Data Sharing Toolkit

Module 7

Ensuring sustainable access to data

Guide

Ensuring sustainable access to data

About this guide

This guide is designed to help **program officers and grantees** ensure data created and used during a project continues to be useful and usable once that project ends. Although this may seem low priority at the start of an investment, planning for sustainable access to data from the beginning can benefit all stages of the work and support good data practices overall.

Investment stage relevance: Start concept | **request proposal** | **refine proposal** | create agreement | request approval | obtain signatures | **active** |

Tools supporting this guide: | Data Access Map | Data Ethics Canvas | Data Ecosystem Mapping Tool

Quick links

- About this guide
- Ensuring data remains accessible
- A checklist for sustainable access to data
  - 1. Assess upfront what data is being created or used
  - 2. Decide if there is a benefit to sustainable data access
  - 3. Plan for data stewardship, management and resources
  - 4. Check data policies, standards and platforms support sustainable access
- Further Resources

Ensuring data remains accessible

Making access to data sustainable means ensuring data remains accessible over time. This maximises the utility of the data by making it possible for the greatest number of people to benefit from it, making better decisions and reducing risk. Sustainable access to data allows governments, businesses, NGOs and others to rely on the data for products, services and research, and increases the impact of investments, avoiding waste down the line. When access to data is unsustainable it cannot be relied upon for investments by others, and may waste resources if it needs to be collected again.

The FAIR data principles state that appropriate data should be findable, accessible, interoperable and reusable. Sustainable access to data supports this ambition by making possible access and reuse of appropriate data for anything anytime. Sustainable access supports data ecosystems, because it gives organisations the confidence to invest in using data to develop new products, services and research.[1]

[1] (2018), Fiona Smith, Leigh Dodds, Charlotte Day, Ruthie Musker, Martin Parr, 'Creating FAIR and open agricultural ecosystems'. Accessed July 2020. https://gatesopenresearch.org/documents/2-42

However, just as benefits increase when they are compounded over time, the impact of harms can grow in the same way. Sustainable access therefore relies on good data stewardship to ensure any harmful impacts of data are minimised in both the short and long term.

Teams and organisations commonly make assumptions about what will happen when a project using data ends, without consulting the other bodies implicated. For instance, an NGO collecting and storing data during a project might assume that a government department can take over stewardship of the data once the project ends, without consulting with the government department about the impact this has on them. Program officers need to address the long term implications for current and future stakeholders as part of the initial planning to ensure these conversations happen in advance.

A checklist for sustainable access to data

When planning for data to be accessible over time consider what will happen to the data created, accessed, used and shared after the project has concluded. Knowing what data is involved and the benefits of making it accessible in the long-term helps inform what sort of stewardship, forward planning and resources are needed. Finally, data standards, policies and platforms need to continue to work over time.

1. Assess upfront what data is being created or used

- Create a data inventory at the start of the grant
- Identify what practical application of data access is needed
- Consider third party rights over data

2. Decide if there is a benefit of sustainable data access

- Consider the purpose and goals of the project
- Identify users and understand what they need from data
- Consider other stakeholders' long term needs
- Consider if the project's funding model supports sustainable access to data

3. Plan for data stewardship, management and resources

- Data stewardship helps ensure sustainable access to data
- Senior accountability for data ensures resources are assigned over time
- Good data management plans include long term data access needs
- Check you have a clear and appropriate data retention schedule

4. Check data policies, standards and platforms support sustainable access

- Data policies should reflect your goals for sustainable access
- FAIR data standards make it easy to preserve useful data over time
- Data platforms support sustainable access

1. Assess upfront what data is being created or used

Planning for sustainable access to data in the initial stage of a project can benefit all subsequent stages and support good data practices overall. At the start of a project, it's not always certain what data you will be created and used, but it is possible to make informed predictions and plan on that basis, updating the assessment and planning over time as needed.

Create a data inventory at the start of the grant

A data inventory records all the datasets created in a project or organisation. It describes the content of each dataset, its sourcing, licencing and other useful information. Creating a data inventory at the start of a project is an important first step to make data accessible over time – it outlines what data the project is likely to have so you can plan ahead for sustainable access. See How to create a Data Inventory.[2]

Identify what is needed for the practical application of data access

For sustainable data access to work in practice, it needs to remain findable, accessible, interoperable and reusable over time. Beyond that, the practical approach largely depends on whether data is static or dynamic:

- **Static data** is data collected during a project that won't be updated after the project, for example, the formulae of patented pesticides.
- **Dynamic data** is data that will continue to be collected and updated after the project, for example crop or market data.

[2] (2018),Tim Beale, Fiona Smith, Leigh Dodds, Pauline L'Henaff, Deborah Yates ,'How to create a data inventory'. Accessed July 2020. https://gatesopenresearch.org/documents/2-45

Although it is necessary to consider where static data will sit, and who is responsible for controlling access if it is not entirely open, it doesn't usually require further investment beyond the life of a project. Dynamic data, on the other hand, is likely to have ongoing costs for data collection, maintenance and the running of any necessary infrastructure to support it. If your project includes dynamic data that needs to be updated beyond the life of the project, you need to ensure this is funded sustainably, and identify where the responsibility for all data will sit.

Consider third party rights

A project typically won't own all the data it creates or uses, and the rights of those third parties who do own or create the data can limit access to it. Some of the data might be licenced from someone else, in whole or in part, or derived from data licenced from someone else. It is important to check the licences for all the data involved in a project to know if it can also be used after it. Many licences, such as Creative Commons and open government licences, support data sustainability. But non-open licences with conditions such as 'no-derivatives' and 'non-commercial use' can limit how useful and usable data can be over time. For more information, see the guides in the Data Sharing Toolkit on How To Choose a Data Licence and 'Considering data rights and permissions in investments'.

2. Decide if there is a benefit to sustainable data access

Sustainable access helps to maximise the utility of data by allowing governments, businesses, NGOs and others to rely on ongoing access to data, so they can invest in products, services and research using that data. At the same time, there could be a risk of harm to groups or individuals to make some types of data accessible beyond its original use. It is important to consider both these perspectives when deciding whether sustainable access to data is required.

Consider the purpose and goals of the project

Start by considering the goals of your project. Do they call for long-term access to data? For example, the Bill & Melinda Gates Foundation's Digital Farming Services Portfolio has a goal that by 2030 at least 50% of smallholder farmers in target geographies will use digitally-enabled services to reduce risk and improve decisions''.[3] To meet this goal, they need to access, use and share data until 2030 and beyond.

Identify users and understand what they need from the data

To assess if there is a benefit to sustainable access to data, you also need to assess who might use it in future, and what they might use it for. Can the data be maintained over time in a way that meets these future needs? Once you have an idea of who might use the data in future, it is important to engage with them and test your assumptions. Useful tools to explore user needs in general include understanding personas in agricultural data ecosystems, the UK's Government Digital Services team's manual on user research and the GSMA AgriTech Toolkit for user-centred design.

[3] (2019) Bill & Melinda Gates Foundation, 'Ag Dev Portfolio Operational Plan – Digital Farming Services ', Internal document. Accessed July 2020.

Consider the long term needs of other stakeholders

Check if any of your stakeholders, such as research partners or service providers, specify how data should be accessed and used over time, or if they have their own strategic goals related to making data accessible and reusable. If your project has government support, this may also bring with it certain commitments or actions to make data accessible over time.

Consider whether the project's funding model supports sustainable access to data

Does the funding model of your project include resources for providing longer term access to data? For instance, if you are a business, does your operating model finance data access and use over time? If your project relies on outside investments or grants, do the investors or grant makers support sustainable access to data? Will they extend their funding to help make data available beyond the life of the project?

For instance, the foundation's new Open Access Policy enables the unrestricted access and reuse of all research and data resulting from the Foundation's funding, including the payment of all reasonable fees required to publish data on these terms.

3. Plan for data stewardship, management and resources

Steps 1 and 2 should help you decide what sort of data stewardship, forward planning and resources are required to support sustainable access to data. Now you can think about the people and processes you need.

Establish data stewardship to ensure sustainable access to data

Sustainable access to data depends on data stewardship. Data stewardship is about people and processes: the people who are responsible and accountable for how data is accessed, used and shared, and the processes that ensure effective oversight of data.

By assigning clear roles it is possible to establish exactly who decides about and invests in datasets so they are accessible and usable in the long term. Data stewards help to ensure data is used in ways that maximise utility and minimise harm by seeing that data principles are followed consistently and ethical issues are addressed. Processes for oversight of data can ensure the integrity of datasets over time, as well as compliance with regulations and legislation, policies including funders' policies and contracts like grant agreements.

It is important to think about how data stewardship will change after the project ends. For instance, will data be handed over to a different organisation or person to be stewarded? If so, who will be impacted by this? What conversations should you have with these people now so you can all plan ahead? And what processes do you have in place for a handover? It is also important to carefully consider oversight, the comfort of repetition or

loss of knowledge through personnel changes can result in lapses in care and attention.

Assign senior accountability for data to ensure resources

Ensure someone at a senior level has responsibility for seeing that the proper resources are invested in data sustainability. This doesn't mean the senior team does the work, it means they can confirm the necessary funds and other resources are allocated for it. Senior responsibility is particularly important to ensure data is managed as an asset, along with other project outputs, and to help make sure processes are kept up over the long term.

Plan for long term data access needs

Data management is part of data stewardship. A data management plan documents the technical processes needed to manage data, and the resources required to support them.

A good data management plan will include:

- How the data will be stored and accessed
- How the data will be described, including naming conventions and metadata
- How the integrity of the data will be maintained over time
- How the security of the data will be maintained over time
- How the data will be shared at the end of the project
- What resources are required.

At this point, it is helpful to revisit Step 1, and the assessment of what data is static or dynamic in your project. Data that is dynamic, i.e. that changes and updates over time, is likely to need a more active data management plan to remain findable, accessible, interoperable and usable in the long term. For more on data management, see Developing a data management plan: a checklist.[4]

Check you have a clear and appropriate data retention schedule

Data retention refers to how long a dataset is, or needs to be, accessed, used and shared. You need to consider retention periods for each type of data created by your project. This includes deciding what data should not be accessible when a project ends. For instance, best practice data protection legislation states that personal data should not be stored 'no longer than is necessary for the purposes for which the personal data are processed'.[5] Implementation requires active review and management over time. You can use the data inventory you did earlier to track the date each dataset should be reviewed or archived. For more on data retention, see Developing a data management plan: a checklist.

4. (2018), Stewart Marshall, Fiona Smith, Tim Beale, Pauline L'Henaff, Deborah Yates, 'Developing a data management plan: a checklist'. Accessed July 2020. https://gatesopenresearch.org/documents/2-46

5. (2020) Information Commissioner's Office UK, 'Guide to the General Data Protection Regulations, Principle (e) – Storage Limitation'. Accessed July 2020. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/

4. Check data policies, standards and platforms support sustainable access

Ensure data policies reflect goals for sustainable access

A data policy sets out expectations about how data should be accessed, used and shared. Policies have a role to play in supporting sustainable data access. It is important to check that ambitions and approaches set out in a data policy can be sustained in the long-term, for example data review and release, data licencing and reuse rights, and data standards.

Some data policies are limited by the life of the project. If this is the case, you need to consider what policy will apply to the data in future, who will be responsible for it, and how this will support longer term data access. For instance, if the data created in your project will be passed to a government department for data stewardship, confirm they have an appropriate data policy to support sustainable access to data. If not, consider how you can help to prepare one for the data they will steward in the future.

For more on data policies in general, see the Data Sharing Toolkit guide ‘What to include in a data policy for multi-stakeholder investments’.

[6] Open Data Institute, ‘Open Standards for Data Guidebook’. Accessed July 2020 https://standards.theodi.org/

Preserve useful data over time by using FAIR and open data standards

Open standards for data are reusable agreements that make it easier for people and organisations to publish, access, share and use better quality data.[6] To support sustainable access, make sure any data that is shared during your project is in well-structured, machine-readable formats, with clear metadata, documentation, protocols and languages, as per the requirements set out in the FAIR principles. For more information on standards in general see Open Standards for Data.

Use data platforms to support sustainable access

There are many ways data can be stored to support future access and use. Open access platforms help to make open data sustainable, but any data can be made more sustainable when stored appropriately, even if it is not open to all. For example, an institution can store data in an internal platform if it is not appropriate to share it with other organisations, and so long as it complies with legislation such as data protection regulation.

The Gates Open Research platform provides guidance on preparing data for publication and a non-exhaustive list of places to store and share data, approved by the foundation, and describing the desirable properties that support sustainable access to data:

- Enable access to the dataset
- Ensure continued access to the dataset
- Ensure dataset stability
- Enable searching and retrieval of datasets

Further Resources

There are a number of external resources available to help support sustainable access to data:

- Bill & Melinda Gates Foundation Open Access Policy
- Gates Open Research: Data Guidelines
- Guide: Developing a data management plan: a checklist
- Guide: How to create a Data Inventory
- Open Standards for Data Guidebook

Data Sharing Toolkit

ACKNOWLEDGEMENTS

This document was authored by the Open Data Institute and CABI as part of a Bill & Melinda Gates Foundation funded investment.

The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or CABI.

datasharingtoolkit.org

DOI: 10.21955/gatesopenres.1116748.1

cabi.org | theodi.org | gatesfoundation.org

CABI Data Sharing Toolkit

open data institute

BILL & MELINDA GATES foundation