How to create FAIR data

1. Documentation

The description of a dataset which is provided as a part of a related publication may often not be enough to provide the full context of data collection and processing. Therefore, a dataset should always be accompanied by rich documentation which provides a detailed description of how the data were created, what instrument was used for taking measurements, who the respondents were, how the data were cleaned and processed, and other relevant information that are necessary for accurate interpretation of the data. Documentation should also include links to other related entities (ideally via persistent identifiers), such as publications stemming from the given dataset, author(s) of the dataset, metadata standards that were use for describing the data and so forth.

2. Standardised vocabulary

Use standardised terminology for your data and their description, which is established in your field. Such practice ensures that everyone will understand your data and also facilitates the potential combination of your data with other datasets (interoperability). If you are using an established standardised vocabulary to describe your data, you should include the information about the used vocabulary in the documentation.

3. Persistent identifiers

Use persistent identifiers both for your data and for referencing other related entities (publications, authors, institutions, other datasets etc.). By using persistent identifiers, you improve the findability of your data and help users to identify the dataset, its authors and other related entities. Repositories typically assign persistent identifiers to datasets.

4. File formats

If possible, you should use open file formats or formats which are established in your field for storing and sharing your data. This way, the possibility of data reuse will not be limited by the availability of a specific software required for accessing them.

5. Access to data

Define who should have access to the data and under what conditions. Under certain circumstances, even data with restricted access can comply with the FAIR principles – typically in cases where data cannot be shared because they contain personal information, sharing would be contrary to intellectual property protection, or because they include data associated with (national) security. If your data include personal information, consider the possibility of anonymisation to enable sharing with a wider community. For sharing research data, use data repositories – either domain specific, where available, or general repositories such as Zenodo , Figshare or Dryad .

6. Licensing

Clearly specify the conditions for reuse. For open sharing, Creative Commons licenses are the most commonly used, but you can also use a custom license that better suits your needs. You can also use multiple licenses for the same dataset, such as free sharing for non-commercial use and available for commercial use at a fee.

How FAIR are your data?

Some data management tools support the FAIR principles and will assist you in evaluating how FAIR your data are. The Data Stewardship Wizard tool for creating data management plans, for instance, incorporates FAIR metrics within relevant questions, guiding users on how to process data to align them as closely with the FAIR principles as possible.

To assess the 'FAIRness' of your data, you can use the FAIR self-assessment tool developed by the Australian initiative ANDS-Nectar-RDS, or you can use this checklist which was created for the EUDAT summer school by Sarah Jones and Marjan Grootveld.