

What are FAIR data?

One of the goals of data management is to facilitate the discovery and reuse of scientific knowledge by humans and computer systems. The FAIR principles (Findable, Accessible, Interoperable, Reusable) serve as a guideline for those wishing to achieve this goal. The European Commission has adopted these principles and requested that all projects funded by the H2020 program follow them.

Findable

Requirements: The data must be easy to find by both humans and computer systems

Note:

- The data must have a **unique permanent identifier**
- The data must be described by detailed **metadata**
- The data must be saved or indexed in a **searchable source**

Accessible

Requirement: the data must be stored in the long term so that they can be readily accessed and/or downloaded. Data can be FAIR without being directly accessible (e.g. restricted access, embargoed) but the metadata must be accessible.

Note:

- The data should be deposited in a **trustworthy data repository**, based on a known and **sustainable** business model, and preferably **certified**
- The **access conditions** and **dissemination license** for both metadata and data should be well defined
- If possible, the data and metadata should be available through **Open Access**
- If the data are restricted or embargoed, the **metadata should be accessible** to draw attention to the existence of the data while securing their protection



This document is made available under the terms of the [Creative Commons CC-BY-NC-SA License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Interoperable

Requirement: the data must be easy to combine with other datasets by both humans and computer systems.

Note:

- **Free and open formats** should be favored (at least for the dissemination and archiving of data)
- If data processing and analysis software is developed, the **source code for the software** should be provided and documented to facilitate its reuse
- **Metadata standards** and **standardized vocabularies** should be preferred
- Data should **refer to other data**, if possible

Reusable

Requirement: the data should be ready for reuse in future research and for processing by computerized methods.

A method for dataset evaluation: Is my dataset FAIR?

This method, proposed by the Dutch Data Archiving and Networked Services (DANS) Institute, aims to evaluate the quality of datasets in terms of FAIR principles. It involves the determination of a score from 1 to 5 for the first three criteria: findable, accessible, and interoperable. From the scores obtained for these three criteria, you can determine whether your dataset is reusable. The quality of your dataset can therefore be represented schematically as shown in the example on the right. The DANS Institute has suggested that this scheme could be appended to each dataset in data repositories. This suggestion has not yet been implemented, but would provide a clear visual indication as to whether the dataset can be considered FAIR.



We provide several tables¹ below for the evaluation of your datasets by this method. You can also use [the online FAIR evaluation tool](#) (as well as the [guide](#) to help you fill it in) offered by the [UK Data Service](#).

¹ **Source:** These tables were constructed with the EUDAT webinar: [FAIR Data in Trustworthy Data Repositories Webinar](#)

The dataset is findable

	Permanent identifier	Metadata
☆	✗	✗
☆☆	✓	✗
☆☆☆	✗	Sufficient*
☆☆☆☆	✓	Sufficient*
☆☆☆☆☆	✓	Rich**

* The metadata are sufficient to understand the dataset.

** The metadata indicate how the data can be reused.

The dataset is accessible

	Reuse license provided	Accessible metadata	Public data	Restricted access
☆	✗	✗	✗	
☆☆	✗	✓	✗	
☆☆☆	✓	✓	✗	Highly restricted access*
☆☆☆☆	✓	✓	✓	Registration required
☆☆☆☆☆	✓	✓	✓	None**

* Examples of restrictions:

- Confidentiality: ethical issues limit access
- Specific software required to access the data
- Commercial interest (data access granted against a payment to the journal)
- Embargo period (access for only 24 h, for example)
- Access limited to a specific group

**Open Access data (e.g. CC0 or CC-BY license)

The dataset is interoperable

	Open format	In one of the preferred formats*	Standardized vocabulary**	Links to other data
☆	✗	✗	✗	✗
☆☆	✓	✗	✗	✗
☆☆☆	✓	✓	✗	✗
☆☆☆☆	✓	✓	✓	✗
☆☆☆☆☆	✓	✓	✓	✓

* Preferred formats are file formats in which the Dutch Data Archiving and Networked Services Institute has confidence and which offer the best long-term guarantees in terms of usability, accessibility, and sustainability: [see the list of preferred formats](#)

** Meaningful connections between datasets from different projects are dependent on the use of common coding methods for data values. Many research communities have defined standard vocabularies, ontologies, coding/marking/labeling schemes, thesauri, classification systems, and other semantic knowledge graphics to make data interoperable.

The dataset is reusable

Reusability is a highly subjective notion, and it is, therefore, difficult to establish a score. You can calculate the reusability of your dataset from the score obtained for the first 3 conditions: $(F+A+I)/3$.