

Module 1

Improving food security through harmonised soil data in South Asia

Case Study



Improving food security through harmonised soil data in South Asia

What the Cereal Systems Initiative for South Asia learned about making data findable, accessible, interoperable and reusable as they developed a soil intelligence system.

Summary

The Cereal Systems Initiative for South Asia (CSISA) improves farmers' access to market information. CSISA, led by CIMMYT in collaboration with IFPRI and IRRI, along with project partners (ICRAF, ISRIC, Cornell University, and CABI) and national stakeholders, planned to develop a Soil Intelligence System for India to support small and marginal farmers with evidence and insights to inform soil agronomy practices. To achieve this, the project intended to make data within the system findable, accessible, interoperable and reusable (FAIR).

¹ Cereal Systems Initiative for South Asia, <https://csisa.org/>, Accessed 1 December, 2020

Key learnings

- Working with partners and stakeholders from the beginning can help to understand and address barriers to sharing data.
- Identifying potential risks and impacts from data sharing early on can help establish the processes, guidance and support required to manage them.
- Engaging with third party data providers is critical to ensuring you have the permissions you need to use the data as you intend within your project.

The challenge

Intensive cropping systems that include rice, wheat and maize are widespread throughout South Asia. These systems constitute the main economic activity in many rural areas and provide staple food for millions of people. The rate of growth of cereal production in South Asia is declining. Simultaneously, issues of resource degradation, declining labor availability and climate variability pose steep challenges for achieving the goals of improving food security and rural livelihoods.¹ The Cereal Systems

Initiative for South Asia (CSISA) aims to improve productivity of cropping systems in South Asia by generating and disseminating new knowledge on cropping system management practices that can withstand the impacts of climate change.

Development of a Soil Information Service (SIS) for India was planned under this program.

The SIS would incorporate existing data assets and use low cost ways to provide soil data for the country. The program planned to make dashboards available to governments, researchers and private sector partners to highlight the status of key soil health indicators. They planned to embrace FAIR data principles to support better decision-making in agriculture and to inform the development of the SIS.

Much time was spent focusing on the technological components of the grant, such as the algorithms needed to map the data. Although a critical component of the project, the focus on this aspect led to reduced development of appropriate processes to access the data needed to fuel these models.

Despite reference to FAIR in the proposal, as the project progressed the soil data was not being produced in a way which was findable, accessible, interoperable and reusable.

Challenges included:

- **Findable** – The data collected for the project was stored on a government database but the key information about the data being held (metadata) was not available to the partners or public, making it difficult for them to find what data was being held and where.

- **Accessible** – There was a lack of awareness for how to manage the risks of publishing the data held by the project, especially as it pertained to the management of personal data and mechanisms for how partners could access what data and when were not considered. As such, data remained inaccessible aside from a small number of government employees. Requests for access were mostly conducted via email and success was dependent on personal relationships.
- **Interoperable** – Although data was being collected via an automated system, it still needed to be processed to make it usable for partners. The standards for (meta)data were developed internally, without reference to external partners, making it more difficult for them, and others, to easily reuse the data. Interoperability, however, was seen as less of an imminent challenge than the other principles which were blocking finding and using the data at all.
- **Reusable** – As data was frequently not published, and never with a licence explaining permissions for use, the conditions for how project partners could reuse the data outputs of the project were unclear

The solution

Working with CABI and the Open Data Institute (ODI), the project team worked to change the processes to maximise successful FAIR data sharing. In order to develop the appropriate frameworks, the data ecosystem for the project was mapped, in order to identify different types of stakeholder who could promote effective data sharing in the system.² A map for the project produced [here](#). The key stakeholder groups can be categorised as:

² (2019) Open Data Institute, 'Data Ecosystem Mapping Tool', <https://theodi.org/article/data-ecosystem-mapping-tool/>. Accessed October 2020.

- **Enforcers** – These are the people who are willing and able to create the policies and frameworks for data sharing in a system. They need to have the ability to follow up and ensure the processes put in place are followed and fit for purpose. In this project, the Senior Leadership team in CSISA were identified as enforcers.
- **Followers** – These are the people who work within the frameworks set by the enforcers. They use data to create information in the form of products, services, analyses or insights. They need to be trained to ensure they can effectively understand and follow any rules set. In this project, the local CSISA team are followers.
- **Advocates** – They need to understand the change that is required and why it is important. They also need to be able and willing to persuade other people in a project or institution to this way of thinking. In this project, this was a combination of the CSISA Senior leadership team and Bill & Melinda Gates Foundation program officers.

The nomenclature of these stakeholder groups is still in development, however, whilst the exact names may change, the roles and responsibilities of the stakeholder groups will remain the same.

Once the appropriate stakeholders were identified, the CSISA team worked with CABI and the ODI to identify the broader challenges which resulted in the specific challenges described above. This was done through a combination of workshops, interviews and informal conversations. High level challenges identified were:

- **Low data quality** – Data quality and collection methods vary from organisation to organisation. Agriculture Extensions Agents are not motivated or incentivised to share quality data.
- **Lack of trust in-country and government silos** – Data sharing was not considered to be an important part of the project at a senior level, and presented a significant change in ways of working. This meant the policies and frameworks to make data findable, accessible, interoperable and reusable were not defined or prioritised.
- **No clear ownership** – Approaches were proposed to mitigate the problems with lack of sharing but without clear ownership for articulation and implementation they never got off the ground.
- **Negative impact on other deliverables** – The ad hoc processes in place were suitable to meet current deliverables, but were not sustainable long term. Data sharing was seen as low priority, and something to be thought of at the end of the project. Changing these processes would take time and effort and may impact on the ability to meet deadlines for other deliverables.
- **Few incentives** – There seemed to be little to gain in the short term from working on data sharing principles.

The main challenges identified were around attitudes and behaviour change, rather than technological difficulties; the project team were very skilled in the technological aspects of the project. Therefore, the priority next steps were to develop relationships with these stakeholders and help build trust, skills and knowledge in and of data sharing. The team achieved this through a range of activities:

- Co-developing engagement strategies to articulate short, medium and long term goals. These were built around the challenges faced by national stakeholders and approaches to overcome them.
- Workshops were found to be useful for bringing all the stakeholders together, helping to develop understanding of the different parts of the project and the challenges each faced.
- Establishing a formal working group for data sharing, known as the Convergence Platform. This was a group made up of enforcers, followers and advocates committed to working together to test approaches and training. This group was also the driving force in the development of new data sharing agreements developed for the project.
- Informal conversations and consistent engagement ('knocking on doors') from the CABI team to keep data sharing on the agenda and to drive home its importance.
- Training sessions to improve data knowledge and skills, tailored to the needs of each stakeholder group. Topics included:
 - Building data sharing agreements
 - Personal data management
 - Minimising harmful impacts from data
 - Improving the quality and interoperability of data
 - Building healthy data ecosystems

- Shadowing the CSISA local team during data collection as a way to meet their networks and understand their challenges, then highlighting these challenges to CSISA leadership through workshops and training sessions.

The impact

By convening domain experts, research institutions, public sector bodies and other key stakeholders CSISA have developed a recognised and well-established community to address challenges around data sharing. This community is now committed to developing policies and establishing a framework for governance of data to ensure the Soil Intelligence System is sustainable beyond the initial funding period. National stakeholders are now more comfortable in sharing their challenges with others and more open to accepting the shortcomings in their existing processes. They feel more able to ask for support in future to bring in the culture change within their organisations.

CSISA have developed new data sharing agreements which ensure that data is managed and shared using a FAIR framework. These include:

- **Findable** – Metadata must now be published on the CSISA Dataverse (an online data store) so partners and the public can find what data is being collected by the project.
- **Accessible** – Gradually introducing tools, such as guides to developing data sharing agreements, to enable understanding of how these can be of benefit to all stakeholders. The project team has developed clear data sharing agreements which dictate how and who can gain access to the data, the use of personal data, and ownership and

accountability across institutions. According to the terms of this data sharing agreement, the data will be published on the CSISA Dataverse no later than 12 months after its collection and will be openly accessible to all. The data will be anonymised in line with the [Government of India's personal data policy](#).

- **Interoperable** – The project team have begun identifying the stakeholders and systems necessary to integrate data with so they can develop standards to include in the agreements. They have also begun training sessions with project partners.
- **Reusable** – Metadata records must now include details of licences which tell people how they can reuse the data.

Lessons Learned

Taking time at the beginning of the project to consider and understand attitudes and levels of comfort with sharing data and potential harmful impacts would have saved the team time later on. Likewise, considering data skills and knowledge of delivery partners is also a key way to identify training or specialist support required to understand suitable mechanisms and approaches for accessing, using and sharing data in the context of the intended use.

Understanding the political, social and regulatory context (the enabling environment), along with any core constraints, can help assess the feasibility of an initiative. This can also help decide whether any additional or parallel work would be helpful to ensure the success of an investment.

The collaborative approach has been key to understanding the cultural and practical challenges each actor faces in accessing, using and sharing data in the Soil Intelligence System. This shift in mindset has also meant that tools are beginning to be used to share data within and between program partners. Gaining a common understanding of the importance of sharing data and what is needed to do this well, has meant the project is able to ensure the right support is in place to achieve their objectives. Confidence with the language around data sharing has been developed through training and stakeholders now have access to the terminology to discuss this in a way which is productive and helpful.

A further major lesson was that it takes a long time to change systems, cultural norms and behaviours. This was further impacted by:

- Political changes – Where there were changes in government at a state level, this sometimes led to a stall in further engagement in that state. This led to a shift in focus from one state to another and represented a delay in the work being conducted. By being aware of and responsive to the political landscape, the project team were able to pivot their efforts to locations where they had the highest likelihood of success.
- Long term vision and incentives – Although included in the project proposal and in Foundation policy, for grantees there was little understanding and no ownership of the long term goals for making data FAIR and safeguarded. This made it difficult to incentivise effective data sharing in the long term. Through consistent engagement and conversations with program officers, this did change but took a long time to do so.

- Gradual approach – The CSISA team wanted to take a slow and pragmatic approach to this intervention and not force significant change on their project partners with no warning. Although this resulted in short term delays, by slowing down the pace and ensuring they were working with people to develop the approach, they have maximised the likelihood of changing the behaviors of their partners and colleagues.
- A global pandemic – In the absence of travel it became difficult to keep ‘knocking on doors’. This made it challenging to maintain the necessary relationships. By focusing on remote training and workshops, and using collaborative remote tools such as **Jamboard** and **Miro**, the project team were able to keep momentum going despite challenging circumstances.

Future Plans

The CSISA team have begun to see the value of improving access to FAIR data and have conducted some important work to embed this in their project. Whilst new data sharing agreements and internal changes within CSISA have begun, the team will be working with more national stakeholders and project partners during the rest of the project to help advance the environment in which the project sits for data sharing. This has been started already, with the CSISA team taking a leading role in organising virtual training sessions with different partners to help develop relationships and begin to build data literacy.

Further resources

If you are planning to make FAIR data a core component of a project in a similar way to this case study there are resources in the Data Sharing Toolkit to help you do this and manage similar challenges:

- Module 2 – Assessing in-country potential for data sharing
- Module 3 – Reusing data from third-party sources
- Module 4 – Protecting individual's rights when sharing data
- Module 5 – Sharing data through data licensing
- Module 6 – Minimising harmful impacts from data sharing
- Module 7 – Ensuring sustainable access to data

Data Sharing Toolkit



ACKNOWLEDGEMENTS

This document was authored by the Open Data Institute and CABI as part of a Bill & Melinda Gates Foundation funded investment.

The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or CABI.

datasharingtoolkit.org

DOI: 10.21955/gatesopenres.1116741.1

cabi.org | theodi.org | gatesfoundation.org

 **CABI** Data Sharing Toolkit



BILL & MELINDA
GATES *foundation*



Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license.