've had some training with IITA on how to use [the] system, the cassava breeding base to design even the field trial before you go to the field. You can design it on the system, collect data using the machine and store it for long-term storage in the system, and it has some friendly, user-friendly facilities for analysing some of this data.* (T_7)

A molecular biology technologist (T_4) described specialised data generated in her laboratory.

> *We do marker-assisted selection—when breeders have a gene of interest, they've been able to do the crosses. So, the offspring, they bring it to us to screen ....[for] the gene of interest... we do disease diagnostics of our mandate crops [such as] cassava mosaic virus and ... yam mottle virus..... We do polymerase chain reaction (PCR), ... reverse transcription PCR as well. We also do qPCR for GMO detection in some grains and end products for some companies.* (T_4)

The technical complexity of molecular data generated—gel electrophoresis images, disease diagnostic reports, PCR results, sequence data—combined with other forms of data create distinct storage requirements and domain-specific interpretive frameworks. The diversity of data types—field measurements, molecular sequences, videos, farmer interviews—creates challenges for standardised institutional data management systems.

Data storage: the personal laptop as default infrastructure

Survey responses reveal almost universal reliance on personal laptops for data storage (Figure 6). Fully 98.2% (54/55) of respondents store research data on their own laptops, with only one researcher reporting use of an institutional repository. Other storage methods include paper files in offices (52.7%, 29/55), reports compiled at research completion (47.3%, 26/55), and institute desktop computers (10.9%, 6/55).

6 PVS trials are typically conducted on farmers' fields or remote locations where internet connectivity is absent or unreliable. Environmental conditions further constrain digital use, as working under direct sunlight can render tablet screens difficult to read due to glare.

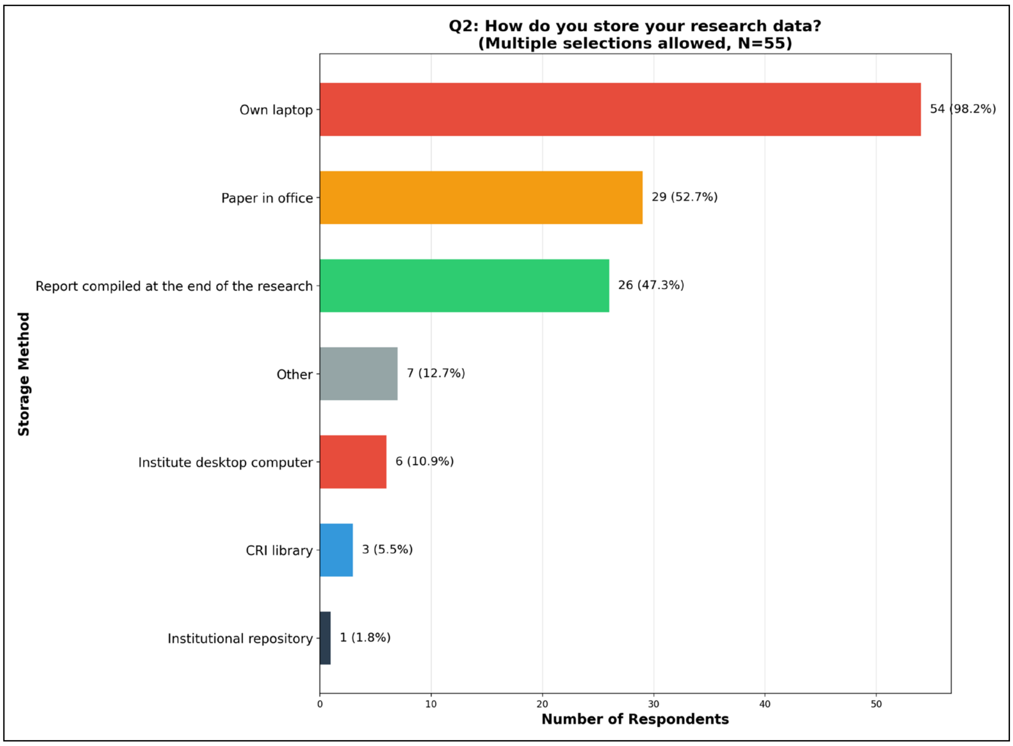

Figure 6 Data storage methods.

Interview data provides context for why personal storage dominates. A researcher (R_11) explained:

> *Unfortunately, we don't have a data policy at the moment here because, you know, elsewhere, there is a portal,... a repository so that not even you alone can have access. Others in the institute can have access, ..., but unfortunately, we don't have such at the moment. It's now that the institute is working on this. Hopefully, from next year, it will be implemented. So, it's on my personal hard drives and computers. Apart from the very few members of the team that I'm working with, nobody else has access to it. And the institute has really realised that it is a big challenge and has made the effort to make sure there will be a repository that we can all save our data and information.* (R_11)

R_11's account illustrates how researchers are aware of the distinction between individual storage arrangements and institutional data infrastructure. His reference to 'a portal … a repository' describes a system through which research data would become accessible beyond individual researchers and research teams, while also providing a mechanism for preserving and archiving research materials over time. This understanding closely corresponds to the definition of a repository above. The challenge identified by R_11 is therefore not a lack of understanding about what a repository should provide, but rather the absence of such infrastructure within the institute. Until this infrastructure is established, researchers continue to rely on personal devices and individually managed storage systems to sustain ongoing research activities, leaving the long-term preservation of research records dependent on personal rather than institutional practices. A technician (T_2) described the current workaround:

> *Most of our documentation [is] through paper documentation, and then we enter data into Excel on our machine for analysis, … then we send it to the researcher in charge of the activity … we use the Google Cloud System to store some of the data. Yes, it's a personnel [Google Cloud System].* (T_2)

T_2's account describes an adaptive response to the absence of shared institutional infrastructure. Using a personal Google Drive account provides a practical way to store data. However, it does not constitute a repository in the sense used in this study. Unlike a managed repository, personal cloud storage remains attached to an individual account and does not necessarily provide the metadata standards, documentation practices, governance arrangements, or long-term preservation mechanisms required for institutional stewardship. T_2 is therefore not describing Google Drive as a repository; rather, he is describing one of the practical solutions researchers have adopted to maintain access to research materials in the absence of a formal institutional system.

The widespread reliance on personal laptops, external drives, and individual cloud accounts therefore reveals how data infrastructure at CRI has become personalised in practice. These tools enable researchers to continue producing and managing data, but they transfer responsibility for preservation and continuity from the institution to individual researchers. Personal devices function as the *de facto* infrastructure of everyday research—not because they were designed to serve institutional archival purposes, but because researchers have adapted them to fill an infrastructural gap. While these arrangements support immediate research needs, they also create vulnerabilities for organisational memory related to potential staff changes or hardware failures.

Ideal versus actual storage: the infrastructure gap

Respondents expressed clear preferences for research data storage that sharply contrast with current practices (Figure 7). The majority (58.2%, 32/55) preferred institutional storage solutions including institutional repositories or CRI-managed storage systems, while 25.5% (14/55) favoured cloud storage or personal devices. This discrepancy between normative commitments (institutional storage) and practical realities (personal laptop storage) indicates not a lack of awareness of best practices but the absence of functional infrastructure to support their implementation. Researchers recognise the value of moving beyond individually controlled storage towards systems that enable collective access, preservation, and continuity, yet everyday research practices remain dependent on personal devices and individually managed storage arrangements.

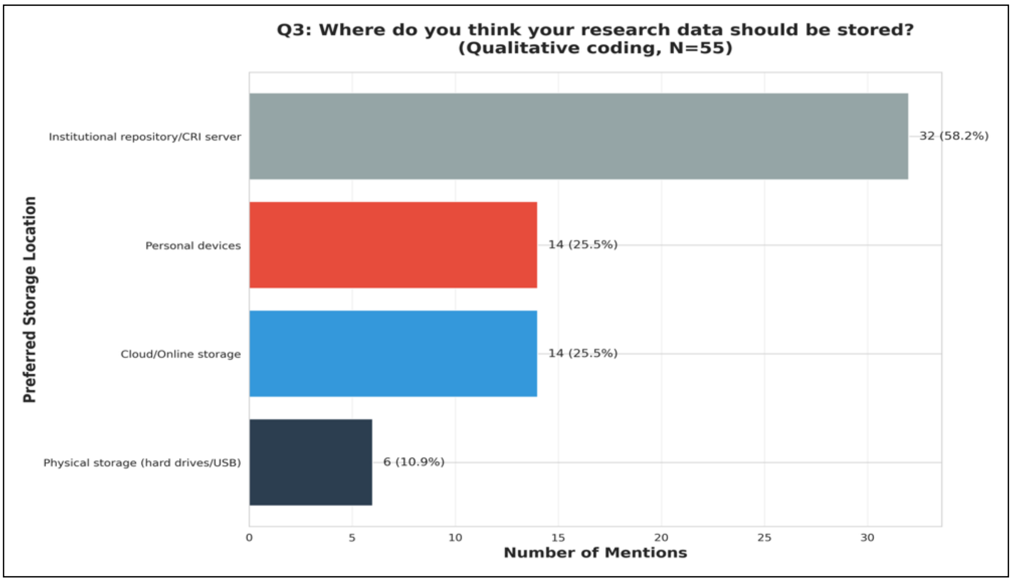

Figure 7 Preferred data storage location.

During an interview about the possibility of establishing an institutional repository, T_2 responded as follows:

> *That's a good idea. Because with that, you'll be able to get the data wherever you are. And then in case you lose your laptop or lose the paperwork, you'll be able to go there and then pick the data. I think that's quite a very good thing.* (T_2)

T_2's response illustrates that researchers associate institutional repositories with more than simply additional storage capacity. His emphasis on being able to retrieve data after losing a laptop or paper records reflects concerns about preservation, accessibility, and continuity beyond the individual researcher. In this sense, the repository is understood as a mechanism through which research data can remain available despite the fragility of personal storage arrangements. Similarly, an administrator (A_1), responsible for knowledge dissemination, described both the potential benefits of digitisation and the challenges created by the absence of reliable institutional systems.

> *For example, if we want information on a particular commodity, … you need to go to the scientists before you get that data. I think that will be helpful because breeders or scientists might go on retirement, and when you are looking for that information, it will be difficult if it is not in the library repository … I think digitisation will help our work. I mean, when you have access to the internet and then access to information … So, if I get money, I will digitise this office so that staff can work effectively and efficiently*

> *... Sometimes getting access, a lack of equipment, our printer got spoiled, but now it is back on track. Sometimes the machines that we used to work like the laptop, it becomes outdated and becomes difficult to get ...* (A_1)

A_1's account further highlights that the concern is not simply where data are stored, but whether institutional knowledge can remain accessible over time—thereby showing the deep link between everyday data practices and long-term expectations and strategies. Rather than reflecting uncertainty about what repositories should provide, these accounts demonstrate an awareness of their intended functions: data security, institutional accessibility beyond individual possession, continuity across personnel changes, asynchronous collaboration, and accumulation of organisational memory. However, this understanding coexists with chronic infrastructural fragility. While respondents articulate expectations for institutional systems, the material conditions needed to sustain them remain limited. When fundamental office equipment cannot be reliably maintained, fully functioning digital repositories become aspirational rather than actionable. The result is a gap between what researchers recognise as necessary for responsible data stewardship and what institutional conditions allow them to implement. This disjuncture produces a form of epistemic precarity: recognition of what good practice requires without the capacity to enact it.

Data stewardship within a fragile research infrastructure

Survey responses identified two primary challenges related to data management: data loss and security, and data access and retrieval, each reported by 20 (36.4%) respondents (Figure 8). However, interviews situated these concerns within a broader infrastructural environment that extends beyond data storage and sharing alone. Researchers' accounts revealed that the ability to preserve, access, and transfer research knowledge is shaped by broader institutional capacity conditions, including equipment maintenance, electricity reliability, project sustainability, and the continuity of technical expertise. These factors do not represent data management failures in themselves; rather, they constitute the infrastructural conditions within which data stewardship practices become possible—or constrained.

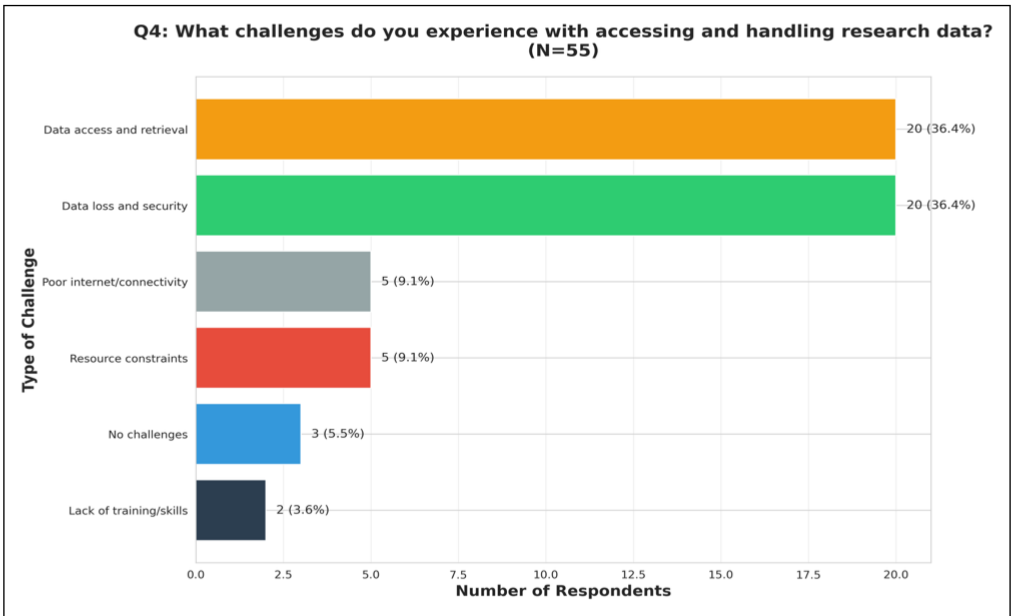

Figure 8 Data management challenges.

A researcher (R_6) described the challenges created by fragile research infrastructure:

> *Some of the chemicals we use are toxic, and so they have been classified as hazardous chemicals. For those ones, the freight charges are higher than the cost of the chemical. And because of that, sometimes it delays and these samples are temperature sensitive. So, if you're not careful, by the time they release it to you, it has gone bad. Again, you know, with the equipment we use, ... a cryogenic freezer, it can go as low as minus 196 degrees Celsius but the amount of power it consumes..... And even with tissue culture, the lights, the air-condition is on 24/7. And the Institute has to pay for electricity. And [a] few months ago, ECG cut off our power. And because of that we have to put off the cryogenic freezer, because we are not getting money to give to the institute to make payments.* (R_6)

This account illustrates the infrastructural conditions that precede and shape data production. Before questions of storage, sharing, or repository access arise, research materials and experimental processes depend on reliable systems of electricity, equipment maintenance, and operational support. When these systems become unstable, the consequences are not only immediate disruptions to ongoing experiments but also limitations on the kinds of knowledge that can be produced and subsequently documented. Data stewardship therefore cannot be separated from the material conditions that make research activities possible in the first place. The same researcher described how externally funded equipment can become difficult to sustain after project completion:

> *This building was provided .... by the one project: West Africa Agricultural Productivity Programme. It will be ten years next year. But you see, have you seen the defects of this building? And when this building was built, most of the equipment, the project purchased it. A DNA analyser, which is a sequencer. It is the old model though ... it came in at a time when the project was almost getting toward the end ... it has not been fully calibrated. And we don't even have the capacity to use it. At that time that it was bought, it was about half a million dollars. Now ... its outmoded and there are new sequencing machines, next generation sequencing machines.* (R_6)

This example illustrates how project-based investments may generate new research capacities without necessarily creating the institutional arrangements required to sustain them. The challenge is not that a piece of equipment becomes obsolete; rather, technological investments can become disconnected from longer-term processes of training, maintenance, calibration, and institutional planning. As a result, potential research trajectories may remain unrealised. The knowledge that could have been generated by such equipment is never produced because the infrastructure required to produce it cannot be sustained. Alongside these material constraints, researchers also highlighted the fragility of institutional memory and the risk that knowledge accumulated through years of practice may leave with individuals. A researcher (R_10) reflected on the consequences of personnel transitions:

> *Over the last few decades, I've learned on the job, ... from senior colleagues, ... from endowed sources. So, I have that repository of knowledge and skill to do certain things that the young one cannot do ... If you have trained me over these 20 years, and I'm not happy about something and I leave, yes, you can replace me but I go with great experience ... But if there is this succession plan like this young lady who came to me, she has joined me for the past two years ... by the time I leave, at least she is somewhere.* (R_10)

R_10's account shifts attention from physical infrastructure to the human infrastructure through which research knowledge is accumulated and transferred. In this setting, expertise functions as a form of organisational memory embedded in individuals and transmitted through apprenticeship, collaboration, and prolonged engagement. This quote resonates with Walsh and Ungson's (1991) notion of organisational memory as distributed across multiple 'storage bins', including individuals and organisational structures. When tacit knowledge and research histories are not systematically documented or retained within institutional infrastructures, staff departures can create discontinuities in organisational memory.

While positions can be refilled, accumulated experience cannot. In the absence of robust documentation and institutional repositories, such departures translate directly into institutional amnesia, as decades of tacit and experiential knowledge leave with the individual. These accounts demonstrate that challenges around data stewardship at CRI cannot be understood solely as technical problems of storage or access. They emerge from a wider infrastructural landscape in which research production, documentation, preservation, and knowledge transfer are interconnected. The absence of robust data infrastructures is therefore part of a broader institutional condition in which researchers recognise the importance of maintaining collective knowledge but must often rely on individual practices and improvised solutions to sustain it.

Accessing one's own historical data: the erosion of personal archives

Survey responses show that even accessing one's own historical data presents challenges. Fully 40% (22/55) of respondents find accessing their own past data difficult (Figure 9).

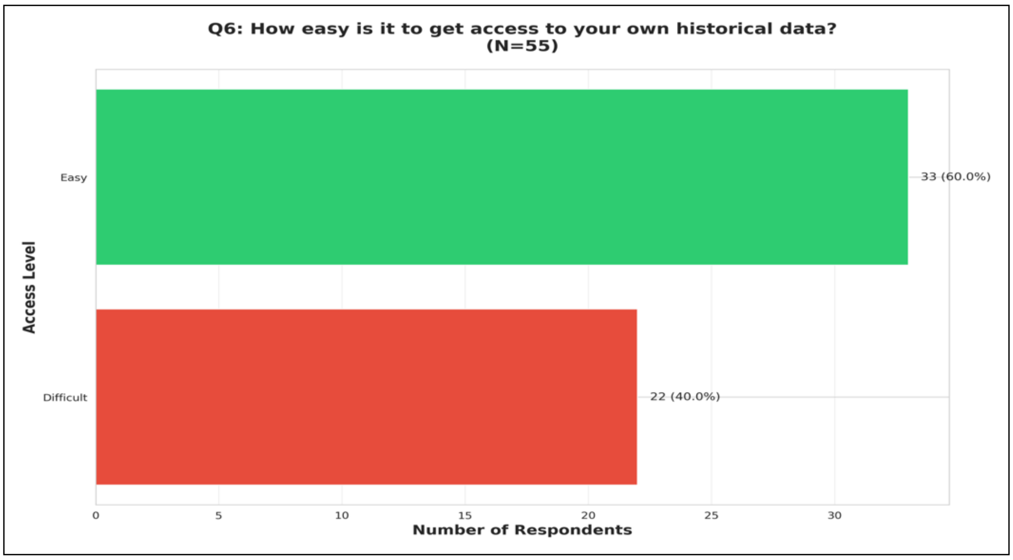

Figure 9 Accessing primary historical data.

While not explicitly stated in the interviews, the survey qualitative responses documented researchers explaining: 'poor data management on my part; I can't even remember where to find what, let alone talk of details and observations relating to the data'; 'Relatively easy. Data are stored in files and ordered chronologically on the cloud', but with caveats about 'if laptop crushes or external drive not compatible with current version laptop'; 'Easily accessible from own laptop but sometimes data cannot be found, either file corrupted or laptop runs into problems'. Even researchers who maintain relatively organised personal systems recognise their fragility. File formats change, storage media become incompatible, organisational schemes are forgotten, and personal memory of data provenance fades. The fact that nearly two thirds of respondents struggle to access their own data indicates that individual memory systems—no matter how well-intentioned—cannot substitute for institutional infrastructure.

Accessing historical data from other scientists: institutional amnesia

Access to colleagues' historical data presents even greater challenges. As shown in Figure 10, a substantial majority of respondents—69.1% (38/55)—reported that accessing other researchers' data is *difficult*, *very difficult*, or *impossible*. Even among those who described access as relatively easy, this ease is qualified: 29.1% (16/55), indicative that obtaining colleagues' data typically requires considerable personal effort.

The survey qualitative responses captured the emblematic phrase: 'Very difficult. Most Scientists retire with their data', or 'Extremely difficult if you're not friends. People don't easily share their data'. Interview data confirm and extend this finding. The R_11 quote above, 'Apart from the very few members of the team that I'm working with, there's nobody else that can have access to it', reveals that even current colleagues cannot access data without personal negotiation.

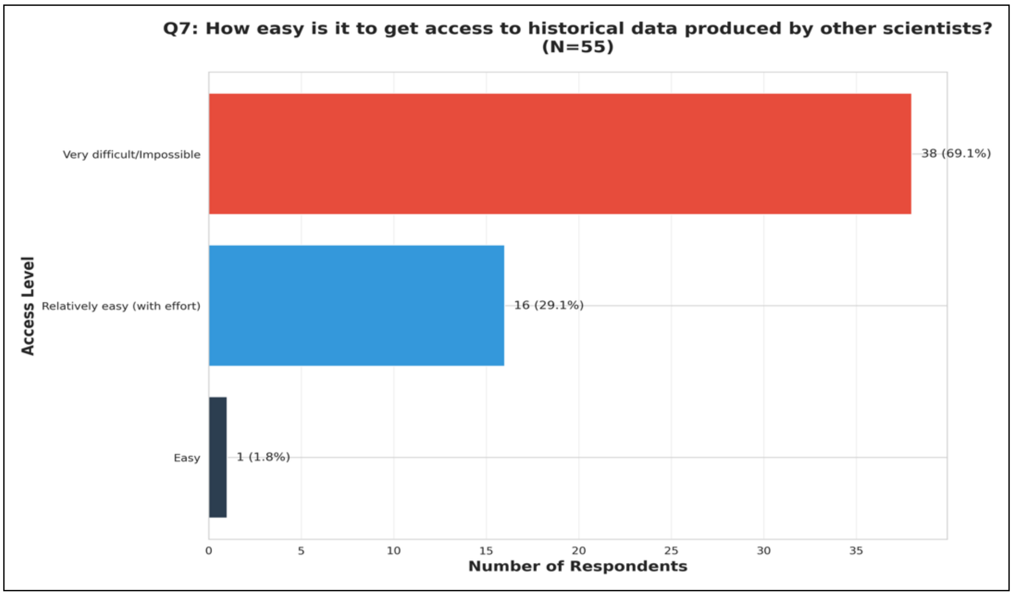

Figure 10 Access to secondary historical data.

This can lead to institutional amnesia. Each generation of researchers must rediscover findings, repeat experiments, and reconstruct knowledge that previous generations already generated.

Support needs

When asked what support would be most helpful for managing and analysing data, researchers identified multiple needs (Figure 11). Storage infrastructure and equipment (47.3%, 26/55) were most frequently mentioned, followed by training and capacity building (34.5%, 19/55), and institutional support (12.7%, 7/55).

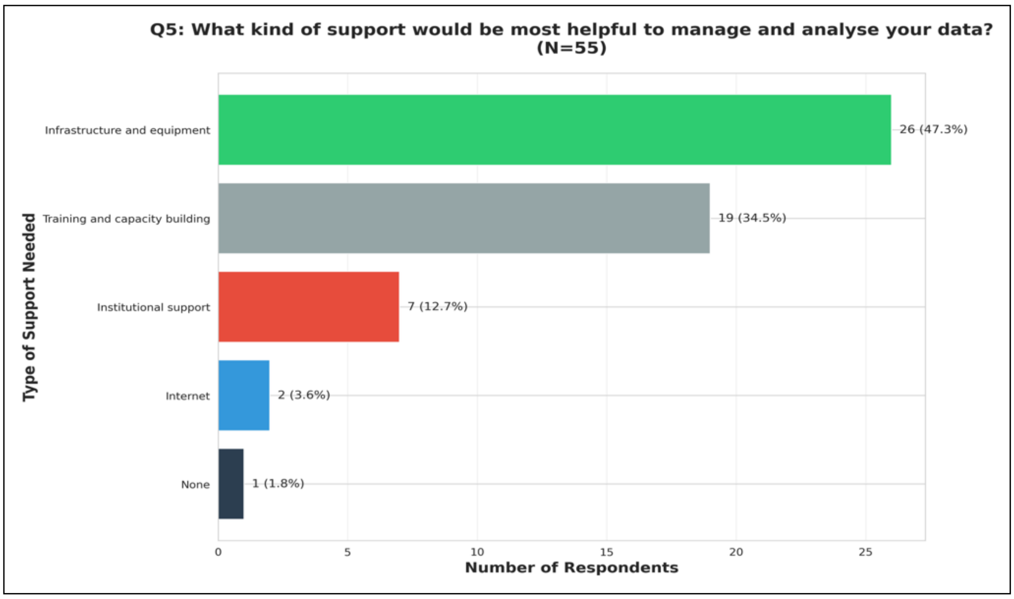

Figure 11 Support Needs.

Interview data reveal that these needs are interrelated rather than independent. R_11 explained equipment and expertise requirements:

> *I can practically do everything … in the field of biosciences. But obviously, …most of the facilities are not here. The transition hasn't been easy at all, because there is a lot of equipment you may need, and then they are not available … for the things that we are doing here: … nutritional analysis, you know, spectrophotometer, HPLC, yes, near-infrared spectroscopy equipment is there. And then the basic laboratory appliances, they're available to carry out, … but obviously, there are times you may need mass spectrometry-based equipment, which we don't have. That is a challenge and because chemicals and reagents are also very, very expensive to come by; and even when the money is there, the procurement process at times will take a long, long, long, long time before you can get access to them.* (R_11)

Support thus encompasses equipment availability, software tools, consumable procurement, and technical training, all of which must function together. Having training without equipment is useless; having equipment without training is equally so. The mention of procurement delays highlights how administrative infrastructure also affects data work: when obtaining reagents takes months, experimental timelines extend, data generation slows, and research productivity suffers. R_6 described the relationship between infrastructure and institutional standards:

> *For me, what I learned through it [ISO certification] is a very good system. And because everything is documented, … it helps you in your day-to-day activities. But then, the costs associated with it should be something that you should be able to generate money from so that you can sustain yourself. Right? Because, for example, you're supposed to do daily calibration, [which] comes with costs, about 10,000 Ghana cedis a year. The certificate itself, you have to renew it every year, and then the amount of money that goes into it. So, … you should be able to get enough money from … whatever you are conducting to sustain it. It came from a project, so when the project was no longer there, it was very difficult to sustain it.* (R_6)

This captures a critical tension: researchers recognise the value of formal quality management systems but cannot sustain them when project funding ends. The annual costs of maintaining standards cannot be absorbed by institutional budgets, creating a cycle in which external projects temporarily introduce best practices, but these practices disappear when projects end because the ongoing costs are unsustainable.

Universal interest in training

When asked directly whether they would be interested in data management training, 100% (55/55) of respondents answered affirmatively (Figure 12).

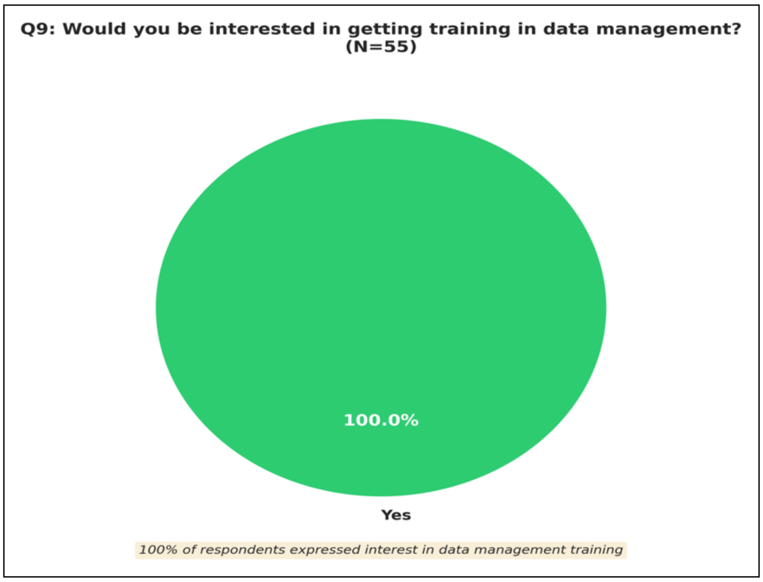

Figure 12 Data management training needs.

The universal training interest must be interpreted in the context of respondents' higher education levels (96.4% with advanced degrees) and extensive experience (63.6% with over 10 years at CRI). These are not early-career researchers seeking basic skills but senior researchers recognising that data management requires specialised knowledge not acquired through disciplinary training.

Attitudes towards centralised data storage: between aspiration and anxiety

When asked about anxieties surrounding centralised, institutional-level data storage, responses were mixed and nuanced (Figure 13). Nearly a third of respondents (30.9%, 17/55) reported no anxiety and expressed general support for institutional repositories. At the same time, 27.3% (15/55) articulated conditional support, typically dependent on how such systems would be designed and governed. Substantive concerns centred on issues of access and reliability (25.5%, 14/55) and security and privacy (25.5%, 14/55), indicating apprehension about both the technical robustness of repositories and the protection of sensitive data. Smaller proportions raised concerns about loss of control or ownership (5.5%, 3/55) and about trust or potential misuse (5.5%, 3/55).

Anxieties about attribution were also prominent. As one researcher noted, 'I support that. My only fear is some scientists publishing my data without my consent', while another emphasised concerns about 'security, third-party usage and acknowledgement'. These statements reflect legitimate fears of appropriation in contexts where career progression is closely tied to publication outputs and where power asymmetries shape who ultimately receives credit. Attribution disputes already occur even without centralised data systems. R_10 described how CRI-developed varieties are disseminated under government ministry branding:

> *We will develop varieties, and all that you hear when you disseminate is 'agricfo abro' [MOFA maize], 'agricfo bankye' [MOFA cassava]. You don't even hear CSIR-CRI unless you're interrogating documents covering maybe that technology before you appreciate that, oh, some people generated it some way, and others too are disseminating it as development.*

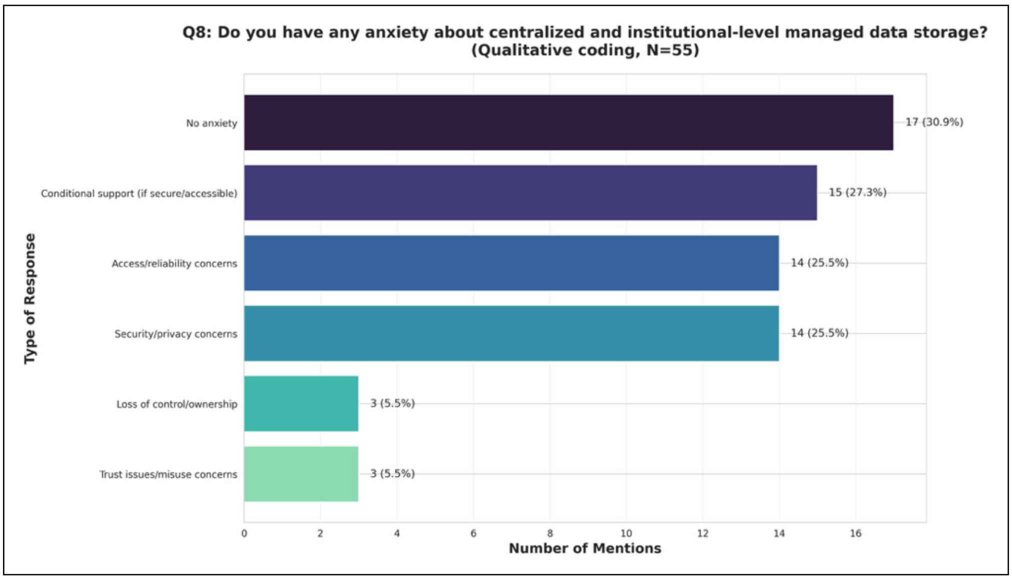

Figure 13 Anxiety towards centralised and institutionalised managed data.

R_10 also recalled the rollout of orange-fleshed sweet potato in partnership with World Vision:

> *The documents did not even mention our name. For me? Um, I'm not hurt. That's me. I say so because I am a patriotic Ghanaian. Whatever effort I put in, and it culminates into Ghana's benefits or developments, that is my quota or that is my contribution to Ghana's development.*

Such concerns are particularly salient within collaborative arrangements that span institutional hierarchies and Global North–South research partnerships, where histories of extractive knowledge practices render data sharing an ethically and politically charged proposition rather than a purely technical one. Additionally, R_6's discussion of ISO certification and GMO testing reveals another dimension: institutional repositories require governance, not just technology. When asked why the ISO certification lapsed, the response was about financial sustainability. Similarly, for repositories to function, they must be resourced, maintained, and governed through clear policies about access, security, attribution, and use rights. Without these elements, centralisation could create new problems rather than solve existing ones.

DISCUSSION

This study examined data management practices at the CRI through a survey of 55 researchers and 11 semi-structured interviews, situating findings within frameworks of institutional memory, continuity, epistemic cultures, and data mobility. What emerges is a picture of translational agricultural research operating under data precarity—knowledge production continues, yet systematic preservation consistently fails. This highlights distinctive challenges that extend existing theoretical frameworks in new directions.

DATA CONTINUITY AS PERPETUAL CRISIS

We know that data continuity requires coordinated effort across technical, social, organisational, and political dimensions (Borgman, 2015; Knorr-Cetina, 1999; Leonelli, 2016). But what happens when all these dimensions falter at once? At CRI, nearly all researchers (98.2%) rely on personal laptops, and most report extreme difficulty accessing colleagues' historical data (69.1%). This creates conditions where institutional memory dissipates with personnel transitions. Walsh and Ungson's (1991) model of organisational memory assumes that storage persists even when access breaks down—you might not be able to find something, but at least it is still there somewhere. CRI demonstrates something more fundamental: storage itself disappears. As one senior researcher put it, 'if you have trained me over these 20 years … I go with great experience' (R_10). You can fill the position again, but what about the knowledge embodied in two decades of experience? That knowledge walks out the door together with the researcher who embodies it.

While Cragin *et al.* (2010) document occasional laboratory-level data loss, CRI shows these dynamics playing out at the institutional scale. And here is what makes it particularly striking:

63.6% of researchers have over 10 years' experience. This represents ongoing haemorrhaging of accumulated knowledge. Reichmann's (2023) observation that high staff turnover impedes research data management and sharing, linked to increasing academic precarity, offers one potential explanation. But CRI presents a somewhat different dynamic. Many researchers have long tenure at the institution, yet knowledge loss still occurs through retirements, deaths, and the gradual erosion of organisational memory in the absence of any systematic way to capture and transfer what people know. The precarity here manifests less as rapid turnover and more as structural inability to preserve knowledge even when staff stay put. When researchers say 'scientists retire with their data'—and they say it repeatedly—they are describing systemic failure, not occasional lapses.

FRUSTRATED EPISTEMIC CULTURES

Epistemic cultures shape how communities produce, curate, validate, and interpret knowledge (Knorr-Cetina, 1999). At CRI, researchers negotiate between experimental rigour (field trials, molecular diagnostics) and applied outcomes (farmer engagement, policy recommendations). They generate heterogeneous data, including field measurements, molecular sequences, soil analyses, interviews, and video documentation, each requiring specialised storage and ways of interpreting it. One researcher emphasised just how comprehensive their documentation efforts are: 'We take video footage, field diaries, record everything … Take pictures, evidence that this is what is being done' (R_10).

Yet these epistemic cultures are frustrated. Here is a telling contradiction: 58.2% of respondents say they support institutional repositories and yet 98.2% rely on personal devices. Interest in data management training is unanimous (100%), showing that researchers absolutely recognise best practices; they just lack the infrastructure capacity to implement them (R_11). CRI functions as a trans-epistemic arena (Knorr-Cetina, 1982; Knorr-Cetina, 1995), where research simultaneously addresses scientific standards, regulatory requirements, policy relevance, and farmer adoption, each of which involves different evaluation criteria. And then there are attribution disputes, such as when CRI-developed varieties are credited to 'agric' rather than to CRI, a clear example of epistemic injustice (R_10). These tensions arise from dual accountability operating in a context where infrastructure fails to adequately meet any of these competing demands. The gap between what researchers know should be done and what they can do reflects not individual failure, but what Reichmann (2023) identifies as the material conditions shaping data work—in this case, conditions characterised by infrastructural absence rather than precarious employment.

DISRUPTED DATA MOBILITY

Latour and Woolgar (1986) famously describe scientific data as immutable mobiles—inscriptions that can travel across contexts while retaining sufficient stability to support scientific claims. At CRI, however, data journeys are frequently interrupted. Survey responses indicate that 40% of researchers experience difficulty accessing their own historical data, while access to colleagues' datasets presents an even greater challenge, with 69.1% reporting difficulties. Practical reasons behind these challenges include changing file formats, obsolete storage media, fragmented storage locations, and the gradual loss of knowledge of how datasets were produced and organised. Researchers rely on publications and final reports when original datasets become difficult to locate. These outputs preserve important findings; nonetheless, they represent only a selected layer of the research process. Field observations, methodological adjustments, experimental decisions, and contextual information that shaped the original dataset frequently remain outside the published record. Over time, this produces a form of data sedimentation: successive layers of processed and summarised knowledge accumulate on top of earlier research records. The accumulated layers preserve traces of previous scientific activity while making the original circumstances of data production increasingly difficult to recover. As with geological strata, recovering what lies beneath becomes harder with each layer that settles.

This process affects the mobility of data. Leonelli's (2016, 2020) conceptualisation of data as context-dependent emphasises the labour required to make datasets travel across research settings. Data become useable elsewhere through documentation, standardisation, metadata annotation, and practices that preserve the connections between datasets and the conditions of their creation. At CRI, these practices are shaped by infrastructure constraints. Researchers described situations involving unreliable electricity, interrupted laboratory operations, ageing equipment, delayed research materials, and limited technical support (R_6, R_11). These conditions influence the continuity of research records and the institutional capacity to maintain the contextual information required for future reuse.

The scale of CRI's research production further highlights this tension. The institute generates long-term datasets through breeding programmes, field trials, and laboratory research that may hold significance beyond individual projects. Yet these datasets often remain embedded within individual researchers' working practices rather than within institutional systems designed for preservation and reuse. In Borgman's (2015) terms, CRI produces data with characteristics associated with 'big data' in terms of accumulated volume, duration, and scientific importance, while managing them through practices closer to 'little data' environments where responsibility for organisation and continuity rests heavily on individuals. The result is disrupted data mobility. Research data may exist, but their movement across time, researchers, and institutional boundaries becomes increasingly difficult. As contextual layers accumulate and connections between datasets and their origins weaken, future researchers encounter records that require significant reconstruction before they can become meaningful again.

INFRASTRUCTURE TIME AND MAINTENANCE CHALLENGES

Sustaining reliable data systems requires long-term investment and ongoing maintenance labour. Karasti *et al.* (2010) describe this temporal dimension as 'infrastructure time', while Ribes and Jackson (2013) show how infrastructures depend on continuous repair work that often remains invisible until systems begin to fail. At CRI, the challenge is shaped by the limited development of institutional data infrastructure. Only one researcher reported using an institutional repository, suggesting that such a system is either unavailable or inaccessible as part of everyday research practice. Survey responses identified data loss and retrieval difficulties (36.4%), poor connectivity (9.1%), and limited resources (9.1%) as key challenges. Interviews situated these concerns within a wider environment where research systems depend heavily on project funding, individual effort, and improvised solutions (R_6, T_4).

Here is a particularly vicious cycle: project-dependent infrastructure, such as ISO certification or equipment calibration, might be feasible during project timelines but collapses afterwards. So instead of gradual development, you find cycles of partial implementation followed by rapid decay (R_6). Build something up, watch it fall apart, and start over. A similar pattern appears in relation to data work. Data management responsibilities are largely incorporated into researchers' existing roles rather than supported through dedicated professional positions. Researchers balance data curation alongside their scientific responsibilities. When experienced researchers retire or leave, the institution loses both accumulated knowledge and the informal systems through which research records have been maintained. Without established procedures for transferring responsibility, continuity depends heavily on individual efforts.

It is important to recognise that some practices observed at CRI should not be interpreted as evidence of infrastructural deficiency in themselves. As discussed earlier, the continued use of paper-based recording reflects researchers' adaptation to the practical conditions of knowledge production. Similar forms of situated record-keeping exist across scientific fields, where researchers develop working practices around the demands of particular research environments (Borgman, 2015; Bowker, 2005). Similarly, research institutions across different national contexts continue to face challenges in preserving access to data after researchers retire or move institutions, partly because dedicated data stewardship capacity remains unevenly developed (Vines *et al.*, 2014).

The distinction at CRI lies in the institutional conditions surrounding these practices. In many well-resourced settings, individual workarounds coexist with additional layers of support that help maintain continuity, including research data services, institutional repositories, professional library expertise, and established procedures for transferring knowledge across personnel changes. At CRI, comparable practices occur within an environment where these buffering mechanisms are limited. The issue is therefore not that CRI researchers engage in unusual data practices; rather, ordinary vulnerabilities become amplified when they are not supported by redundant institutional infrastructures.

RECOGNITION ECONOMIES AND EPISTEMIC JUSTICE

Credit allocation in data management is not just about fairness to individual contributors; it reflects broader epistemic and political-economic dynamics. Merton's (1968) seminal work on the Matthew effect and the latest research by Ross-Hellauer *et al.* (2022) show, among other studies, that well-resourced actors navigate mandates far more easily than those without resources. Our survey data reveal something interesting: 20% of respondents expressed security concerns about centralised storage, while 5.5% expressed concern about attribution. Personal storage might be dysfunctional for institutional continuity, but it protects individual contributions in the absence of governance mechanisms.

Fricker's (2007) and Dotson's (2014) frameworks help us see forms of epistemic injustice operating here: contributory injustice, in the form of lack of recognition for individual contributions, appears when outputs get misattributed (R_10), while distributional injustice, in the form of insufficient ability to take advantage of existing resources, arises when PhD researchers generate scientifically valuable data yet lack basic infrastructure that Global North institutions take for granted. Such injustice generates a frustration so profound as to challenge institutional legitimacy; some researchers even suggested dismantling the institute entirely if proper support cannot be provided for their work and data-sharing requirements (R_10). Meanwhile, donor-driven, short-term funding continues to shape research priorities, in some cases inhibiting—rather than sustaining—long-term capacity building and infrastructure development (Bezuidenhout *et al.*, 2017a).

IMPLICATIONS: STRUCTURAL INTERVENTIONS

When formal policies or infrastructure is absent, researchers develop workarounds—selective documentation and informal sharing networks (Neff *et al.*, 2017). But the CRI case demonstrates that training alone cannot address structural gaps. You cannot train your way out of having no electricity, and institutional infrastructures, credit systems, and clear acknowledgement of the significance of responsible practice are required to support the activities of individual researchers or research groups (Odebrecht, 2025). Interventions must operate across multiple, interdependent levels:

- Technical: There is a need for reliable power and connectivity, well-maintained equipment, functional procurement systems that consider open-source options, and robust repositories that work in the long term.
- Epistemic: Standards are required that can accommodate both experimental rigour and local context, as well as protocols that balance flexibility with enough standardisation for data to travel.
- Social: Communities of practice must be fostered to enable supportive mentorship relationships and collaborative networks that sustain knowledge exchange.
- Organisational: There is a need for clear governance structures, data policies, and most significantly, the integration of data responsibilities into formal roles rather than treating them as add-ons (which means shifting credit systems within institutions and funding organisations to recognise this form of research work).
- Political-economic: Sustained government funding is required to support the above changes, together with some levels of institutional autonomy to foster flexible arrangements mindful of local constraints, and data management issues need to be prioritised in national development agendas.

These interventions need to be aligned. Training without equipment or governance produces knowledge that may not prove applicable or usable in new contexts. Policies without resources create mindless forms of compliance and bureaucracy: everyone going through the motions of following rules that cannot be implemented.

LIMITATIONS AND FUTURE RESEARCH

These findings are based on a single institution during 2023 and may not generalise to other West African or Global South research institutes. Response rates may underrepresent junior or marginalised perspectives, and we did not systematically examine gender or career-stage effects. The survey item underlying Figure 6 ('how do you store your research data') did not distinguish between active, working storage and long-term archival storage, since these are hard to disentangle in practice, as discussed above, particularly given the crucial importance of long-term expectation in shaping everyday data practices; where possible we have cross-checked this figure against interview data that fit this distinction, and the discrepancy between reported storage practices and expressed repository preferences should be read with this concern in mind. Future work should explore the longitudinal evolution of data practices, comparative studies across institutions, the intersection of data management with local knowledge systems and epistemic pluralism to assess how infrastructure can support multiple forms of knowledge. This said, these findings are valuable in providing a detailed perspective on the challenges faced by individual institutions when confronted with data-sharing and management issues and, in parallel, broader examinations and discussions of the situation of open data and open science activities in Africa as a whole (Leonelli *et al.*, 2018).

CONCLUSION

CRI's translational agricultural research operates under data precarity and related infrastructural fragility, in which the technical, epistemic, social, organisational, and political dimensions of continuity may fail simultaneously. When nearly everyone relies on personal devices, when most researchers struggle to access colleagues' data, and when everyone recognises the need for training, the challenge is not one individuals can solve on their own. These are structural challenges that need to be addressed systematically by all stakeholders involved and taken seriously by all relevant institutions (private or public). In this respect, this study's focus on the issues faced by one institution provides insight into the broader system of data sharing in Africa and beyond and points to the fundamental importance of collaboration and investment in this area.

This study uncovered problems such as institutional-scale research functioning within fragmented and typically insufficient 'small science' infrastructure; frustrated epistemic cultures; the failure of data to achieve either immutability or mobility; and infrastructure that may be absent or subject to inadequate upgrades and maintenance, thereby risking decay. The refrain 'scientists retire with their data' captures something profound about the systemic knowledge loss caused by the lack of relevant infrastructure, training, and institutional arrangements. Addressing these challenges requires integrated interventions across technical, social, organisational, epistemic, and political dimensions, alongside sustained government commitment. It is well recognised that knowledge production depends on collective material conditions rather than merely individual responsibility (though the latter, of course, continues to play a significant role).

One cannot expect researchers to solve institutional infrastructure problems through personal heroics, a situation too often encountered on the ground but clearly unstable and unsatisfactory in its outcomes. The question raised by this study is not whether CRI's researchers are dedicated or skilled, for they clearly are; nor is it whether CRI management is doing enough to support its researchers, which it is certainly attempting to do within the limits of local conditions. The key question continues to be: what would it take to build research infrastructure that serves the systematic, long-term needs of a large community and does not disappear when the people who created it do?

DATA ACCESSIBILITY STATEMENT

The processed dataset in the form of graphs used to write this paper can be found at [ANONYMISED FOR REVIEW].

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendices.** Appendix A to B. DOI: https://doi.org/10.5334/dsj-2026-035.s1

ACKNOWLEDGEMENTS

We express our gratitude to the CSIR-CRI management and staff for their invaluable support and contributions, as well as our PHIL_OS team members at the University of Exeter in the UK and the Technical University of Munich in Germany for feedback on early drafts, particularly Paola Castaño.

FUNDING INFORMATION

This work has been supported by the PHIL_OS project ('A Philosophy of Open Science in Diverse Research Environments', www.opensciencestudies.eu), funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101001145).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Joyce Koranteng-Acquah orcid.org/0000-0003-0380-1287
Technical University of Munich, DE

Sabina Leonelli orcid.org/0000-0002-7815-6609
Technical University of Munich, DE

REFERENCES

Baker, M. (2016) '1,500 scientists lift the lid on reproducibility', *Nature*, 533, pp. 452–454. Available at: https://doi.org/10.1038/533452a

Baruch, Y. and **Holtom, B.C.** (2008) 'Survey response rate levels and trends in organisational research', *Human Relations*, 61(8), pp. 1139–1160. Available at: https://doi.org/10.1177/0018726708094863

Belarde-Lewis, M., Littletree, S., Braine, I.S., Srader, K., Guerrero, N. and **Palmer, C.L.** (2024) 'Centering relationality and CARE for stewardship of indigenous research data', *Data Science Journal*, 23:32, pp. 1–16. Available at: https://doi.org/10.5334/dsj-2024-032

Bezuidenhout, L., Kelly, A.H., Leonelli, S. and **Rappert, B.** (2017a). ''$100 Is Not Much to You': open science and neglected accessibilities for scientific research in Africa', *Critical Public Health*, 27(1), pp. 39–49. Available at: https://doi.org/10.1080/09581596.2016.1252832

Bezuidenhout, L., Leonelli, S., Kelly, A.H. and **Rappert, B.** (2017b). 'Beyond the digital divide: towards a situated approach to open data', *Science and Public Policy*, 44(4), pp. 464––475. Available at: https://doi.org/10.1093/scipol/scw036

Birnholtz, J.P. and **Bietz, M.J.** (2003) 'Data at work: supporting sharing in science and engineering', *Proceedings of the 2003 international ACM SIGGROUP conference on supporting group work*, pp. 339–348. Available at: https://doi.org/10.1145/958160.958215

Borgman, C.L. (2015) *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA: MIT Press. Available at: https://doi.org/10.7551/mitpress/9963.001.0001

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

Bowker, G.C. (2005) *Memory practices in the sciences*. Cambridge, MA: MIT Press.

Briney, K. (2015) *Data management for researchers: organize, maintain and share your data for research success*. London: Pelagic Publishing. Available at: https://doi.org/10.2307/jj.28833749

Charmaz, K. (2014) *Constructing grounded theory*. 2nd ed. Beverly Hills, CA: Sage.

Cragin, M.H., Palmer, C.L., Carlson, J.R. and **Witt, M.** (2010) 'Data sharing, small science and institutional repositories', *Philosophical Transactions of the Royal Society A*, 368(1926), pp. 4023–4038. Available at: https://doi.org/10.1098/rsta.2010.0165

Crosas, M. (2011) 'The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data', *D-Lib Magazine*, 17(1/2). Available at: https://doi.org/10.1045/january2011-crosas

Diaba-Nuhoho, P. and **Amponsah-Offeh, M.** (2021) 'Reproducibility and research integrity: the role of scientists and institutions', *BMC Res Notes*, 14(451). Available at: https://doi.org/10.1186/s13104-021-05875-3

Dotson, K. (2014) 'Conceptualizing epistemic oppression', *Social Epistemology*, 28(2), pp. 115–138. Available at: https://doi.org/10.1080/02691728.2013.782585

Douglas, H. and **Branch, T.Y.** (2024) 'The social contract for science and the value-free ideal', *Synthese*, 203, 40. Available at: https://doi.org/10.1007/s11229-023-04477-9

Edwards, P.N. (2010) *A vast machine: computer models, climate data, and the politics of global warming.* Cambridge, MA: MIT Press.

Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C. and **Borgman, C.L.** (2011) 'Science friction: data, metadata, and collaboration', *Social Studies of Science*, 41(5), pp. 667–690. Available at: https://doi.org/10.1177/0306312711413314

Errington, T.M. (2024) 'Building reproducibility bridges to cross the "Valley of Death"', *The Journal of Clinical Investigation*, 134(1), p. e177383. Available at: https://doi.org/10.1172/JCI177383

Faniel, I.M. and **Jacobsen, T.E.** (2010) 'Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data', *Computer Supported Cooperative Work*, 19(3–4), pp. 355–375. Available at: https://doi.org/10.1007/s10606-010-9117-8

Fricker, M. (2007) *Epistemic injustice: power and the ethics of knowing*. Oxford: Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Jackson, S.J., Edwards, P.N., Bowker, G.C. and **Knobel, C.P.** (2007) 'Understanding infrastructure: history, heuristics, and cyberinfrastructure policy', *First Monday*, 12(6). Available at: https://doi.org/10.5210/fm.v12i6.1904

Karasti, H., Baker, K.S. and **Millerand, F.** (2010) 'Infrastructure time: long-term matters in collaborative development', *Computer Supported Cooperative Work*, 19, pp. 377–415. Available at: https://doi.org/10.1007/s10606-010-9113-z

Knorr-Cetina, K. (1982) 'Scientific communities or trans epistemic arenas of research? A critique of quasi-economic models of science', *Social Studies of Science*, 12(1), pp. 101–130. Available at: https://doi.org/10.1177/030631282012001006

Knorr-Cetina, K. (1995) 'Laboratory studies: the cultural approach to the study of science', in S. Jasanoff *et al*. (eds.) *Handbook of science and technology studies.* Beverly Hills, CA: Sage. pp 140–166. Available at: https://doi.org/10.4135/9781412990127.d12

Knorr-Cetina, K. (1999) *Epistemic cultures: how the sciences make knowledge*. Cambridge, MA: Harvard University Press. Available at: https://doi.org/10.4159/9780674039681

Latour, B. and **Woolgar, S.** (1986) *Laboratory life: the construction of scientific facts*. Princeton, NJ: Princeton University Press. Available at: https://doi.org/10.1515/9781400820412

Leonelli, S. (2016) *Data-centric biology: a philosophical study*. Chicago, IL: University of Chicago Press. Available at: https://doi.org/10.7208/chicago/9780226416502.001.0001

Leonelli, S. (2018) 'Rethinking reproducibility as a criterion for research quality, research in the history of economic thought and methodology', in *Including a symposium on Mary Morgan: curiosity, imagination, and surprise*. Emerald Group Publishing Limited. Vol. 36. pp. 129–146. Available at: https://doi.org/10.1108/S0743-41542018000036B009

Leonelli, S. (2020) 'Learning from data journeys', in S. Leonelli and N. Tempini (eds.) *Data journeys in the sciences*. Cham: Springer International Publishing. pp. 1–24. Available at: https://doi.org/10.1007/978-3-030-37177-7_1

Leonelli, S. (2025) 'What to do about data distance? Responsible alternatives to data sharing', *Harvard Data Science Review,* 7(2), pp. 1–10. Available at: https://doi.org/10.1162/99608f92.e95b5c26

Leonelli, S. *et al*. (2018) 'Introduction: open data and Africa', *Data Science Journal*, 17, p. 5. Available at: https://doi.org/10.5334/dsj-2018-005

Lord, P. and **Macdonald, A.** (2003) *e-Science curation report: data curation for e-science in the UK*. London: The Digital Archiving Consultancy Ltd.

Lynch, C.A. (2003) 'Institutional repositories: essential infrastructure for scholarship in the digital age', *Libraries and the Academy*, 3(2), pp. 327–336. Available at: https://doi.org/10.1353/pla.2003.0039

Merton, R.K. (1968) 'The Matthew effect in science: the reward and communication systems of science are considered', *Science,* 159(3810), pp. 56–63. Available at: https://doi.org/10.1126/science.159.3810.56

Mittal, D., Mease, R., Kuner, T., Flor, H., Kuner, R. and **Andoh, J.** (2023) 'Data management strategy for a collaborative research centre', *Gigascience*, 12, giad049. Available at: https://doi.org/10.1093/gigascience/giad049

Mwelva, J. *et al.* (2020) 'Developing open science in Africa: barriers, solutions and opportunities', *Data Science Journal,* 19, p. 31. Available at: https://doi.org/10.5334/dsj-2020-031

Neff, G., Tanweer, A., Fiore-Gartland, B. and **Osburn, L.** (2017) 'Critique and contribute: a practice-based framework for improving critical data studies and data science', *Big Data*, 5(2), pp. 85–97. Available at: https://doi.org/10.1089/big.2016.0050

Nulty, D.D. (2008) 'The adequacy of response rates to online and paper surveys: what can be done?' *Assessment & Evaluation in Higher Education*, 33(3), pp. 301–314. Available at: https://doi.org/10.1080/02602930701293231

Odebrecht, C. (2025) 'Research data governance. The need for a system of cross-organisational responsibility for the researcher's data domain', *Data Science Journal,* 24, p. 12. Available at: https://doi.org/10.5334/dsj-2025-012

Peng, R.D. (2011) 'Reproducible research in computational science', *Science*, 334(6060), pp. 1226–1227. Available at: https://doi.org/10.1126/science.1213847

Pinamang Acheampong, P., Owusu Danquah, E., Agyeman, K., Obeng Dankwa, K. and **Addison, M.** (2021) 'Research and development for improved cassava varieties in Ghana: farmers' adoption and effects on livelihoods', in *Cassava - Biology, Production, and Use*. IntechOpen. Available at: https://doi.org/10.5772/intechopen.97588

Piwowar, H.A. and **Vision, T.J.** (2013) 'Data reuse and the open data citation advantage', *PeerJ*, 1, p. 1:e175. Available at: https://doi.org/10.7717/peerj.175

Reichmann, S. (2023) 'Mobile researchers, immobile data: managing data (producers)', *Social Studies of Science*, 53(3), pp. 341–357. Available at: https://doi.org/10.1177/03063127231156862

Ribes, D. and **Jackson, S.J.** (2013) **'**Data bite man: the work of sustaining a long-term study', in Lisa Gitelman (ed.) *"Raw Data" Is an Oxymoron.* Cambridge, MA: MIT Press. pp. 147–166. Available at: https://doi.org/10.7551/mitpress/9302.003.0010

Rolland, B. and **Lee, C.P.** (2013) 'Beyond trust and reliability: reusing data in collaborative cancer epidemiology research', *Proceedings of the 2013 conference on computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA. pp. 435–444. Available at: https://doi.org/10.1145/2441776.2441826

Ross-Hellauer, T., Reichmann, S., Cole, N.L., Fessl, A., Klebel, T. and **Pontika, N.** (2022) 'Dynamics of cumulative advantage and threats to equity in open science: a scoping review', *Royal Society Open Science*, 9(1), p. 211032. Available at: https://doi.org/10.1098/rsos.211032

Soundararajan, S. and **Mishra, S.** (2023) 'Data management: the first step in reproducible research', *Indian Journal of Occupational and Environmental Medicine*, 27, pp. 359–363. Available at: https://doi.org/10.4103/ijoem.ijoem_342_22

Star, S.L. (1999) 'The ethnography of infrastructure', *American Behavioral Scientist*, 43(3), pp. 377–391. Available at: https://doi.org/10.1177/00027649921955326

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E. *et al.* (2011) 'Data sharing by scientists: practices and perceptions', *PLoS One*, 6(6), p. e21101. Available at: https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B. *et al.* (2015) 'Changes in data sharing and data reuse practices and perceptions among scientists worldwide', *PLoS One*, 10(8), p. e0134826. Available at: https://doi.org/10.1371/journal.pone.0134826

Tenopir, C., Rice, N.M., Allard, S., Baird, L., Borycz, J., Christian, L. *et al.* (2020) 'data sharing, management, use, and reuse: practices and perceptions of scientists worldwide', *PLoS One*, 15(3), p. e0229003. Available at: https://doi.org/10.1371/journal.pone.0229003

Vines, T.H. *et al.* (2014) 'The availability of research data declines rapidly with article age', *Current Biology*, 24(1), pp. 94–97. Available at: https://doi.org/10.1016/j.cub.2013.11.014

Wallis, J.C., Rolando, E. and **Borgman, C.L.** (2013) 'If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology', *PLoS One*, 8(7), p. e67332. Available at: https://doi.org/10.1371/journal.pone.0067332

Walsh, J.P. and **Ungson, G.R.** (1991) 'Organisational memory', *The Academy of Management Review*, 16(1), pp. 57–91. Available at: https://doi.org/10.2307/258607

Whyte, A. and **Tedds, J.** (2011) *Making the case for research data management*. Edinburgh: Digital Curation Centre, Jisc Briefing Paper.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016) 'The FAIR guiding principles for scientific data management and stewardship', *Scientific Data*, 3, p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18

Williamson, H.F. and **Leonelli, S.** (eds.) (2022) *Towards responsible plant data linkage: data challenges for agricultural research and development*. Cham: Springer Verlag. Available at: https://doi.org/10.1007/978-3-031-13276-6

Wu, M.-J., Zhao, K. and **Fil-Aime, F.** (2022) 'Response rates of online surveys in published research: a meta-analysis', *Computers in Human Behaviour Reports*, 7. Available at: https://doi.org/10.1016/j.chbr.2022.100206

Wu, M. *et al.* (2026) 'Bridging the data discovery gap: user-centric recommendations for research data repositories', *Data Science Journal*, 25, p. 6. Available at: https://doi.org/10.5334/dsj-2026-006

Wyatt, S. (2025) 'Distance and other metaphors about reuse of research data', *Harvard Data Science Review*, 7(2). Available at: https://doi.org/10.1162/99608f92.1d109ab1

TO CITE THIS ARTICLE:
Koranteng-Acquah, J. and Leonelli, S. 2026 Data Management Challenges for Translational Agricultural Research in West Africa: The Case of the Crops Research Institute, Ghana. *Data Science Journal*, 25: 35, pp. 1–27. DOI: https://doi.org/10.5334/dsj-2026-035

Submitted: 11 September 2025
Accepted: 05 August 2026
Published: 26 August 2026

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.