3.1 Plan the inventory

Deciding on the purpose, scope and granularity of your data inventory

Why should I do this?

To gain alignment among all the stakeholders involved. A properly developed data inventory will enable you to better communicate with other stakeholders—including grantees and local partners—on what data assets mean to the investment, and gain a clear understanding of the different types of data assets required and created, and how data flows.

The process of determining these factors also creates an opportunity for the relevant stakeholders to come together to collaboratively discuss and decide what is required for the investment.

In this activity you will:

Decide on the **purpose**, **scope** and **granularity** of your data inventory.

What is a data inventory?

A data inventory is a list of 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 in which it will be made available for others to use. It is a useful tool for any organization or project dealing with multiple types and sources of data.

A data inventory provides critical insights into the location and quality of data, and the technical and legal parameters guiding its management, usage and sharing.

Developing and updating a data inventory is crucial for recognizing data as a valuable asset, ensuring it is utilized and shared to its fullest potential.

Deciding the type of data assets to be included, and determining the metadata attributes, frequency of updating and level of detail, can help ensure all parties using the inventory are aligned, and that the information they are providing can be used by others.

When performed correctly, a data inventory process can have several benefits, including:

Mapping (or identifying) data assets that can be used by the investment.

Facilitating the analysis of transformed, normalized and enriched data.

Identifying gaps and areas where additional data may need to be collected.

Facilitating the identification of data outputs produced by the investment, and why they were produced.

Enhancing compliance with regulatory requirements using industry-specific compliance tools.

Enabling identification of any potential permission issues that could prevent the outputs being shared.

Facilitating conversation to bring stakeholders into alignment regarding what is meant by data (input and output), and various aspects across the data lifecycle (data sources/ownership/sharing/reuse).

When a data inventory process is not conducted correctly, there can be negative side-effects, including:

Greater difficulty in updating data as volumes increase

Impeding the utilization of unstructured informational assets

Increasing storage costs

Creating challenges in data cleaning, leading to higher data processing costs

Complicating the management of growing data volumes

Requiring more time to locate data, which can delay investment readiness

Inhibiting or complicating compliance and governance efforts

Reduced data discoverability, making it difficult for users to find and access relevant datasets

Poor alignment among stakeholders, leading to misunderstandings about data availability, ownership, and sharing protocols

Loss of valuable data insights, as overlooked datasets may prevent informed decision-making

Potential data security risks due to untracked data sources or gaps in security measures.

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 (an 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 questions below to create a plan for your inventory:

What are you trying to achieve with the inventory?

Start by clearly defining the objectives of the inventory. This will guide the selection of the necessary metadata attributes.

For example,
>   If your goal is to improve data management, consider gathering technical details about data storage, format, access rights, and maintenance costs.
>
>   To identify business opportunities, focus on metadata that reveals data value and usage frequency.
>
>   For compliance, prioritize identifying data containing sensitive personal information and access restrictions.

By defining your objectives, you ensure the inventory supports your organization's needs and adds strategic value.

Are you all on the same page?

Clarify the project definitions such as 'data' and agree the project's scope. Decide whether to include physical documents, all types of information assets, or solely data assets, including those produced by third parties. Agree on what dataset structures and format you will use (data can be unstructured, semi-structured and structured).

How detailed does it have to be?

Decide whether you will document individual data assets or provide a higher-level overview – for example, by grouping them by subject, time frame, or creator.

Who are the different personas who would use and contribute to the data assets?

Revisit the work done in Step 2.1 and build on it if required. It is also important to consider which personas need to be involved in deciding various aspects of your inventory.

Is your inventory sustainable?

Have a plan to keep it current, and integrate inventory updates into regular workflows. Consider whether you employ project managers or data stewards to update when the project is completed. Plan for the inventory's hosting, access, and maintenance responsibilities.

Efficient data gathering

To efficiently conduct your data inventory process, consider the data volume and available information sources. Employing the right strategies will ensure comprehensive coverage and stakeholder engagement. Here are some recommended approaches:

Organize workshops and/or trainings for stakeholders
Address data-sharing concerns by organizing training sessions or workshops with key stakeholders. Discuss best practices, the need for data inventories, and how data will be used, cataloged and protected. Workshops are a positive way to build wider awareness of FAIR and responsible data practices, and to bolster data literacy across a data ecosystem.

Engage 'Data Custodians'
Have the people responsible for managing and safeguarding specific datasets (referred to as

data custodians) in the process of gathering information for the data inventory.

Conduct interviews with stakeholders
Arrange discussions with data custodians and other relevant personas to understand the context of their data and to gather detailed insights. Interviews provide valuable context and are very feasible, especially if a few targeted discussions can provide deep insights.

Distribute surveys
Use online surveys to efficiently reach multiple respondents and gather consistent responses.

Implement automation
Set up systems to prompt users for essential metadata when new items are added to content management systems, ensuring long-term sustainability.

A **benefit-risk assessment** is a valuable step to help you maximize the benefits of data use while identifying and mitigating potential risks. This assessment helps you clarify who the beneficiaries are, define the expected benefits, and assess how likely they are to be achieved. It also articulates any risks that might prevent these benefits from being realized. For example, if data includes sensitive information, a benefit-risk assessment can guide you in managing privacy and security issues.

The USAID Using Data Responsibly document provides helpful guidance for conducting a benefit-risk assessment, especially for handling sensitive data assets (see pages 11 and 17). This assessment also serves as a prompt for discussions with data custodians, ensuring shared understanding and preparedness for managing data responsibly.

Investment types
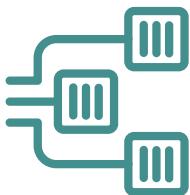
Overview

©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: Planning for the data inventory

AgriConnect is building a data platform that provides farmers with real-time insights on soil health, crop prices and weather conditions. Their inventory plan is driven by the need to make agricultural data accessible and useful for smallholder farmers.

What are you trying to achieve with the inventory?

AgriConnect's main goal is to support farmers in making informed decisions. It wants an inventory that focuses on real-time data access, data interoperability, and usability for non-technical users. By organizing data assets by relevance to the growing season, it ensures that farmers receive timely, actionable information.

Are all stakeholders aligned?

AgriConnect holds a planning session with farmers, data providers and tech developers to define data needs and formats. They align on including only digital data assets, such as soil quality data from sensors, climate data from government sources, and market prices from cooperatives. Structured data formats are prioritized to facilitate integration into the platform.

What level of detail is needed?

To support decision-making, AgriConnect plans a detailed inventory that tracks specific variables like crop type, geolocation and current price. It organizes data by region and crop season, allowing farmers to search for the information that is most relevant to their location and growing schedule.

Who are the personas that will use and contribute to the data assets?

AgriConnect identifies key personas: farmers who need practical, easy-to-understand data, data stewards to manage updates, and tech developers who integrate the data into the platform. Each persona's role is defined, ensuring data updates are maintained consistently for reliable access.

Is your inventory sustainable?

The AgriConnect team plans to incorporate regular updates, especially for weather and market price data, as these change frequently. They assign data stewardship responsibilities to the platform team, who will update the inventory daily. Hosting will be centralized, with cloud access for secure, scalable data management.

AgroThrive: a policy and advocacy investment

©Gates Archive/Thomas Omondi

AgroThrive: Planning for the data inventory

AgroThrive aims to provide evidence-based policy recommendations for agricultural reform. Its inventory focuses on data assets that support government decision-making and advocate for sustainable agricultural policies.

What are you trying to achieve with the inventory?
AgroThrive wants to develop a policy-focused inventory that supports data-driven decision-making. Its objective is to centralize policy-relevant data on topics like agricultural financing, climate resilience, and smallholder livelihoods, to streamline policy analysis and collaboration with stakeholders.

Are all stakeholders aligned?
AgroThrive holds alignment sessions with policymakers, local government representatives and NGO partners to agree on the inventory's purpose and scope. Together, they decide to include both digital data and physical reports for reference. The focus is on structuring data by theme—such as financing or climate resilience—to support policy development efficiently.

What level of detail is needed?
AgroThrive opts for a moderate level of detail, ensuring that their inventory captures essential variables like policy impact, regulatory alignment and demographic reach. Data is organized by policy area and jurisdiction, enabling targeted analyses for regional policy planning.

Who are the personas that will use and contribute to the data assets?
AgroThrive identifies key personas, including policy analysts, who perform data analysis; government officials, who review findings; and advocates, who use the data to push for reform.

Each role contributes to the inventory by providing updates or using data for specific policy recommendations.

Is your inventory sustainable?
To ensure sustainability, AgroThrive integrates the inventory into routine policy review cycles, establishing biannual updates to reflect the latest data and assign roles to policy analysts for maintaining data accuracy. Hosting will be secured on a government-shared server, with access controls to protect sensitive data.

NGBT: a field research investment

©Gates Archive/Esther Mbabazi

NGBT: Planning for the data inventory

NourishGen BioTech (NGBT) is conducting field research to develop resilient crop varieties. Its data inventory is designed to support in-depth scientific analysis and facilitate data sharing among research partners.

What are you trying to achieve with the inventory?
NGBT aims to build an inventory that tracks research data from field trials. Its objective is to document data on crop yields, pest resistance, and climate adaptability, which will support crop improvement and collaboration with other research institutions.

Are all stakeholders aligned?
NGBT aligns with crop scientists, field technicians and research partners to define the inventory's scope. They agree on including data collected from digital sources like field sensors and lab results,

as well as geospatial data for environmental analysis. Structured formats are chosen to facilitate interoperability with other research databases.

What level of detail is needed?

NGBT decides on a highly detailed inventory to capture specific attributes like crop variety, soil conditions, and experimental treatments. Data is organized by experiment ID and trial location, allowing for precise tracking of research activities and environmental conditions.

Who are the personas that will use and contribute to the data assets?

Key personas include crop scientists, who analyze data; data managers, who ensure proper cataloging; and research partners, who access and share findings. These roles are critical for maintaining a robust data lifecycle and ensuring all trial data is accurately documented.

Is your inventory sustainable?

NGBT establishes a routine for data entry and updates, assigning data management tasks to the research team and setting up quarterly updates to reflect the latest trial data. It also chooses an open-access repository to share findings with the broader scientific community. Access is restricted to certain fields to protect proprietary data, while published findings remain open for peer review and collaboration.

FAIR principles also foster greater collaboration, boost citations, maximize publications, enhance reputation, and show leadership by example.

Arun Jadhav, Manager Digital Development, CABI, at International Conference on Open and FAIR Data Ecosystem Principles, Policies, and Platforms 2023

Learn more

Acknowledgements

FAQs

Glossary

Accessibility

Privacy & cookies

T&Cs