5.0 Developing a FAIR data strategy

To develop a considered strategy that will increase the likelihood of successfully implementing FAIR principles

Why should I do this?

To define the processes, technologies and governance measures needed to help you ensure data quality, availability, accessibility and security within your investment and to improve the chances of successful FAIR implementation.

Articulating this strategy will also help you outline how your investment will collect, manage, analyze and use data to achieve its objectives.

A well-implemented FAIR data strategy can help your projects maximize the value of their data while minimizing risks associated with data misuse or mismanagement.

Download this factsheet for more insights.

In this step you will:

Engage in three activities to create the core components of your FAIR data strategy:

- 5.1 Data governance policy: Developing a policy that establishes clear principles and guidelines for the collection, management, protection, and ethical use of data within your investment.
- 5.2 Data management and access plan (DMAP): Developing a living document that outlines how data will be collected and handled through the life of your investment.
- 5.3 Data sharing agreements: Establishing agreements between two or more entities that define the terms for how data will be shared and used. This ensures transparency, compliance, and proper data stewardship.

Unlike earlier steps that involved documenting activities in workbooks, these activities will result in the creation of formal documents for your investment. Each template will contribute to your FAIR data strategy:

Download Data Governance Policy Template
Download Data Management and Access Plan Template
Download Data Sharing Agreement Template

* The template is just a guide, and you will need to seek your own legal advice before using this or any Data Sharing Agreement.

If you would like to understand what inspired the approaches and methods used to develop each step, view background research. This is not required to complete any activities

When should I do this?

During proposal development:
Developing a data governance policy and data management plan may not be possible at this stage. However, POs should initiate early thinking around data governance so that a grantee plans for it as a future activity in the proposal. The proposal should also clearly identify who is best positioned to lead this activity, and whose support will be required.

Sample text to include in proposal
In co-design sessions with local stakeholders, [a grantee] will work in tandem with local stakeholders to determine data policies and best practices that align with rules and regulations within each specific geography. While the policies and practices will be locally specific, we will provide input on how to adhere to data privacy best practices. We may need to support local advocacy for regulations that directly impact our ability to provide advanced analytical services and an enabling data ecosystem, but this will not be a primary focal area for the team. If this emerges as the top priority for counterparts, we will consult with Bill & Melinda Gates Foundation counterparts to determine if or how we may operate in this space, and the resultant impact on work planning or resource allocation.

When the investment is live:
POs should consider initiating this activity with the grantee as early as possible. Having a FAIR data strategy is integral to every stage of the data lifecycle in an investment. It ensures that the people who collect, manage, access and use data understand their responsibilities and it also promote the most effective use of data whilst considering responsible and ethical data practices that are aligned to existing policies and norms.

Do this sooner rather than later. Leaving this activity until toward the end of the investment would diminish its value.

Top tip

Deciding when to use a DMAP or a data governance policy?

Think of a data governance policy as your high-level strategy. It sets the principles and guidelines for how data should be handled within your organization or project. It is like the blueprint—outlining what is important, defining roles, and stating the commitment to data quality, privacy, and security.

In contrast, a DMAP is the detailed implementation plan. It outlines the day-to-day specifics of how data will be collected, stored, shared (made FAIR), and preserved throughout the project's life cycle. A DMAP puts the data governance principles into practical steps and tasks, ensuring alignment with the strategy while focusing on the management of specific datasets.

In short, use a data governance policy to guide your overall approach to data, and a DMAP when you need to get specific about managing and accessing project-level data.

Consider the size and complexity of the project

For larger, more complex projects, having both is ideal. The data governance policy provides strategic direction, while the DMAP ensures operational consistency. However, for smaller projects, a DMAP may include essential governance/strategic elements, especially if no organizational-level policy exists.

Working with existing organizational policies

If an existing organizational data governance policy is in place, align your DMAP with it, referencing the policy to ensure consistency. This helps avoid redundancy and maintains coherence in how data is managed.

Collaborating partners with different policies

When multiple partners with different organizational data governance policies collaborate, use the DMAP to harmonize approaches. Firstly, ensure all stakeholders agree on a shared vision and goal for FAIR data in the project. Then establish clear guidelines within the DMAP for addressing differences or conflicts between the policies, creating a cohesive plan for the project's specific needs.

Key concepts

Below is an introduction to the key concepts you will come across in this step. As you complete an activity, you may need to refer back to some of these key concepts.

Data governance

The set of processes, policies and procedures that ensure the effective management, protection, and ethical use of of data within an investment project.

A **data governance policy** is a formal document that establishes guidelines for the collection, management, protection and use of data to ensure its quality, security, and compliance with legal and ethical standards. This is a comprehensive approach to managing data assets, ensuring that data is accurate, accessible, secure, and compliant with regulations.

Data management and access plan (DMAP)

A more operationally-focused plan that outlines the specific methods and procedures for handling data throughout the lifecycle of a project, from data collection to sharing and preservation.

Data sharing agreement

A formal, legally binding agreement between two or more parties that outlines the terms and conditions for sharing data. It defines the roles and responsibilities of the data provider(s) and recipient(s), and the purpose of sharing it; defines authorized uses of the data; and it outlines data security measures, and the legal and ethical considerations for the use and storage of, and access to, the shared data. The agreement aims to ensure transparency, protect privacy, and maintain compliance with relevant laws and regulations, while facilitating responsible data sharing.

Data lifecycle

The stages through which data passes, from creation and collection to storage, sharing, analysis, and archiving or disposal.

Data quality

The condition of data based on factors such as accuracy, completeness, reliability and relevance, which affect its usability for a given purpose.

Data quality management encompasses strategies to measure, improve and maintain the quality of data across its life cycle.

Data security

Implementation of technology defenses that prevent data from being lost due to internal actions or external hacking attempts. It ensures the protection of data from unauthorized access, use, disclosure, disruption, modification or destruction.

Data misuse

The unauthorized or inappropriate use of data, which can lead to breaches of privacy, compliance issues, and harm to stakeholders.

API (Application Programming Interface)

A set of definitions, protocols and tools for building software and applications. It acts as a software intermediary that allows two applications to communicate with each other, enabling data exchange and functionality sharing. APIs simplify and speed up the integration of systems and technologies, allowing applications to leverage functionalities without having to build them from scratch. They are fundamental in the creation of interconnected systems where software components can interact seamlessly, enhancing user experiences and enabling developers to create complex functionalities efficiently.

Interoperability

The ability of different data systems, applications and formats to work together to exchange data, and make use of it in a seamless manner.

Ethical data use

The practice of handling data in ways that respect privacy, security, and the rights of individuals and organizations.

Privacy protection

Measures taken to ensure the confidentiality and security of personal and sensitive data, preventing unauthorized access or disclosure.

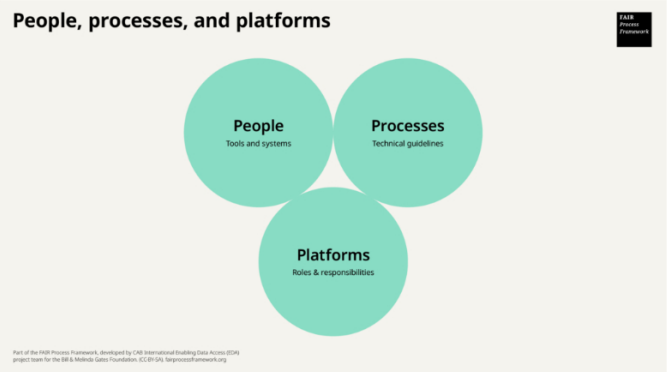
The 3Ps: People, Process, Platform

The 3Ps is a framework that highlights people, processes and platforms as three main components of a FAIR data strategy. These have been adapted from the data governance industry-standard concept of 'people, processes, and technology', to include 'platform' as a more specific component for the work you are undertaking.

According to this methodology:

- **People** are defined as the organizational structure of those involved in data governance, including the roles and responsibilities of those who own, collect, store, manage and use data.
- **Processes** refers to guidelines for using, protecting and managing data, ensuring consistency and compliance.
- **Platforms** (or technology) is indicative of the tools and systems that support data governance, and can include

security protocols or data management platforms.



Investment types

Overview AgriConnect: a digital solutions investment AgroThrive: A policy and advocacy inv

©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

©Gates Archive/Alissa Everett

Recently, a large African-led organization, AgriConnect, has decided to make its data processes FAIR. Its work focuses on scaling agricultural innovations to improve smallholder livelihoods, and ultimately increase food security across the continent.

AgriConnect is determined to apply FAIR principles to avoid inefficiencies and the economic costs of efforts duplication, legal risks, and missed opportunities that might hamper its mission's accomplishment. To aid decision-making and collaboration, stakeholders need to be able to easily locate data. This platform must be accessible for stakeholders with different levels of digital literacy, meaning that usability considerations must be ensured.

Moreover, as the platform combines diverse data sources and formats, data must follow the interoperability principle, which facilitates collaboration and data sharing.

Finally, data must be reusable, so that they continue to add to the richness of new work, rather than stagnating after collection or publication.

To adhere to the FAIR principles, AgriConnect plans to create a system that is open access, thereby increasing its utility to end users, and making the collected data available (when possible) for other researchers to find, access and reuse. The organization will include a technical perspective through data formatting, hosting infrastructure, and sharing agreements and permissions.
This commitment, from a large, respected organization, should act as an institutional and cultural guidepost for other organizations to adhere to the FAIR principles within their own work, thereby increasing FAIR visibility and utility for future work within Dataland.

AgroThrive: A policy and advocacy inv

©Gates Archive/Thomas Omondi

The well-established policy and strategy organization AgroThrive works to improve enabling conditions for people across the AgDev ecosystem (including smallholder farmers), with the goal of improving smallholder livelihoods, and ultimately increasing food security.

Read less

AgroThrive is launching a project in the lower-middle income nation of Datapur to bolster the local government's capacity to scale

AgroThrive is launching a project in the low- [illegible] income nation of Datapur to bolster the local government's capacity to scale inclusive agriculture and attract private sector investment. It aims to create a more enabling policy environment for Datapur's agricultural sector by providing evidence-based policy recommendations.

Its work specifically includes access to credit and affordable financing options for agricultural inputs, infrastructure development (transportation and storage facilities), educational opportunities for smallholders, fostering technological advancements, climate resilience, and land tenure security.

AgroThrive believes that all of these conditions should also be viewed through an intersectional lens, with special attention given to including traditionally marginalized communities.

Additionally, AgroThrive will work with Datapur officials to increase the state's implementation capacity by helping to improve government planning, accountability, delivery of services and sector coordination.

The AgroThrive team has determined that the FAIR Principles are an integral component of the project's long-term success. The principles are transversal to all the project's phases, from data gathering to findings dissemination. This will allow AgroThrive to leverage existing data and analyses, speeding up the policy development process and optimizing resource use.

Once a final report is prepared and shared with the government of Datapur, AgroThrive will (to the best of its capabilities) adhere to the FAIR Principles. It will do this by making the data used, and the resulting policy recommendations, findable and accessible to those working in the broader AgDev sector, so they can be built on or reused for similar initiatives, with interoperability as an enabling condition for transparency, collaboration, and data sharing.

This will allow the recommendations to become a part of the evolving ecosystem around the country's policy environment, rather than stagnating after they have been shared with the government.

As the provision of policy recommendations is largely based on the analysis of existing data sources, applying FAIR to the project will necessitate data sharing agreements to be put in place with relevant data owners. As much of this data may be sensitive (or proprietary, in the case of private sector data), these agreements will need to include privacy provisions and data safeguarding measures, and explain how the data sources will be used and why they are necessary to the project's outputs.

Trust-building with stakeholders is also a vital component of the project, both to facilitate data sharing and to create buy-in for the targeted recommendations (thereby employing a 'bottom-up' approach to the recommendations).
This will be done through open-source publication, and the appropriate use of metadata, in addition to sharing the recommendations in widely accessible journals and platforms. As a large, well-respected organization, this adherence and promotion of FAIR will act as a guidepost for other organizations to consider FAIR in their own work.

NGBT: A field research investment

©Gates Archive/Esther Mbabazi

NourishGen BioTech (NGBT) is a multinational research organization committed to combating global hunger, addressing gender disparities, and mitigating the impact of climate shocks on vulnerable populations. Its lab-to-field approach has already improved nutritional outcomes for millions of people by optimizing crops for widespread planting.

Read less

NGBT is launching a project in the low-income nation of Datastan, where increases in crop productivity and yields will provide

tangible benefits to a population overwhelmingly made up of smallholder farmers, and to the country as a whole, for which agriculture is a main driver of economic growth. NGBT is cognizant of the fact that women play a crucial role in this domain.

In Datastan, NGBT is developing a more nutritious varietal of barley that is more resistant to climate change and produces greater yields. The project not only seeks to optimize crop varieties capable of thriving in challenging environmental conditions, but also indirectly addresses the exacerbation of child nutrition gaps due to climate-related disruptions. Simultaneously, it empowers women, who are often disproportionately excluded from field-research projects, by enhancing their access to resources and opportunities in agriculture.

If successful, this would allow farmers to use less farmland without decreasing the quality of their outputs. This work will involve both small and large data points, including (but not limited to) historical and current weather data, data on current crop yields and production, land use for agricultural purposes, and government data on the population involved in the production of the target crop.

Planning for the application of the FAIR data principles is vital from the project's inception. As this project is focused on the creation and eventual sharing of newly generated data, NGBT will need to plan for how the data will be published. This may include data, methodologies, research materials, results, budgets, or other useful information within the project's outputs for greater findability, accessibility, interoperability and reusability of results. The main objective is to ensure the research process is transparent, so that others can review, critique and build on their work.

Additionally, NGBT must consideration how to communicate results with relevant stakeholders (through in-country meetings or workshops), and, after the results are published, collect stories about how this positively impacted the community.

It must also ensure that the outputs of the project contribute to the overall scientific community, and allow results to be built on in future research. The outputs will need to be appropriately labeled with metadata and stored in an accessible platform, which will facilitate their findability and accessibility for researchers working in similar spaces.

The NGBT team will also, through mapping of the data ecosystem and data assets, decide on the most appropriate storage system for research outputs, with an eye toward greater interoperability.

As a well-respected and influential organization, the consideration of FAIR principles will influence other researchers and organizations on the utility of FAIR, as the scientific community as a whole (including NGBT) benefits from findable, accessible, interoperable and reusable research outputs, which can help to advance discovery and reduce repetition of research.

A key outcome of the data ecosystem mapping exercise in Step 2 was a widespread acknowledgement of the importance of data sharing agreements and a data sharing policy. This meant we were also able to put Step 5 into action.

FAIR *Process Framework*

CABI

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.