



6.6 Data licenses

Choosing an approach to structure the inventory of your data assets

Why should I do this?

To enable the responsible and ethical use of data. It is a crucial aspect in the evolving landscape of information exchange.

How do data licenses relate to FAIR?

Reusability: Licensing your own dataset, if it is being released as part of a product, provides legal conditions for the ways other people can re-use it.

Download this data licenses factsheet for more insights.

What is a data license?

A data license is the legal framework for granting permission to access, use and share data, as well as to monetize it.

- 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) If you are a grantee, ensure you have technical team members involved in this process. While the content is accessible to both technical and non-technical members, technical expertise will be required to make decisions for the investment in this step.
- 3) If you have not already downloaded 'Project SIS' or 'Waterways', the illustrative scenarios provide examples on how each theme is navigated. These scenarios are frequently referred to across the content in Step 6 to help you understand how different aspects within a theme are applied.

Things to consider for your investment:

Determine the scope of license agreements

Which data licenses are being used, or are planned to be used in the project? A range of agreements are available to protect data and related intellectual property rights. Where a research project is seeking to access data or license a database it has created, an appropriate agreement is necessary. The first step is to work out where data is to be entered into the value chain from external providers, and where the results of the data product are to be made available to

data consumers. In general, even when data is apparently freely offered, parties always have different perspectives on data that must be negotiated in one of these ways:

Some data will be subject to financial agreements—for instance, paying a satellite provider for specific crop health data for an area over a specific time period. Some data might be clearly supplied free from license or under a Creative Commons agreement—for instance, where data can be freely used if credited.

Some other relationships between data providers and users will require individual agreements, outlining the terms of use under which data is being supplied.

You may have already started answering these questions under Step 3. Now you can review those decisions and build on them.

Put license agreements in place

Once the different license requirements are identified, they must be acquired and put in place:

Paid subscription or copyright payments need to be made to access necessary data where required, and the details of the license agreement should be stored within the data catalog.

Creative Commons agreements need to be chosen to meet requirements about distribution and accreditation of sources. A Creative Commons license permits users to access, distribute and download data according to the provisions specified in the different licenses. The authorizer can see which licenses are on offer, and what they allow the user to do.

A third form of license arises ad hoc because of the differing ways in which the data is seen by different parties. For instance, a grey area might exist between:

the individual farmers who allow their fields to be surveyed, their practices to be monitored, and their methods to be questioned, as the basis for the data about what is taking place; and

companies, such as sellers of fertilizers or crop types, who want access to such data from a point of collective overview, to justify and better target the products they are selling.

Where farmers are surveyed on the impact of their fertilizer use, as in Project SIS, or on the relationship of their crop growing to the provision of water, as in Waterways, a license agreement at once clarifies this relationship between data provision and data consumption. A license agreement between the farmer providing the data and the consumers who get to see the results can specify how the data of crop outputs and water use is targeted primarily for research purposes. A license agreement with the farmer can detail how the data will be used and specify how data providers (such as the farmers) might themselves access the results as data consumers (benefiting from the research).

Identify restrictions in license agreements

Once the license agreements are in place, the metadata needs to keep in view the applicable restrictions for all users:

The commercially licensed data might have restrictions on the original satellite or field data not being transferred to other users. In this case, only the results of the analysis of the data through the data product are to be communicated.

The Creative Commons licenses typically list restrictions on the free use of data. This might require that one acknowledges the source, and is not looking for financial gain in using the data. Such restrictions need to be held in the metadata, to ensure that use of the dataset is clear to all users.

Individual agreements, for instance with farmers providing access to their land to record the data, also impose restrictions on how this data is used. Some of these restrictions will be generic on not including Personally Identifiable Information (PII), but others might be individual agreements between parties.

Process and integrate license agreements in the Data Value Chain (DVC)

The DVC should be circular in the journey from data provision to data product. Considering the DVC adds value in the case of the scenarios, because it brings the results of the research back to the farmers' original questions.

When this activity is properly planned in advance, and the license agreements are held in the data catalog, there is no friction in the DVC over who has access to which dataset and data product:

The paid licenses are properly documented and cover the necessary external data.

The Creative Commons licenses are clearly visible and credited where necessary.

Individual license agreements ensure that the circle of data provision and data product is indeed rounded, so that all those who feed into the DVC are also the beneficiaries of the research.

A successful DVC, as illustrated in the scenarios, naturally and clearly honors all points of data exchange between parties.

License composite data products

Once the research has been through the DVC, the next issue is how the data product is to be licensed.

For example:

Are there any restrictions involved in the license of the external datasets that one needs to heed when distributing the data product?

Does one require a special agreement, subscription or license for others to access the data product results?

In ESRI, the data can be published as a story map, where one can allow for all the edited content to be freely viewed without any restriction or needing a subscription. Journal papers might require subscription or payment to view, and there are restrictions on how one can distribute the material outside of this paywall.

Acknowledge sources in data products

When distributing the data product, any license requirements for acknowledging the data sources should be honored. Normally, one would acknowledge the main sources of data on which the analysis and results are based. Naming the source gives a validation that the data is trustworthy.

Illustrative scenarios

Overview



©Gates Archive/Mansi Midha

Refer to the illustrative scenario that you have downloaded to see how this has been considered.

Ensure any work notes or decisions taken are being documented, as this would be useful to refer to at later stages or for someone new joining the team.

Project SIS

<https://www.fairprocessframework.org/steps/step-6-6/>



Only the specific theme related content has been highlighted here. To get a feel for the scenario, read [here](#).

1. Data onboarding

Onboarding TPP data like that received from Visual Crossing must be used in a legal way, as dictated by the datasets' licenses. We have examined the license for use of Visual Crossing data (in its terms of use [[link to be inserted-https://www.visualcrossing.com/weather-services-terms](https://www.visualcrossing.com/weather-services-terms)]) to ensure the data can be used for our purposes.

In doing so, we identified the type of membership we require and the cost this would incur. We are in the process of exploring whether information on the Normalized Difference Vegetation Index (NDVI) of regions in Dataland's highlands would be valuable to include in the SIS. If so, this data would be bought as an asset from a commercial satellite source, in which case data licensing problems may arise: sharing proprietary data of satellite companies with the final users of SIS may require a specific legal contract between SoilScience and said companies. If that is not possible, NDVI data will not be onboarded to the project at all. Similarly, consent is required from farmers to collect data from their land (with regard to soil samples) and their experiences (with regard to interviews).

We will draft a data-sharing agreement for farmers to sign in order to provide consent, with its terms published on the project website as part of the license/terms of use for SIS.

2. Data analysis

Preliminary analysis on the SIS' coverage of Dataland as well as the identification of general trends will be done with Python or R notebooks. Given these are third-party tools, the licenses for TPP datasets must be consulted to understand their restrictions for use. For Visual Crossing data, analysis with third-party software is allowed in its terms of use. Proprietary commercial satellite data, if used, will likely have different restrictions.

3. Data products

We will need to investigate how to license the SIS data to researchers beyond Dataland's MOA and our project partners, both of whom we will have data-sharing agreements with. We could use an open license like Creative Commons 4.0, in which case we will reserve attribution rights, but otherwise the data is free to use for any sort of project, although another option would be a non-commercial license (Creative Commons BY-NC 4.0).

The latter may be preferable to fit the overall community aims of SoilScience, but user research and some further stakeholder interviews would be required to judge what license to choose.

Waterways



Only the specific theme related content has been highlighted here. To get a feel for the scenario, read here.

1.Data onboarding

Satellite data that we collect from TPPS must be used in a legal way, as dictated by the datasets' licenses. Depending on the license, we may not have the authority to use data in certain ways. In the best case, we might only have to credit the original data authors (as is the case with CC BY 4.0 licenses); whereas in the worst case, licenses may have 'No Derivatives' clauses that could prohibit the publishing of the topographical map as its own resource (as is the case with CC BY-NC-ND 4.0s). Similar to licensing, we must understand the means of authorization to get access to TPP satellite data. Some datasets will require fees for access, which must be budgeted for. Thankfully, SoilScience has a professional network that may be able to help with fee waivers. Finally, consent is required from LFs to collect their data via interviews.

2. Data processing

Work with the satellite data will most likely be done with third-party tools. Besides the standardization script in Python, we will also use GIS software for the map presentation. Sometimes, data licenses can prohibit the use of third-party tools on the data, so our terms of use for the satellite data must be considered.

3. Data analysis

Data will be analyzed using R or Python notebooks. These are both third-party tools, so licenses for the satellite data must be checked. This should be done in the 'Data processing' stage, but is worth considering here too.

4. Data products

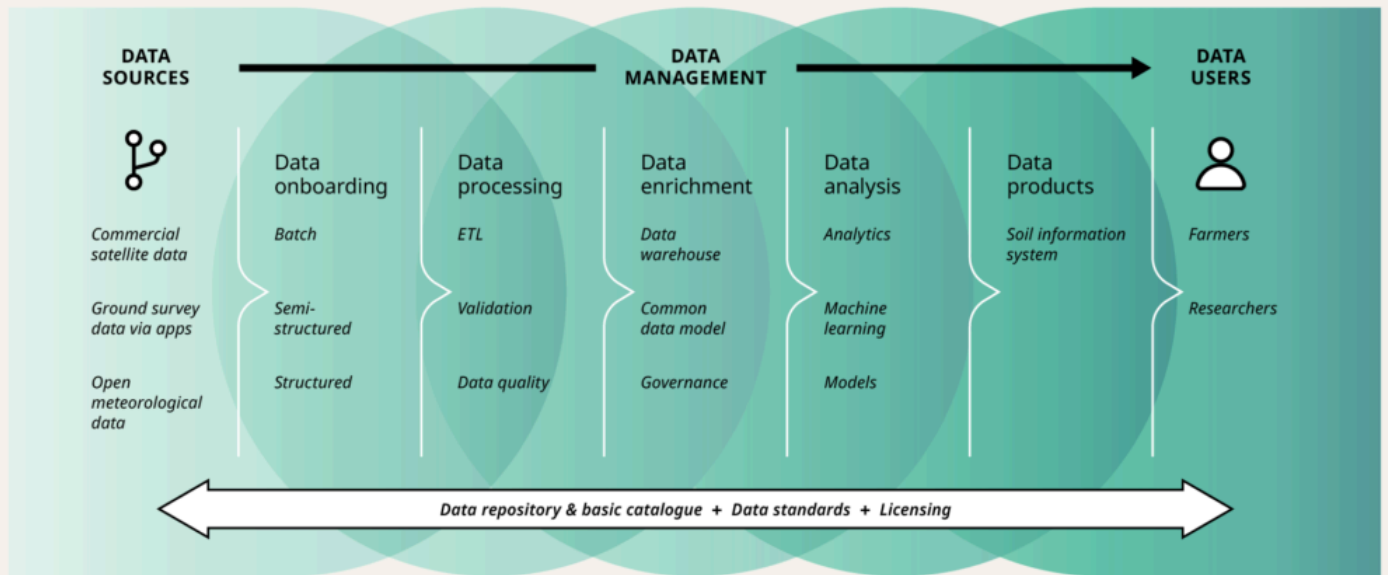
We will need to make sure we are allowed to publish the satellite data in the topographical map. If the data is licensed with a non-derivative component, that would be impossible. This should have been addressed in the data onboarding, so it is hoped that it would not be a concern by the time of product release.

The theme of data licenses can be important at different stages of your project, whether or not you expect that to be the case. To help you incorporate them into your project planning, this section provides suggestions about where you should think about the theme, structured using the stages from the Data Value Chain (DVC).

The DVC is a way of viewing the process of running a project from the point of view of the data, thereby identifying how it is onboarded, processed, enriched, analyzed and released in a product. In doing so, the DVC shows the moving parts in project implementations, making it a useful framework regarding the general steps of any project working with data.

Data value chain: Project SIS

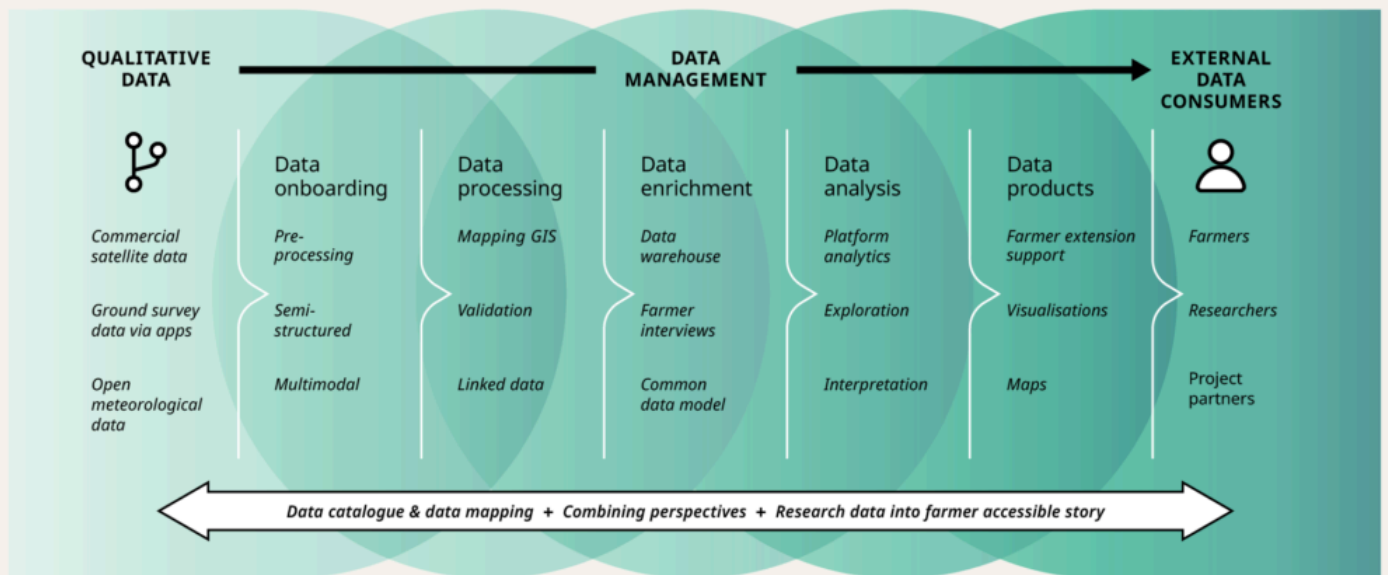
An illustrative scenario for implementing FAIR and responsible data practices



Part of the FAIR Process Framework, developed by CAB International Enabling Data Access (EDA) project team for the Bill & Melinda Gates Foundation. (CC-BY-SA) fairprocessframework.org

Data value chain: Waterways

An illustrative scenario for implementing FAIR and responsible data practices



Part of the FAIR Process Framework, developed by CAB International Enabling Data Access (EDA) project team for the Bill & Melinda Gates Foundation. (CC-BY-SA) fairprocessframework.org

For planning and implementation of projects and programs, data sharing among various stakeholders is very key and also prior to data sharing, data quality issues on FAIR are paramount.

[Learn more](#)

[Acknowledgements](#)

[FAQs](#)

[Glossary](#)

[Accessibility](#)

[Privacy & cookies](#)

[T&Cs](#)

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. Except where otherwise noted, the content on this website is licensed under a Creative Commons Attribution 4.0 International License.