

## 5.2 Data Management and Access Plan (DMAP)

Developing a living document that outlines how data will be collected and handled through your investment's life

### Why should I do this?

To guarantee that your data-handling practices are not only efficient and regulatory-compliant, but also safeguard against unauthorized access and potential security breaches.

By developing a DMAP, you create a clear plan for managing the lifecycle of your data from collection to usage, ensuring that all stakeholders understand their roles and responsibilities.

---

## In this activity you will:

Develop a DMAP for your investment.

### What is data management?

Data management involves the operational activities required to treat data as a valuable asset throughout its entire lifecycle. Some key activities include:

Defining technical standards and architecture to describe how data assets will be managed within the project.

Data modeling and design, to ensure that data is interoperable and well structured.

Storing, managing and archiving data, including the creation of metadata and supporting documentation.

Implementing data access controls to ensure data is only used by those with the appropriate rights, in accordance with privacy regulations and security policies.

Assessing and improving the quality of data, to ensure it is fit for use.

Providing the tools and technologies that support the access and use of data, in accordance with the processes described in the data governance policy.

Publishing and sharing data using technical processes that promote openness and accessibility.

---

## The DMAP will contain sections covering:

A description of the data to be collected or used during a project. This might include reuse of existing data assets as well as collection of new data.

The data formats in which data will be stored, and why those formats and standards are appropriate.

Notes on the sources, volume, storage, sharing and archiving of data.

The processes by which data will be collected and processed.

Notes on data management, including naming conventions, version control, etc.

A description of how the quality of the data will be assessed, documented and maintained.

A description of the metadata and documentation that will be produced.

Notes on the standards used to create and format the metadata.

Ethics and privacy implications of collecting, storing and sharing data.

Notes on any intellectual property rights or licensing issues that are relevant to the data.

---

## **Risks of not having a clear plan**

Data may be stored in inconsistent formats, making it harder to analyse and share.

Data may be modeled or structured in ways that make it harder to use or reuse.

Data may be of unknown provenance, making it harder to use.

Data may not be published or shared, as there are no processes for doing that.

Data may not be archived, as there are no plans for where to store data in the long term.

Duplicated data might be collected, as previously collected assets are not discoverable.

As well as many other possible challenges.

---

1) If you are a Program Officer (PO), you may want to share this page directly with your grantee, so they can act on it.

- 2) Use the DMAP for Step 5, to populate the relevant sections.
- 3) Consult collaborating partners within the project to gather expertise.
- 4) Refer to the investment type examples to help you with this activity.
- 5) If you have created a data governance policy, ensure the DMAP is aligned to that policy.
- 6) Refer to your previous completed worksheets from Steps 1 to 4. Details from those worksheets will be useful while completing your DMAP.
- 7) Complete the various sections in your data management access plan, and share it with all the relevant parties.

If you have not created a separate data governance policy, remember to include essential governance elements in your DMAP. For example, defining clear roles and responsibilities is crucial, and must be documented in one of your project's core documents.

# Question guidance

This section provides guidance in section two of the DMAP.

---

## Data Summary

**Question 11** – *Provide a concise summary of the objectives and reasons for data collection or generation within the project. This should align with the project's goals and objectives and could mention specific research questions or hypotheses that the data will address.*

Things to consider:

- What are the primary objectives of your project that require data collection, usage or generation?
  - How will the data contribute to achieving the overall goals of your project?
- 

## Making data findable

**Question 19** – *Outline how data will be made discoverable by stakeholders during the project and internally, afterwards. If there is a plan to publish the data openly, after the project, outline how it will be made discoverable for that, too.*

Things to consider:

- How will your data be made discoverable to users, how will they be able to search and find it?
  - What tooling (e.g. repositories, or catalogues) will you use to list your data? For information around repositories and catalogues see the 'Resources' section in the recipe.
  - What recognized metadata standards will you employ to enhance visibility?
- 

**Question 20** – *Describe the use of persistent identifiers for your data types. These unique ID's ensure datasets can be reliably found overtime and avoid confusion and duplication.*

Things to consider:

- What persistent identifiers will you assign to your datasets (e.g. DOIs, GUIDs)?
  - How will you ensure that each dataset has a unique identifier?
  - How will you maintain and manage these identifiers over time?
- 

**Question 21** – *Describe the practice of naming conventions for datasets within this project.*

Things to consider:

- What naming conventions will you use for your datasets?
  - How will you ensure that dataset names are both consistent and descriptive?
  - Are there standard naming practices in your field that you need to follow?
- 

**Question 22** – *Specify the metadata that will be captured to describe the data.*

Things to consider:

- How will you ensure that project metadata is comprehensive and includes all necessary information? For guidance on metadata to capture, see the 'Resource' section in the recipe.
-

**Question 23** – Describe how the data will be documented, and in what detail and format it will be provided.

Things to consider:

- What documentation will you provide to accompany your data, think about explaining methodologies, schemas, and any relevant context?
  - How will you ensure that users can understand and use your data effectively, consider terms of use and guidance and how your data should be cited by others?
  - How will you document changes and versions of your datasets?
- 

## Making data accessible

**Question 24** – Outline how project data will be made accessible, if data is not to be made open, provide reasoning as to why.

*Aim to make data openly accessible unless there are specific reasons to restrict access such as personal or sensitive information.*

Things to consider:

- How will data be accessed during the project and after the project? Will the data be openly accessible to all users, or will access be restricted?
  - How will users access the data once they've found it? (e.g. through a specific repository or platform, download, email, API)? You can refer to the **"Directory: Repositories and other data platforms and services"** to see which repositories is most suited to the needs of your investment project.
  - What procedures will be in place to manage access to restricted data, think about the process of managing access requests?
- 

**Question 25** – Are there any embargo periods applicable to project data, and if so, provide the details and a rationale explaining why it is necessary.

Things to consider:

- Is there an embargo period before the data can be publicly accessed?
  - Why is an embargo period necessary for your data?
  - How long will the embargo period last, and when will the data be released?
- 

**Question 26** – Describe the strategies and infrastructure in place to preserve data beyond the project lifecycle.

Things to consider:

- How will you ensure that data remains accessible in the long term, specify the retention period?
  - What strategies are in place for the long-term preservation of data, how will you go about ensuring its tenure?
  - Who will be responsible for maintaining access to the data after the project ends including reviewing and updating retention policies?
- 

## Making data interoperable

**Question 27** – Outline what software and tools are required to support the interoperability of the project data?

Things to consider:

- What software or tools are required to use or analyze your data?
  - Will you provide documentation on how to use the software and tools and provide data in an interoperable way?
-

**Question 28** – *Where it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies, if so, outline your approach.*

Things to consider:

- Has this already been decided, if not, how will you decide what mappings to use
  - What vocabularies are already in use within the investment and is it necessary to create new ones
  - Will you consult the Data and Innovation Team on whether they can offer guidance?
- 

**Question 29** – *Describe the end to end infrastructure in place to support data management (e.g., institutional repositories, cloud storage, data analysis tools). It might help to refer to the data lifecycle diagram in the recipe and address each step in turn.*

Things to consider:

- For each step in the data lifecycle, what infrastructure is being used?
  - Which tools are recommended by CGIAR and could more be used?
  - Is each tool fit for purpose?
- 

**Question 30** – *Indicate which domain specific data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?*

- Place a cross in the 'In Use' column for each vocabulary in use. If you are using a vocabulary not present, add it to the list.
- 

**Question 31** – *Indicate which domain agnostic standard vocabularies for all data types present in your data set to allow interdisciplinary interoperability?*

- Place a cross in the 'In Use' column for each vocabulary in use. If you are using a vocabulary not present, add it to the list.
- 

## **Making data reusable**

**Question 32** – *Specify the licenses under which data will be made available (e.g., Creative Commons, GPL) and what conditions or restrictions that has on data reuse. Also specify where you will make the guidance for this available.*

Things to consider:

- What licenses will you use to clarify the terms of data reuse? For more guidance on choosing the correct licence, see the 'Resources' section in the recipe
  - Are there any conditions or restrictions on how the data can be reused?
  - Where can users find the full text of the license?
- 

**Question 33** – *Indicate how long datasets will remain reusable.*

Things to consider:

- How long will project data be available for reuse? Some datasets may remain usable longer than others
  - Will the data be updated over time, and if so, how?
  - How will you ensure that the data remains available for long-term reuse (reference the above question on long term access)?
-

**Question 34** – *Outline the actions that will be taken to ensure data quality.*

Things to consider:

- What quality control measures will you implement to ensure data quality (e.g. peer review)?
  - How will you validate and error check the data before sharing?
  - Will you document the quality assurance process? If so, how, and where will it be documented?
- 
- 

## **FAIR assessments**

**Question 35** – *Describe how you will track the project's FAIR potential maturity level throughout the investment.*

Things to consider:

- What will be the process, triggers or cadence for reassessing the project's FAIR maturity level using the FAIR Potential tool?
  - Will you keep a record of each assessment and if so, where will these be kept, alongside what documentation?
- 

**Question 36** – *Describe how you will monitor the FAIRness of the project's data assets through the use of the FAIR Data Assessment Tool.*

Things to consider:

- What will be the process for monitoring and assessing the project's data assets using the FAIR Data Assessment Tool?
  - What will you do with the information received from the tool? (i.e. record it, action it)
  - Will you create any documentation to go alongside the assessments and where will this be kept?
- 

## **Data security**

**Question 37** – *Outline measures in place for data protection throughout the data lifecycle.*

Things to consider:

- What measures will you implement to protect your data?
  - How will personal and/or sensitive data be handled to ensure it is collected and stored correctly?
  - What access control measures will you use to restrict access to authorized personnel and collaborators?
- 

**Question 38** – *Describe the planned use of access control for project data.*

Things to consider:

- What are the different levels of sensitivity in your data and what levels of access are there to each?
  - How will you manage the different access levels to your data?
  - How will you monitor access to your data?
-

**Question 39** – *Detail the backup and recovery procedures in place.*

Things to consider:

- Where will data be backed up to and what is the retrieval process in the case of an incident?
- Which data will be backed up and how long will it be backed up for?
- Who is responsible for completing the back up procedure and how often will they do it, is this a manual or automated process that needs verifying?

---

**Question 40** – *Outline your approach to managing risk for your sensitive and personal data. More guidance on this can be found in the 'Resources' section of the recipe.*

Things to consider:

- Data minimization – are any extra, unlikely to be used data being collected? If there is no need to collect the data, it should not be collected
- What measures will you take to protect confidential data? For instance, making use of anonymization techniques to retain the utility of the data but reduce the risk
- What will your retention policy be for personal/sensitive data?

---

## Ethics

**Question 41** – *Outline any ethical issues that have been identified and need managing throughout the duration of the investment.*

Things to consider:

- What ethical issues have been identified? If you're unsure what ethical issues might be present, see the 'Resource' section of the recipe, which discusses the use of a data ethics canvas
- Have you obtained approval from relevant ethics committees if applicable?

---

**Question 42** – *Describe how informed consent will be obtained if applicable to this project.*

Things to consider:

- How will you ensure participants are fully informed about the study and data use?
- What methods will you use to obtain and document informed consent?
- How will you handle situations where participants withdraw consent?

---

**Question 43** – *Outline the approach to anonymizing data where applicable.*

Things to consider:

- Identify which data needs to be anonymized and where it will be recorded
- What process will happen to ensure specified data is anonymized?

---

## Roles and responsibilities

**Question 44** – *Outline any skill gaps within the team related to the data lifecycle and how this will be addressed.*

Things to consider:

- Are there any gaps in skills that can't be covered by the current project team?
- How will these gaps be addressed?
- How will you support any necessary learning and skill development throughout the project?

## Data sharing agreements

**Question 45** – Briefly outline your data sharing agreement (if applicable) and any recipients of it, alongside any restrictions, constraints or challenges observed.

Things to consider:

- Will you need to put one or more Data Sharing Agreements in place?
  - If so, who with and in regards to which data?
  - What challenges, if any, are there in regards to the sharing agreement or data shared, and how will those be managed?
- 

## Compliance and support

**Question 46** – Describe how you will monitor compliance with the Data Management and Access Plan throughout the investment.

Things to consider:

- How will you monitor compliance with your data management plan?
  - What procedures will you implement for regular audits?
  - How will you report and address compliance issues?
- 

**Question 47** – Outline the mechanisms for supporting the use of Data Management and Access Plan including training for those who need it.

Things to consider:

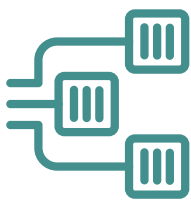
- What mechanisms will you implement to enforce the data management plan?
  - What are the consequences for non-compliance?
- 

## Remember



If you have not created a separate data governance policy, remember to include essential governance elements in your DMAP. For example, defining clear roles and responsibilities is crucial, and must be documented in one of your project's core documents.

## Investment types



## Overview



©Gates Archive/Mansi Midha

## Every investment project is unique

The application of the six steps will vary accordingly. To provide examples that align with your project, common characteristics of AgDev investments were researched and three 'investment types' were developed.

# AgriConnect: a digital solutions investment



# AgriConnect: Creating a data management and access plan (DMAP)

Rashima, the lead of AgriConnect, recognized that a comprehensive DMAP was essential for ensuring data collection, management, sharing, and governance activities align with their FAIR principles and the organization's mission. Rashima initiated a team meeting to outline the key aspects of their DMAP, using the provided template as a guide.

## Aligning with the data governance policy

Since AgriConnect had already developed a data governance policy (as per Step 5.1), Rashima began by referencing this policy to fill Section One of the DMAP. She reviewed the foundational principles with the team, reaffirming their commitment to data integrity, privacy, security, and transparency.

Chris, the third-party publisher responsible for managing the national repository of agricultural indicators, emphasized the importance of aligning AgriConnect's internal data handling practices with external regulations and standards. He provided insight into compliance with Dataland's data privacy laws and the practicalities of integrating derived data with the national repository.

## Establishing roles and responsibilities

Rashima and Marie, the UX consultant, collaborated to assign clear roles. They appointed John, the data steward, as the primary contact for data-related queries. His responsibilities included overseeing the execution of the DMAP, ensuring compliance with internal and external policies, and liaising with data providers such as Chima, the smallholder farmer, and Faisel, the researcher.

## Data summary

Using the data inventory already created during Step 3 of the FAIR Process Framework, AgriConnect cataloged the various datasets they planned to collect or reuse. This inventory was essential to answering questions about data types, sources, and third-party data policies, in section two of the DMAP.

Chima provided details about the raw agricultural data collected through his village's cooperative. Faisel categorized the types of data they would generate, including survey results, sensor readings, and observational records. They also identified any third-party data AgriConnect planned to utilize, such as national agricultural indicators shared by Chris's agency.

## Data management planning and compliance

Rashima facilitated discussions with Nemy, the cybersecurity consultant, to outline security measures for protecting sensitive and personal data. Together, they identified datasets containing

farmer names, GPS locations, and financial details, determining the necessary encryption protocols and anonymization techniques.

Marie and Noora, the technical consultant, developed plans for using persistent identifiers (e.g, DOIs) and naming conventions to ensure the findability of datasets. They also defined metadata standards and documentation practices, referencing domain-specific vocabularies (e.g., AGROVOC and BrAPI) to promote interoperability.

### **Data accessibility and long-term preservation**

The team reviewed long-term access and preservation strategies, prioritizing open data access where possible. Marie worked on identifying repositories where data could be hosted, ensuring that users could easily search for, and access, AgriConnect's data assets. Rashima and Chris agreed on implementing an embargo period for sensitive data collected from smallholder farmers to protect their privacy while adhering to national guidelines.

### **Fair assessments and quality control**

AgriConnect planned periodic reviews of their datasets' FAIRness, as Rashima introduced a biannual evaluation of the project's data assets. This included monitoring improvements, conducting quality control checks, and updating the data inventory as needed.

### **Risk management and ethical considerations**

In collaboration with Marie and Nemy, Rashima outlined the risk management approach in the DMAP, focusing on data minimization and anonymization. They also defined the backup and recovery protocols for key datasets and described how informed consent would be obtained and documented when collecting sensitive farmer data.

### **Outcome**

By following this structured approach, AgriConnect successfully developed a comprehensive DMAP that not only met their internal governance requirements but also aligned with Dataland's external policies and regulations. This plan provided a clear pathway for managing data effectively throughout the project's lifecycle, ensuring that all stakeholders, from smallholder farmers to national agencies, could trust in the integrity and accessibility of AgriConnect's data outputs.

# **AgroThrive: a policy and advocacy investment**



©Gates Archive/Thomas Omondi

## **AgroThrive: Creating a data management and access plan (DMAP)**

Kaira, the lead of AgroThrive, recognized that to provide comprehensive policy recommendations for Datapur's government, their team needed a clear and actionable DMAP. The project focused on delivering evidence-based agricultural policy advice on credit, infrastructure, climate resilience, and land tenure issues.

### **Integrating data from multiple sources**

Kaira began by consulting with partners like Saanvi, the technical consultant on climate resilience, and Imamu, the third-party publisher managing the national repository. AgroThrive's data sources included farmer surveys, climate sensor readings, soil health assessments, and policy studies. Kaira decided to use common data formats like NetCDF for climate data and CSV for survey results, allowing seamless integration with the national repository.

### **Defining roles and responsibilities in the DMAP**

Kaira appointed Anna, the project partner responsible for developing AgDev training modules, as the data lead, overseeing data collection and quality assurance. Haben, the smallholder farmer, played a key role in providing on-the-ground data and validating findings. Aziz, the NGO head, helped align the DMAP with existing governance policy, avoiding policy conflicts.

### **Designing data collection and security protocols**

Lata, the gender consultant, worked closely with Kaira to develop data collection protocols that ensured inclusivity and sensitivity to gender issues. The team established guidelines for anonymizing sensitive data and obtaining informed consent. Nemy, the cybersecurity consultant,

supported by defining data security protocols, including encrypted storage and role-based access control.

### Planning for data quality, sharing, and archiving

Kaira and Imamu collaborated to outline the protocols for data validation, using automated quality checks and manual reviews. The DMAP also detailed the use of standardized metadata formats and FAIR-aligned data sharing protocols to facilitate open access where possible. Imamu provided guidance on archiving policies to ensure the long-term preservation of AgroThrive's data assets.

### Outcome

AgroThrive's DMAP provided a comprehensive approach to managing their diverse datasets, aligning data handling practices with project goals and external policies. The plan ensured that data from multiple partners and sources remained consistent, secure, and accessible.

## NGBT: a field research investment



©Gates Archive/Esther Mbabazi

### NGBT: Creating a data management and access plan (DMAP)

Farah, the lead of the NGBT project, knew that developing a climate-resistant and nutritious barley varietal required collaboration among multiple researchers and specialists. The project's complexity necessitated a DMAP to ensure consistency and transparency in handling data.

### Coordinating data from diverse disciplines

Farah brought together key partners, including Nasser, the researcher collecting agricultural data, and Charlotte, the climate scientist. NGBT's data sources ranged from genetic research data and

soil analysis to climate impact models. They standardized the formats, using VCF for genetic sequences and GeoTIFF for spatial climate data, ensuring interoperability across disciplines.

### **Setting clear roles and responsibilities**

Farah appointed Cali, the agricultural geneticist, as the data manager responsible for overseeing data integrity and compliance. Davu, the smallholder farmer, contributed ground-level data and played a key role in validating field research findings. Jaya, the gender and children specialist, ensured that data collection protocols were sensitive to gender issues and children's needs.

### **Establishing data collection, privacy, and quality protocols**

Charlotte collaborated with Chanda, the soil scientist, to design robust data collection protocols, focusing on consistency in methodology and integration of field observations. Nemy, the cybersecurity consultant, developed protocols for data security and access management, ensuring that sensitive and personal information was protected. Cali implemented version control and quality checks to maintain high data standards.

### **Planning for data sharing, archiving, and long-term access**

Farah, with Joe's support as the third-party publisher, outlined detailed plans for making data FAIR. They planned to publish research findings and datasets through open-access repositories, using DOIs for persistent identification. The DMAP also included a comprehensive archiving strategy, outlining storage infrastructure and long-term access policies to preserve data beyond the project lifecycle.

### **Outcome**

The NGBT project's DMAP enabled seamless collaboration among researchers and ensured data integrity and accessibility throughout the project. It aligned with external standards, while addressing the project's specific challenges and needs.



Researchers are willing to share data if robust governance and regulation frameworks are in place.

Learn more

**Acknowledgements**

**FAQs**

**Glossary**

**Accessibility**

**Privacy & cookies**

**T&Cs**

FAIR Process Framework has been developed by the Enabling Data Access (EDA) project team at CABI and is funded by the Bill & Melinda Gates Foundation to support the foundation's Open Access Policy. The FAIR Process Framework is a tool to assist partners in developing data access and management plans (DMAPs) that incorporate FAIR and responsible data practices. Except where otherwise noted, the content on this website is licensed under a Creative Commons Attribution 4.0 International License.