FAIR principles

The main goal of FAIR principles is to provide the scientific community with a framework in which data, methodologies and tools are used for validating (and, in certain cases, for replicating) the conclusions of a research project.

The first step is **finding** the data. Then, a user needs to know how to **access** them (either through a free download or through an authentication and authorisation mechanism). The data must be in the condition to be **integrated** with other data and **interoperate** with different applications or workflows.

The goal of FAIR principles is to **optimise the reuse** of data. They describe how research results should be organised to make it easier for anyone to **access, understand, exchange and reuse** the data. Not all FAIR data is open data.

Data and FAIR principles

We have seen that repositories are archives that preserve data and make it findable online. The data deposited by a researcher in a repository is **more likely to abide by FAIR principles**. In fact, in order to be findable, data and datasets must be accompanied by a Persistent Identifier (PID), i.e. a long-term, unique and unambiguous reference. Examples include DOIs, Handles and URNs. Data and datasets must also be accompanied by meaningful metadata and keywords (containing the PID).

Accessible data is not necessarily open data; rather, it is **data associated with clear access conditions**. Data may be openly accessible (usually the default) or accessible through an authentication and authorisation system, when their nature prevents them from being open. Privacy and intellectual

property protection are only two of the reasons for restricting data access. Remember that the metadata associated with closed data should be kept openly accessible via standard protocols.

In order to be interoperable, data need to be combinable and usable with other data or tools. This is achieved by using **common and preferably open formats** as well as **standardised languages** that are shared internationally by the various indexing services. Repositories support **interoperability standards** and may recommend using **specific ontologies or vocabularies** **Repositories** **Using different types of standards**.

To guarantee maximum reusability, data must also be **described and documented** thoroughly, to ensure compliance with the standards adopted by the relevant scientific communities and allow reuse and combination across different contexts.

Last but not least, accompanying a dataset with a **clear, accessible, possibly open licence** is key to establish and declare how the dataset can be reused **Copyright**.

Metadata and FAIR principles

FAIR principles apply to both data and metadata.

Rich, descriptive metadata make a significant impact on data utility by improving findability, accessibility, interoperability and reusability. To comply with FAIR principles, metadata must be always accessible and associated with a permissive licence (CC0 or equivalent), even when the data is not **Respecting privacy**.

Sul campo!

How do I know if my data is FAIR?

To easily self-assess compliance with FAIR principles in research data management, please find below a checklist prepared by EUDAT and reworded by AlmaDL

Findable

- ☐ Has a persistent identifier (e.g. DOI, Handle, URN) been assigned to your dataset?
- ☐ Are there rich metadata, describing your dataset?
- ☐ Are your metadata recorded in an online searchable resource e.g. a catalogue or data repository
- ☐ Does your metadata record specify the persistent identifier?

Accessible

- ☐ Does the persistent identifier take you directly to the dataset or associated metadata?
- ☐ Does the protocol by which data can be retrieved follow recognised standards?
- ☐ Are metadata public, wherever possible, even if the data are not?

Interoperable

- ☐ Have your data been provided in commonly understood and preferably open formats?
- ☐ Do your metadata follow relevant standards?
- ☐ Are controlled vocabularies, keywords, thesauri or ontologies used where possible?
- ☐ Are qualified references and links provided to other related data, such as publications, technical reports or software applications?

Reusable

- ☐ Are your data accurate and well described with many relevant attributes?
- ☐ Has your dataset been assigned a clear and accessible data usage licence?
- ☐ Have the scientific responsibility for and purpose of the data been made clear in the metadata and attached documentation?
- ☐ Do your data and metadata meet relevant domain standards?

Useful links

Wilkinson et al, The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Video “UGent Open Science. Knowledge clip: FAIR data principles”. https://www.youtube.com/watch?v=2uZxFu9SFi8

More on the FAIR principles:

- GOFAIR. FAIR Principles https://www.go-fair.org/fair-principles
- FAIRsFAIR Fostering Fair Data Practices in Europe https://www.fairsfair.eu/
- How to FAIR https://howtofair.dk/

Video “Dati della ricerca: la European Open Science Cloud e i principi FAIR” (in Italian) https://www.youtube.com/watch?v=eNiHNaU6MrQ