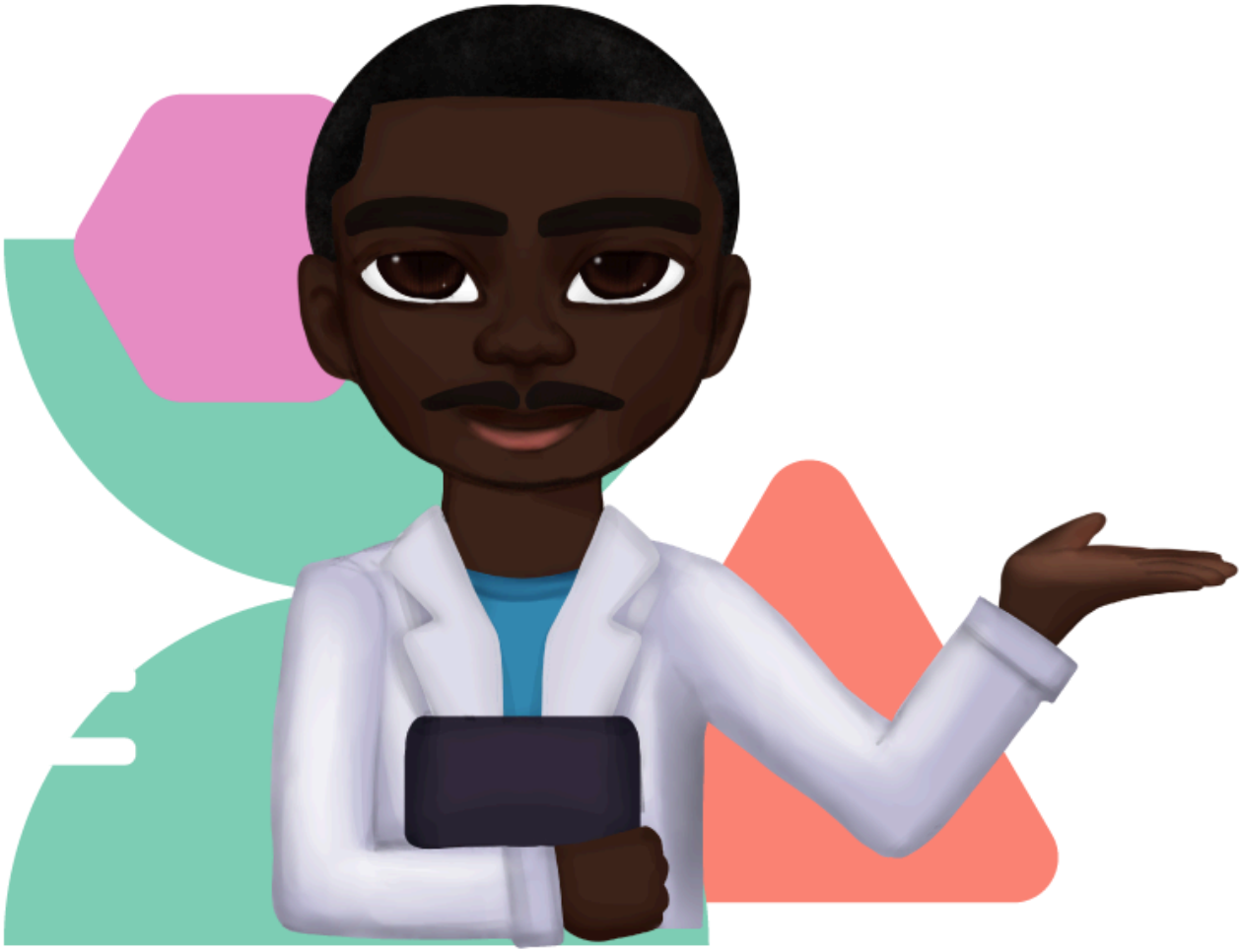

5.1 Data Governance Policy

Developing a policy that establishes clear principles and guidelines for the collection, management, protection, and ethical use of data

Why should I do this?

To set clear expectation and accountabilities for the collection, management, protection, and ethical use of data within your investment.

By developing a data governance policy, you ensure consistency, compliance, accountability, and strategic alignment with your data goals, ultimately fostering trust and enhancing data quality and security.

In this activity you will:

Develop a data governance policy for your investment. It involves using people, processes and technology to manage data to maximise its value while minimizing risks and harms. Data governance involves managing data as an asset, e.g., to enable decision-making, support research, or drive innovation.

A data governance policy will provide:

Clarity in roles and responsibilities

Accountability: Defined roles ensure that specific individuals or teams are responsible for different aspects of data management, from data quality to data security. This accountability helps ensure that tasks are completed effectively and that there is a clear point of contact for any issues that arise.

Consistency: With clearly defined roles, procedures and processes can be standardized, which helps to maintain data integrity, ensure compliance with regulations, and facilitate smooth operations.

Efficiency: When roles and responsibilities are clearly delineated, duplication of effort and gaps in coverage are reduced, leading to more efficient operations.

Compliance: Many regulations and standards require organizations to have formal data governance practices, including clear role definitions. Compliance with these regulations is critical to avoid legal penalties and reputational damage.

Risk management: Clearly defined roles help to identify and mitigate risks associated with data management, as well as designating individuals or teams to handle risk assessment, mitigation, and response.

Improved decision-making: With clear roles, data governance activities such as data stewardship, data quality management, and data privacy are better managed,

leading to more reliable and accurate data, which in turn supports better decision-making processes.

Alignment on policies and processes

Compliance and privacy: for handling personal data in compliance with relevant regulations and ethical standards, and the specific data security measures that will be utilized.

Data sharing: for collaborative efforts and informed decision-making processes where the value of data can be maximized while adhering to legal and ethical standards.

Data quality: for reliable and accurate data insights, which can significantly influence the success of interventions and its sustainability.

Data access: for data handling practices—encompassing collection, storage, security and access—are not only efficient and regulatory compliant, but also safeguarded against unauthorized access and potential security breaches.

Better decision-making

With consistent governance, data is better managed, which leads to higher quality data. This, in turn, supports more reliable, data-driven decision-making, giving you and your stakeholders a clearer understanding of the investment's progress and impact.

What will the policy cover?

The policy will address the following key areas:

Purpose and scope: Define why the policy is needed and the boundaries it applies to within your investment.

Data governance principles: Establish high-level principles that guide how data is handled, focusing on integrity, security, transparency, and FAIR principles.

Roles and responsibilities: Clearly outline the roles involved, including data stewards, custodians and users, and their responsibilities for ensuring data governance.

Compliance and privacy standards: Describe the approach to data privacy, informed consent, and adherence to legal frameworks like GDPR.

Data quality standards: Define how data accuracy, completeness, and consistency will be ensured.

Data access and control: Detail the methods for securing and controlling data access.

Data security and risk management: Outline how data will be protected against breaches, including encryption and access controls.

Data sharing and usage policy: Explain the rules for secure and compliant data sharing within the project and with external stakeholders.

Monitoring and review: Set up a process for regular reviews of the policy and compliance checks.

Training and awareness: Ensure all stakeholders understand their roles and responsibilities and are equipped with the knowledge to maintain high standards in data governance.

---

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) You can use the Data Governance Policy template for Step 5, to populate the relevant sections.

3) Refer to the investment type examples to help you with this activity.

4) The other collaborating organisations within the investment should be consulted on areas of their expertise, and also for review.

5) Complete the various sections in your data governance policy template. You may use the sample text provided as guidance in the template, which can then be deleted.

If you require additional guidance, consider the below sections:

Roles and responsibilities

Data stewards: A data steward is an individual or entity that has the ultimate responsibility and authority over a specific set of data within an organization, and is typically accountable for making decisions regarding the use, access, security, and overall management of data. Data stewards also ensure the appropriate handling of data, aligned to organizational policies and goals.

Data custodians: The data steward, while ultimately responsible for creating linkages, will not act alone in facilitating the flow of data to those who need access. Data custodians are responsible for facilitating data access.

Data users: Those needing to use (or access) data, or 'data users', are a fluid group, as many people involved in the project will be stewards of certain data who also need to access and use other data.

Data subjects: Those whose personal data has been collected and is the subject of processing.

Below is a list of various key members of a data governance team, and their responsibilities and roles in maintaining an investment's alignment with the FAIR data principles. The list is exhaustive and is meant to represent the full breadth of positions that may be required for the most complex project. Remember that your investment's needs may vary, and not all roles are necessarily held by distinct individuals, if at all. For example, should your investment include fewer data sources and entries, a single person could hold the role of both data governance manager and database administrator, among others.

Grantee (steward, custodian and user): A grantee is likely to be the most familiar with the data (and outputs) of a project, and can share this knowledge with the data management team, making them the de-facto 'data steward' of the organization's data related to this project. They will also be working directly with the stakeholders inputting data or accessing the organization's data, and will need to be aware of the procedures that must be followed to inform those stakeholders.

Data manager (custodian): A data manager is responsible for designing and implementing data architectures, as well as defining data structures to ensure data integrity and quality.

Data engineer (custodian): A data engineer builds and maintains the systems and architecture for collecting, storing and analyzing data.

Data analyst (user): A data analyst works with raw data to extract insights to inform decision-making, and to decide which data will be included in the final output.

Data scientist (user): A data scientist applies statistical and machine learning techniques to analyze and interpret complex datasets, in addition to developing predictive models and algorithms.

Database administrator (custodian): A database administrator manages and maintains databases to ensure performance, security, and availability, as well as performing routine recovery and backup operations.

Data governance manager (custodian): A data governance manager develops and enforces data governance policies and procedures, in addition to ensuring internal compliance with data privacy regulations that have been developed by the compliance officer.

Data quality analyst (custodian): A data quality analyst monitors and accesses the quality of data, while also identifying any issues with data quality.

Security analyst (custodian): A security analyst focuses on securing and protecting data from unauthorized access or breaches, as well as implementing security measures for transmission and storage.

Compliance officer (custodian): A compliance officer ensures that an organization's data practices comply with the legal and regulatory requirements specific to the country or region in which they operate. Consequently, it is vital for them to stay informed about changes and updates in the regulatory environment affecting their organization.

Compliance and privacy standards

Informed consent: Obtain explicit consent from individuals before collecting their personal or sensitive data, clearly outlining the data collection's purpose. Practice data minimization by collecting only what is essential for specified purposes, and maintain transparency by informing individuals about the data collector, the purpose of collection, and any relevant details concerning their data's use

Purpose limitation: Use personal data only for its original, consented purposes and avoid using it for anything else without extra consent. Keep personal data accurate and current by regularly updating it and quickly fixing any mistakes.

User rights: Enable individuals to access their personal data and request it in a portable format when relevant. Offer options for correcting inaccurate data and erasing unnecessary information on request or withdrawal of consent. Additionally,

establish processes for individuals to object to specific uses of their personal data, especially for direct marketing purposes. Should a party feel that their data has been misused, or if a privacy violation occurs (such as a data leak or hack), there should be grievance redress mechanisms in place for them to register these concerns. Policies around grievance redressal should designate a responsible person to address these issues. There should be a clear mechanism for complaint submission and a procedure for reviewing complaints. A timeline should be established for the swift resolution of complaints, along with a structure for corrective measures, depending on the type of complaint. This can include rectification of the grievance, policy updates, or even compensation in some instances. Additionally, there should be a mechanism for appeals if the party feels that the resolution was unjust.

Data quality standards

Define data quality dimensions: Ensure that all collected data accurately reflects the real-world scenarios it intends to represent, maintain consistency across various datasets and over time, and capture all necessary information while minimizing missing values. Additionally, keep data timely and readily available, and adopt data collection and processing methods that guarantee stability and consistency of results.

Incorporate FAIR principles: Revisit the shared FAIR goal (developed in Step 4) and ensure data quality requirements are aligned to that goal.

Establish data acquisition standards: Before collecting data, evaluate the reliability and credibility of sources to ensure the integrity of the data collected. Establish standardized data collection methods for consistent and reliable data gathering, and simultaneously collect detailed metadata to enhance the data's findability and support its future reusability.

Data access and control policies

Establish data collection protocols: Ensure you have a comprehensive understanding of data provenance. Then select the most appropriate methods for collecting data, whether through APIs, manual input, or other forms of integration, while rigorously ensuring that all practices comply with relevant legal and regulatory standards, such as GDPR or HIPAA.

Incorporate FAIR principles: Revisit the shared FAIR goal (developed in Step 4) and ensure data access requirements are aligned to that goal

Establish data security measures: Ensure data security by encrypting it and implementing access control mechanisms such as Role-Based Access Control (RBAC) to restrict data access to authorized personnel only. Conduct regular security audits and monitor access logs to promptly identify and address any unauthorized access or potential security risks.

Facilitate data access: Set clear guidelines for accessing data through APIs and query interfaces for efficient access, and ensure users are strictly authenticated and authorized to access only the data relevant to their roles.

Data security and risk management

Data storage systems: Outline the guidelines for storage solutions to ensure optimal security and accessibility, and establish clear data retention policies that dictate how long data is stored, including procedures for its secure archiving or deletion under legal obligations.

Establish data security measures: Outline the high-level measures to be followed to ensure data security.

Data sharing and usage policy

Establish data sharing guidelines: Clear rules for when and how you can share the data with others, who those others are, and what agreements or rules you follow to do this safely and legally (sharing).

Foster transparency and accountability: Enhance transparency and accountability in your data-sharing policy by committing to regularly publishing transparency reports detailing your data practices, including the types of data shared, the purposes for sharing, and the parties with whom the data is shared. You can also establish accountability measures by setting up mechanisms to address any issues or breaches related to data sharing. This includes creating clear procedures for reporting incidents such as data breaches, and outlining steps for their rectification.

Consider data licenses: Consider licenses to ensure data is used legally and ethically. The type of license attached to the data being shared—whether open (e.g., Creative Commons), commercial or proprietary—dictates how the data can be used, shared and distributed. Consider usage rights specifying how the data can be used, and attribution requirements, which may require citing the original source. Attention to licensing is especially important for commercial or proprietary data, which often has stricter requirements than data intended for educational or research purposes.

A data governance policy is a high-level strategic document that outlines how data is managed, ensuring compliance, security, and quality within your investment. Depending on the size and complexity of your investment, you may not need a separate data governance policy. You may just choose to include some of the elements from this section in your data management and access plan (DMAP).

Investment types

Overview

©Gates Archive/Mansi Midha

Every investment project is unique

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

AgriConnect: developing a data governance policy

AgriConnect aims to enhance food security by collaborating with organizations to collect and analyze information on crop production levels, soil quality and climatic conditions in areas. Key stakeholders include Rashima (lead grantee), Chima (smallholder farmer), Faisel (researcher), and Chris (third-party publisher).

As the project advanced, it became apparent that each organization had its methods for handling and exchanging data, which resulted in variances in data quality and difficulties in safeguarding information such as farmers' details. To tackle these issues, AgriConnect implemented a policy on data governance.

This policy had advantages.

The responsibilities were clearly defined for each individual involved in handling the data, to ensure that everyone understood their role—whether gathering data from farms or storing and securing

it properly. For example, the regional teams were assigned the role of data stewards, making them responsible for ensuring that data gathered from local farms was accurate and safe.

AgriConnect ensured consistency among all partners by establishing a procedure for gathering and presenting data. This uniformity was crucial for evaluating data consistency across regions and for making trustworthy assessments regarding crop wellbeing and productivity.

With defined responsibilities and processes, in place the exchange of information among collaborators was more seamless. Each team member understood their tasks clearly resulting in a decrease, in tasks and enabling teams to focus on data analysis and using it to enhance their work.

Because the project covered multiple regions, each with its own privacy rules, the policy also ensured that every partner followed local data protection laws. This not only kept the project compliant but also helped build trust with the communities they were working with.

By assigning responsibility for different aspects of data management, AgriConnect was better equipped to handle risks, like data breaches or the loss of important information. Regular security checks were implemented, and each organization had clear guidelines for protecting the data, including using encryption and controlling who could access it.

AgriConnect and its partners were able to make decisions due, to data quality and consistency. For instance they could pinpoint regions experiencing decreased crop yields. Promptly allocate resources to tackle soil health problems.

This case shows how a well-implemented data governance policy helped AgriConnect keep everything on track, ensuring that all the partners were aligned. The policy played a big part in the project's success by improving data quality, which in turn led to more effective decision-making and better outcomes overall.

AgroThrive: a policy and advocacy investment

©Gates Archive/Thomas Omondi

AgroThrive: developing a data governance policy

As the project's scale and complexity became clear, Kaira needed to decide whether a data governance policy or a data management and access plan (DMAP) was more suitable—or if both were necessary. AgroThrive's objective was to provide data-driven agricultural policy recommendations to governments and stakeholders. These recommendations depend heavily on integrating and analyzing a wide array of existing data sources, including soil health databases, crop yield statistics from local governments, climate data from meteorological services, and market price trends from various third-party providers.

The project collaborates with key partners such as Zora (private sector liaison), Saanvi (technical consultant), and Lata (gender consultant), among others.
Kaira explained to her team: "With such a complex project, we need clear principles to guide us, but we also need an actionable plan for managing the different types of data day-to-day. A data governance policy will give us that strategic direction, while a DMAP will put these principles into practical tasks."

There was a possibility that the data they collated would have varied data standards, formats, levels of detail, and gaps in quality documentation, thereby making it difficult to draw accurate and reliable insights. There would also likely be some concerns about privacy and intellectual property rights from the organisations providing the data.

During the data ecosystem mapping phase of the project, Kaira had earlier discovered that one of the project's key partners, the national NGO headed by Aziz, already had a comprehensive organizational-level data governance policy.

While that existing policy provided strategic direction on issues like data privacy and ethical considerations, which aligned closely with AgroThrive's goals, it did not cover all of AgroThrive's unique requirements, such as integrating gender-sensitive data and ensuring inclusive participation in policy recommendations. Therefore, the team created a more targeted policy.

The team first outlined specific roles, appointing key members like Adnan, the government liaison, to ensure compliance with Datapur's policies, and Zora, the private sector liaison, to align with industry needs. Each role included detailed responsibilities for data management, protection, and sharing.

By establishing data quality checks, they would be better able to evaluate the completeness, accuracy and relevance of incoming third-party data. This would also help them to define and communicate clear data standards to external partners, ensuring consistency in formatting and reporting.

Having a data governance policy would enable them to outline policies for respecting privacy laws and third-party agreements, ensuring that any personally identifiable information (PII) and confidential datasets are handled in compliance with regional regulations and contractual terms. They were also able to define specific guidelines for sharing and accessing restricted datasets, maintaining compliance with external data providers' usage conditions.

Lastly, the policy ensured that the data they received was only accessed by authorized staff, especially the sensitive datasets. This would help protect intellectual property and maintain accountability.

By focusing on these key elements, AgroThrive was able to ensure that its reliance on diverse secondary and third-party data sources did not compromise the quality and reliability of its policy recommendations. The data governance policy helped AgroThrive manage incoming data more effectively, maintain transparency with stakeholders, and improve overall data integrity in its analyses.

NGBT: a field research investment

©Gates Archive/Esther Mbabazi

NGBT: developing a data governance policy

The NGBT research project is focused on developing a new varietal of barley that not only offers enhanced nutritional value, but is also more resilient to the impacts of climate change. The project aims to increase agricultural yields and sustainability, contributing to improved food security in vulnerable communities. This initiative involved extensive collaboration among experts such as Farah (lead grantee), Nasser (researcher), Kama (private sector liaison), Charlotte (climate scientist), and Jaya (gender and children specialist).

At the project's conclusion, the NGBT team planned to publish a comprehensive set of data assets responsibly. These assets would include detailed datasets, research methodologies, outcomes, experimental materials, and budget information that can be used for future research and development efforts.

Given the complexity and scale of the project, the diversity of the information and the sensitive nature of some of the data, the team realized the need for a customized data governance policy to ensure consistency, accuracy, privacy and security in the collection, management and sharing of valuable data assets. Their policy would apply to all data collected, processed and shared throughout the project, ensuring alignment with the project's goals of transparency and responsible data management. The policy clearly assigned roles of project data steward, custodians and users.

The NGBT project committed to respecting data privacy by adhering to relevant regulations and ethical standards.

It made a commitment always to obtain informed consent, maintain confidentiality of participants' information, and respect the intellectual property of all contributors.

Its policy emphasized the need for high standards for data quality, focusing on accuracy, consistency, completeness and timeliness. Regular quality checks and metadata documentation practices would be essential to ensuring reliable datasets for project analysis and publication.

It clarified that there would be controlled data sharing within the project and with external stakeholders, governed by formal agreements that define confidentiality, usage terms, and security requirements.

The policy also indicated that NBGT would undergo regular reviews to maintain alignment with legal standards and evolving project requirements.

Lastly, the policy stated that all team members would receive training on data governance principles, privacy standards, and security protocols.

This tailored data governance policy ultimately enabled NGBT to responsibly manage and share its diverse data assets. By focusing on data integrity, privacy and access, the project not only safeguarded its participants and stakeholders but also laid the groundwork for future research.

Delivering FAIR data practices is as much about people as it is about technology.

Dr Negussie Efa, Country Programmes Manager, Africa, CABI

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.