

Answer: FAIRness of a dataset

1 A. The **assessment** using the FAIR self-assessment tool for **dataset A** is as follows:

Findable

1. Does the dataset have any identifiers assigned?
(d) Globally unique identifier, citable and persistent, i.e. DOI
2. Is the dataset identifier included in all metadata records/files describing the data?
(a) No
3. How is the data described with metadata?
(d) Comprehensively using a recognised formal machine-readable metadata schema
4. What type of repository or registry is the metadata record in?
(e) Data are in one place, i.e. UK Data Service, but discoverable through several registries, e.g. EUDAT B2Find, CESSDA Data Catalogue, DataCite

Accessible

5. How accessible is the data
(f) Fully accessible to persons who meet explicitly stated conditions, i.e. accessible to people registered with the UK Data Service and agreeing to the UKDS End User Licence
6. Is the data available online without requiring specialised protocols or tools once access has been approved?
(c) File download from online location
7. Will the metadata record be available even if the data are no longer available?
(c) Yes

Interoperable

8. What (file) formats is the data available in?
(a) Mostly in a proprietary format, i.e. RTF
9. What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?
(b) No standards have been applied in the description of data elements, i.e. data elements in this dataset are interview transcripts that have been described with textual documentation
10. How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?
(a) There are no links to other metadata

Reusable

11. Which of the following best described the licence/usage rights attached to the data?
(d) Standard text-based licence, i.e. the UK Data Service End User Licence
12. How much provenance information has been captured to facilitate data reuse?
(c) Fully recorded in a text format, i.e. recorded in various documentation files



1 B. The **assessment** using the FAIR self-assessment tool for **dataset B** is as follows:

Findable

1. Does the dataset have any identifiers assigned?
(d) Globally unique identifier, citable and persistent, i.e DOI
2. Is the dataset identifier included in all metadata records/files describing the data?
(b) Yes
3. How is the data described with metadata?
(b) Brief title and description
4. What type of repository or registry is the metadata record in?
(e) Data are in one place, Zenodo, but discoverable through several registries, e.g. Google Data Search

Accessible

5. How accessible is the data
(g) Publicly accessible
6. Is the data available online without requiring specialised protocols or tools once access has been approved?
(c) File download from online location
7. Will the metadata record be available even if the data are no longer available?
(c) Yes

Interoperable

8. What (file) formats is the data available in?
(c) In a structured, open standard, machine-readable format, i.e. csv
9. What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?
(b) No standards have been applied in the description of data elements
10. How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?
(a) There are no links to other metadata

Reusable

11. Which of the following best described the licence/usage rights attached to the data?
(e) Standard machine-readable licence, i.e. Creative Commons Attribution 4.0 International
12. How much provenance information has been captured to facilitate data reuse?
(b) Partially recorded

2 A. **Dataset A could be made more FAIR** by:

- including the dataset identifier in all metadata files (e.g. in file names)
- improving interoperability by converting to a structured open standard machine-readable file format such as text or PDF/A
- tagging the data elements with standardised ontologies
- representing metadata in RDF

Improving accessibility may be difficult due to the sensitive nature of the data and the potential disclosure risk that prohibit making these data available openly.

2B. **Dataset B could be made more FAIR** by:

- improving interoperability by tagging the data elements with standardised ontologies
- representing metadata in RDF
- reusability can be improved by enhancing the provenance information.

