Results of an Analysis of Existing FAIR Assessment Tools

Authors: Christophe Bahim, Makx Dekkers, Brecht Wyns.

Abstract: This document is a first output of the FAIR Data Assessment Model WG. As a landscaping exercise, the editorial team of the WG analysed current and existing approaches related to FAIR self- assessment tools. The analysis was made based on publicly available documentation and an online survey. Questions and options stemming from theses different approaches were classified according to the FAIR principles/facets. Comments were collected and incorporated. This resulted in five slide decks, combined in this pdf document, that make up this preliminary analysis.

Results of an Analysis of Existing FAIR assessment tools

V0.03, created 9 January 2020, by the Editorial team

Background

This document was created as an initial product for the Research Data Allliance FAIR data maturity model working group. Comments were welcomed and collected through Github: https://github.com/RDA-FAIR/FAIR-data-maturity-model-WG/tree/master/results%20of%20preliminary%20analysis/v0.02

These comments were fed into the next phase of the development of a set of indicators.

Index

This document consists of slides listing the aspects broken down according to the dimensions:

- Findable
- Accessible
- Interoperable
- Reusable
- Beyond FAIR

FAIR Principles

–

Findable

Analysis of existing approaches
v0.03

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

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
- Globally Unique, citable and persisten (e.g. DOI, PURL, ARK or handle)

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

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 standard protocols
- 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 users)

> 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 prioritization in place.
> 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; Retrospective QC.

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?   *A2*
    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 explicity 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-based discovery metadata and published to discovery catalogs
Previous + Web services supporting product are described with standards-based rich metadata and published to discovery catalogs a searchable resource; -
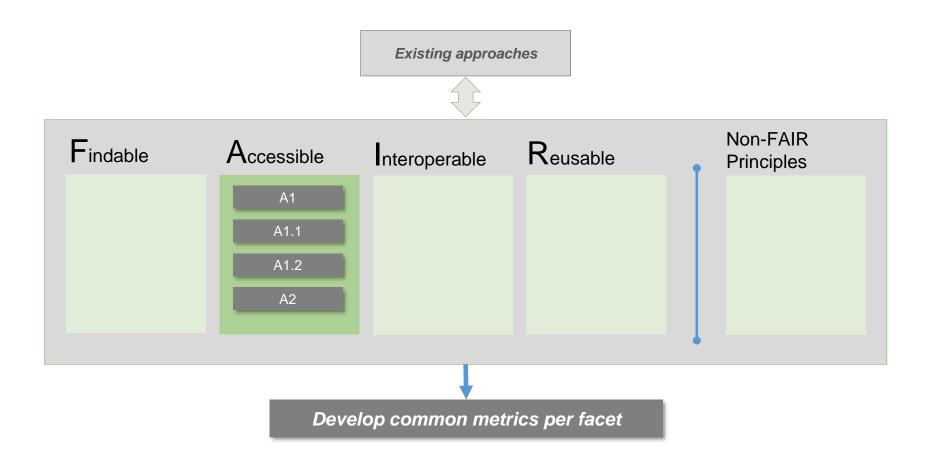 Product relationships (e.g., lineage or provenance) described with standards-based metadata and published to discovery catalogs

F**A**IR Principles

–

Accessible

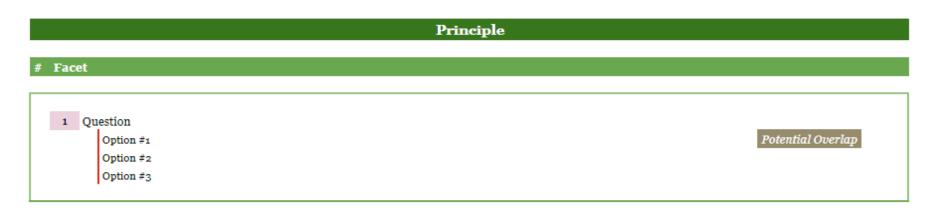*Analysis of existing approaches*
v0.03

FAIR Principles

To be accessible

A1. (meta)data is retrievable by their identifier using a standardised communications protocol

A1.1. the protocol is open, free, and universally implementable

A1.2. the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata is accessible, even when the data are no longer available

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, RDA FAIRsharing, 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

1 Is the data available online without requiring specialised protocols or tools once access has been approved?
No access to data
By individual arrengement
File download from online location
Non standard web service
Standard web service API

4 Published -Is the data accessible to users other than the creator or owner ?
No
By individual arrangement
File download
Institutional or community repository
Bespoke web service (informal API)
Bespoke web service (OpenAPI/Swagger)
Standard web service API (e.g. OGC)

7 The nature and use limitations of the access protocol

6 Accessibility
Not publicly available / Person-to-person
Publicly available / Direct file download (e.g., via anonymous FTP server) / Collection/dataset level searchable online
Non-standard data service Limited data server performance Granule/file level searchable Limited search metrics
Community-standard data service / Enhanced data server performance / Conforming to community search metrics / -
    Dissemination report metrics defined and implemented internally
Dissemination reports available online / Future technology and standard changes planned

11 Accessibility
Data not available publicly; Person-to-person contact needed.
Basic online services available for data access (e.g. FTP/HTTP direct download).
Non-standard data services.
Standard-based interoperability data services.
Previous + Full capability of sub-setting, aggregation and visualization.F7

A1.1 the protocol is open, free, and universally implementable

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)   *F4*
- No
- Yes

5 Please provide a URL to the description of the protocol.

5 Is the propotocol open (technical details are provided)?
- FALSE
- TRUE

5 Is the protocol free?
- FALSE
- TRUE

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

5 Authorization is required to access the content of my RESOURCE ID
- No
- Yes

5 Please provide a IRI that resolves to a description of the process to obtain access to restricted content

7 Specification of a protocol to access restricted content.

10 In case of a non legal restricted access, is the restriction properly justified by the researcher ?
- Never /NA
- If mandatory
- Sometimes
- Always

A2   metadata is accessible, even when the data are no longer available

1 Will the metadata record be available even if the data is no longer available?
- No
- Unsure
- Yes

5 Please provide the URL to a metadata longevity plan

7 The existence of metadata even in the absence/removal of data

FA**I**R Principles

–

Interoperable

Analysis of existing approaches
v0.03

Existing approaches

Findable	Accessible	Interoperable	Reusable	Non-FAIR Principles
		I1 / I2 / I3		

Develop common metrics per facet

FAIR Principles

To be interoperable :

I1. (meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation

I2. (meta)data uses vocabularies that follow FAIR principles

I3. (meta)data includes qualified references to other (meta)data

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, RDA FAIRsharing, 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

I1   (meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation

1 What (file) format(s) is the data available in?
> No access to data
> By individual arrengement
> File download from online location
> Non standard web service
> Standard Web Service API

1 What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?
> Data elements not described
> No standards have been applied in the description of data elements.
> Standardised vocabularies/ontologies/tagging schemas without global indentifiers
> Standardised open and universal using resolvable global identifiers linking to explanations

2 Is the data file in a proprietary format?
> No
> Yes

2 Are all of the data files in a proprietary format?
> No
> Yes

2 Please indicate which of these statements is the most applicable to the dataset:
> Most of the data files are proprietary
> Around half of the data files are proprietary
> Few of the data files are proprietary
> None of the data files are proprietary, they are all in a preferred format

3 Are the data stored and archived in preferred archival formats?   *R1.3*
> No
> Yes

5 Please provide the URL to the specification of the language

7 Use of a formal, accessible, shared, and broadly applicable language for knowledge representation.

8 Will you be using common ontologies?
- No
- Yes

9 metadata includes community accepted keywords and/or terms associated with relevant standards or terminologies
- No
- Somewhat
- Yes

10 Are standard vocabularies, thesaurus or ontologies used for all data types present in datasets, to enable interdisciplinary interoperability between well defined domains? If not, is a well-defined open data dictionary provided?
- Never /NA
- If mandatory
- Sometimes
- Always

10 Are the interoperability criteria explained?
- Never /NA
- If mandatory
- Sometimes
- Always

11 Data Portability
- Non-machine readable
- Basic machine readable
- Standards-based machine readable
- Machine independent, self-describing, interoperable format
- Previous + capability of providing user required format

I2   (meta)data uses vocabularies that follow FAIR principles

3 Did you use standardized vocabulary?
>   No
>   Yes

4 Comprehensible - supported with unambiguous definitions for all internal elements
>   Local field codes or labels
>   Labels with full text explanations
>   Community standard labels (e.g. CF Conventions, UCUM units)
>   Some fields linked to externally managed definitions
>   All fields linked to standard, externally managed definitions

5 Please provide one or more (max 3) IRIs representing the vocabularies used within the (meta)data that is returned by resolving the RESOURCE ID

7 The metadata values and qualified relations should themselves be FAIR

I3 (meta)data includes qualified references to other (meta)data

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

3 How is the metadata linked to other data and metadata (to enhance context and clearly indicate relationships)?
- There are no links to metadata
- The meta data records includes URI links to related metadata, data, definitions
- Metadata is represented in a machine readable format e.g. in a linked data format such as RDF

4 Linked - to other data and definitions using public identifiers (e.g URIs)
- No links
- In-bound links from a catalogue or landing-page
- Out-bound links to related data and definitions

5 Please provide the URL to a formal Linkset or copy/paste the content of a formal linkset that describes *R1.2*
at least a portion of the content at RESOURCE ID

7 Relationships within (meta)data, and between local and third-party data, have explicit and 'useful' semantic meaning

11 Usage
- No or weak citations in scientific publication in peer-review journal or as institutional reports.
- Intermediate citations + referenced in institutional climate assessment reports (e.g., by NOAA).
- Strong citations + referenced in national climate assessment reports (e.g., by USGCRP).
- Previous + referenced in international climate assessment reports (e.g., by IPCC).
- Previous + referenced in international decision/policy making published reports (e.g., by UNFCCC, UN-ISDR, World Bank, etc.).

FAI**R** Principles

–

Reusable

*Analysis of existing approaches
v0.03*

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

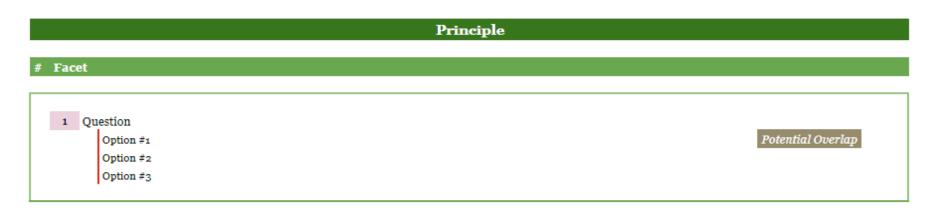|			
1	ANDS-NECTAR-RDS-FAIR data assessment tool	ARDC	Link to the methodology
2	DANS-Fairdat	DANS	Link to the methodology
3	DANS-Fair enough?	DANS	Link to the methodology
4	The CSIRO 5-star Data Rating tool	CSIRO	Link to the methodology
5	FAIR Metrics Questionnaire	The FAIR Metrics Group	Link to the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

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 long-term ? (e.g. versioning, archival and long term reuse issue for protocols, softwares, required methods and contexts to create, read and understand data)
  - 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, documentation, source code) / Data citation tracked (e.g., utilizing Digital Object Identifier (DOI) system)
   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

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 recongized authority in your community or domain, -
  indicating that the content of RESOURCE ID is compliant with the standards of your community

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 -
  recommended by community-approved or appropriate authorities specified? If no standards exist, has the researcher created a well described data dictionary
  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.
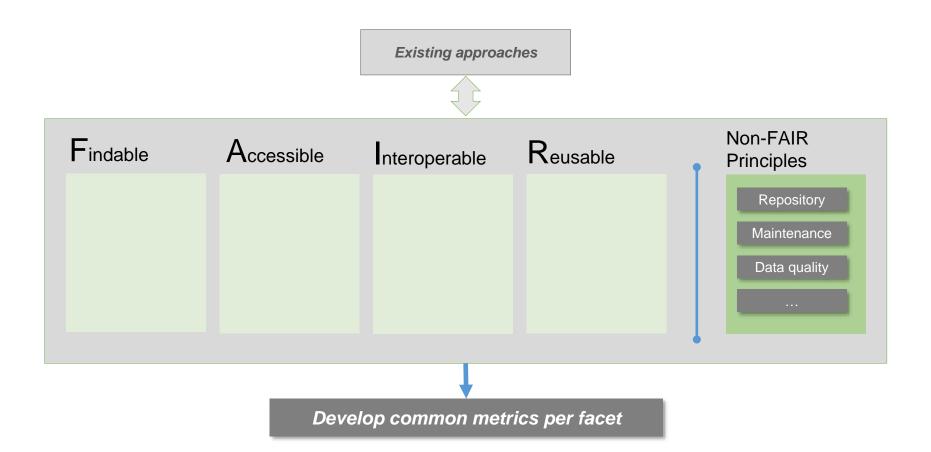Previous + complete granular-level metadata; Metadata QC-ed and Regularly updated

FAIR Principles

–

Beyond FAIR

Analysis of existing approaches v0.03

Existing approaches

Findable	Accessible	Interoperable	Reusable	Non-FAIR Principles

Non-FAIR Principles:
- Repository
- Maintenance
- Data quality
- …

Develop common metrics per facet

Non-FAIR Principles

Data repository

Curation and maintenance

Open data

Data quality

Others

www.rd-alliance.org -   @resdatall

LEGEND

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

Principle

Facet

1 Question
  Option #1
  Option #2
  Option #3

Potential Overlap

X1   Data repositiory

1 What type of repository or registry is the metadata record in ?
  The data is not described in any repository
  Local institutional repository
  Domain-specific repository
  Generalist public repository
  Data is in one place but discoverable through several registries

3 Is the data repository you have chosen trustworthy
  No
  Yes

9 Is dataset located within a CoreTrustSeal-certified repository ?
  No
  Yes

9 Is dataset located within a World Data System or Data Seal of Approval certified repository?
  No
  Yes

9 Repository representative stipulates that structure, harmonization, completeness, and correctness of the dataset comports -
  with typical data curation activities conducted by this (CoreTrustSeal certified) repository
  No
  Somewhat
  Yes

10 Does the researcher use data repositories for the storage of data ?
- Never /NA
- If mandatory
- Sometimes
- Always

6 Preservability
- Any storage location / Data only
- Non-designated repository / Redundancy / Limited archiving metadata
- Designated archive / Redundancy / Community-standard archiving metadata / Conforming to limited archiving standards
- Conforming to community archiving standards
- Archiving process / Performance controlled, measured and audited / Future archiving standard changes planned

11 Preservation
- Any storage location; Data only; Data not backed up.
- Non-designated repository; A backup copy of electronic data is made.
- Designated archive; Basic retention policy publicly defined. Routine backups made, including offsite copy.
- Previous + Conforming to community archiving standards. Comprehensive retention policy defined and implemented.
- Previous + Archiving process performance controlled, measured, and audited; Future archiving standard changes planned.

X2   Curation and maintenance

4 Updated - part of a regular data collection program or series, with clear maintenance arrangements and update schedule
one-time dataset
part of series - occasional/irregular update
part of series - regular scheduled updates

4 Curated - commitment to ensuring the data is available long term
Once-off dump, no ongoing commitment
Best effort, project website
Public or institutional repository (e.g. CKAN, GitHub)
Certified repository

11 Governance
Responsibility is not defined; No person is assigned.
Responsible entity is identified; Accountability and competency are not well-defined.
Responsibility/accountability and compliance mechanisms are defined; Good competency; Processes established conforming to community standards
Previous + Competency defined; Conforming to international standards; auditable.
Previous + Accountability and responsibility well defined and fully compliant with international standards; transparent; monitored and audited.

X3   Open data

3 FAIR enough, but also Open? Will your data be published as open as possible and as protected as necessary?
No
Yes

8 Will you be working with the philosophy 'as open as possible' for your data?
No
Yes

6 Data Quality Control / Monitoring

None or Sampling unknown or spotty / Analysis unknown or random in time

Sampling and analysis are regular in time and space / Limited productspecific metrics defined & implemented

Sampling and analysis are frequent and systematic but not automatic / Community metrics defined and partially implemented / Procedure documented and available online

Anomaly detection procedure well-documented and fully implemented using community metrics, automatic, tracked and reported / Limited quality monitoring metadata

6 Data Quality Assessment

Algorithm/method/model theoretical basis assessed (methods and results online)

Research product assessed (methods and results online)

Operational product assessed (methods and results online)

Quality metadata assessed / Limited quality assessment metadata

Assessment performed on a recurring basis Conforming to community quality metadata & standards External ranking

11 Quality Assurance & Control   *F2*

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 prioritization in place.

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; Retrospective QC.

11 Quality Assessment

Product quality assessment not done or done internally and information not available.

Assessed by Principal Investigator (PI) or data producer; Assessment results available online.

Previous + Product validation and evaluation done by PI published in peer-reviewed journal.

Previous + Independent product validation and evaluation published in peer-reviewed journal.

Previous + The complete product provenance is captured and publicly available.

X5   Others

8 Will data interpretation and modeling require significant compute infrastructure capacity?

No
Yes

8 Will you be doing (automated) knowledge discovery?

No
Yes

11 Uncertainty Analysis

Uncertainty estimates not available.
Uncertainty estimates presented without explanation.
Uncertainty estimates presented with partial explanation.
Full uncertainty budget available with all assumptions; Estimates of accuracy of trend available.
Full uncertainty assessment published in peer reviewed journal.