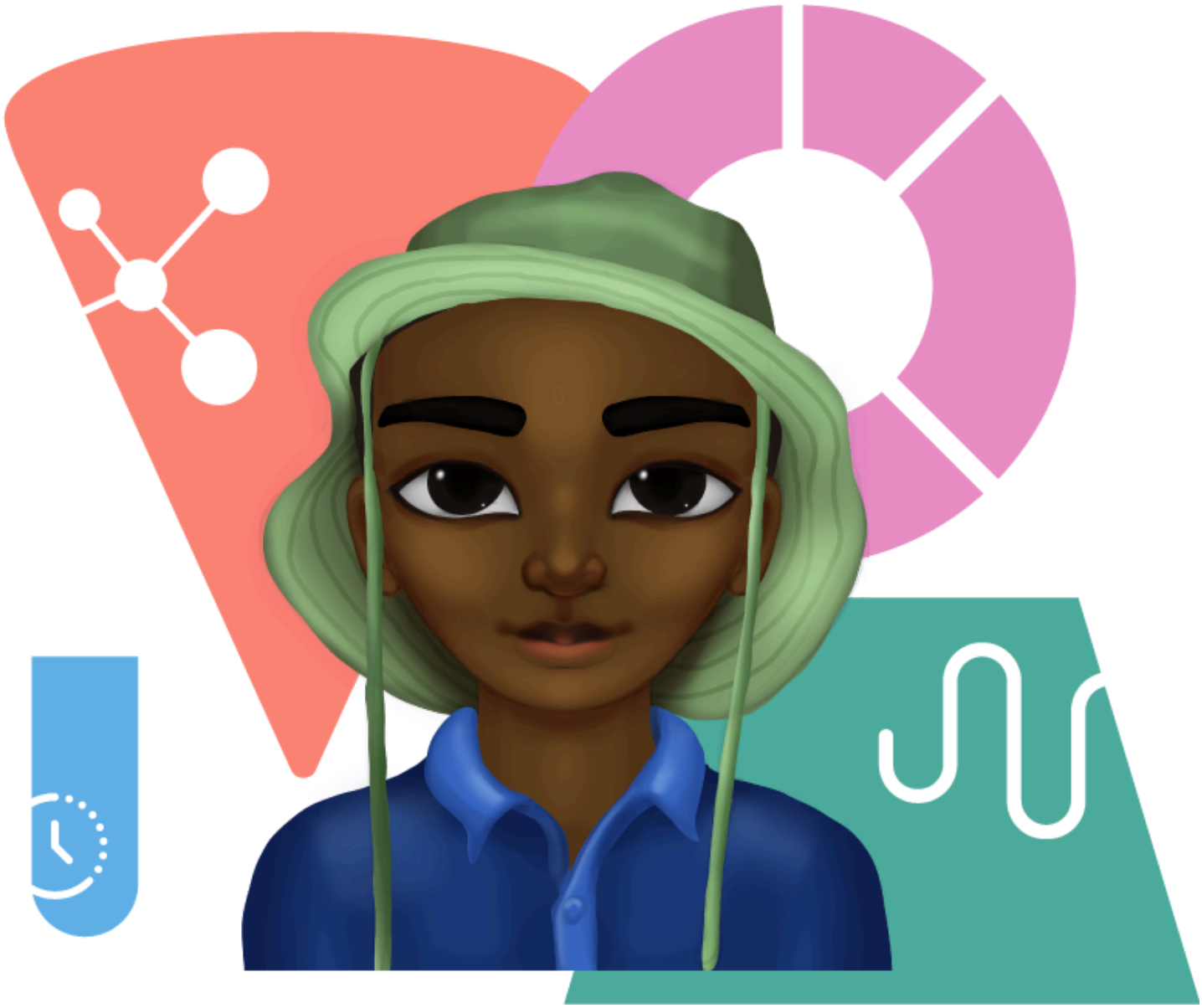


3.3 Establish attributes and metadata

Establishing the important metadata to be added to data inventories

Why should I do this?

To help users searching for this data to find it and, by using controlled vocabularies for attributes, improving future search and data linking potential—e.g., finding related data.

Using well-chosen metadata attributes improves data discoverability and interoperability, enabling others to locate and integrate data more easily. Applying a controlled vocabulary—a predefined set of terms for each attribute—enhances consistency, making searches more reliable and allowing data from different sources to be more easily linked. Classification, which involves grouping data by relevance or value, further refines data management by highlighting assets that are critical to strategic goals and operational success.

In this activity you will:

Understand what metadata is and how it provides essential information about data assets, offering context that allows users to understand, find, and effectively use the data.

Learn how attributes are specific characteristics or pieces of information that define aspects of each data asset. These could include technical details, such as file format or data size, or descriptive elements, such as subject matter and geographic area.

Define the metadata attributes for your data assets and consider classification based on data's impact and value, as this can further organize assets by strategic goals and value, which is particularly useful for prioritizing efforts.

Use a controlled vocabulary for consistency, and work with technical experts if needed.

Consider which attributes will best serve end users and enhance the data's findability.

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 workbook (and supporting factsheet) for Step 3 here. We recommend using the same document throughout this step, so you have a single document that captures all your workings.

3) Consider the below tips to help you decide on your attributes and metadata:

Tips to help you decide on your attributes and metadata:

Use easy-to-search vocabulary.

Use a controlled vocabulary for each attribute to improve future search and data linking potential—e.g., finding related data.

Include more attributes based on your particular investment needs.

Think about your end users and potential future users of your outputs, and consider attributes that would be useful to these groups.

While you do not need to be technical to plan for FAIR, part of your plan is to identify which technical people and skills need to be involved to deliver FAIR and responsible data for the project. Determining your vocabulary may be more efficient if you work with someone experienced in assigning attributes.

Examples:

Attributes and MetadataID: A unique identifier for the data asset. Once correctly assigned, this ID should never alter.

Title: The name of the data

Description: A description of the data asset

Purpose: Why was the data collected or produced?

Data Manager: Who created the data, or manages it?

Language: In which language(s) has the data been created?

Location: Where is the data is located or stored?

Type: What type of data it is—e.g., text, numbers, statistics, database...

Format: The file format of the data—e.g., csv, jpg, sql, json, geojson,...

Creation date: When was the data was created?

There are a lot of tools online to help deliver FAIR and responsible data, some of which you can find in Step 6.

You may find there is benefit in classifying your data assets. Classification based on data assets' direct impact on strategic goals, operational efficiency, and long-term sustainability, ensures that resources and attention are allocated effectively, focusing on data that drives the most value. This structured approach to prioritizing data assets allows your investment to focus on what is most critical to its immediate success and operational efficiency, while still maintaining an eye on long-term strategic goals and sustainability. If you think your investment will benefit from such classification, review the classification examples provided below, and start classifying your investment.

Priority levels

High Priority: e.g., Core operational data and financial data

This data has an immediate and significant impact on the operational efficiency and financial health of agricultural activities. These data assets are critical for day-to-day decision-making, risk management, and ensuring the economic viability of operations.

Medium Priority: e.g., Supply chain data, market and consumer data, and regulatory and compliance data

This data is considered mid-priority. These categories are essential for strategic planning, compliance, and market responsiveness, but might not require immediate daily action and may not change daily.

Low Priority: e.g., Research and development data, environmental and sustainability data and human resource data

This data is categorized as low priority, not because these assets are unimportant, but because they are often used in long-term planning and strategy rather than immediate operational decisions. This data does not affect immediate operational decisions.

Confidentiality levels

Public: Anyone can access the data and it can be sent to anyone. For example, open government data.

Internal use only: Only employees can access the data and it cannot be sent outside the company.

Confidential: The data can be shared only if it is needed for a specific task, and it cannot be sent outside the company without a non-disclosure agreement.

Sensitive (personal data): The data contains private information, which must be masked and shared only on a need-to-know basis for a limited time. The data cannot be sent to unauthorized personnel or outside the company. The data can be shared only with named individuals who are accountable for its protection. For example, legal documents or trade secrets.

Restricted: The data can be shared only with named individuals who are accountable for its protection. For example, legal documents or trade secrets.

Data retention

None: Data can be deleted at any time.

Temporary: Keep data for a short period of time. For example, keep Twitter/X data for a week.

Fixed-period: Keep data for a set number of years, after which it can be deleted. For example, keep tax records for seven years to comply with government laws.

Permanent: Never delete data. For example, legal correspondence.

Governance levels

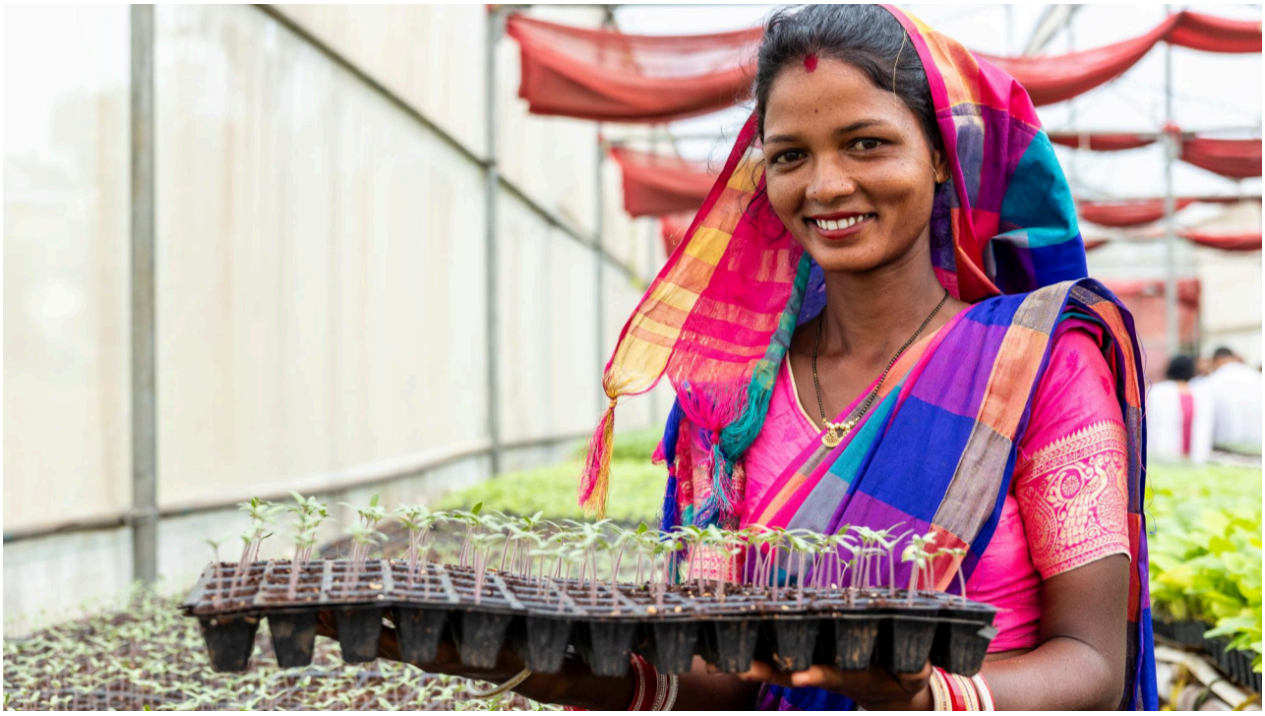
Ungoverned: No policies and rules to govern data.

Somehow governed: Some policies and rules to govern data.

Fully Governed: Policies and rules to govern data consolidated and enforced everywhere.

Investment types

Overview

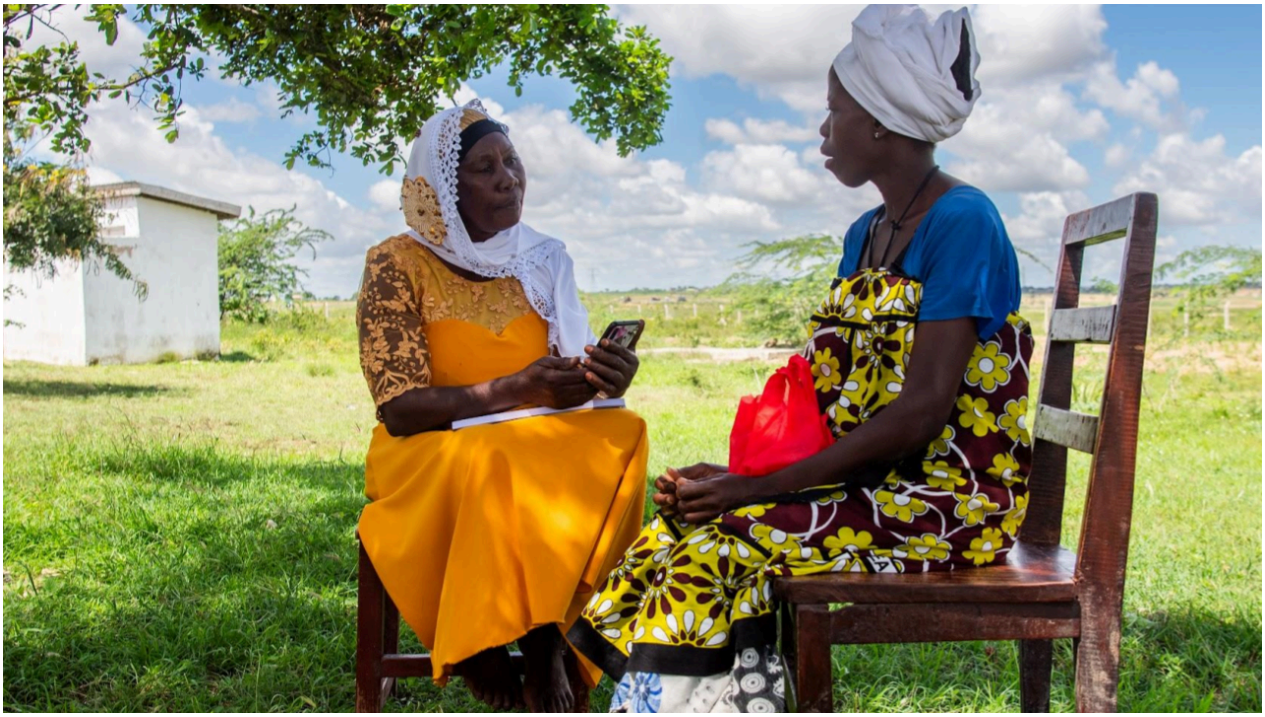


©Gates Archive/Mansi Midha

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



©Gates Archive/Alissa Everett

The AgriConnect team established a standardized format for data collection, ensuring compatibility with FAIR principles.

Metadata attributes includes source, collection date, and relevance to specific farming practices.

The digital solution was designed with an intuitive interface, offering multilingual support and visual data presentations to cater to farmers with varying literacy levels.

	Attribute	Description
AgriConnect's attributes and metadata	ID	Unique identifier for the data asset
	Title	The name of the data asset
	Description	A description of the data asset
	Purpose	Why was the data collected or produced?
	Data creator	Who created the data?
	Data Manager	Who manages the data?
	Language	In which language(s) is the data
	Location	Where is the data located or stored
	Type	What type of data is it? Text, numbers, statistics, images, a database?
	Format	What format is the data in? For example, XLS, XSLX, CSV, JPEG, SQL DB, ODS, JSON, GEOJSON
	Creation date	When was the data created?

AgroThrive: a policy and advocacy investment

©Gates Archive/Thomas Omondi

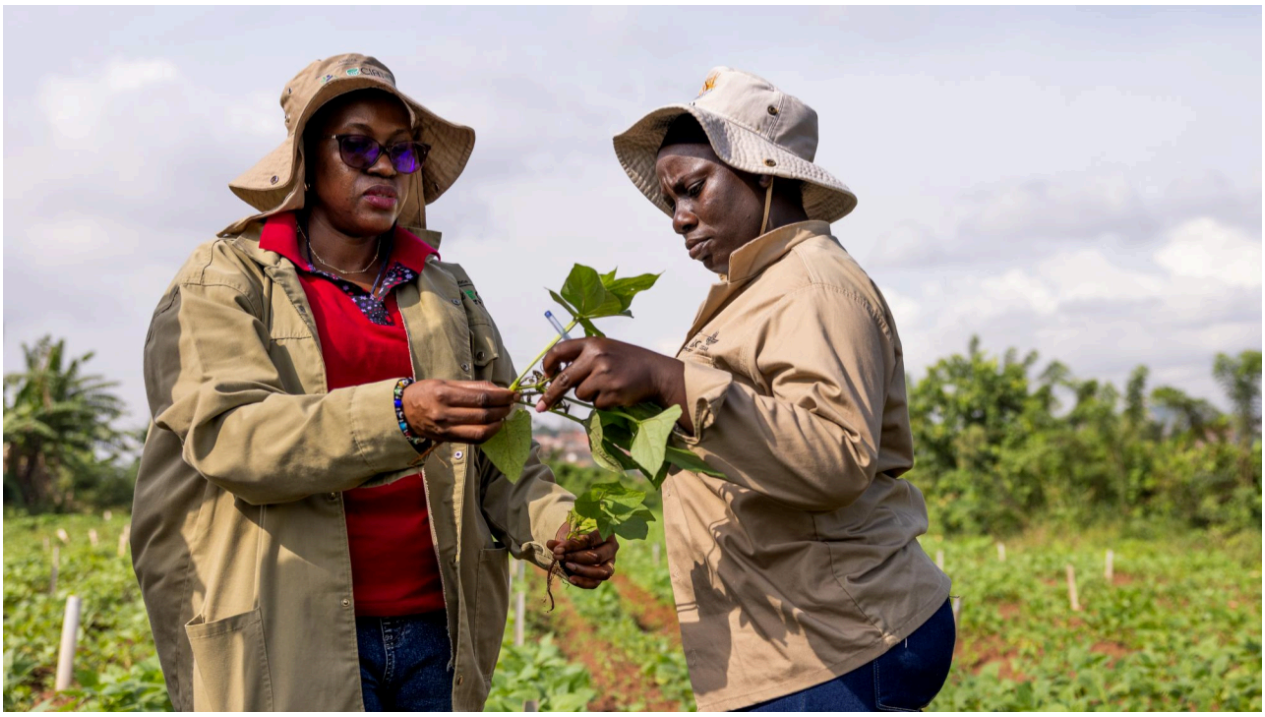
The AgroThrive team established standards for metadata and attributes of the data assets, ensuring that they aligned with FAIR principles.

Prioritized metadata that supports the findability and accessibility of data, such as clear descriptions of the data source, collection methods, and usage rights.

With a wealth of information, they developed standards that would make their findings easily discoverable and accessible, ensuring that every piece of data was meticulously cataloged and described.

	Attribute	**Description**
AgroThrive's attributes and metadata	ID	Unique identifier for the data asset
	Title	The name of the data asset
	Description	A description of the data asset
	Purpose	Why was the data collected or produced?
	Data creator	Who created the data?
	Data Manager	Who manages the data?
	Language	In which language(s) is the data
	Location	Where is the data located or stored
	Type	What type of data is it? Text, numbers, statistics, images, a database?
	Format	What format is the data in? For example, XLS, XSLX, CSV, JPEG, SQL DB, ODS, JSON, GEOJSON
	Creation date	When was the data created?

NGBT: a field research investment



©Gates Archive/Esther Mbabazi

As the project unfolded, NGBT carefully documented each data point with detailed metadata attributes.

This meticulous documentation ensured that future researchers could understand the context, methods, and significance of the data, fostering an environment of transparency and reproducibility.

	Attribute	Description
NGBT's attributes and metadata	ID	Unique identifier for the data asset
	Title	The name of the data asset
	Description	A description of the data asset
	Purpose	Why was the data collected or produced?
	Data creator	Who created the data?
	Data Manager	Who manages the data?
	Language	In which language(s) is the data
	Location	Where is the data located or stored
	Type	What type of data is it? Text, numbers, statistics, images, a database?
	Format	What format is the data in? For example, XLS, XSLX, CSV, JPEG, SQL DB, ODS, JSON, GEOJSON
	Creation date	When was the data created?

The EU estimates the true opportunity cost of not having FAIR data in research to be at €10.2 Bn each year.

European Commission, 'Cost-benefit analysis for FAIR research data—Cost of not having FAIR research data' (2018)

Learn more

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.