

## The FAIR data principles

By Noortje Dijkstra, UiT The Arctic University of Norway ORCID:  
0000-0003-2596-7691

Date: February 2022



Have you come across the FAIR data principles? No matter what kind of research you do, sooner or later you will. What are the FAIR data principles and what is FAIR data?

FAIR is an acronym that stands for Findable, Accessible, Interoperable and Reusable. The FAIR data principles can be considered as a set of guidelines that enable and increase the reuse of data by humans, but also computers. In other words, computational systems should be able to find, access, interoperate with and reuse your data with none or minimal human intervention.

The FAIR data principles are increasingly gaining attention by the scientific community. In addition, research funders and universities often require that you make your data as FAIR as possible.

It is important to realise that FAIR data is not necessarily the same as open data, data can be FAIR and closed or open but not FAIR.

Open data is data that can be freely used and shared by anyone and for any purpose. The FAIR principles, on the other hand, provide a set of guidelines for sharing data and at the same time respecting any ethical, legal, or commercial restrictions.

So how can you make your data FAIR? Unfortunately, there's no one size fits all answer. However, note that an important step towards FAIR data is taken when you deposit your data in a trustworthy repository for research data.

But making your data FAIR is not only about archiving your data. Remember to apply the FAIR principles throughout the entire life cycle of research data management. The earlier on in the project you start to collect and manage your data in a FAIR way. The easier it will be to keep the data FAIR. That is why planning for research data management is so important.

Now let's have a look at what each of the four principles means.

**Findable** means that others can discover your data. When uploading your data to a repository for research data, you typically need to provide metadata.

Metadata is data that describes your data. By equipping your data with metadata, using a scheme that complies to a specific standard, your metadata will become machine actionable and searchable in discovery tools. In addition, data becomes findable by archiving it in a

repository that assigns a persistent identifier to your data, such as a DOI or a Handle. And finally, it is important that the repository is indexed in a search engine for research data. Such as DataCite or Google Datasets Search. If you choose a reliable archive for your data, you do not have to worry about this. A reliable archive will make sure that the metadata of your dataset is indexed in the search engine.

**Accessible** means that your data can be made available to others and is always obtainable. It is about archiving your data for the long term in an archive that can be easily accessed and downloaded from by humans, but also computers. Metadata and data should therefore be accessible via the persistent identifier, using a standard communication protocol such as HTTP or HTTPS. However, accessible does not necessarily mean that there are no access restrictions. It rather means that there are sufficient authentication and authorisation procedures in place to make sure that the access conditions for the data are respected. If such access restrictions are needed, this should be clear to both humans and also computers. In addition, your metadata should be accessible even if the data themselves are restricted and not available anymore.

**Interoperable** means the data can be integrated with other data and can be easily used by computers. Data becomes interoperable when it can be compared and processed across archives and systems. Whenever possible, metadata and data should use standard formats, terms or vocabularies that the research community has agreed upon. This will allow metadata to be exchanged between researchers, institutions and countries, and also across computational systems. Another way to make your data more interoperable is providing context to your dataset by including references to relevant metadata or other datasets on which you have built your own datasets.

**Reusable** means that your data can be reused. Providing others with access to your data facilitates data reuse for new research and innovation projects, and it will also increase research transparency. To achieve this, data should be sufficiently documented and described. The documentation accompanying your data should help others to understand the provenance of your data. In other words, where did the data come from and what happened to it along the way? This will help others to understand the context of the data and also helps them to judge how relevant your data is for them. It is also important to let others know how they can reuse your data and what kind of reuse is allowed, by equipping the data with a clear user license. Trustworthy repositories will make sure that you can choose a machine-readable license.

So now that we have discussed the four components of the FAIR principles in more detail, it is important to say that data is not either FAIR or unFAIR. Data can be FAIR to a greater or lesser extent. And the four aspects are closely connected.

Data repositories are key infrastructure, enabling FAIR data, but they won't do all the work for you, you know your own data best, and therefore you are responsible for equipping your data with rich metadata and good documentation to make sure that your data is reusable and understandable.

Following best practice for making your data FAIR can result in maximising research output and impact, which in turn can result in enhancing your recognition as a researcher.