Exercise: FAIRness of a dataset

1. Assess how Findable, Accessible, Interoperable and Reusable (FAIR) an existing dataset is using the FAIR self-assessment tool[1]. Note down details of how each aspect has been achieved (or not). You can use the question list below, or use the online tool. Use the following datasets for this exercise:

    A. Qualitative dataset: Kovacheva, S. and Demireva, N. (2018). The lived experiences of migration 1996-2017. [Data Collection]. Colchester, Essex: UK Data Service. DOI: 10.5255/UKDA-SN-853333

    B. Quantitative dataset: Damián, J., Pastor-Barriuso, R., García-López, F., Ruigómez , A., Martínez-Martín, P. and de Pedro-Cuesta, J. (2018) Facility ownership and mortality among older adults residing in care homes [Data set], Zenodo. DOI: 10.5281/zenodo.2549444

2. Consider what could be done to make these datasets more FAIR.

FAIR self-assessment tool

Findable

1. Does the dataset have any identifiers assigned?
    (a) No identifier
    (b) Local identifier
    (c) Web address (URL)
    (d) Globally unique identifier, citable and persistent (e.g. DOI, PURL, ARK or Handle)
2. Is the dataset identifier included in all metadata records/files describing the data?
    (a) No
    (b) Yes
3. How is the data described with metadata?
    (a) The data are not described
    (b) Brief title and description
    (c) Comprehensive, but in a text-based, non-standard format
    (d) Comprehensively using a recognised formal machine-readable metadata schema
4. What type of repository or registry is the metadata record in?
    (a) The data are not described in any repository
    (b) Local institutional repository
    (c) Domain-specific repository
    (d) Generalist public repository
    (e) Data are in one place but discoverable through several registries

Accessible

5. How accessible is the data
    (a) No access to data or metadata
    (b) Access to metadata only
    (c) Unspecified conditional access, e.g. contact the data custodian

---

[1] ARDC (2018) FAIR self-assessment tool, Australian Research Data Commons: https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/

(d) Embargoed access after a specified date

(e) A de-identified / modified subset of the data is publicly accessible

(f) Fully accessible to persons who meet explicitly stated conditions, e.g. ethics approval for sensitive data

(g) Publicly accessible

6. Is the data available online without requiring specialised protocols or tools once access has been approved?

(a) No access to data

(b) By individual arrangement

(c) File download from online location

(d) Non-standard web service (e.g. OpenAPI, Swagger, Informal API)

(e) Standard web service API (e.g. OGC)

7. Will the metadata record be available even if the data are no longer available?

(a) Unsure

(b) No

(c) Yes

Interoperable

8. What (file) formats is the data available in?

(a) Mostly in a proprietary format

(b) In a structured, open standard, non-machine-readable format

(c) In a structured, open standard, machine-readable format

9. What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?

(a) Data elements not described

(b) No standards have been applied in the description of data elements

(c) Standardised vocabularies/ontologies/schemas without global identifiers

(d) Standardised, open and universal, using resolvable global identifiers linking to explanations

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

(b) The metadata record includes URI links to related metadata, data and definitions

(c) Metadata are represented in machine-readable format, e.g. in a linked format such as Resource Description Framework (RDF)

Reusable

11. Which of the following best described the licence/usage rights attached to the data?

(a) No licence

(b) Non-standard text-based licence

(c) Non-standard machine-readable licence (clearly indicating under what conditions the data may be reused)

(d) Standard text-based licence

(e) Standard machine-readable licence (e.g. Creative Commons)

12. How much provenance information has been captured to facilitate data reuse?

(a) No provenance information is recorded

(b) Partially recorded

(c) Fully recorded in a text format

(d) Fully recorded in a machine-readable format