3.2 Identify output and input data assets

Identifying the output data assets your investment will yield, and the input data assets that produce them

Why should I do this?

To start exploring the availability, quality and terms of reuse for the input data assets that your investment's output data assets depend on.

In this activity you will:

Gain clarity on what the key output data assets are, and what inputs are required. This is essential for mapping dependencies, ensuring data quality, and understanding the licensing requirements to meet project goals.

Identify what output data assets your investment is producing.

Identify what input data assets are required to produce your output data assets.

Populate your data inventory with your input and output data asset.

For each data asset, consider questions like: Does the input data exist? Is it accessible and reusable? What are the licensing terms? By identifying these assets, you can assess any challenges in data availability and quality, as well as compliance needs.

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 questions to identify your input and output data assets, and populate your data inventory:

What are the output data assets your investment is producing?

As it may not be feasible to make a note of all data assets being produced by your investment, start with the most important or 'key' data assets. Activities that will help you

identify these key data assets are:

- Discussion between the PO and grantee.
- Discussion with other partners (e.g. those you identified in Step 2.1, if you followed these steps sequentially).
- Revisiting or completing the activity on data ecosystem mapping in Step 2.3.

What input data assets are being used to produce the key output data assets identified?

Make a note of the input data assets required for your output data assets in the same table in your workbook.

What are the sources of the input data assets?

Determine whether the data comes from internal operations, public data assets, or third-party providers. This helps in assessing the reliability and availability of the data.

How is the data licensed, and does the license support its intended use, including reuse and sharing?

When thinking about the input data, consider the following questions:

- Do you have the required license?
- Does the license allow for it to be used as per the investment's plan (including reuse and data sharing)?

Has the data been collected following good practice (e.g., have you considered bias/informed consent)?

If your activities related to your intervention and data levers are dependent on input data assets, and if those input data assets are not available, you may need to reassess

activities and identify new data levers. Data levers can be determined using the activities in Step 1.

What types of data are being collected to produce the identified output data assets?

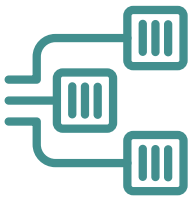
Identify the variety of data being gathered, such as surveys, market research reports, sensor data, and financial transactions. Understanding these types helps in recognizing the raw materials available for processing and analysis.

How is the data required for the output data assets being captured and at what frequency?

Understanding the collection mechanism and frequency is vital for ensuring data relevance and timeliness.

Refer to the investment type examples to help you complete your plan.

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 populating the data inventory

Rashima (lead/assigned grantee) organized a data alignment workshop with key stakeholders, including Faisel (researcher) and Chris (third-party publisher). During the session, they began mapping out essential output data assets, such as digital records of soil quality assessments, crop trend analysis reports, and market access insights. They also captured the corresponding input data that would feed into AgriConnect's digital platform. Initially, the team assumed that accessing government climate data and economic indicators would be straightforward, as these were part of open data portals.

However, when Rashima approached Chris (third-party publisher), she discovered that some of the most detailed government data, particularly on farmer income levels and private sector agricultural investments, was restricted due to data privacy regulations. Chris explained that these indicators were tied to confidential economic strategies, and special permissions were needed for access. Rashima realized their assumptions had overlooked these licensing constraints.

To adapt, Rashima updated the data inventory workbook to include notes on restricted datasets and highlighted those that required special access. She also decided to explore alternative data sources, such as summaries from agricultural reports or aggregated regional-level data to fill these gaps. This process revealed to Rashima the importance of conducting thorough data audits and checking licensing conditions for all external datasets.

Next, Rashima reached out to Chima, the smallholder farmer, who had already contributed initial raw data about crop conditions and input usage from his community. She used Chima's feedback to better understand the gaps in the existing data. Chima also identified other farmers who could contribute localized data, ensuring the platform's relevance.

With this information, Rashima organized a meeting with two of Dataland's foremost experts on precision agriculture (Noora and Marie), who were invited to review and refine the inventory. Noora provided insights on sensor data, highlighting the importance of setting standards for incoming soil and climate data, while Marie discussed how best to display these insights on the platform. Noora flagged that detailed sensor data from private farms could reveal sensitive information about farmers' land conditions or cultivation methods. Rashima took note of this in the inventory, marking it as 'Sensitive Data' to ensure secure storage and limited access within the digital platform.

Rashima used the feedback to populate the final inventory, linking input and output assets and clearly mapping dependencies, accessibility, and licensing requirements. By the end of this iterative process, AgriConnect had a robust inventory that reflected contributions and dependencies from each partner and participant. Rashima distributed the final inventory document to all stakeholders, inviting ongoing feedback to keep it a living document.

AgroThrive: a policy and advocacy investment

©Gates Archive/Thomas Omondi

AgroThrive populating the data inventory

For AgroThrive, building a comprehensive data inventory was a complex task, given the wide range of stakeholders and the diversity of the data involved. Kaira (lead/assigned grantee) started the inventory-building process with a central meeting, inviting key project partners: Aziz (project partner), Anna (project partner), and Zora (private sector liaison).

During this meeting, Kaira guided the team through a brainstorming session to identify the critical policy output data assets, such as infrastructure improvement reports, financing models for smallholder farmers, and recommendations for gender-inclusive AgDev initiatives.

The session was productive, but as they outlined the inputs for these outputs, they realized there were assumptions about accessibility that had not been fully validated.

During a follow-up meeting with Imamu (third-party publisher), Kaira discovered that accessing key government data on agricultural trends was not as straightforward as they had assumed. Imamu explained that certain economic and demographic indicators were restricted due to privacy laws and required special permissions or anonymization protocols to be used in policy recommendations. This revelation meant that Kaira and her team needed to revisit their data inventory and identify alternative sources or approaches for obtaining sensitive information.

Next, Kaira reached out to Haben (smallholder farmer) to discuss the real-life challenges faced by smallholders and the data that could support solutions. During this conversation, Haben provided detailed information on loan repayment rates and crop yield fluctuations, and gave feedback on existing government support schemes. Kaira reassured Haben that any sensitive data collected, such as financial and land ownership records, would be anonymized and only shared with explicit permission. The team created a separate section in their data inventory workbook to flag datasets requiring ethical considerations or specific licensing permissions.

Later, Kaira met with Saanvi (technical consultant) to review sensor data capturing climate variations and soil conditions. Saanvi explained the significance of soil moisture and temperature measurements, which were essential for designing climate resilience strategies. Together, they outlined the required input data assets and standardized the data collection process for consistency and quality.

Kaira asked Saanvi about the potential sensitivity of geospatial climate data. Saanvi flagged this as an important issue, noting that precise location data could inadvertently reveal private information about specific farm plots or practices. The team agreed to anonymize location-specific details and documented this decision in their inventory under a 'Sensitive data' category, ensuring compliance with AgroThrive's data governance policy.

Aziz (project partner) and Anna (project partner) took the lead in refining the inventory by contributing training data and financial outreach insights from their respective programs. They discussed the accessibility and licensing of their data, ensuring it aligned with AgroThrive's objectives. Lata (gender consultant) also provided her input, adding gender-specific metrics that needed to be tracked to support inclusive policy formulation.

Once all input and output assets were mapped, Kaira arranged a final meeting with Adnan (government liaison) to review the preliminary inventory. Adnan provided valuable feedback on compliance with government policies and identified gaps in some of the datasets.

To wrap up the process, Kaira hosted a virtual 'data validation' session with Imamu (third-party publisher). Imamu gave feedback on accessibility and licensing terms, helping Kaira make necessary adjustments to ensure all external data could be used according to the project's data governance policies.

NGBT: a field research investment

©Gates Archive/Esther Mbabazi

NGBT populating the data inventory

Farah, the lead grantee on the NGBT project, started the inventory-building process with a sense of urgency. The goal of the project was to develop a climate-resistant and more nutritious barley varietal, which required gathering insights from multiple disciplines, each contributing different datasets. Knowing this, Farah scheduled a planning meeting with her core team, including Nasser, Cali, and Chanda, at the research center.

The team began by discussing key output data assets like varietal development reports, nutrient analysis results, and soil health improvement findings. As each member outlined their expected outputs, it became clear that understanding and documenting the required input data was the next step. Nasser, the researcher, emphasized the importance of combining on-the-ground data with long-term climate trends and genetic analyses.

To get a comprehensive view, Farah organized field visits with Davu, the smallholder farmer. During these visits, Davu shared details on current crop yields, soil erosion issues, and coping mechanisms against climate variability. This raw, community-level data was crucial for the project's output. During the visit, Farah also learned of the sensitivity around data on soil health and cropping techniques. Davu expressed concerns that revealing specific practices could give competitors an advantage or expose their village to external scrutiny. Farah assured Davu that such information would be anonymized, and she made sure this was clearly documented in the inventory's 'Sensitive data' section, which included protocols for handling confidential or location-specific datasets.

Farah then met with Joe, the third-party publisher at the government agency, to access historical data on agricultural indicators. Joe's open data repository provided additional economic, social, and

climate trend insights that the team needed. Back at the research center, Farah worked with the team to document all input and output assets in a detailed inventory.

Charlotte, the climate scientist, and Cali, the geneticist, helped cross-reference this inventory against their own datasets on climate impacts and genetic variations. Together, they identified gaps in existing data and proposed solutions to fill them. Farah set up a separate meeting with Jaya, the gender specialist, who provided essential context on how to track the intersectional benefits of the new barley variety for women and children.

By the end of this iterative process, the team had built a comprehensive data inventory. Farah convened a final meeting where each partner reviewed their contributions, ensuring everything was well-documented and compliant with open-access requirements. This inventory became the backbone of NGBT's research, providing a clear path forward to responsibly publish and share their findings at the project's conclusion.

Data and data ecosystems foster a diverse range of benefits for individuals and communities.

Investment Case - Multiplying progress through data ecosystems

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.