FAI**R** Principles

–

Reusable

Analysis of existing approaches v0.02

Existing approaches

Findable   Accessible   Interoperable   Reusable   Non-FAIR Principles

R1
R1.1
R1.2
R1.3

Develop common metrics per facet

FAIR Principles

To be reusable :

R1. meta(data) is richly described with a plurality of accurate and relevant attributes

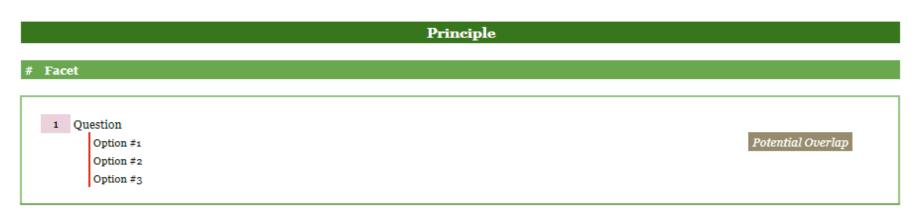R1.1. (meta)data is released with a clear and accessible data usage licence

R1.2. (meta)data is associated with detailed provenance

R1.3. (meta)data meets domain-relevant community standards

LEGEND

1	ANDS-NECTAR-RDS-FAIR data assessment tool	ARDC	[Link]
2	DANS-Fairdat	DANS	[Link]
3	DANS-Fair enough?	DANS	[Link]
4	The CSIRO 5-star Data Rating tool	CSIRO	[Link]
5	FAIR Metrics Questionnaire	The FAIR Metrics Group	[Link]
6	Stewardship Maturity Mix	NOAA's CICS-NC, NOAA's NCDC	[Link]
7	FAIR Evaluator	GO FAIR, LUMC CBGP, IDS, OeRC, IQSS	[Link]
8	Data Stewardship Wizard	ELIXIR NL/CZ	[Link]
9	Checklist for Evaluation of Dataset Fitness for Use	Assessment of Data Fitness for Use WG (WDS/RDA)	[Link]
10	RDA-SHARC Evaluation	SHARC IG (RDA)	[Link]
11	WMO-Wide Stewardship Maturity Matrix for Climate Data	The SMM-CD WG	[Link]
12	Data Use and Services Maturity Matrix	The MM-Serv WG	[Link]

Principle

 # Facet

1 Question
   Option #1
   Option #2
   Option #3

Potential Overlap

R1	meta(data) is richly described with a plurality of accurate and relevant attributes

2 Is there sufficient metadata available?
- No
- Yes

2 How is the data described with metadata? — *F2*
- The data is not described
- Brief title and description
- Comprehensively, but in a text-based, non standard format
- Comprehensively, using a recognized formal machine readable metadata schema

2 Is there extensive metadata and rich additional documentation available? — *R1.2, I3*
- No
- Yes

3 Did you provide enough information (metadata) about your data for others to understand and reuse your data?
- No
- Yes

9 Granularity of data entities in dataset is appropriate in Respect of Meta-Data Granularity
- No
- Somewhat
- Yes

9 Structure, size and MIME type of the dataset agrees with description of the dataset content
- No
- Somewhat
- Yes

• • •

9 Content of the dataset agrees with description of the dataset content

No

Somewhat

Yes

9 Coverage (spatial, temporal, or other dimensions) adequate

No

Somewhat

Yes

10 Which relevant actions have been undertaken by the researcher to enhance the data reuse potential

Never /NA

If mandatory

Sometimes

Always

10 Does the researcher provide information on methods and tools that permit the understanding, integrity, value and readability of data intended to be kept on the

Never /NA

Sometimes

Always

11 Documentation

Product information not publicly available online.

Limited online documentation (e.g., User Guide).

Document on how the data product was created and how to use it, is available online.

Full documentation based on a standard template and available online.

Previous + Online tutorial on using and analyzing the dataset; Complete production system information available online.

• • •

· · ·

12 Data Use

No use or usability metadata/documentation is available to help users understand and use the data

Use or usability metadata/documentation is available from local systems (e.g., product website)

Standard-based use/usability metadata/documentation is available from enterprise systems

Enterprise systems include online use/usability support services (online help, hints, etc.)

Enterprise systems include advanced use/usability support service such as interactive visualizations of relationships (e.g., to papers, other products, researchers, etc.)

11 Metadata   *R1.2, R1.3*

Metadata not publicly available and/or not usable.

Limited Metadata publicly available; Conforming to community-standard; Basic characteristics of dataset.

Previous + Conforming to international standards in most aspects; limited quality and provenance metadata.

Fully compliant with international standards; Rich metadata content; Basic granular-level metadata; Support dataset provenance.

Previous + complete granular-level metadata; Metadata QC-ed and Regularly updated

R1.1 (meta)data is released with a clear and accessible data usage licence

1 Which of the following best describes the license/usage rights attached to the data?
No licence
Non standard text based licence
Non standard machine readable licence (e.g.clearly indicatingunder what conditions the data may be used)
Standard text based licence
Standard machine-readable licence (e.g. Creative Commons)

2 Does the user license have any user restrictions for accessing the data?
No
Yes

2 Does the dataset have a user license?
No
Yes

3 Does the dataset have a usage licence?
No
Yes

4 Licensed - conditions for re-use are available and clearly expressed
No
License described in text
Link to a standard license (e.g. Creative Commons)

5 Please provide the IRI for you usage license regarding the content returned from RESOURCE ID

7 The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents

9 Terms of usage (licenses, other conditions of reuse, data protection, ethical issues)
No
Somewhat
Yes

R1.2 (meta)data is associated with detailed provenance

1 How much provenance information has been captured to facilitate data reuse

No provenance information is recorded

Partially recorded

Fully recorded in a text format

Fully recorded in a machine readable format

2 Is there extensive metadata and rich additional documentation available? *R1, I3*

No

Yes

3 Did you give detailed provenance information for the data?

No

Yes

4 Trusted - accompanied by, or linked to, information about how the data has been used, by whom, and how many times

No information about usage

Usage statistics available

Clearly endorsed by reputable organization or framework

4 Assessable - accompanied by, or linked to, a data-quality assessment and description of the origin and workflow that produced the data

No quality or lineage information

Text lineage statement

Formal provenance trace (e.g. PROV-O)

5 Please provide the IRIs (maximum 3) for the vocabularies being used to describe the provenance of the content resolved from RESOURCE ID

5 Please provide the IRIs (maximum 3) for the vocabularies being used to describe the domain information of the content resolved from RESOURCE ID

• • •

7 That there is provenance information associated with the data,

9 Citation exists, including authorship, year, comprehensive title, persistent identifier (e.g. DOI) `F1`

  - No
  - Somewhat
  - Yes

10 Are the provenance and type of all data properly specified (origin of raw, primary, transformed, secondary..)

  - Never /NA
  - If mandatory
  - Sometimes
  - Always

8 How will you be making sure there is good provenance of the data analysis?

  - We use lab notebooks
  - We use an electronic lab notebook
  - We use other arrangements

6 Transparency / traceability

  - Limited product information available / Person-to-person
  - Product information available in literature
  - Algorithm Theoretical Basis Document (ATBD) & source code online / Dataset configuration managed (CM) / Unique Object Identifier (OID) assigned (dataset, documentati
  - Operational Algorithm Description (OAD) online, OID assigned, and under CM
  - System information online / Complete data provenance online

11 Metadata `R1, R1.3`

  - Metadata not publicly available and/or not usable.
  - Limited Metadata publicly available; Conforming to community-standard; Basic characteristics of dataset.
  - Previous + Conforming to international standards in most aspects; limited quality and provenance metadata.
  - Fully compliant with international standards; Rich metadata content; Basic granular-level metadata; Support dataset provenance.
  - Previous + complete granular-level metadata; Metadata QC-ed and Regularly updated

R1.3 (meta)data meets domain-relevant community standards

3 Are the data stored and archived in preferred archival formats?

I₁

    No

    Yes

3 Do you make use of relevant community standards?

    No

    Yes

4 Loadable - represented using a common or community-endorsed (i.e. standard) format

    Bespoke format (text, binary)

    One standard format, denoted by a MIME-type

    Multiple standard formats

4 Usable - structured using a discoverable, community-endorsed (standard?) schema or data model

    No formal schema

    Explicit schema or data model, formalized in DDL, XSD, DDI, RDFS, JSON-Schema, data-package or similar

    Community-shared schema or data model , available from a standard location

5 Please provide the IRI that represents the certification from a recognized authority in your community or domain, indicating that the content of RESOURCE ID

7 Certification, from a recognized body, of the resource meeting community standards.

10 Do the data reuse control and data sharing arrangements meet the data protection and "local/national ethics requirements?

    Never /NA

    If mandatory

    Sometimes

    Always

• • •

· · ·

10 If relevant, has the researcher used valid and updated standards for data describing ? If so, are the data standards and particularly versioning data standards re
  Never /NA
  If mandatory
  Sometimes
  Always

9 Additional metadata adequate to respective research domain (if applicable)
  No
  Somewhat
  Yes

6 Usability
  Extensive product-specific knowledge required / No documentation online
  Non-standard data format / Limited documentation (e.g., user's guide) online
  Community standard-based interoperable format & metadata / Documentation (e.g., source code, product algorithm document, processing or/and data flow diagram) online
  Basic capability (e.g., subsetting, aggregating) & data characterization (overall/global, e.g., climatology, error estimates) available online
  Enhanced online capability (e.g., visualization, multiple data formats) / Community metrics of data characterization (regional/cell) online / External ranking

11 Metadata   *R1, R1.2*
  Metadata not publicly available and/or not usable.
  Limited Metadata publicly available; Conforming to community-standard; Basic characteristics of dataset.
  Previous + Conforming to international standards in most aspects; limited quality and provenance metadata.
  Fully compliant with international standards; Rich metadata content; Basic granular-level metadata; Support dataset provenance.
  Previous + complete granular-level metadata; Metadata QC-ed and Regularly updated