FAIR Data

Purpose: Guidance on applying FAIR Data principles to research data

Audience: UKERC researchers who create research data

Summary

Enabling your research data to be FAIR (Findable, Accessible, Interoperable, Reusable) is good research practice. FAIR is not the same as Open Data. FAIR Data can have restricted access.

As a research data creator you can make your data FAIR by **depositing data** in a repository**,** providing **good quality descriptive metadata** , using **domain information standards,** providing a **data license** and **contextual information.**

Introduction

The FAIR principles were created in 2016 as supporting the process to enable the sharing of research data. They apply to humans and machines equally.

Findable

To be able to use or reuse data it must be able to be found.

- **Persistence:** the data should have a unique and persistent identifier. *This can be achieved by depositing the data into a repository*
- **Rich description:** the data should have good metadata associated with it which includes the persistent identifier. *This can be achieved by depositing the data into a repository. Additional metadata can also be stored in files with the data*
- **Indexed or registered in a searchable resource:** the data, or its metadata should be searchable. *This can be achieved by depositing the data into a repository*

Accessible

It must be clear how to access the data and what, if any, authorisation/authentication steps needed.

- **Persistence:** the metadata about the data should be available even if the data is no longer available. *This can be achieved by assigning a DOI (usually though deposit in a repository)*
- **Authentication and authorisation:** it should be possible to restrict who is able to access the data using a standard protocol *This can be achieved by depositing the data into a repository. Additional metadata can also be stored in files with the data*
- **Retrievability:** the data, or its metadata should be able to be retrieved using standard ICT protocols. *This can be achieved by depositing the data into a repository or placing it on a website.*

Interoperable

Being able to combine data from multiple sources means that a common understanding of the data needs to be identified.

- **Descriptive standards:** the data and metadata should use formal and shared schema to describe the data which is both human and machine readable. *This can be achieved by using Dublin Core for the basic descriptive metadata and standards for domain specific content. Depositing data in a repository will also achieve this for descriptive metadata.*
- **Vocabulary standards** the data and metadata should use FAIR vocabularies to add information in a consistent manner to add a common understanding which is both human and machine readable. *This can be achieved by depositing the data into a repository and using standard vocabularies such as international agreed date formats.*
- **Referencing other objects:** the data, or metadata should include references to other objects such as vocabularies used. *This can be achieved by using persistent identifiers such as DOIs (digital objects) or ORCIDs (people) in the data and metadata. Depositing data in a repository may also enable linking between digital objects.*

Reusable

To re-use data, it should be clear what it is, what rights have been assigned and how it was generated.

- **Description:** the data should have a sufficient and accurate attributes identified for it to be able to be used by someone who was not involved in the creation/generation of the data. *This can be achieved by good data documentation and recording.*
- **License:** the data should have a license assigned to it which identifies the rights and responsibilities of any re-user. *This can be achieved by assigning a license. There may be domain expectations or your employer may have a policy on this.*
- **Provenance:** the data, or its metadata should have information on how it was generated/collected and any processing together with anything of significance (such as gaps in coverage/collection) needed to understand the data. *This can be achieved by good data documentation and recording.*
- **Community standards:** the data, or its metadata should use domain standards. This is a challenge for energy researchers as they come from many domains. If in doubt follow the standards set out by the repository the data is to be deposited in. *This can be achieved by seeking advice from appropriate domain repositories and community organisations.*

Further Reading

- More EDC advice from our service website About Data Management pages
- Wilkinson, M., Dumontier, M., Aalbersberg, I. 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
- Training course on applying the FAIR principles What is FAIR?
- Information on standards and policies: https://fairsharing.org/