Data Management at the Crops Research Institute of the Council for Scientific and Industrial Research, Ghana:

A Scoping Report

Authorship and citation information

The report was jointly developed and written by Joyce Koranteng-Acquah and Sabina Leonelli, with input from Patricia Acheampong, Maxwell Asante and Ernest Baafi concerning data access, study design and feedback on data analysis. Please cite as follows: **Koranteng-Acquah J., Acheampong, P., Asante M.D., Baafi E. & Leonelli, S.* (2024)** ***Data Management at the Crops Research Institute of the Council for Scientific and Industrial Research, Ghana: A Scoping Report*****. PHIL_OS Report, June 2024. DOI:** 10.5281/zenodo.11479135**.**

*corresponding author: Sabina Leonelli, Exeter Centre for the Study of the Life Sciences (Egenis), University of Exeter, UK; s.leonelli@exeter.ac.uk

Acknowledgements

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). 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 for feedback on early drafts.

Table of Contents

Executive Summary

This report evidences the crucial significance of data management practices and infrastructures for research conducted at the CSIR-Crops Research Institute (CSIR-CRI). In a context where scientific endeavours are embedded in the pursuit of sustainable agricultural practices, meticulous data management makes it possible for research findings to endure over time and thereby laying the groundwork for future innovation. As Ghana aspires to achieve food self-sufficiency and contribute to global food security, the role of data-driven insights becomes increasingly pivotal. The adaptability of agriculture to emerging challenges hinges on a robust data ecosystem that preserves and propagates valuable insights. The transition to digital platforms and cloud storage aligns with the global trend toward collaborative and open science.

This survey found that while researchers agree on the significance and implications of data management strategies for their research and its impact, the ways in which data are backed-up, stored, shared and accessed at CSIR-CRI are varied and inconsistent. There are considerable generational differences in how researchers rely on digital, machine-readable data management strategies, as well as their perception and use of data storage facilities, institutional repositories and international databases. For instance, we noted a tendency for senior researchers to use Excel spreadsheets and paper, and to ensure that the data are stored in common repositories, while early career researchers are more likely to resort to machine-readable storage, but less likely to do this in places other than their personal laptop. Most researchers encountered obstacles in accessing and re-using data produced by others, particularly legacy data derived from past experiments, which is detrimental to research on novel crop varieties, methods for technology transfers, and the effects of climate change. The majority of respondents demanded institutional support and centralised training to ensure more consistent and effective data management strategies, yet they also exhibited anxiety around the use of centralised data storage facilities, due to security and privacy concerns as well as uncertainty around how such platforms may be implemented to respond to the specific needs of their research.

We conclude that CSIR-CRI's commitment to excellence in crop research is tied to its ability to embrace and champion cutting-edge data management strategies, which includes utilising both analogue and digital tools, increasing investment in centralised storage and training facilities, and ensuring the secure storage and sharing of sensitive data. However, the Institute's efforts to prioritise safe data storage and sharing are impeded by limited funding, prompting

CSIR-CRI to actively seek support from partners and donors. The Institute believes that collaborative research efforts, underpinned by robust and responsible data practices, can catalyse the development and adoption of innovative farming techniques, ensuring sustainable livelihoods for local communities and fostering economic growth; and is committed to continue making strides towards improving its data management policies, tools and training resources.

1. Introduction

In the dynamic landscape of scientific research, the importance of effective data sharing and management strategies cannot be overstated. Data, in their myriad forms, serve as the foundational element for knowledge creation and dissemination. The ability to effectively and responsibly share and manage data accelerates the pace of scientific discovery, enabling researchers to collectively build upon existing knowledge, validate findings, and contribute to the continuous evolution of their respective fields. This holds particularly true in the realm of crop research, where advancements in technology and a growing emphasis on collaborative, transdisciplinary initiatives are reshaping the way knowledge is generated and disseminated.

The CSIR-Crops Research Institute (CSIR-CRI), the largest of the thirteen Institutes under the Council for Scientific and Industrial Research (CSIR), stands at the intersection of this transformative shift. Established in 1964 by an ACT of Parliament, CSIR-CRI is dedicated to enhancing food security, agricultural productivity and rural livelihoods in Ghana, through the development and dissemination of improved crop varieties, sustainable farming practices and innovative agricultural technologies tailored to diverse agro-ecological zones of Ghana. Through collaborative research partnerships with national and international organisations, the Institute conducts research on a wide range of staple crops, addressing challenges such as pests and disease management, soil fertility, and climate resilience. The CSIR-CRI also plays a crucial role in capacity building and knowledge dissemination to empower farmers and strengthen the agricultural sector in Ghana.

This report is based on a survey sent out to CSIR-CRI staff between October and December 2023 with the goal to assess the current state of data practices in the Institute and the data-related needs and aspirations of researchers going forward. The survey is part of the research conducted within the European project 'A Philosophy of Open Science for Diverse Research Environments' (www.opensciencestudies.eu), which explores the cultural, social, political, and institutional dynamics influencing knowledge co-production, translation, and utilisation. This research seeks to construct both a philosophical framework and empirical investigations into how open science can promote sound research methodologies while accommodating the diversity of research approaches and settings globally. One part of this project was conducted by Joyce Koranteng-Acquah and Sabina Leonelli in collaboration with CSIR-CRI staff, and consisted of an ethnographic study of research conducted at the CSIR-CRI between August

and December 2023, with an emphasis on data management practices. The survey detailed in this report provides further empirical materials towards this research, and vice versa our analysis of survey results is informed by ethnographic research findings.

As scientific questions become more complex, drawing insights from multiple domains and publics becomes imperative. The significance of data sharing becomes more pronounced in an era where collaborative and interdisciplinary research is increasingly essential. Responsible and effective forms of data sharing allow for scrutiny and validation by peers, contributing to the collective trust in scientific findings. The transparency afforded by robust data management practices therefore becomes crucial to fostering and maintaining the credibility of scientific research. Moreover, meticulous data management is fundamental for maintaining the integrity and reproducibility of research. It ensures that valuable insights – and related methods and materials - are not lost over time, forming the foundation upon which future breakthroughs can be built. The agricultural community's ability to adapt to emerging challenges, such as changing weather patterns and evolving pest threats, relies on a robust data ecosystem.

The strategies employed in data sharing and management not only shape the trajectory of individual research endeavours, but also contribute significantly to the collective knowledge base of the scientific community. Crop research, inherently multidisciplinary, involves aspects of genetics, agronomy, environmental science, and beyond. Effective data sharing facilitates the cross-pollination of ideas, accelerates scientific discoveries, and promotes innovation. This integration is particularly relevant where outcomes have far-reaching implications for food security, agricultural sustainability, and adaptation to environmental changes. Researchers rely on comprehensive datasets to identify genetic traits associated with desirable agricultural characteristics, understand the impact of environmental factors on crop performance, foster agro ecological interventions and optimise cultivation practices. As the agricultural landscape evolves to meet the challenges of a growing population and climate variability, the ability to leverage shared knowledge is indispensable.

This report evidences how CSIR-CRI researchers acknowledge the critical role that institutional platforms play in safeguarding and disseminating their work. At CSIR-CRI, where scientific endeavours are deeply embedded in the pursuit of sustainable agricultural practices, meticulous data management is not merely a technical prerequisite but a cornerstone for scientific integrity. It is the safeguard that ensures research findings are not ephemeral but

endure over time, laying the groundwork for future innovation. The adaptability of agriculture to emerging challenges hinges on a robust data ecosystem that preserves and propagates invaluable insights. The transition to digital platforms and cloud storage aligns with the global trend toward collaborative and open science. CSIR-CRI's commitment to excellence in crop research is tied to its ability to embrace and champion cutting-edge data management strategies, and the management of the Institute is making strides in improving such strategies to enhance investigation and impact. A new CSIR-CRI data management policy is now under development, and will be informed by the findings of this report. As Ghana aspires to achieve food self-sufficiency and contribute to global food security, the role of data-driven insights becomes increasingly pivotal. Collaborative research efforts, underpinned by robust data practices, can catalyse the development and adoption of innovative farming techniques, ensuring sustainable livelihoods for local communities and fostering economic growth.

2. Methods and Sample Analysis

2.1 Design and Dissemination

This survey was developed through a collaborative effort between the management of CSIR-CRI and the PHIL_OS team based at the University of Exeter in the UK. Data collection rested on a survey designed to elicit insights into the prevailing data management practices among researchers at CSIR-CRI. Leveraging a mix of closed-ended and open-ended questions, we sought to capture both quantitative metrics and qualitative narratives by posing thirteen (13) questions tailored to gauge researchers' experiences, challenges, and perspectives on data storage, accessibility, security and researchers' receptiveness to training in data management. The survey was disseminated electronically via Google forms to technologists and scientists across various designations within CSIR-CRI, ensuring a diverse representation of perspectives.

Researchers were contacted through official CSIR-CRI communication channels, including email and institutional messaging systems. The digital format was designed to facilitate widespread participation, offering respondents the flexibility to engage at their convenience. Emphasising the anonymity of participants and ensuring the confidentiality of their feedback, the survey was created on September 11, 2023 and shared with management of CSIR-CRI on

September 13, 2023. It was subsequently made available and circulated from October 10 to December 15, 2023. The final response was received on December 4, 2023 at 14:20:28 GMT.

2.2 Accessibility, Inclusivity and Representation

Recognizing the diverse roles and hierarchical structures within CSIR-CRI, our data collection strategy prioritised inclusivity. Instructions and questions were presented in a clear and straightforward manner. The survey was designed to be user-friendly and accessible across various devices to accommodate the diverse technological preferences of the target audience. By targeting researchers across different designations and divisions, we aimed to capture a comprehensive view of the Institute's data landscape. Additionally, the survey was intentionally designed to span varying levels of experience, acknowledging that data practices may evolve with the researcher's tenure. This approach facilitated a nuanced examination of perspectives, acknowledging the unique challenges faced by both seasoned principal scientists and early-career researchers.

2.3 Analysis and Synthesis

Upon the completion of the survey phase, the collected data underwent a rigorous process of analysis and synthesis. Quantitative responses were subjected to statistical scrutiny to reveal key trends and patterns, offering a numerical lens into the prevalent sentiments. Simultaneously, qualitative responses were subjected to thematic analysis, identifying recurrent themes and nuanced narratives. For the multiple-choice questions, Google forms generated pie charts and bar graphs displaying percentages, while thematic categorisation and clustering of free-text responses were done manually. Word clouds (using Python) were generated to visually represent the most frequently occurring words in the dataset. In the case of multiple-choice questions, the percentages may not total 100% because respondents were permitted to select multiple answers. The same applies to free text responses, as participants could provide more than one answer.

Excel spreadsheet was used to generate bar graphs and pie charts from the categorisation. By triangulating quantitative metrics with qualitative insights, we aimed not only to present a snapshot of the current landscape of data management practices at CSIR-CRI, but to offer insights that can inform strategic decisions and interventions. The synthesis phase laid the groundwork for subsequent discussions, interpretations, and recommendations within the

overarching context of data sharing and management strategies in crop research. The recommendations derived from the analysis aim to guide CSIR-CRI in enhancing its data management infrastructure, fostering a culture of collaboration, and ensuring that the Institute remains at the forefront of transformative crop research.

2.4 Sample Size

The survey targeted a population of 127 researchers at the CSIR-Crops Research Institute, seeking their perspectives on data management practices. Out of this population, a total of 55 researchers responded to the survey, representing a response rate of approximately 43%. While the response rate is moderate, it provides a substantial dataset for analysis and insight into the opinions of a significant portion of the target population. Note that the demographics presented here are based on the responses received, and are not necessarily representative of the actual composition of the entire CSIR-CRI researcher population.

2.5 Limitation of Study

As discussed above, the study garnered responses from 55 out of 127 researchers, representing a 43% participation rate. This is a modest response rate, as typical of surveys of this kind, yet the range of responses obtained provides a robust snapshot of current viewpoints and data practices within CSIR-CRI (see demographics section). The reliance on self-reported information introduces the potential for self-reporting bias such as respondents providing answers that they perceive as socially desirable or aligned with institutional expectations. Some individuals may not fully disclose their activities due to the complexity of articulating them, as practices can be intricate and varied. By acknowledging these limitations, the study sets a foundation for future research to explore complementary methods. For instance, integrating survey data with performance metrics, such as tracking actual data management practices and conducting in-depth, systematic interviews with researchers, could enhance the robustness and reliability of findings.

The survey captures a snapshot of data management practices at a specific point in time. The rapidly evolving landscape of technology and institutional policies means that the findings may become outdated, particularly with CSIR-CRI on the verge of instituting their data management policy. Changes in infrastructure, policies, or the introduction of new tools after the survey period might not be reflected in the study. Therefore, the study's temporal scope is limited, and

interpretations would consider the dynamic nature of data management practices. Although open-ended questions were included to capture qualitative insights, the depth of qualitative analysis may be limited. Thematic analysis provides a broad understanding, but a more in-depth qualitative investigation, such as interviews or focus groups, could offer richer insights into researchers' experiences and perceptions.

3 Results

3.1 Demographics

The respondents exhibited diverse demographic characteristics, including varying designations, years of service at CSIR-CRI, gender distribution, and educational backgrounds.

3.1.1 Designation

In terms of designation, participants spanned various roles within the Institute, including 5 Principal Research Scientists (9.1%), 17 Senior Research Scientists (30.9%), 12 Research Scientists (21.8%), 3 Assistant Research Scientists (5.5%), 16 Principal Technologists (29.1%) and 2 technical officers (3.6%). This broad representation across different designations enables a comprehensive exploration of data management perspectives, encompassing expertise levels from seasoned researchers to those at earlier stages of their careers. Such diversity in roles fosters a richer understanding of data management practices, as perspectives may vary based on experience, responsibilities, and areas of specialisation. By integrating perspectives from individuals across various designations, the survey outcomes offer insights that can inform targeted strategies aimed at enhancing data management effectiveness within the CSIR-CRI research community.

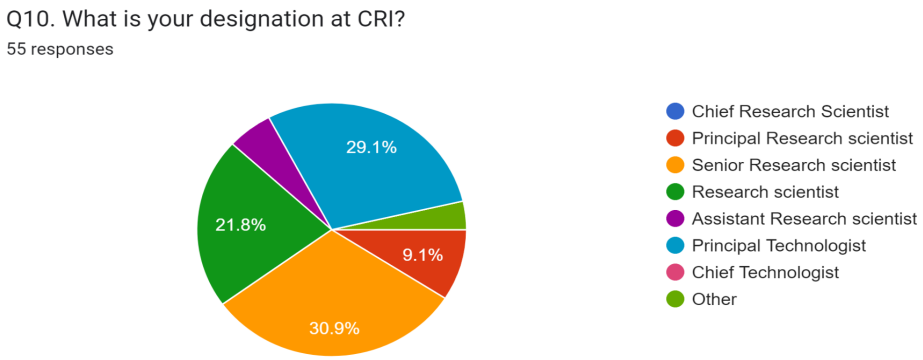


Fig. 1: A pie chart of participant designation at CSRI-CRI.

3.1.2 Years of Service

The survey revealed a prevalence of experienced individuals within the respondent pool. Majority of respondents (35; 63.6%) possessed over 10 years of experience at the Institute, highlighting a significant presence of seasoned professionals with extensive institutional knowledge. Furthermore, 7 respondents representing 12.7% had a tenure of 5 to 10 years at CSIR-CRI, while 20% (11 respondents) fell into the 1 to 5 years category. A smaller proportion, comprising 3.6% (2 respondents), reported having less than 1 year of experience at the Institute.

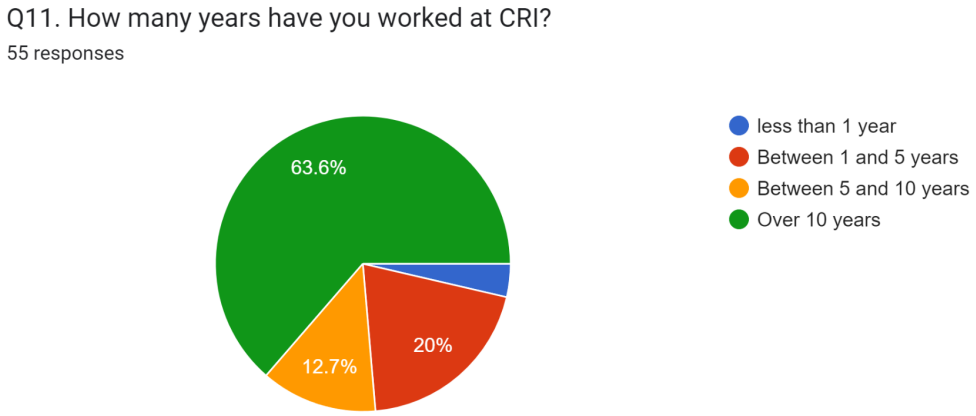


Fig. 2: A pie chart of experience levels of staff at CSIR-CRI.

3.1.3 Gender Distribution

The responses exhibited a gender distribution, with both male and female researchers participating in the survey. Males constituted (37), representing 67.3% of the respondents, while females (18) represented 32.7% of the participants.

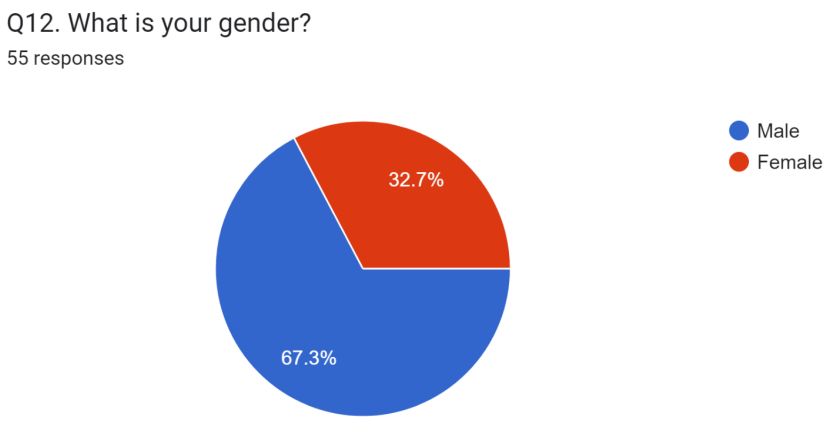


Fig. 3: A pie chart of gender distribution of staff of CSIR-CRI.

3.1.4 Educational Background

The respondents held diverse educational qualifications, with the majority possessing Ph.D. degrees. Respondents with PhD degrees were 32 (58.2%), masters degrees were 21 (38.2%) and undergraduate degrees were 2 (3.6%). This educational diversity is indicative of a well-qualified and specialised research community, essential for the complexities of crop research. It also ensures a broad spectrum of insights, considering that researchers with different academic backgrounds may approach data management differently.

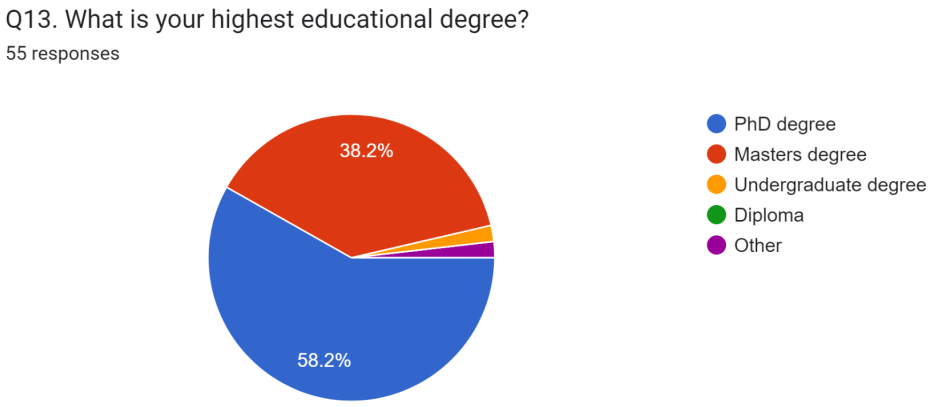


Fig. 4: A pie chart of educational background of survey participants at CSIR-CRI.

3.2 Data Collection Methods

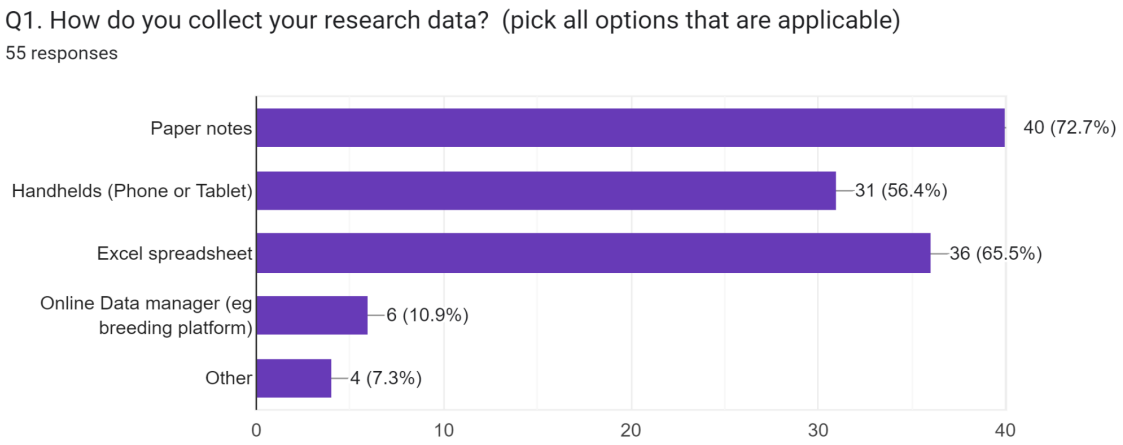


Figure 5: A bar chart of research data collection methods by staff of CSIR-CRI

The responses indicated that multiple methods were utilised for collecting research data. Paper notes emerged as a primary method, with 40 responses (72.7%). This approach suggests a continued reliance on tangible mediums for recording observations and data. Handheld devices such as phones or tablets were also cited, with 31 responses (56.4%). This highlights a significant integration of technology into research practices, facilitating more convenient and efficient data collection processes. Excel spreadsheets were reported as a prevalent tool for data collection, with 36 responses (65.5%) indicating their usage. The widespread use of Excel underscores the importance of organised data management and analysis in research endeavours.

Online data management platforms, such as breeding platforms, were utilised by a smaller proportion of respondents, with only 6 (10.9%) indicating their use. This suggests that while digital solutions are gaining traction, there is still room for improvement in the adoption of online data management tools within the research community. Additionally, a small percentage of respondents reported using "Other" methods for data collection, indicating the presence of diverse approaches beyond the options provided in the survey. These findings underscore the importance of flexibility and adaptation in research methodologies to meet the diverse needs and preferences of researchers across different disciplines and contexts.

3.3 Data Storage Practices

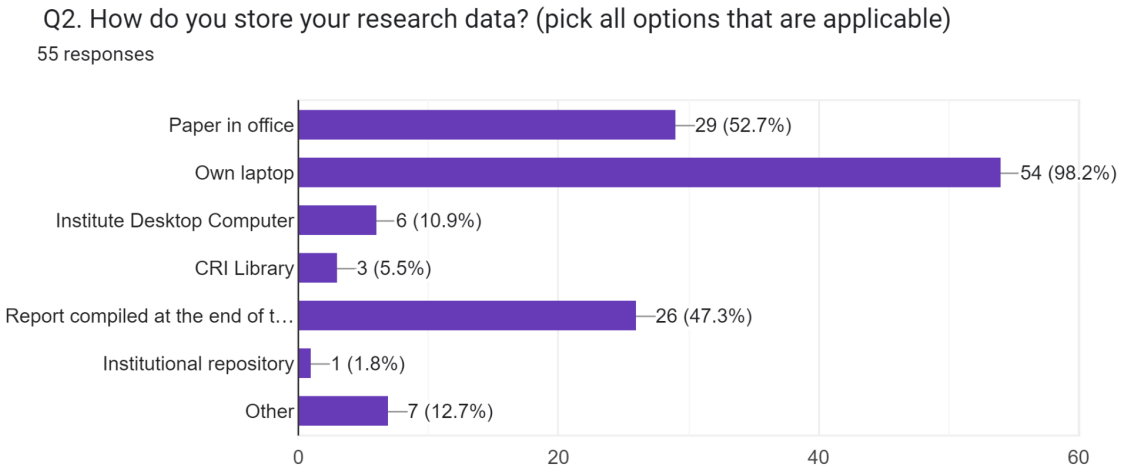


Figure 6: A bar chart of research data storage practices of CSIR-CRI staff

The selection of storage methods for research data was predominantly focused on personal devices, with 54 responses (98.2%) indicating the use of laptops for this purpose. This widespread practice highlights researchers maintaining control and accessibility over their data. Paper-based storage methods, such as office filing followed with 29 responses (52.7%). This approach suggests a preference for physical copies of data for backup or reference purposes, despite the risks in terms of data security, accessibility, and longevity compared to digital storage solutions. A significant portion of respondents (26; 47.3%), reported compiling reports at the end of their research, indicating a structured approach to archiving research findings, albeit potentially limiting the accessibility and comprehensiveness of the stored data.

Institutional resources for data storage such as desktop computers, institutional repositories or CSIR-CRI library were less commonly utilised with only 6 responses (10.9%), 1 response (1.8%) and 3 responses (5.5%), respectively. This suggests a lesser reliance on institutional infrastructure for data storage compared to personal devices or physical storage. Lastly, 12.7% representing 7 responses mentioned "Other" methods for storing research data, suggesting a diversity of approaches beyond the options provided in the survey.

3.4 Data Storage Preferences

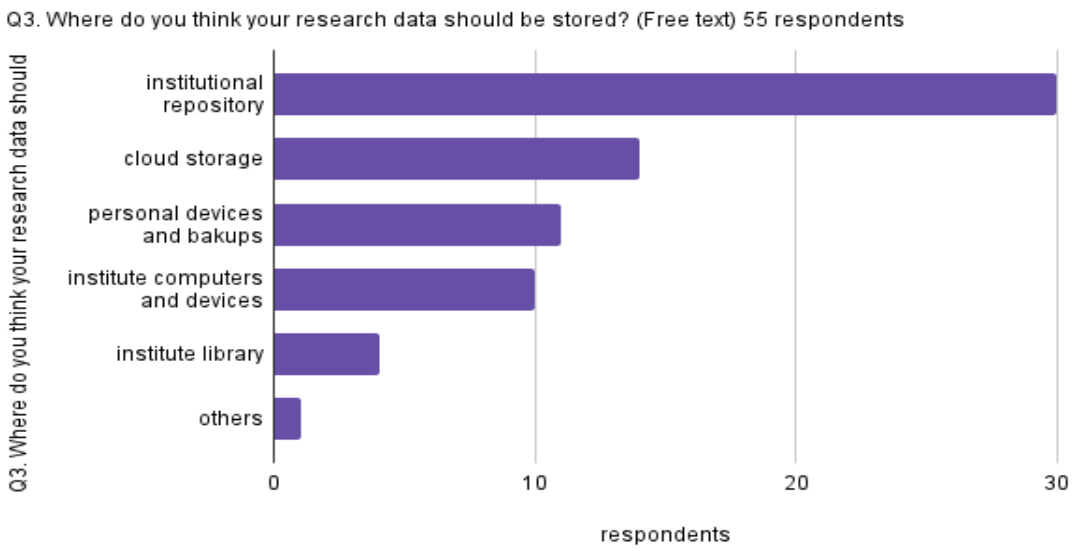


Fig. 7: A bar chart of research data storage preferences of CSIR-CRI staff

A significant portion (54.45%) indicated a tendency to store data in an institutional repository. This choice may reflect adherence to perceived norms or expectations with the research community. Institutional repositories offer centralised and standardised platforms facilitating transparent collaboration and long-term data preservation, but could also have limitations in access and scalability. Cloud storage was identified as another favoured option, with 25.45% responses indicating it as a preferred storage location. This method ensures flexibility, scalability, and accessibility, enabling researchers to access data remotely and ensuring redundancy against data loss.

Additionally, 20.0% of responses indicated an inclination towards storing research data themselves, reflecting a desire for control and autonomy. While local storage provides immediate access and control, it may pose risks in terms of data security, backup, and long-term preservation. A minority of responses mentioned storing data in institute libraries (7.27%) or on institute computers (18.18%). These options suggest a reliance on institutional resources for data storage, which may offer central management and support but could also have limitations in access and scalability.

3.5 Challenges with Data Access and Handling

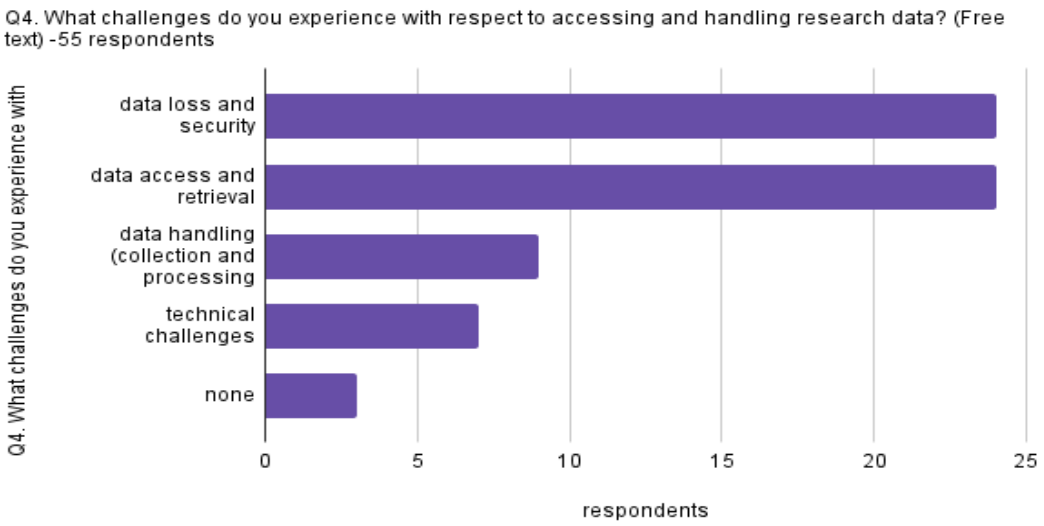


Fig. 8: A bar chart showing data access and handling challenges experienced by respondents.

Based on the responses regarding challenges experienced with accessing and handling research data, two main issues emerge as the most prevalent–data loss and security; and data access and retrieval, each mentioned by 24 respondents representing 43.6%.

Data loss and security concerns encompass a range of issues, including the risk of losing valuable research data due to hardware failures, accidental deletion, inadequate backup procedures, staff turnover or unavailability of institutional storage infrastructure. Ensuring data security and implementing robust backup procedures are essential to mitigate these risks and safeguard research integrity. Similarly, data access and retrieval challenges highlight difficulties in efficiently locating and retrieving research data when needed. This includes issues related to data organisation, internet connectivity, storage formats, or access permissions. Establishing clear data management protocols, implementing effective retrieval systems, and providing appropriate access controls can help address these challenges and streamline data access processes.

A smaller proportion of responses (9; 16.4%) identified challenges related to data handling and processing, particularly the transition from hard copies to soft copies. This involves digitising and organising physical documents or adapting to new software tools for data processing and analysis. Providing training and support for researchers to effectively navigate these changes

can facilitate a smoother transition and improve overall data handling efficiency. Additionally, a few responses (7; 12.7%) mentioned facing technical limitations, which include hardware or software constraints that hinder data access, processing, or analysis. Addressing technical limitations may require investing in updated infrastructure, software tools, or training to empower researchers with the necessary resources and skills to overcome these obstacles.

3.6 Data Management Support

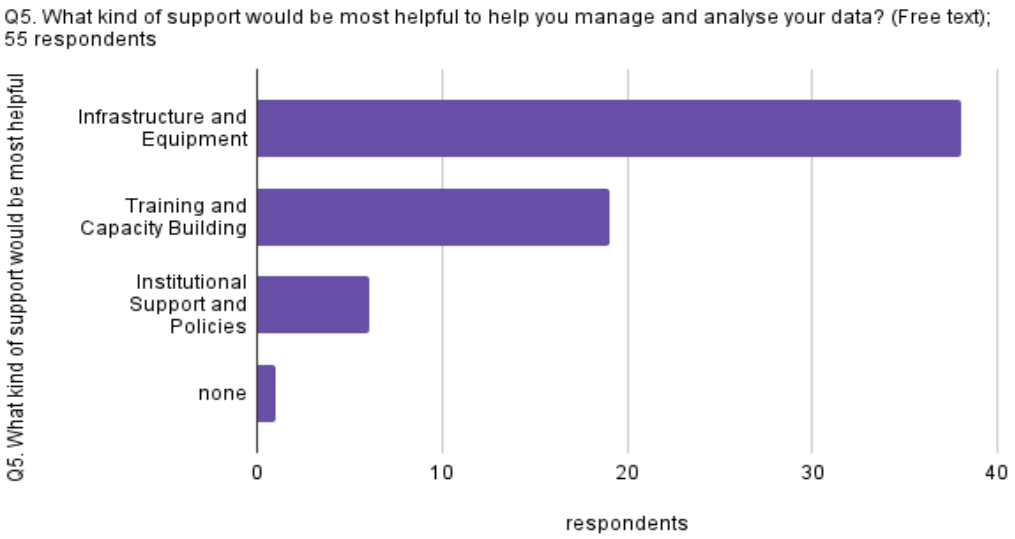


Fig. 9: A bar chart depicting training needs assessment of survey respondents of CSIR-CRI staff.

The responses provided highlight several key areas where support would be most beneficial for researchers to effectively manage and analyse their data. Firstly, many participants (38; 69.09%) emphasised the importance of having easy access to institutional repositories or centralised data storage platforms. These repositories should be user-friendly and provide seamless access to stored data for analysis. Additionally, support for acquiring and using digital tools for data collection, such as tablets with strong batteries and handheld devices, was identified as crucial for efficient data gathering in the field. Some participants (6; 10.91%) also highlight the need for specialised support, such as access to biometricians or data curators, to assist with complex data analysis tasks.

Training and capacity building emerged as another significant area of support needed. Nineteen researchers (34.55%) expressed a desire for training in various aspects of data management and

analysis, including the use of software tools for statistical analysis. Regular workshops or seminars on data collection, storage, and analysis would help researchers stay updated on best practices and modern techniques in data management. Furthermore, participants highlighted the importance of having access to reliable internet connectivity and sufficient storage space, whether through institutional desktop systems or cloud-based storage solutions. This infrastructure support is essential for ensuring that researchers can store and access their data securely and efficiently.

3.7 Access to Primary Historical Data

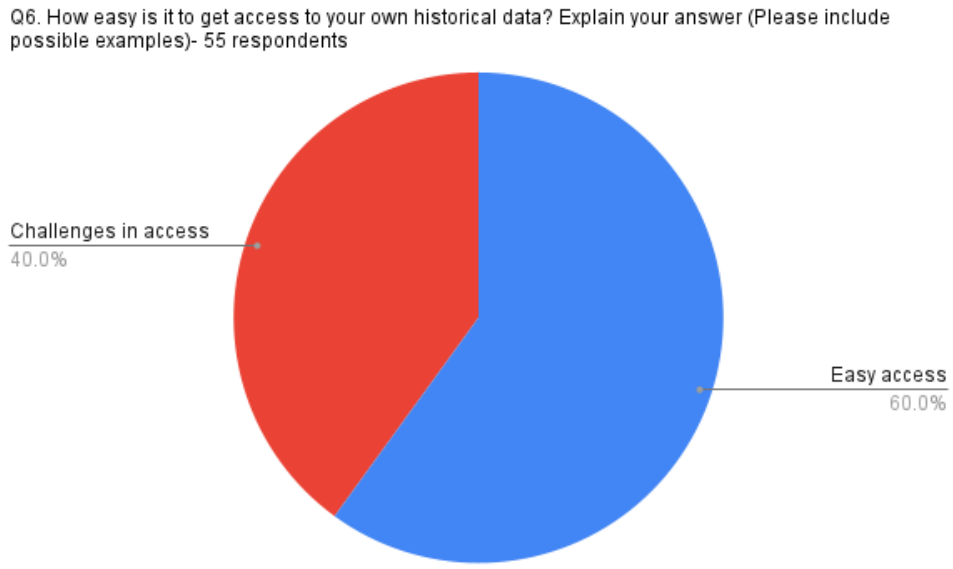


Fig. 10: A pie chart of responses from survey participants on challenges in access to own historical data

Accessing historical data varies greatly depending on how it's stored and managed. Thirty-three (33) respondents representing 60% find it relatively easy, especially if they utilise cloud storage solutions like Google Drive or OneDrive, as data can be accessed from any device with an internet connection. Others (22; 40%) find it challenging, particularly if data is stored on personal devices like laptops or external drives, as there may be issues with file corruption, hardware failures, or compatibility. Additionally, reliance on paper-based records poses its own challenges, including difficulties in organisation, retrieval, and potential loss due to wear and tear or misplacement.

The accessibility of historical data can be influenced by factors such as data organisation, backup practices, and individual familiarity with storage systems. Those who maintain well-organised digital records on their personal devices or in the cloud generally find it easier to access their historical data compared to those who rely on paper records or have less structured storage systems. Moreover, challenges arise when data management responsibilities are transferred between individuals or when staff turnover occurs, leading to potential gaps in data accessibility. Although advancements in digital storage and cloud computing have made accessing historical data more convenient for many researchers, challenges still exist, particularly regarding data organisation, backup strategies, and technological dependencies. Establishing robust data management practices and utilising reliable backup solutions can help mitigate these challenges and ensure the continued accessibility and integrity of historical research data.

3.8 Access to Secondary Historical Data

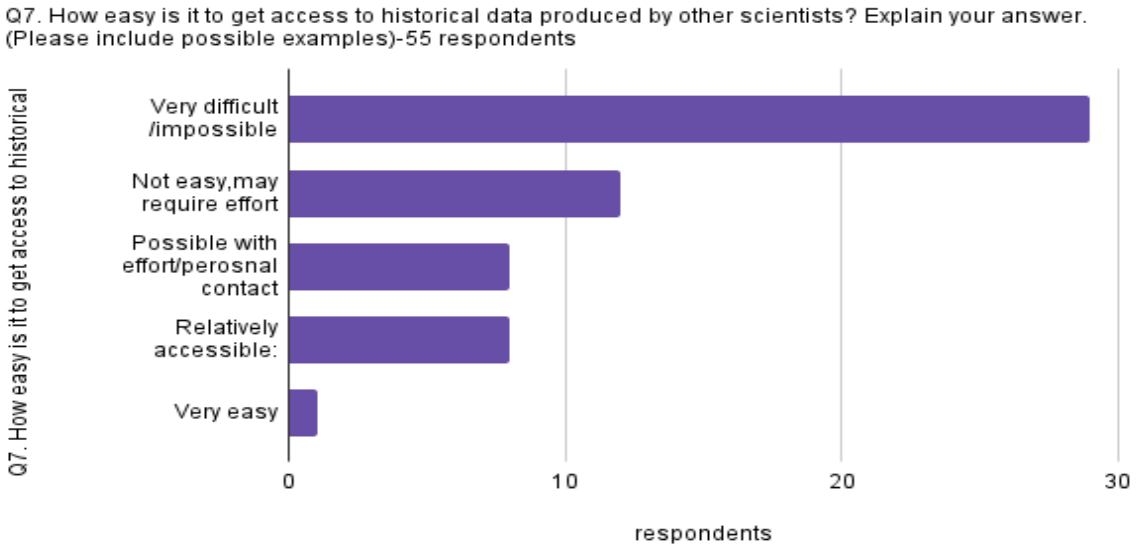


Fig. 11: A bar chart of access to historical data produced by other researchers

In contrast, accessing historical data produced by other scientists, 41 (74.55%) respondents found it impossible or quite challenging for several reasons. Firstly, data may be stored on personal computers or devices, making it difficult to access without direct communication with the scientist who generated the data. This lack of centralised storage can lead to issues with data availability, especially if the scientist has retired, resigned, or is no longer affiliated with

the institution where the data was generated. Furthermore, there may be concerns about data security and privacy, as scientists may be hesitant to share their data with others, particularly if it has not been published or made publicly available.

However, 16 (29.09%) respondents found accessing historical data produced by others relatively easier when stored in online repositories or made available through open access platforms. Personal contacts or collaborations with other scientists can facilitate access to specific datasets, especially if there is a mutual interest in sharing knowledge and fostering collaboration. One respondent found it very easy to access data because it is stored on backup drives. Accessibility of historical data produced by other scientists is often hindered by factors such as the decentralised nature of data storage, concerns about data security and privacy, and the absence of standardised practices for data sharing and preservation. Addressing these challenges may require the development of centralised data repositories, standardised data management practices, and improved communication and collaboration among researchers and institutions.

3.9 Anxiety Towards Centralised Data

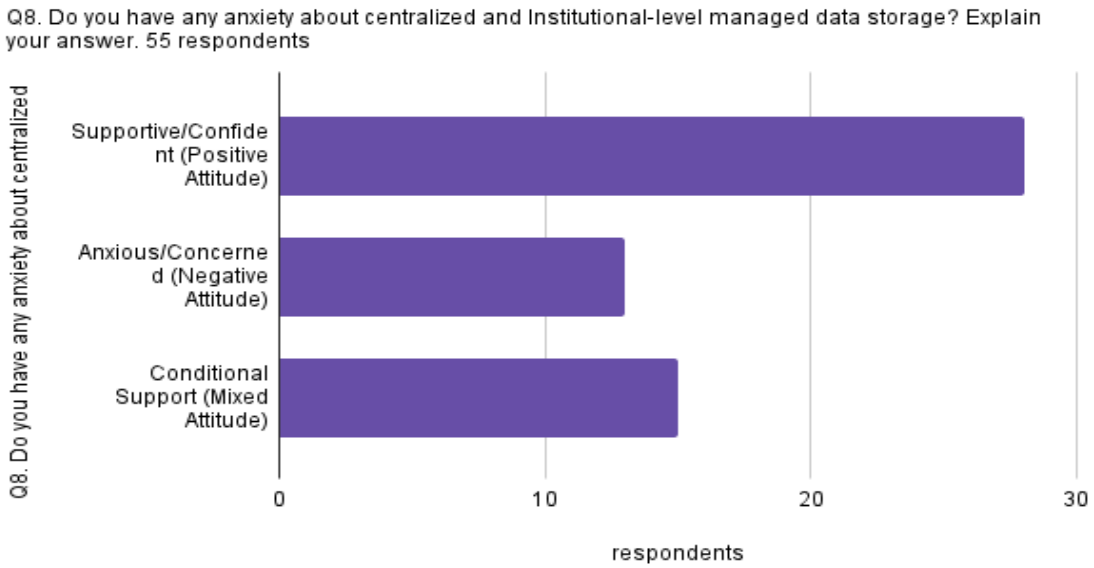


Fig. 12: A bar chart depicting attitudes towards centralised data storage

Responses to anxiety about centralised and institutional-level managed data storage vary among participants. Some express little to no anxiety, stating that as long as data remains secure

and protected from hacking, they have no concerns. Others emphasise the importance of data security and accessibility, suggesting that proper management and cybersecurity measures are crucial to alleviate any worries. Quite a number of respondents express concerns about data theft, unauthorised access, and potential misuse of their research data by other scientists or third parties. They highlight the importance of privacy protection and the need for strict authorization processes to prevent unauthorised usage or reproduction of data. Some also stress the significance of ensuring confidentiality and control over their data, particularly in cases where data may be used for personal gain without consent.

However, several respondents expressed support for centralised and institutional-level managed data storage, citing benefits such as continuity, easy access to historical data, reduced duplication of work, and enhanced collaboration among researchers. They believe that with proper management and security measures in place, centralised storage can facilitate knowledge sharing, streamline research processes, and preserve institutional memory effectively. While concerns about data security, privacy, and unauthorised usage are evident among some respondents, many acknowledged the potential benefits of centralised data storage and emphasise the importance of proper management and security protocols to address these concerns effectively. This highlights the need for CSIR-CRI to implement robust data management practices and cybersecurity measures to ensure the integrity, confidentiality, and accessibility of research data.

3.10 Data Management Training

Fig. 13: A pie chart of interest in data management training

The consensus among participants is a resounding "Yes" to the question of whether they would be interested in receiving training in data management. This widespread interest reflects a recognition of the importance of effectively managing and utilising research data in scientific endeavours. By expressing a willingness to undergo training, participants demonstrate a proactive approach to enhancing their skills and knowledge in data management, which is crucial for optimising research processes, ensuring data integrity, and facilitating collaboration and knowledge sharing within the research community. This collective enthusiasm for training underscores the value of investing in educational initiatives focused on data management, which can ultimately contribute to the advancement of research outcomes and scientific innovation.

4 Discussion

4.1 Data Collection Methods

The fact that both male and female researchers participated in the survey indicates some level of gender diversity within the research community at CSIR-CRI, which suggests that both genders are engaged in the discussion surrounding data management practices. There seems to be a relationship between the designation of researchers and their preferred methods of data collection, as well as the influence of years of experience on these preferences. *Principal research scientists, senior research scientists, and research scientists, especially those with over 10 years of experience, tend to utilise a combination of traditional and digital methods for data collection.* These methods include paper notes, handheld devices (phones or tablets), Excel spreadsheets, and online data management platforms–an indication of a level of versatility and adaptability in their approach to data collection, likely developed through years of experience and exposure to various research methodologies.

The contrasting preference for modern, digital methods among assistant research scientists and those with fewer years of experience, particularly within the 1 to 5-year range, highlights a generational shift towards technology adoption in data collection practices. This preference may stem from their familiarity with digital tools and a desire for efficiency in managing and analysing data. The reliance on handheld devices and Excel spreadsheets reflects an acknowledgment of the benefits of digitalization, such as real-time data entry, organisation, and easy sharing or collaboration.

Principal technologists, regardless of their years of experience, demonstrate a consistent preference for Excel spreadsheets as their primary data collection tool. This preference may be attributed to the nature of their role, which involves a greater emphasis on data organisation, analysis, and reporting. Excel's versatility and capability to handle large datasets make it a suitable choice for principal technologists who are tasked with conducting and overseeing laboratory and field research work, managing research projects, technical or technological initiatives within the institute. The differing preferences for digital tools among assistant research scientists and principal technologists reflect both generational trends and the specific demands of their respective roles within the research environment.

The consistent reliance on paper notes for data collection across all levels of experience within the Institute suggests that this method remains popular and trusted among researchers. This enduring preference for paper notes may be influenced by a variety of factors, including personal preferences, the specific nature of the research being conducted, and established institutional practices. Some researchers may simply prefer the tactile experience and convenience of jotting down notes by hand, while others may find it easier to quickly sketch diagrams or record observations on paper. Certain types of research, such as fieldwork where electronic devices may be impractical or unreliable, may lend themselves better to paper-based data collection. However, institutional norms and practices, such as the unavailability of resources and training in digital data collection methods, can influence the continued use of paper notes.

4.2 Data Storage Practices

The designation and years of experience of researchers within CSIR-CRI appear to influence their preferred methods of data storage, with more senior researchers tending towards personalised and multifaceted approaches, while younger researchers exhibit a greater inclination towards digital and streamlined storage methods. *Principal Research Scientists and Senior Research sScientists with over 10 years of experience demonstrate a predilection for storing their research data primarily on their own laptops, often supplemented with paper-based records or other resources such as institutional desktop computers or the CSIR-CRI library*. The preference for a dual storage approach may reflect extensive experience and familiarity with traditional methods, coupled with the convenience and efficiency offered by digital storage solutions.

Assistant Research Scientists and Researcher Scientists with fewer years of experience, particularly those with between 1 and 5 years, exhibited a preference for storing data on their own laptops, without as much reliance on additional resources like paper-based records or institutional desktop computers. This inclination towards digital storage may be attributed to their relatively shorter tenure at the institute and a greater comfort with modern technology. Principal Technologists, regardless of their years of experience, demonstrate a varied approach to data storage, with some favouring personal laptops and others opting for a combination of paper-based records, institutional desktop computers, and other resources. This could be

attributed to the diverse nature of their roles within the institute, which may involve a range of responsibilities spanning data organisation, analysis, and reporting, thus necessitating flexible storage solutions tailored to their specific needs.

4.3 Data Storage Preferences

The responses reveal a *consensus among researchers regarding the importance of centralised, secure, and accessible storage solutions for research data*. Commonly advocated options include institutional repositories and cloud-based storage solutions. However, variation in preferences stem from factors such as institutional policies, technological infrastructure, individual preferences for data management and security. Institutional repositories are perceived as secure and centralised locations for storing research data, ensuring accessibility and organisation. Cloud-based platforms are favoured for their convenience, accessibility, and potential for data backup and synchronisation.

Some researchers prioritise *data security and confidentiality*, recommending the use of secure databases or encrypted repositories to protect sensitive research data from unauthorised access or breaches. There is also a notable *emphasis on redundancy and backup strategies,* with many suggesting *storing data on multiple devices or platforms*, such as personal laptops, external hard drives, or institutional databases, to ensure data integrity and availability.

While there is a consensus on the importance of centralised and secure storage solutions, preferences vary based on designation and years of experience. Research Scientists and Technologists, especially those with over 10 years of experience or 5-10 years tend to favour storing data in institutional repositories or designated cloud platforms, reflecting a trust in established institutional systems and recognition of the importance of secure storage.

Conversely, researchers and technologists between 1- and 5-years' experience demonstrated a more diverse range of preferences. While some prefer cloud-based platforms or institutional repositories, others mention storing data on personal laptops or external drives. This variability may indicate a greater reliance on personal devices and a less ingrained adherence to institutional storage practices among younger researchers.

4.4 Challenges with Data Access and Handling

The challenges encountered by researchers in accessing and handling research data reveal underlying dynamics within CSIR-CRI, encompassing power differentials, knowledge transmission, technology access and ethical considerations. *More experienced researchers confront issues such as data loss, difficulties in accessing historical data, and concerns about data security and privacy. Researchers with fewer years of experience grapple with challenges like corrupted files, limited proficiency in statistical software and difficulties accessing data stored in physical formats when away from the office, alongside ensuring data accuracy, and managing technical issues in data collection tools.*

These disparities based on designation and experience levels highlight existing hierarchies within research organisations, with senior researchers often exercising greater control over data management processes. The challenges in accessing historical data underscore the importance of knowledge transfer and continuity across generations of researchers, while differences in technology proficiency contribute to a digital divide, necessitating efforts to provide equitable access to technology and training. Moreover, concerns about data security and ethics emphasise the need for transparent and accountable data management practices.

Addressing these implications requires fostering a supportive and inclusive research environment that promote knowledge sharing, provides access to technology access and upholds ethical standards to ensure the integrity and transparency of research practices (Williamson and Leonelli 2022).

4.5 Data Management Support

Responses regarding the desired support for managing and analysing research data reflect diverse needs and preferences among participants. Emphasis is placed on institutional support, including user-friendly institutional repositories, accessible cloud storage and reliable data repositories and software. Additionally, there is a shared recognition of the importance of training and capacity building in data management and analysis, as well as the need for institutional involvement in providing centralised data storage solutions and ensuring data security. Furthermore, there is consistent demand across all designations and experience levels for institutional support highlighting the significance of organisational infrastructure in facilitating effective data management practices. Experienced researchers prioritise access to

user-friendly digital tools, advanced training in data analysis software and institutional data storage platforms, whereas those with fewer years of experience stress the importance of accessible cloud storage, digital tools for data collection, and training in modern statistical software.

4.6 Access to Primary Historical Data

Access to historical data among individuals of various designations and experience levels reveal patterns in data management strategies. Overall, reliance on technology for storing historical data is common, with mentions of laptops, external drives, cloud storage and email. There is a preference for digital storage solutions such as cloud storage and online platforms like Google Scholar, among individuals seeking ease of access to their data. Additionally, the emphasis on personal responsibility for data management, as seen in statements like "I consciously put this info online" and "I keep it myself," aligns with traditional notions of self-reliance and control. However, challenges such as computer crashes, data corruption and the importance of backups are noted, reflecting a cultural inclination toward valuing efficiency and convenience in managing information.

Researchers with longer tenure at CSIR-CRI often encounter challenges including difficulties in finding files, data loss due to computer issues and reliance on paper-based records. Strategies to facilitate access such as storing data online or in cloud storage devices are mentioned. Newer employees reported difficulties attributed to unfamiliarity with data management systems and reliance on outdated storage methods like paper sheets or external hard drives. Principal Technologists, regardless of their experience, seem to have relatively better access to historical data, possibly due to their technological proficiency and organisational skills in utilising cloud-based storage solutions and labelling data across multiple devices and platforms.

The emphasis on technology in data management practices is evident, with researchers utilising cloud storage or online platforms finding it easier to access historical data compared to those relying solely on physical storage methods. This underscores the importance of technological infrastructure and digital literacy in shaping individual experiences with data management.

4.7 Access to Secondary Historical Data

The ease of accessing historical data produced by other scientists unveils notable dynamics within the scientific community. Some respondents expressed difficulty in accessing such data,

attributing it to the individualistic data management practices, or challenges of retrieving information from retired scientists. Comments suggest a sense of exclusivity surrounding the data, with indications that scientists may withhold their data for personal gain or that accessing it requires a personal connection with the scientist. This hints at a cultural norm among scientists to closely guard their data, possibly stemming from competitiveness within the research community or viewing data as personal intellectual property, thus prioritising its preservation over collaboration. The emphasis on the challenge of accessing data unless it's published aligns with traditional notions of academic competition and the pressure to produce original research, potentially perpetuating barriers to collaboration and knowledge sharing (Fecher et al, 2015; Tenopir, et al, 2011).

Others also noted challenges in accessing historical data but emphasise strategies like direct communication, lack of proper organisation, or reliance on published reports. However, there's an added layer of complexity such as instances where Division Head retire without proper handover, leading to difficulties in accessing crucial data. This highlights potential systemic issues within the research institution, where processes for data management and knowledge transfer may lack formalisation or rely on interpersonal relationships. The emphasis on interpersonal communication and collaboration suggests a more communal approach to knowledge exchange, where building relationships and leveraging networks are essential tools for accessing valuable data.

These differences in navigating the accessibility of historical data underscore the complex interplay between individual agency, institutional dynamics, and broader cultural norms within the scientific community. The use of personal connections, informal communication channels and individual devices further reinforces stratification within the organisation, where data access is influenced by professional standing, interpersonal networks and technological resources. These challenges reflect broader issues within the research institution including data management, knowledge transfer and institutional memory, and hinder scientific advancement. Addressing these challenges requires a multifaceted approach considering both structural and cultural factors to foster a more equitable and collaborative research environment at CSIR-CRI.

4.8 Anxiety Over Centralised Data

While centralised and institutional-level managed data storage is seen as a promising avenue for enhancing data accessibility and continuity, concerns about security, privacy, and unauthorised use persist among respondents, with variations influenced by individual

perspectives on data protection and control. Some expressed confidence in centralised data storage, believing that as long as the data is well-protected and secured, there is no cause for anxiety. They perceive centralised storage as beneficial for ensuring continuity, easy access to historical data, and preventing data loss due to staff turnover or unforeseen events. However, others harbour reservations, particularly regarding potential data breaches, theft, or misuse. They worry about unauthorised access by competing scientists or researchers, as well as the risk of someone using their data without consent for personal publication or benefit. Additionally, concerns about data privacy and confidentiality emerged, with respondents emphasising the importance of stringent security measures and policies to safeguard sensitive information.

It is noticeable that respondents across different designations and tenure levels exhibit varying levels of anxiety regarding centralised data storage. Those in higher positions, such as Principal and Senior Research Scientists, generally expressed less anxiety or no anxiety at all. They often emphasise the benefits of centralised storage, such as continuity, accessibility, and the prevention of data loss due to staff turnover. Respondents in lower positions, such as technologists or those with fewer years of experience, tend to express more concerns about data security and privacy. They highlighted potential risks such as data theft, unauthorised use, or misuse of data, particularly if proper management protocols are not in place. This suggests that anxiety about centralised data storage may be influenced by factors such as organisational hierarchy and level of responsibility.

There are nuanced perspectives within each designation and tenure category. Some respondents express conditional support for centralised data storage, contingent upon robust cybersecurity measures and strict policies to safeguard data integrity and privacy. Others raised specific concerns, such as the possibility of unauthorised data usage or the risk of data theft by competing scientists or third parties. Addressing these concerns requires comprehensive strategies that balance accessibility with robust security measures and clear policies to ensure responsible data management and usage.

References

Fecher, B., Friesike, S. & Hebing, M. (2015). What drives academic data sharing? PLoS One. 25;10(2):e0118053. doi: 10.1371/journal.pone.0118053. PMID: 25714752; PMCID: PMC4340811.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, AU, Wu L, et al. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6: e21101 10.1371/journal.pone.0021101 [PMC free article][PubMed] [CrossRef] [Google Scholar].

Williamson, Hugh F. & Leonelli, Sabina (eds.) (2022). Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development. Springer Verlag.

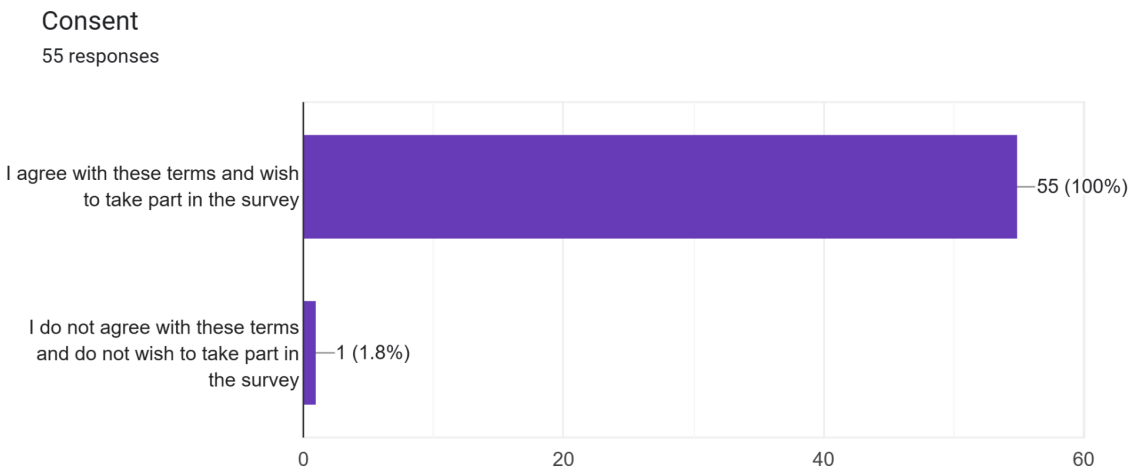
Appendix 1

Information Section

This survey is carried out by the PHIL_OS research team, in collaboration with the management of the CSIR-Crops Research Institute (CRI), for the purpose of understanding the data management habits and needs of CSIR-CRI staff. Results will be used to inform the future Data Policy of CRI, as well as the ongoing study of open research practices by the PHIL_OS team (www.opensciencestudies.eu). The survey is intended to be anonymous and all efforts have been made to ensure anonymity. Responses to the survey are automatically anonymized and aggregated in a manner neither allowing for the identification of individual respondents nor for the attribution of individual responses to a respondent. Although your email address may be sent along with the answers, your specific responses will not be linked to you in any way once the responses are posted online or on the Institutes' website. For any further information, please contact Joyce Koranteng-Acquah (jk677@exeter.ac.uk) or Sabina Leonelli (s.leonelli@exeter.ac.uk).

Consent

This section sought consent from the respondents.



Appendix 2: Results

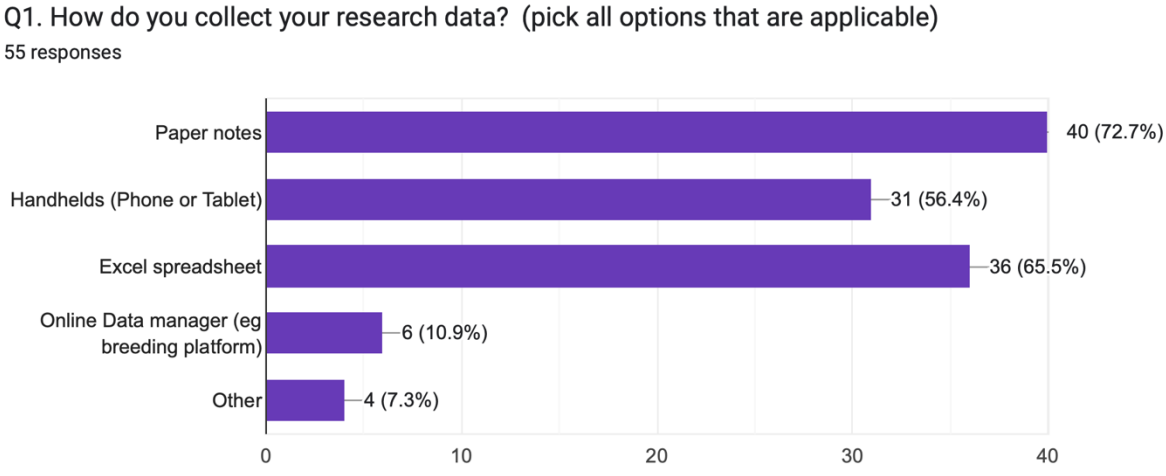
Demographic Data

Demographics	**Gender**		**Sub-total**
	Male	**Female**	
Designation			
Chief research scientist	0	0	
Principal research scientist	3 (5.5%)	1 (1.8%)	4 (7.3%)
Senior research scientist	12 (21.8%)	5* (9.1%)	17* (30.9%)
Research scientist	8 (14.6%)	4* (7.3%)	12* (21.9%)
Assistant research scientist	2 (3.6%)	1 (1.8%)	3 (5.4%)
Chief technologist	0	0	0
Principal technologist	10 (18.2%)	7 (12.7%)	17 (30.9%)
Other	2 (3.6%)	0	2 (3.6%)
Total	**37 (67.3%)**	**18 (32.7%)**	**55 (100%)**
Educational Background			
PhD	23 (41.8%)	9 (16.4%)	32 (58.18%)
Masters	12 (21.8%)	9 (16.4%)	21 (38.2%)
Undergraduate	2 (3.6%)	0	2 (3.6%)
Diploma	0	0	0
Other	0	0	0
Total	**37 (67.2%)**	**18 (32.8%)**	**55 (100%)**
Years of Service			
>10	21 (38.2%)	14 (25.5%)	35 (63.7%)
5-10	6 (10.9%)	1 (1.8%)	7 (12.7%)
1-5	9 (16.4%)	2 (3.6%)	11 (20%)

<1	1 (1.8%)	1 (1.8%)	2 (3.6%)
Total	**37 (67.3%)**	**18 (32.7%)**	**55 (100%)**

Table 1:

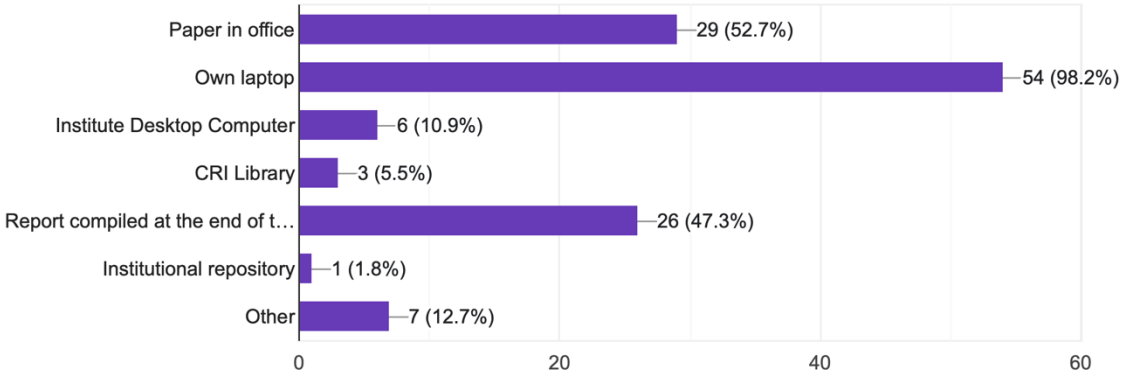
Results



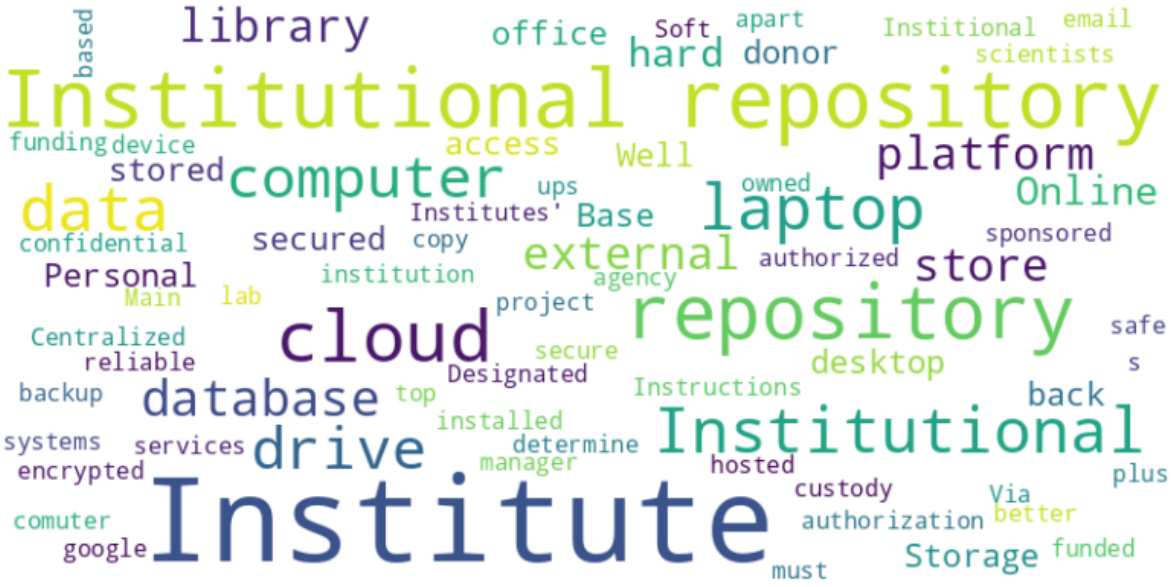
Word Cloud for Q2. How do you store your research data? (pick all options that are applicable)

Q2. How do you store your research data? (pick all options that are applicable)

55 responses



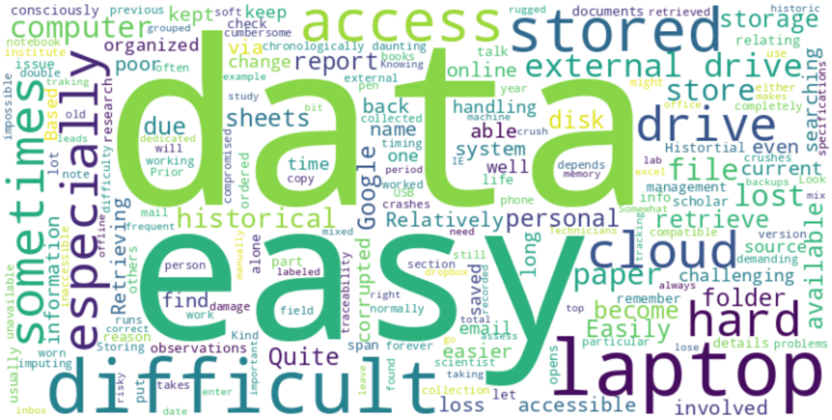
Word Cloud for Q3. Where do you think your research data should be stored? (Free text)



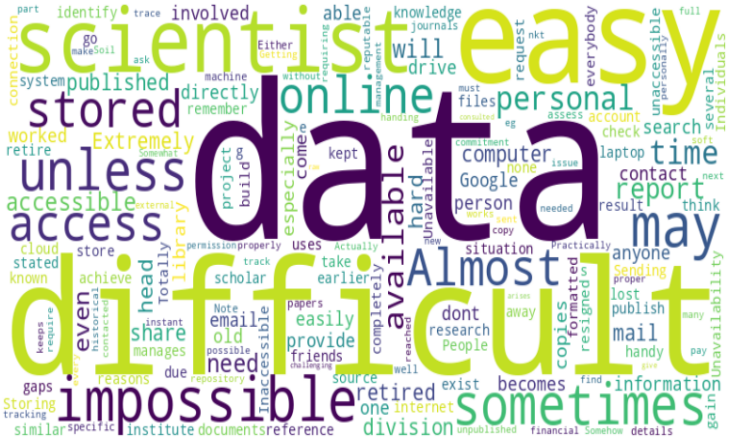
Word Cloud for Q4. What challenges do you experience with respect to accessing and handling research data? (Free text)

Word Cloud for Q5. What kind of support would be most helpful to help you manage and analyse your data? (Free text)

Word Cloud for Q6. How easy is it to get access to your own historical data? Explain your answer (Please include possible examples)



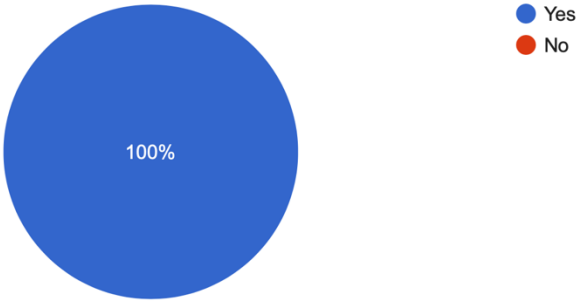
Word Cloud for Q7. How easy is it to get access to historical data produced by other scientists? Explain your answer. (Please include possible examples)



Word Cloud for Q8. Do you have any anxiety about centralized and Institutional-level managed data storage? Explain your answer.

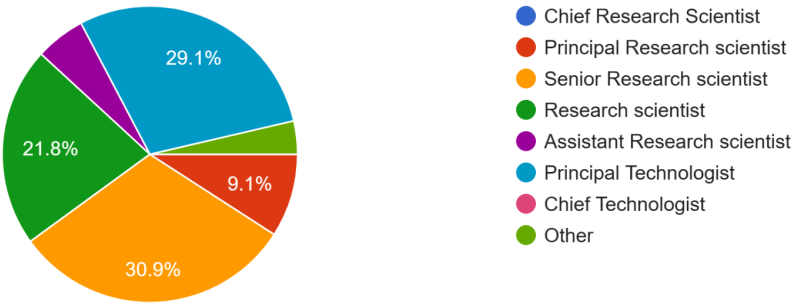
Q9. Would you be interested in getting training in data management?

55 responses



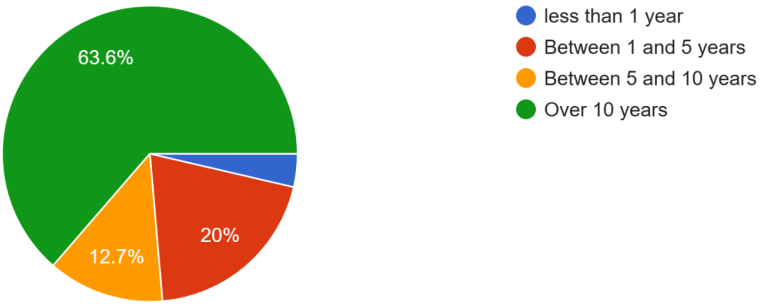
Q10. What is your designation at CRI?

55 responses



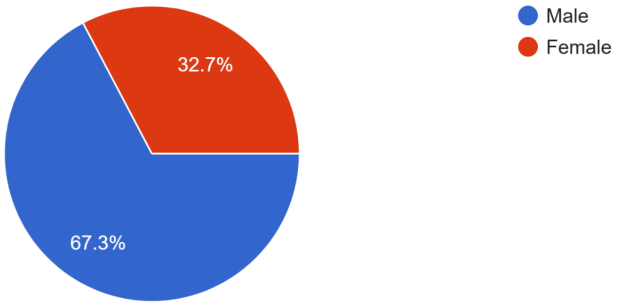
Q11. How many years have you worked at CRI?

55 responses



Q12. What is your gender?

55 responses



Q13. What is your highest educational degree?

55 responses