

Is your Scientific Data FAIR

For many years, we have seen the proliferation of data as we increasingly instrument our scientific processes. We have developed a diverse landscape of tools and processes, making significant leaps from paper based documentation, but created a new nightmare of integration and complex analysis. The FAIR initiative or set of principles is a framework to reduce that complexity through the application of a core set of principles outlined below, making data machine readable across sources. This unlocks the data from the proprietary structure and system walls, and offers a foundation to build interconnected analysis and insights.

Reference this excerpt from the abstract here that summarizes quite nicely what the objective is:

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.

[HTTPS://WWW.NATURE.COM/ARTICLES/SDATA201618#ABS1](https://www.nature.com/articles/sdata201618#ABS1)

The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

There are an increasing number of resources targeted at supporting the movement to FAIR data, a couple of which are included here to get you started. There is much to cover on this topic, but these links and materials are a start on the conversation.

[How to GO FAIR](#)

[FAIR Implementation Project](#)

References

- NIST:
<https://www.nist.gov/itl/ssd/information-systems-group/configurable-data-curation-system-cdcs/cdcs-help-and-resources-1>
- scientific data (Nature.com): <https://www.nature.com/articles/sdata201618#Abs1>
 - Local link in case above link fails: [The FAIR Guiding Principles for scientific data management and stewardship](#)