

## Module 3

# Understanding personas in agricultural data ecosystems

# Guide

# Understanding personas in agricultural data ecosystems



## User personas

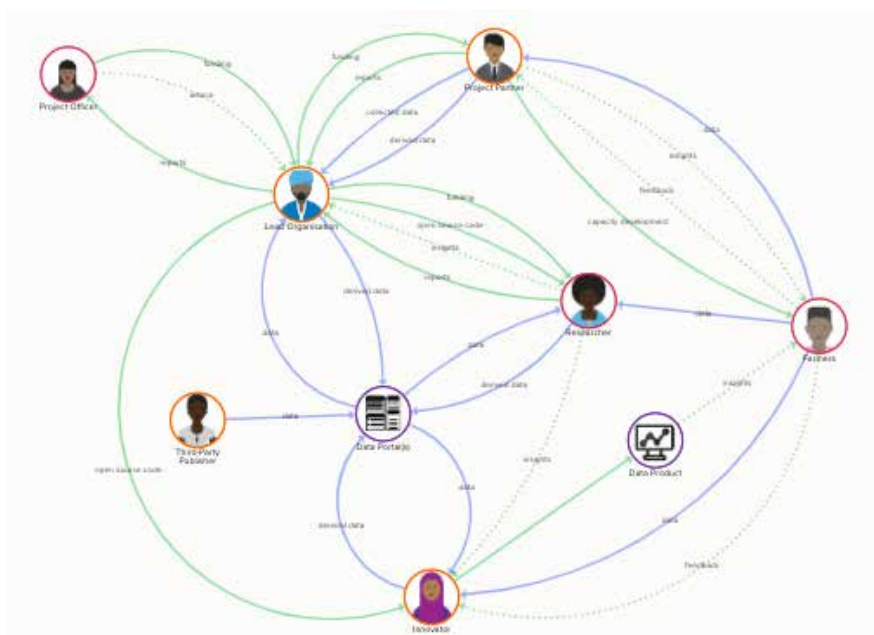
Personas are imaginary characters who can help us to understand the learning needs, barriers, motivations and goals of real people. They develop as we learn more about the real people we are trying to help.

This document presents the personas we are using to help us understand what support is needed, and to help signpost people towards relevant methodologies, guidance, and tools they can use in their projects.

These personas are based on interviews conducted in January-June 2018 with a range of stakeholders connected with the Gates Foundation programmes in Seattle, Tanzania, Ethiopia, and India (Andhra Pradesh and Odisha). Each persona is a composite based on several interview respondents.

## **Data ecosystem**

Each persona is positioned within the 'data ecosystem' of a typical Gates Foundation agriculture programme. A data ecosystem is a visual technique for mapping the actors, relationships, and data flows around a particular project or service. It helps us highlight where important value exchanges happen (such as transfers of funds, supply of data from one party to another, use of open data, etc). This can also tell us where there are pain points for different actors, and what kind of strategies or support they might need.



## FAIR and open data ecosystems

The above diagram illustrates a **generalised data ecosystem**, produced from our analysis of sample projects within the Gates Foundation agriculture decision-making portfolio.

The ecosystem shows some value exchanges common to a number of projects. In this example:

- a funded organisation is **publishing openly licensed datasets and some supporting code** that can be used to analyse that data;
- **open data is created by combining together existing third-party sources**, which are also published under an open licence, with new data that has been collected by a project partner;
- **software and data is being used by researchers** as part of other research projects; while
- **open data is also being used to create a new product** for farmers by an innovator.

## Supporting FAIR and open data

For data to be both FAIR and as open as possible, each of the data flows and exchanges in an ecosystem should be governed by **appropriate policies, contracts, guidance** e.g. on data sharing.

For example, in order to publish a FAIR and openly licensed dataset the lead organisation will need to ensure that:

- **the data used from third-party publishers are available under a compatible open licence.** If this is not the case, then alternative sources might be selected, or an alternate data licensing arrangement agreed
- **the procurement of data collection and/or analysis services from the project partner ensures that the lead organisation has the necessary rights to use it to create new derived datasets**
- **the datasets and software are published with appropriate metadata and documentation** to ensure it can be found, accessed and used by the innovators and researchers
- **the datasets and metadata are published according to relevant standards** that will ensure they can be reused
- **the published datasets and software are published under a licence that gives permission for the anticipated reuses,** including where necessary, permission to archive and distribute the resources to ensure they can be sustainably accessed

Explore the example data ecosystem.  
Can you see the different types of value exchanged between the personas?  
How does this compare with your projects.

## Persona interactions within the data ecosystem

The personas described in this document interact with one another within the data ecosystem.

Here are some scenarios illustrating how the different persons may interact:

Which persona describes the type of work you do? Which persona best describes the person you want to support or influence?

### How to read the personas

In this document, **goals** relate to the persona's overall work objectives. **Pain points, needs, data collected, and key resources** relate to activities and tools relevant to apply the FAIR principles when managing data related to Gates Foundation funded projects.



**Sara – Innovator:** Sara works in an organisation that offers a tailored, information service built on open data that helps farmers make better decisions.



**Alan – Lead organisation:** Alan works for an agriculture organisation that manages a consortium of actors to produce a platform combining data from multiple sources, which innovators and researchers can access to develop data-enabled services, and products.



**Joelle – Researcher:** Joelle works for a university research centre which collects raw data, processes it, and combines it with available sources of open data to produce insights, which may be published via an open access repository.















**Chris – Project partner:** Chris works in an organisation that develops software tools to make it easier to analyse and visualise insights on data about farming practices.



**Ida – Third party publisher:** Ida works for a central government agency in charge of managing the national repository of soil data, which is made available to researchers and innovators via an open data portal.



**Anna – Project officer:** Anna works for a grant-making organisation responsible for managing multiple investments in data-intensive agriculture programmes. She needs to oversee that formal M&E requirements for reporting are being met, and activities are being implemented to a high quality.

-  **Anna** is awarding funding to  **Alan**'s organisation in order to establish an innovation platform for soil data. This involves hosting a data repository, comprised of raw data and secondary analysis produced by researchers like  **Joelle**
-  **Alan** needs to make sure that in the terms of the subcontracting agreement with  **Joelle**'s research institution that he has the necessary rights to use the raw data and create derived datasets from it, and that it complies with the funding policy of  **Anna**'s organisation.
-  **Sara** wants to combine data which has been published by  **Ida** to  **Alan**'s platform in order to develop an SMS-based farm extension information service. For this,  **Sara** needs to ensure that she has permission from all of these data publishers to reuse this data under an open licence.
- At the same time,  **Chris** has an idea for a software-based data analysis tool that would make it easier for farmers and policymakers from  **Ida**'s agency to make decisions. He wants to publish the software under an open licence for others to access and use. He needs to make sure the right permissions are in place.



## Sara

### Innovator

*“I want to make data more accessible to farmers so they can improve their productivity”*

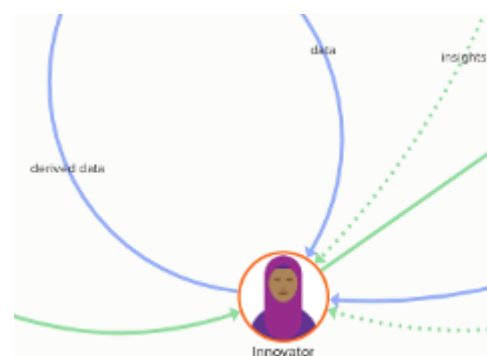
Sara works in an organisation that helps farmers benefit from data by offering an advice service that helps them make better decisions about when and where to plant.

### Goals

- Develop a mobile phone-based advisory service for farmers
- Integrate government open data with data from farmer profiles
- Provide accurate predictive analysis and advice to Improve soil health, income, and crop yield.

### Pain points

- Data quality unreliable (e.g regularly updated, complete, current)
- Data not available in an easy to process format (i.e machine readable, curated, clear metadata)
- Unclear or no license applied to the data
- Negotiating access to, and permission to use personal data while respecting privacy
- Internal capabilities – attracting, training and retaining agronomists with sufficient data literacy



**Explore Sara's role in the ecosystem**

### Key exchanges she is involved with

- Using open data to create a product
- Procurement of data analysis services
- Negotiating access to personal data



## Needs

- Access to historical data from the Ministry of Agriculture's open data portal and soil data repository (e.g crop yields by year, temperature, precipitation records by area, soil types and coverage, demographic data, market information)
- Assurance over rights to reuse government data to produce a new service
- Permission to access and use farmer profile data (containing personal data) from the government agriculture extension service
- Subcontract a team of trained agronomists to analyse the multiple data sources, and produce tailored advice
- Subcontract developers to design the mobile-based technology to deliver the information to farmers

## Data collected

- Farmer contact details and preferences



## Alan

### Lead Organisation

*“I want to streamline how my partners access, share and publish data”*

Alan’s organisation has received funding to develop an online platform with the aim of improving agriculture decision making for researchers, policymakers and innovators. He needs to cooperate with various stakeholders involved in collecting and processing data to make it accessible via the platform.

### Goals

- Create a Soil Intelligence Service (SIS) allowing stakeholders in multiple jurisdictions to access raw data, research and software code
- Combine soil data with crop, agronomy, and geospatial data using machine learning models to make it more useful
- Develop a sustainable business model based on a mixture of open (free) and subscription-based products/services
- Create measurable impact which he can document for his funders

### Pain points

- Difficult to discover where all the data needed for the platform is held, and who has rights to the data
- Harmonising data produced using different formats, metadata, and process flows
- Negotiating access to data with different stakeholders, with different sensitivities
- Government data stewards unwilling to share data outside of their jurisdiction
- Tracking downstream use of data assets to report impact/return on investment



**Explore Alan’s role in the ecosystem**

### Key exchanges he is involved with

- Publish openly licensed datasets and supporting code
- Publish openly licensed software;
- Procurement of data collection services;
- Recipient of funding

## Needs

- Access to high quality, standardised data from government, research institutions, and industry eg crop and agronomy
- Shared data management strategy outlining approach to data governance, access, sharing, publication, rights to use, responsibilities, and sustainability
- Buy-in from government data stewards
- Metrics to monitor progress and measure impact over time from data use
- Subcontract soil data collection services from research institutions, to collect and validate data in standardised way

## Data collected

- Secondary research data (e.g soil grids)
- Lab data (e.g spectral data)
- Geospatial data (e.g remote sensing covariates)
- Software code



## Joelle

### Researcher

*“I want to process and format data so it is useful for other agriculture researchers to use”*

Joelle is a researcher based in a university laboratory that receives funding to collect and process soil data samples. She oversees data collection in the field, while producing analysis of the samples that she wants to make available to other researchers.

### Goals

- Deliver high quality research backed by robust evidence base
- Receive recognition for her/her institution’s research, and promote research uptake
- Contribute to research breakthroughs and policy decisions leading to improved agricultural productivity and soil health.

### Pain points

- Field researchers record data in inconsistent formats – difficult or impossible to know data quality (no validation mechanisms). Increasing need to reformat the data into a common structure
- Time pressure to produce secondary research analysis, while balancing the need to improve the quality of the underlying data
- Budgeting adequate time and resource for publishing research (including underlying datasets)
- Navigating competing funder and organisation’s policy on open access and data sharing
- Low recognitions or incentives for quality data collection and good data management



**Explore Joelle’s role in the ecosystem**

### Key exchanges she is involved with

- Collecting data
- Publish openly licensed secondary research and underlying datasets
- Recipient of funding

## Needs

- Access to openly licensed software tools to assist her with processing large volumes of data
- Consistent policy and guidelines to follow on meeting open access and open data requirements
- Assurance that field researchers are collecting high quality data eg standards or tools for easily validating data quality
- Flexible funding and rewards to support quality data collection and publication efforts
- Evidence to demonstrate research uptake and data reuse, to report back to funders and support future research proposals.

## Data collected

- Physical soil samples
- Metadata on soil samples (e.g unique ID, geolocation, depth, date of collection)
- Lab notebooks



## Chris

### Project partner

*“I want to be make it easier for agriculture extension workers to visualise insights from data”*

Chris works in an organisation. He has designed a software tool that helps agriculture extension workers or researchers, to draw insights from the vast amount of data available on the soil information platform.

### Goals

- Develop effective, sustainable software tool to allow users to visualise insights from data
- Publish under an open license for anyone to access, use and share

### Pain points

Needs high quality and frequent data inputs from third parties, sometimes unsure of data accuracy

Ensuring anonymisation of personally identifiable information (PII), and protecting against re-identification of individual farmers

Using SIS data, which is difficulty to access and understand for a non-scientist

### Key exchanges he is involved in

- Publish openly licensed software
- Negotiating access to data with lead organisation
- Recipient of funding



**Explore Chris's role in the ecosystem**

## Needs

- A data-sharing agreement with the lead organisation and other project partners providing data
- Regularly updated, accurate, complete data inputs which are supported by clear metadata to enable traceability and verification
- Assurance over rights to reuse data to create a new software, and then to publish that software under an open license

## Data collected

- farmers group profiles
- demographic/context data e.g village, farm, community details
- M&E/impact data e.g adoption rates of new farming techniques



## Ida

### Third party publisher

*“I want to make sure my agency publishes good quality data, that can be used to benefit my country”*

Ida works for a central government agency in charge of managing the national repository of soil data.

### Goals

- Design and maintain a central repository of soil data, which can be accessed externally
- Define and implement data quality control standards
- Would like to share data for other researchers and countries to access

### Pain points

- Lack of open data policy means there is low confidence in whether she can make data available to others and under what conditions
- No consistent license applied to the data stipulating permissions and accepted uses
- Responding to multiple requests for data held in the repository (time consuming to process)
- Low buy-in from her boss and other agency heads to share data in an open format



**Explore Ida's role in the ecosystem**

### Key exchanges she is involved in

- Publish data on a central portal
- Requests for data access



## Needs

- Strengthened data infrastructure to facilitate data sharing (common quality standards and metadata, open data policy)
- Help making the case for making data in the repository available for others to access, use and share
- Identify use cases for soil data held in the repository from industry and researchers
- Understanding of global principles and standards to support cross-border interoperability

## Data collected

- Soil & agronomy data
- Geospatial data



## Anna

### Project officer

*“I want to make sure the projects I fund are following best practice around FAIR and open data to maximise the value of our investments”*

Anna manages several agriculture grants in multiple jurisdictions. She needs to understand best practice around data management, and to make sure they are applied by organisations receiving funding (and their project partners)

### Goals

- Make grantees’ life easier while being still accountable for their investments
- Be a transparent funder
- Ensure internal policy around data management and open access is being followed
- Promote maximum impact and value for money from investment

### Pain points

- Data management is not prioritised or well budgeted for by grantees
- Mixed data literacy skills within grantee and subgrantee organisations
- How to incentivise and reward good data management within grantees
- Conflict between own organisation’s open access policy, and policy of grantees (more restrictive)
- Low visibility of all data assets being collected and published by her portfolio of grantees – difficult to know where to prioritise investments



**Explore Anna’s role in the ecosystem**

### Key exchanges she is involved in

- Providing funding to the lead grantee organisation

## Needs

- Understand the source, quality and volume of the data being generated through her investments
- An overview of the entire data ecosystem and different actors involved to understand capacity and funding needs
- Capacity building/learning materials to roll out with grantees and sub-grantees
- Develop a harmonised system/process to promote and monitor implementation of FAIR and open data across projects

## Data collected

- M&E information (KPIs, impact stories)
- Data management plans from grantees

## Methodology

### Research methodology

This research was produced by interviewing Gates Foundation Decision Agriculture Grantees across 4 different regions (Tanzania, Ethiopia, India: Odisha & Andhra Pradesh) and two specific Soil and Agronomy investments in Africa. We asked interviewees about their role, the nature of data collection and use within their programmes, challenges faced, and their specific learning needs. Prior to this we spoke with the Gates Foundation programme officers based in their Headquarters to understand the funder perspective.

Following the interviews, we analysed the data flows, interactions between actors, pain points, and learning needs across all regions using the elito method. This helped us to synthesise common value exchanges, pain points, and learning needs according to type of actor.

In parallel, we attempted to map data value exchanges within each regional programme using an 'ecosystem mapping' methodology. This helped us to visualise the nature of data flows, where pain points occur, and what kind of guidance is needed to overcome barriers to FAIR and open data within specific regions. The data ecosystem maps will also be used to help each region develop a 'future state' strategy for data sharing and publication within their programmes.

Finally, we compared the individual data ecosystem maps to develop a common model of data value exchanges within the Gates Foundation agriculture programmes (known as a 'generalised reference model' illustrated on pg 2). This was designed to help the Gates Foundation and funded organisations see how and where they can turn the general FAIR data principles into specific actions.

## **Acknowledgements**

This research was conducted by the ODI, CABI and GODAN as part of a joint project commissioned by the Gates Foundation, to advise on how to enhance FAIR and open data within the agriculture decision-making portfolio.

The research methodology and output is informed by work the ODI has done on understanding data learning needs of local public services providers, which was supported by the UK government and can be found here.

This is a redesigned version of the same document included on Gates Open Research.  
[10.21955/gatesopenres.1114887.1](https://doi.org/10.21955/gatesopenres.1114887.1)

# 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](http://datasharingtoolkit.org)

DOI: [10.21955/gatesopenres.1116755.1](https://doi.org/10.21955/gatesopenres.1116755.1)

[cabi.org](http://cabi.org) | [theodi.org](http://theodi.org) | [gatesfoundation.org](http://gatesfoundation.org)

 **CABI Data Sharing Toolkit**



BILL & MELINDA  
GATES *foundation*



Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license.