Sustainable and FAIR data sharing in practice in agriculture contexts in Malawi – Successes and challenges

Background of the project

[The Malawi Digital Plant Health Service](#) (MaDiPHS) is a 5-year project to develop a tool for targeted and efficient pest and disease management of five crops (maize, tomato, cassava, groundnuts and banana) and their pests in Malawi. The service will combine national, regional, and global input data with international web platforms to feed into and create a national client adapted to meet the local user needs. The project leverages existing data assets first before collecting new data for the development of the national digital plant health service. [Our first report](#) explored how CABI implemented FAIR in MaDiPHS, and this case study captured learning from partners regarding FAIR implementation in the project.

Where is the data?

In many data-rich development projects, effective data collection and sharing are key components of a successful initiative. The types of data collected may include but are not limited to research data, advisory materials, reports, and monitoring and evaluation data. However, not all projects consider the sustainability of the datasets collected, meaning the data shared only has limited scope in a particular project. The data is, therefore, usually kept (if indeed it is kept at all) by an institution or an individual, until someone is interested in the insights of the research and asks about the availability of the data.

This challenge is evident in MaDiPHS, as its National Coordinator, Dr Jonathan Mkumbira, outlined: "Data is usually just data and is not supposed to be shared with others, but the information and insights generated from the data are most important. Only when people are interested in the insights do they then ask if the researcher can share the data."

Figure 1: Dr Jonathan Mkumbira (Credit: CABI)

Expanding on this, Dr Mkumbira explained that data is usually kept on an individual's laptop and may not even be shared across individual institutions. There was no data server to centralize data for sharing on an institutional level either.

"There hasn't been much data sharing because it is protected by individual organizations"

Dr Jonathan Mkumbira

Malawi Digital Plant Health Service (MaDiPHS)

In government and academic institutions, many research activities are supported by donor-funded projects. One of the interviewees noted that donors have different requirements on intellectual property rights, making it unclear what the terms are for data sharing. This observation chimes with our findings regarding the lack of a shared definition covering key terms ('open data') and publishing principles in donor open data policy, which result in fears around data sharing among partners (Smith, Fawcett, Musker, & CABI, 2017).

Haswell Dambolachepa, Plant Pathologist (MSc.) of the Department of Agricultural Research Services, agreed that there has been resistance to data sharing. Sharing data collected under an ongoing donor-funded project is even more challenging.

"Some scientists thought that it was not possible to share data because there is sensitive data, or they were collected from a different project", he said. Sensitive data includes personal details or pest surveillance data on a pest that has not become a threat in Malawi, which may affect the export value of the crop of concern.

These are the issues that FAIR – Findable, Accessible, Interoperable, and Reusable – data principles can help address in order to facilitate a dialogue on what and how data can be shared.

Why the FAIR data principles matter

Since the inception phase of MaDiPHS in 2022, FAIR data management has been emphasized in the project, which has laid the foundation for effective data sharing across partners.

CABI has organized eight data-sharing and management capacity-strengthening workshops[1] to explain the values and benefits of FAIR data principles, and how to operationalize them using the FAIR Process Framework and the tools and resources associated with it.

Malawi national partners had notable input into the development and signing of the programme Data Sharing Agreement (DSA), an exercise that has built trust between data holders and users. Each agreement provides a space to discuss the extent to which datasets should be open or closed to sharing and usage. This has enabled partners such as the International Centre of Insect Physiology and Ecology (icipe) to use the data in pest risk modelling even when the data holders have not yet published it.

Figure 2: Mr Dambolachepa (middle) and Mr Blessing (right) in a key informant interview in Malawi (Credit: CABI)

"[Before data sharing] we were depriving people from getting information that they were supposed to get; we didn't want to release it, because we were afraid that if they got access to this data and information, they might change it", Mr Blessing Susuwele, National Coordinator of the Department of Agricultural Extension Services (DAES), said. "With the training and signing of DSAs, some sections of the datasets that we don't want to release probably we can put conditions [on]", he added.

---

[1] The workshops are part of Work Package 2 of the MaDiPHS project. For more information about how the workshops implemented the FAIR Process Framework step by step, please visit: https://gatesopenresearch.org/documents/8-83

Seven DSAs have been signed between Norwegian Institute of Bioeconomy Research (NIBIO) and the seven partners. Datasets are now being widely shared and used for pest modelling.

Mr Dambolachepa said that, with the help of DSAs, he can explain to fellow scientists how data sharing will help farmers.

Data modelling and the interoperability of data

Brian Kanji, Machine & Deep Learning, Spatial Data Science Research Consultant of icipe, explained that metadata is the key to securing the 'Interoperability' and 'Reusability' of data. "Metadata provides clear descriptions of data attributes to facilitate understanding of the datasets", he said.

He explained that since the existing datasets were collected for different purposes, it is vital for data users to understand the original data collection methodology to enable the proper use of the data in modelling.

Dr Elfatih M. Abdel-Rahman, Research Scientist of the Data Management, Modelling and Geo-Information Unit of icipe, reiterated that requirements for the usability of the data can be very specific to the modelling.

"We go through the datasets and see the usability of it, and we come up with questions […] in terms of metadata and variables that were not well explained. Once all these questions were answered we can identify which part of the data is useful or not", Dr Elfatih said.

Where the metadata is clear, the icipe's modelling team has used the datasets shared to generate four pest risk models for the MaDiPHS project (the target is risk modelling of nine pests), which will be integrated into the national digital plant health service and what has been termed the 'International Platform'. The International Platform combines national, regional, and global input data intended for future use, such as in the development of agricultural decision-support tools and data modelling in the country and beyond.

"The data is now FAIR", Dr Elfatih said. He agreed that the datasets shared would be FAIR and sustainable for the benefit of the people of Malawi.

His colleague Dr Komi Mensah Agboka, Geospatial Data Science and Artificial Intelligence Expert at icipe said: "FAIR data sharing is very good because some

of those pests, for instance, are transboundary. When we talk about fall armyworm [it] doesn't have any boundary."

Additional capacity strengthening within national partners

Implementing the FAIR data principles motivates national partners to not only share data for the success of this project but has also catalysed an understanding of the need for long-term data management capacity strengthening on a national level. For instance, the DAES is now digitalizing advisory resources. It feels obliged to share them and improve their accessibility well beyond the MaDiPHS project.

"Because we [the Ministry of Agriculture] are like the hub for most of the advisories, for most of the data, we should keep the data very well and [make it] easier for data users to access them", Mr Blessing Susuwele said.

Figure 3: Idah Mwato, Deputy Director of DCD, in an interview with CABI. (Credit: CABI)

Idah Mwato, Deputy Director of the Department of Crop Development (DCD), told us: "As an organization, data sharing is good, which does not mean that we always disclose everything but to open it to let others understand what we are doing."

Ms Mwato pointed out the importance of cataloguing the datasets on the Comprehensive Knowledge Archive Network (CKAN), a data repository platform that has been developed to locate where the datasets are: "We now know what other departments are doing, and we can respond to queries more effectively even at home", she said.

"People from other countries can use our data to help people in Malawi."

Idah Mwato

Deputy Director, Department of Crop Development

Emmanuel Mwanaleza, Principal Statistician at the Department of Agricultural Planning Services, also agreed with his colleagues in DCD. "I don't have to make calls and send emails to request or share datasets anymore", he said, and this has increased his work efficiency.

All of the national partners we interviewed who have shared datasets or advisory materials agreed that the MaDiPHS data catalogue should be maintained and expanded for sustainable FAIR data sharing and management beyond the project.

Sustainable "FAIRness" of data

Despite evidence of additional skills developed and data being reused in modelling, there are still challenges to data sharing.

The institutions we interviewed often rely on donor-funded projects to conduct research. Data needed for MaDiPHS modelling comes from several different projects. Some scientists think that sharing data between different projects could be seen as redistributing resources funded by one donor to benefit another donor's project, while others said they would not share data until after it had been processed or published. These worries have been addressed through the signing of DSAs, but they raise the question of what more we, as donors, could do – or do differently – to alleviate these concerns (Smith, Fawcett, Musker, & CABI, 2017).

Another pre-condition for effective FAIR data sharing is data management capacity, ranging from technical skills to ensure the interoperability of data between systems to the availability of data management infrastructure such as a data storage server. The discussion in Malawi has been focusing on uploading data and resources to the government's data storage server[2], which is considered to be secure in terms of 'ownership' and 'control' of data. Although

---

[2] https://www.datacenterdynamics.com/en/news/malawi-launches-national-data-center/

cloud computing has received much attention in development contexts, especially given its ubiquity in AI applications, many national contexts including Malawi do not accept it as a go-to solution. This is in part because of culture and concerns about ownership, but also because policy does not easily facilitate its use in government systems.

As partners from icipe said, leveraging existing datasets is not just about sharing what data exists, but what should or might in the future. There are challenges around missing data variables such as geographic coordinates, or the data methodology may not be clear.

Effective communication between data users and data providers is needed to clarify which part of the data can be used for which specific purposes. Thus, questions arise around how to ensure the reusability of the data at the point of starting planning for data collection.

Dr Jonathan Mkumbira said, "Training on FAIR data principles and data quality and standards should start in the early career of researchers, or even in secondary schools for sustainable FAIR data sharing."

However, the ongoing major concern from national partners is data storage at this stage. If the data collected has no obvious place to go in the long term, one cannot start the conversation on sustainable data management, let alone FAIR data sharing. Data storage and other infrastructural issues should be given more attention earlier in any future similar initiatives, through conducting a needs assessment before proposal development to ensure all relevant government departments are involved from the outset. In the case of MaDiPHS, it is recognized that the project team should have engaged with the Department of E-Government much earlier.

Conclusion

The positive experience of MaDiPHS partners in regard to FAIR data sharing has demonstrated the importance of integrating FAIR data management at the outset of project proposal development. This engagement approach has enhanced partners' understanding of the value and benefits of FAIR and responsible data management, trust building between data holders and users, and knowledge transfer of FAIR data at an institutional level.

While issues of 'Findability' and 'Accessibility' have been largely addressed, there are technical and contextual challenges to advancing 'Interoperability' and 'Reusability'. The FAIR Process Framework usefully emphasizes a human-centred approach to data management to help address such issues. Successful and sustainable data sharing requires collaboration and learning throughout the FAIR implementation journey between donors, grantees, and partners. This story seeks to support donors to rethink their grant-making approach in data-rich investments and help other data ecosystem actors to reflect on their roles in advancing FAIR data sharing for sustainable agricultural transformation. Such challenges could best be addressed at a donor level, as advocated by CGIAR (CABI, 2024), as well as via a bottom-up approach at a national level, such as the case in Ethiopia (CABI, 2024).

Recommendations

There are four major recommendations for improved data management and sharing within and beyond the MaDiPHS project arising from insights obtained from interviews and field observations:

- **Engagement with senior decision-makers** in the Ministry of Agriculture, the Department of E-Government, and ICT specialists should be undertaken **to discuss the way forward for storing the datasets collected in this project on a national level** (e.g. the National Data Centre[3]), and a plan for maintaining the MaDiPHS data catalogue should be formulated. This will ensure ownership of the catalogue across relevant departments in government.
- **Leverage existing policy** (in draft or enacted) **related to digitalization/data sharing** on a national level to support the provision of data storage infrastructure, capacity strengthening on FAIR, and build ownership of the data catalogue and the FAIR data-sharing agenda at senior levels of the

---

[3] https://www.datacenterdynamics.com/en/news/malawi-launches-national-data-center/

Ministry of Agriculture, supporting infrastructures that are foundational components of the International Platform.

- **Strategic partnerships should be formed with intergovernmental bodies to secure long-term investment** in the capacity strengthening of researchers, scientists, and students in basic data management, FAIR data implementation, and technical capacity in maintaining ICT systems and data repositories such as CKAN. This will enable data storage and data sharing for future research, data modelling, and the development of user-friendly digital decision support tools.
- **Improve the data management planning process** and **include data users in the planning process** before commencing new data collection to better meet the needs of data users.

Acknowledgements

---

[4] CABI's Data Policy and Practice team launched the FAIR Process Framework in November 2024 at the Bill & Melinda Gates Foundation and will continue to integrate its use in different projects to enhance the applicability of the Framework in different projects and contexts.

References

Bill & Melinda Gates Foundation (2024). *Fair Process Framework*. Developed by CABI. Retrieved 2024, from https://www.fairprocessframework.org/

CABI (2024). *The lasting impact of advocating for and facilitating FAIR-informed practice in soil and agronomy – the case of the Coalition of the Willing (CoW) and Ethiopia's Soil and Agronomy Data Sharing (SADS) directive. [version 1; not peer reviewed]*. Gates Open Res 2024, 8:26 (document). doi:https://doi.org/10.21955/gatesopenres.1117085.1

CABI (2024). *Understanding the enablers and disablers of mainstreaming FAIR – the Case of CGIAR [version 1; not peer reviewed]*. Gates Open Res 2024, 8:25 (document). doi:https://doi.org/10.21955/gatesopenres.1117084.1

Smith, F., Fawcett, J., Musker, R., & CABI (2017). *Donor open data policy and practice: an analysis of five agriculture programmes [version 1; not peer reviewed]*. F1000Research 2017, 6:1900 (document). doi:https://doi.org/10.7490/f1000research.1115013.1