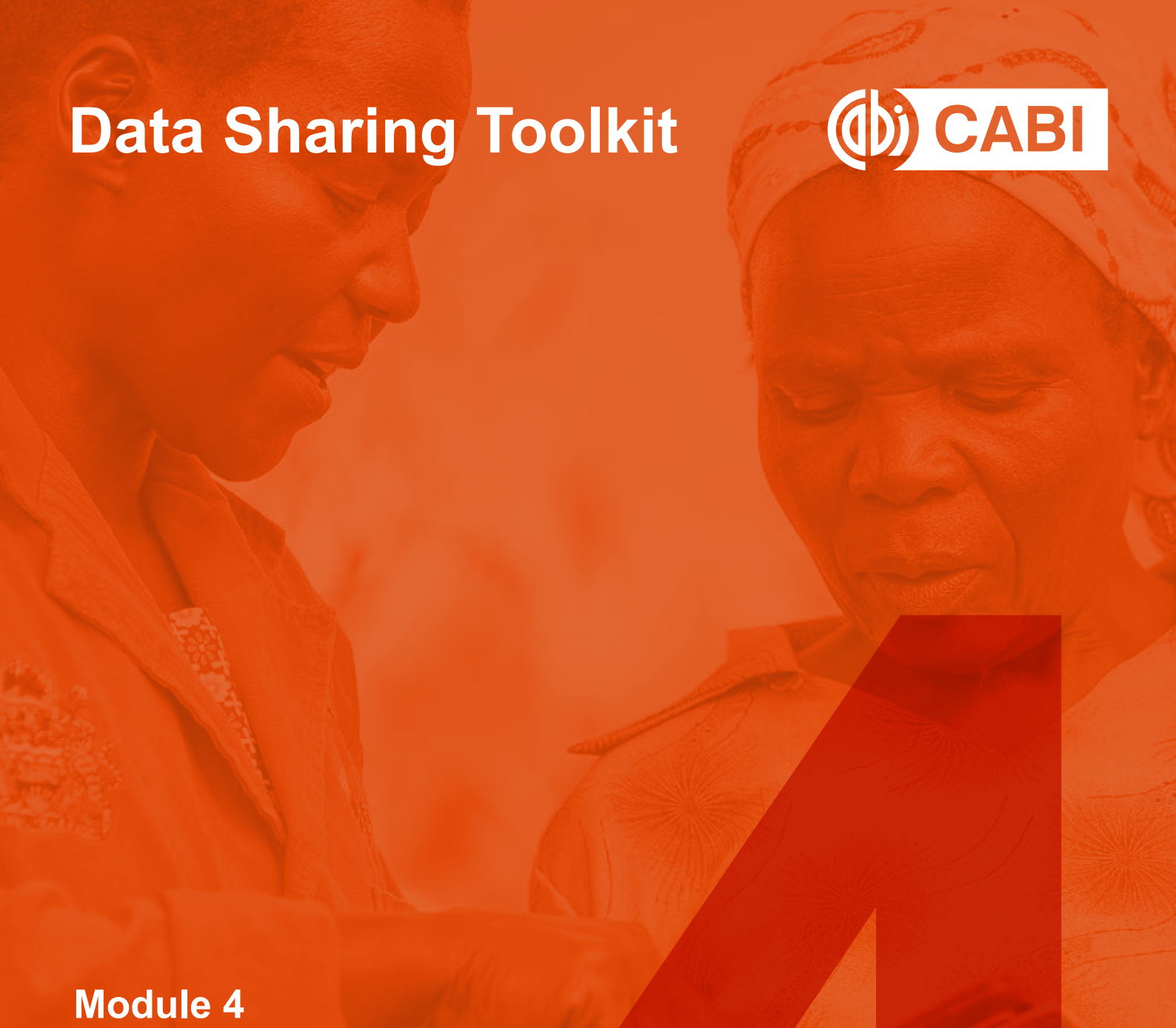
Data Sharing Toolkit



Module 4

Protecting people – managing risk when handling personal data

Guide

Protecting people – managing risk when handling personal data

Purpose of this guide

One of the biggest barriers to sharing data is legitimate concerns over breaking legal obligations, security breaches or sharing data that could cause harm to individuals, communities or society. This guide will help Bill & Melinda Gates Foundation **grantees** to navigate these concerns.

Grantees can use this guide to identify whether their project will collect personal data, and consider techniques to maximise utility of data while protecting the rights of individuals.

Specifically it will help to:

- recognise personal data in agricultural development projects
- identify where data is suitable for sharing as openly and widely as possible
- comply with data protection and data rights legislation
- minimise risk by identifying relevant mitigating actions (e.g. anonymisation)

This guide is a subset of a larger guide on identifying and managing risks when accessing, using and sharing agricultural data. The larger guide includes information on commercial and personal data, rights and permissions to use data, and encouraging best practices.

*This document is **not legal advice** and if you are uncertain you should seek guidance from a legal professional.*

This guide includes

- The definition of personal data
- The role of data protection regulations
- Possible negative and positive impacts to consider
- Practical actions to mitigate risk

When to use this guide

- At the beginning of a project involving data
- Whenever data is being shared as part of a project
- When grantees are unsure whether they can share data
- When grantees have concerns over privacy or confidentiality

Applicable grant stages:
Concept | Proposal | Agreement | **Active**

Protecting people – assessing risk

What is data about people?

Data about us comes in many forms and can have a range of impacts, not always for the individual, but often for society as a whole. The graphic below helps define the different types of data about us: personal, sensitive, behavioural and societal.

It is important to think about the different types of data about people and take into consideration both the opportunities and risks when collecting, using and sharing this type of data.

The role of data protection regulations

Data protection regulations across the world are designed to enable personal data to be used while minimising the risk of harmful impacts.

These regulations typically outline three key things:

1. The lawful basis for using and sharing personal data
2. The rights of the data subject (the person the data is about)
3. Liabilities and penalties

Different countries have their own data protection legislation, which enables personal data to be used while minimising the risk of harmful impacts.

Data Protection laws in some countries include a 'right to erasure' – this means that when data is published online an individual has the right to have personal information about them deleted. In the case of the details of legal proceedings or convictions being included within a dataset, this could be after a certain time period.[1]

Find out about local data policy and legislation via the 'agriculture data country briefing' example guides and template in the Data Sharing Toolkit.

[1] Information Commissioners Office UK (2017), 'Right to erasure', https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/. Accessed September 2020.

Defining personal data

Personal data is defined by the United Nations as "information relating to an identified or identifiable natural person".[2] Different countries have their own data protection legislation covering personal data, which needs to be adhered to. To build trust, grantees and organisations using personal data should also be open with people about how they use and share that data.

Data can be used to identify an individual either directly or indirectly. Directly identifiable data might include a person's name, social media or application log-in information, or identity card or social security numbers. Indirectly identifiable information is data that could reasonably be expected to identify an individual in isolation or when combined with other information, for example geolocation, IP address or physical address.

Personal data in agricultural development projects

In the context of agricultural development projects, personal data or data that could identify individuals, could include:

- **Information about farm employees** – for example name, address, bank details.
- **Mobile phone numbers** – often these will have associated geo-location that could identify an individual.
- **Socio-economic data** – such as household income, ages and education levels of each household member . This is often collected by M&E teams.

[2] United Nations (2018), 'Principles on personal data protection and privacy', https://www.unsystem.org/privacy-principles. Accessed September 2020.

- Operational machine data – for example information collected by "smart tractors", which could contain farming practice and location data, and be linked to a user profile that may contain both personal and sensitive information.[3]
- User profiles in apps providing tailored advice to farmers may contain both personal and sensitive information.
- Free text or comments fields often contain notes from the data inputter. By definition they are not restricted in value so could contain personal, sensitive or confidential information that might otherwise be missed – for example notes from the data inputter about individuals, information about the person who has collected the data or locations.

[3] Wiseman et al (2019), 'Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming', https://www.sciencedirect.com/science/article/pii/S1573521418302616. Accessed September 2020.

Could the data in this project directly, or indirectly, identify individuals?

If the answer is yes, refer to the section on minimising risks, below.

Negative or positive impacts on people from use of data

When considering risk, it is important to consider broader harmful impacts as well as legal risks – for example on individuals, sections of society or whole nations. Data ethics are relevant when data activities have the potential to directly or indirectly impact people and society.

When considering broader harmful impacts, think about the people the data is about, people impacted by its use and organisations using the data. For example could this data create bias in decisions drawn from it?

In the context of agricultural development projects, examples of potentially harmful impacts through unethical use of data include:

- An automated data model might make decisions about whether smallholder farmers are eligible for a subsidy, a financial incentive or mortgage. The decisions the model makes will be based on the data collected – and excluded – which could adversely affect groups in a society.
- Big data is used in new technologies for 'prescriptive planting' which can automatically plant fields. There are reports of mistrust that prescriptive-planting firms could use the data to buy farms known to be underperforming and directly compete with smaller farmers.[4]
- Data published from different sources about, for example, field extents and soil quality plus satellite imagery, could enable competitor farmers or financiers to get a lot of information they have historically not had access to. This could drive up lease pricing or affect a farmer's ability to get loans, for example.

Ask yourself, does the data contain or omit any information that could adversely impact people or society?

If the answer is yes, see the section on minimising risk below.

[4] Irish Times (2014), 'Farmers up in arms over potential misuse of data'. Accessed September 2020. https://www.irishtimes.com/business/farmers-up-in-arms-over-potential-misuse-of-data-1.1863181

Minimising risk – practical actions

This section outlines a number of techniques to apply to maximise the utility of data while protecting rights and individuals from harm.

1. **Data minimisation**[5] – Do you really need to collect the data? If you don't need personal details or commercial information, don't collect them. This can help avoid navigating data protection laws and make it more likely the data could be shared with others. This could also help overcome any ethical issues.

2. **Educate data inputters** – Educating those collecting or curating data to consider the information they collect and how they can help to avoid adversely impacting people and/ or society. For example by thinking about potential bias in data collection, or minimising the options for notes and free text fields.
   - The Data Ethics Canvas can help identify potential harmful or positive impacts, how to manage these and how to communicate your use of data effectively.
   - Removing any free text comments fields or redacting parts of the dataset can reduce the risk. If free text fields are important to retain, you could instruct data inputters to follow strict guidelines for their content or it could be an option to share the data under contract.
   - Finally, you could remove the risk at source and design your data in a way that has no need for free text fields, using restrictive values (e.g. date only, or pick lists) is one way to do this.

[5] Information Commissioner's Office UK (2017), 'Principle (c): Data minimisation'. Accessed September 2020. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/

3. **Anonymisation and suppression** – It is possible to process data into a modified form that can be shared or made open while significantly reducing the possibility of anyone recovering sensitive or personal information from it. "The anonymisation of personal data is possible and can help service society's information needs in a privacy-friendly way".[6] For sensitive data in general, this process is called suppression; for personal data it is called anonymisation.
   - More detail, including a worked example, is included in this guide: An introduction to managing the risk of re-identification.
   - You can also refer to the UKAN Anonymisation Decision Making Framework (ADF) which provides step-by-step guidance, a step-by-step interactive guide to the ADF built by the Open Data Institute, and a Risk, Harms and Benefits checklist tool created by Global Pulse.
   - Finally if you need expert input there is a register of actors that can help with anonymisation.

4. **Use synthetic data** – This data is created by an automated process and contains many of the statistical patterns of an original dataset.[7] It is also sometimes used as a way to release data that has no personal information in it, even if the original did contain lots of data that could identify people.
   - This hands-on Python tutorial demonstrates how to create a synthetic dataset.

[6] Information Commissioner's Office UK (2012), 'Anonymisation: Managing data protection risk code of practice', Accessed September 2020. https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

[7] The Open Data Institute (2019), 'Anonymisation and synthetic data: towards trustworthy data', https://theodi.org/article/anonymisation-and-synthetic-data-towards-trustworthy-data/. Accessed September 2020.

Further information and help

Understanding and assessing risks when sharing personal data may require specialist input. You may need to consult colleagues, partners or legal specialists. including;

- Scientific/or policy specialists to inform the data required to be shared, how the data is to be used, as well as the validity of data exchanged.
- Data and information specialists that understand:
  - technical aspects of the data to be shared;
  - how the data may be integrated with other data sources that could raise other data protection or legal issues not inherent within the data alone
- Legal support from Gates Foundation lawyers, and in some instances external legal guidance.

In all cases, this document is **not legal advice** and if you are uncertain you should seek guidance from legal professionals.

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.1116766.1

cabi.org | theodi.org | gatesfoundation.org

CABI Data Sharing Toolkit

open data institute

BILL & MELINDA GATES foundation