3.0 Identifying data assets

To manage data assets effectively, create an inventory to clarify data types, ownership, and future use, supporting data sharing and long-term impact

Why should I do this?

To understand the diverse kinds of data assets in your investment—both input and output data assets. Over the course of many months or years, an investment will utilize and produce a vast array of different kinds of data. Yet in the course of performing this work, it may not always be clear precisely what these components or data assets are, who has ownership over them, and what will become of them once an investment concludes.

By creating an inventory of assets, you will improve how data are utilized in your investment, plan for data-sharing agreements and licenses that may be required, and ultimately improve the longevity and utility of your investment's mission.

Creating a data inventory is also an important part of designing a data management plan.

While it may appear complex, anyone can conduct this process, but specialist advice may be needed for some legal or technical nuances. If you are a Bill & Melinda Gates Foundation Program Officer, requiring legal support, contact AgDev's legal relationship manager.

In this step you will:

3.1 Plan the data inventory: Deciding on the purpose, scope and granularity of your data inventory.

3.2 Identify output and input data assets: Identifying the output data assets your investment is producing and identifying the input data assets that are used to produce the outputs.

3.3 Establish attributes and metadata: Establishing the important metadata to be added to data inventories.

3.4 Compile the inventory: Gathering information to start putting your inventory together and deciding how best make it accessible to others.

As you complete these activities, document your outputs in the provided workbook. We recommend using the same document as you go through each step, so you have a single document that captures all your workings. Each workbook is segmented by activity, allowing you to complete either the entire workbook or only the specific sections relevant to your current project stage. Download workbook (and supporting factsheet).

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:

In the proposal, grantees should have a basic list of their likely data outputs and data inputs (and any assumptions about the data). The proposal should include an activity to develop a more comprehensive key data asset inventory. The proposal should also clearly identify who is best positioned to lead this activity, and whose support will be required.

When the investment is live:

For an investment that is live and that has not inventoried its data assets, investment Program Officers (POs) should consider initiating this activity with the grantee. To make data FAIR in an investment, knowing what assets are being produced, and what input data assets are being used, is imperative.

Do this sooner rather than later! Leaving this activity to the end of the investment and filling in data assets retrospectively can be burdensome. It can also be inaccurate, as team members involved at

different stages of the data lifecycle may no longer be present, and detailed documentation may not have been maintained.

Refer to the data interventions types you identified in Step 1 in your workbook. Knowing your data intervention types will help define the parameters of your data asset inventory.

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

Data assets

A piece, or collection, of data that holds significant value and contributes to the strategic objectives of a community. A valuable resource that can be leveraged to drive insights, make informed decisions, improve processes and, ultimately, improve people's lives. In every investment, there are data assets and key data assets. Data assets are valuable for general use within the project. Key data assets are crucial for strategic decision-making and project success.

Metadata | (Meta)data

There is a difference between the definitions of metadata and (meta)data within and outside the context of the FAIR Principles and FAIR investments.

In the context of the FAIR Principles:

Metadata: Data providing information about other data. Essential for understanding, accessing and using primary data. It includes details like creator, date, format and source. It must be machine-readable to enable automated systems to find, access, interoperate and reuse data efficiently.

(Meta)data: Encompasses both primary data and its metadata. Ensures data and descriptive information adhere to standards. Must be findable, accessible, interoperable and reusable. Both data and metadata should be machine-readable to facilitate automated processes and enhance data sharing and long-term stewardship.

Outside the context of FAIR Principles:

Metadata: Data providing information about other data. Essential for understanding, accessing and using primary data. It includes details like creator, date, format and source. Machine readability can be beneficial, but is not mandatory.

(Meta)data: Encompasses both primary data and its metadata. Highlights the importance of data and descriptive information. The emphasis is on usability and integration, but not necessarily strict standards. Machine readability is less pronounced, but can enhance usability.

Key differences:

FAIR Context: Emphasizes machine readability and adherence to standards for findability, accessibility, interoperability and reusability.

General context: Less stringent on machine readability and FAIR attributes, focusing more on basic descriptive roles and data usability.

Input data assets

Any data that is used as an input for processes, analyses or decision-making within an investment. These data assets serve as the raw materials that feed into systems, models or workflows to generate insights, support operations or drive strategic initiatives. Input data can come from a variety of sources, both internal and external to the investment, and can be structured or unstructured.

Some examples include: surveys, market research reports, sensor data, financial transactions, and social media feeds.

Output data assets

The results or outputs generated from processing input data through an investment's operational, analytical or decision-making processes. These assets are valuable for their

insights, conclusions or actionable information that can inform future actions, strategies or operational adjustments. Output data assets often take the form of reports, dashboards, predictive models, and strategic recommendations.

Examples include: sales forecasts, customer segmentation reports, agricultural yield predictions, financial performance dashboards, and operational efficiency recommendations.

Data inventory

A structured and comprehensive list of priority data assets, annotated with important information (known as metadata) that can help users understand why data has been collected, what it contains, how it is managed and the ways it will be made available for others to use. It is a useful tool for any project dealing with multiple types and sources of data.

Data sharing

The practice of making data available to others for viewing or use, beyond the original collecting entity or individual.

Data license

The ways that data can be used, granting generalized permission to use and reuse data, provided the user complies with a specified set of conditions.

Risks

The potential for loss, damage, or undesirable outcomes resulting from internal or external vulnerabilities within data management practices. This encompasses a wide range of issues, including but not limited to data breaches, unauthorized access, data corruption, loss of data integrity, privacy violations, and non-compliance with data protection law.

Data rights

The legal rights and protections related to the collection, use, storage and sharing of data, particularly personal data about individuals.

Personal data

As defined by the United Nations, information relating to an identified or identifiable natural person.

Permission

The authorization granted to users, systems or applications to access, manipulate or manage data within a specific environment or framework.

Access control

The processes and technologies used to manage and regulate who can view or use data stored in a computer system or network.

Datasets

Typically collections of data, each on a common topic, with consistent format and structure. These may include raw data, analyzed results, or derived information.

Investment type

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

©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 investment

©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.

AgroThrive is launching a project in the lower-middle 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 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.

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.

Co-developing tools that meet the needs of the stakeholders

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.