



FAIR Principles

—

Findable

Analysis of existing approaches

Existing approaches



F_{indable}

F1

F2

F3

F4

A_{ccessible}

I_{nteroperable}

R_{eusable}

Non-FAIR
Principles

Develop common metrics per facet

FAIR Principles

To be findable :

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

LEGEND

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

Principle

Facet

1 Question

- Option #1
- Option #2
- Option #3

Potential Overlap

F1 (meta)data is assigned a globally unique and eternally persistent identifier

- 1 Does the dataset have any identifiers assigned?
 - No identifier
 - Web address (URL)
 - Local identifier
- 2 Does the dataset have a persistent identifier (PID)?
 - No
 - Yes
- 3 Will your dataset have a Persistent Identifier after deposit?
 - No
 - Yes
- 4 Citable - denoted using a formal identifier
 - Not citeable
 - Local identifier
 - Web address (URL - not guaranteed stable)
 - Persistent web identifier (URI)
- 4 Citable - denoted using a formal identifier
 - Not citeable
 - Local identifier
 - Web address (URL - not guaranteed stable)
 - Persistent web identifier (URI)
- 5 Please provide the IRI for a registered identifier schema for your resource's IRI (e.g. DOI, HTTP)
- 5 Please provide the IRI to the document describing the persistence policy for the identifier of this (meta)data
- 7 Whether there is a scheme to uniquely identify the digital resource.
- 7 Whether there is a policy that describes what the provider will do in the event an identifier scheme becomes deprecated.

A2

...

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

F2

No
Somewhat
Yes

9 Persistent identification of the dataset and related work (related literature and data, authors, projects, terms)

No
Somewhat
Yes

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

R1.2

No
Somewhat
Yes

10 Are each data/dataset identified by an indexed and independant identifier ?

Never /NA
If mandatory
Sometimes
Always

10 Are the data identifiers unique, global and persistent ?

Never /NA
If mandatory
Sometimes
Always

10 Has any identifying schema been used for data (e.g. DOI)

Never /NA
If mandatory
Sometimes
Always

10 Are all datasets linked to an authority (legal entity) through a unique and persistent identifier over time (e.g. institution, association or established body)?

Never /NA
If mandatory
Sometimes
Always

10 Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?

Never /NA
If mandatory
Sometimes
Always

F2 data is described with rich metadata

2 How is the data described with metadata?

R1

- 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 Please provide the IRI to a document that contains machine-readable metadata for the digital resource

2 Are the metadata accessible?

F4, A2

- No
- Yes

4 Described - tagged with metadata ?

- No metadata
- Abstract and keywords
- Basic metadata (e.g. Dublin Core)
- Specialized metadata (e.g. Darwin Core, ISO 19115/19139, schema.org scientific data profile)
- Rich metadata using multiple standard RDF vocabularies (e.g. DCAT, PROV, ADMS, GeoDCAT, FOAF, ORG, GeoSPARQL)

7 The availability of machine-readable metadata that describes a digital resource.

9 Dataset is provided in a widely-used or community-accepted machine-readable format and using standard terminologies for nominal data and available

- No
- Somewhat
- Yes

10 Are the types and formats of data generated / collected well described?

- Never /NA
- If mandatory
- Sometimes
- Always

3 Did you provide rich additional documentation?

- No
- Yes

...

...

9 Description of methods used to create this dataset are appropriate for the context and discipline

No
Somewhat
Yes

10 Does the researcher use efficient and rich services to access data (various formats, visualisations, practical tools and systems adapted to different types of use and

Never /NA
If mandatory
Sometimes
Always

6 Data Quality Assurance

Data quality assurance (DQA) procedure unknown or none
Ad Hoc and random / DQA procedure not defined and documented
DQA procedure defined and documented and partially implemented
DQA procedure well documented, fully implemented and available online with master reference data / Limited data quality assurance metadata
DQA procedure monitored and reported / Conforming to community quality metadata & standards / External review

6 Data Integrity

Unknown or no data ingest integrity check
Data ingest integrity verifiable (e.g., checksum technology)
Data archive integrity verifiable
Data access integrity verifiable / Conforming to community data integrity technology standard
Data authenticity verifiable (e.g., data signature technology) / Performance of data integrity check monitored and reported

...

...

11 Quality Assurance & Control

X4

Ad hoc or no data quality assurance (QA) & control (QC) procedure or information unknown.

QA/QC procedure are defined, documented, and partially implemented.

QA/QC procedure are well-defined according to community best practices, documented and fully applied.

Previous + provision of error statistics published or tracked with results made available online and communicated to data providers; Procedure for user feedback, improvement

Previous + detailed analysis of errors and gaps at space-time unit level: (Station, grid-points, daily, monthly and or annual time-scale, etc.); QA/QC procedure monitored; Re

11 Data Integrity

Unknown or no data integrity check.

Random data integrity check.

Data integrity verified systematically but methodology not commonly known.

Data integrity systematically verified and following well known practices but not necessarily consistent across platforms.

All steps in data integrity check systematically verified and adhering to well-known practices.

F3 metadata clearly and explicitly includes the identifier of the data it describes

1 Is the dataset identifier included in all metadata records/files describing the data?

No
Yes

5 Please provide the IRI of the metadata

7 Whether the metadata document contains the globally unique and persistent identifier for the digital resource

10 Are the metadata linked to the dataset through a persistent identifier?

Never /NA
If mandatory
Sometimes
Always

F4 (meta)data is registered or indexed in a searchable resource

1 How accessible is the data?

No access to metadata or data

Access to metadata only

Unspecified conditional access (e.g contact the data custodian for access)

Embargoed access after a specified date

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

Publicly accessible

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

2 Are the metadata accessible?

No

Yes

2 Is the dataset available for public access? (i.e. the restriction is only registration on a website before the person has access the data)

A1.1

No

Yes

3 Is the metadata publicly accessible?

No

Yes

...

4 Findable - Indexed in a discovery system

- No
- Local or internal system only
- Community wide or jurisdictional system
- Highly ranked in general purpose index (Google, Bing etc)

5 Please provide the URL to a search engine, and the query that will be executed to discover your RESOURCE ID

7 The degree to which the digital resource can be found using web-based search engines

11 Discoverability

- By personal contact only; Dataset information not discoverable.
- Limited dataset information, such as scientific description of the methodology, in the literature.
- Minimal catalogue-level metadata; Dataset searchable online.
- Complete set of collection-level discovery metadata + minimal granular metadata.
- Previous + available on an international catalogue, prominently displayed online and routinely updated.

12 Data Discoverability

- Information not published for public discovery; Internal or person-to-person sharing information exchange only
- Minimal product information published for public users; Product findable on local product website
- Product described with standards-based discovery metadata and published to discovery catalogs
- Previous + Metadata attributes included in HTML/other objects for indexing by web search engines (e.g., schema.org metadata); Product granules described with standards-
- Previous + Web services supporting product are described with standards-based rich metadata and published to discovery catalogs a searchable resource; Product relationshi