

RDA-SHARC fairness assessment tools for crediting/rewarding scientists data sharing activities

CONTEXT: the **rda-SHARing Reward & Credit ig**, *Corresponding authors : R. David, L. Mabile, A. Cambon-Thomsen*

What for? to foster data sharing by improving recognition of the work required

How? by providing a set of recommendations to guide researchers and other relevant stakeholders (research institutions administrators, funders, policy makers and publishers/editors) in moving through the necessary steps towards crediting and rewarding in the data/resources- sharing process (in progress); and to encourage the adoption of data sharing activities- related criteria in the research evaluation process at the institutional, national and European/international levels.

As part of it, **3 human readable assessment tools are under development** that will assess semi-quantitatively the fairness knowledge & practices of scientists:

1.1 extensive FAIRness external assessment grid

52 criteria so far

1.2 simplified FAIRness external assessment grid (can be used as a quick self-assessment grid)

18 essential criteria

<https://zenodo.org/record/2551500#.XGK4llxKg2w>

2.2 extensive checklist for fairness self-assessment (adapted from the 2 previous grids)

FINDABLE (8 essential criteria)

Indexed identifier ?
 Identification
 Never/NA If Mandatory Sometimes Always
Are each dataset identified by an indexed and by standard identifier ?

Unique, global, persistent ID ?
 Identification
 Never/NA If Mandatory Sometimes Always
Are the data identifiers unique, global and persistent ?

ID scheme ?
 Identification
 Never/NA If Mandatory Sometimes Always
Are any identifier schemes been used for data (e.g. DOI) ?

Persistent metadata / data link ?
 Metadata availability
 Never/NA If Mandatory Sometimes Always
Are the metadata linked to the dataset through a persistent identifier ?

Metadata & authority linked ?
 Metadata availability
 Never/NA If Mandatory Sometimes Always
Are the metadata of each dataset linked to unique authority (responsible for the dataset) at given time ?

Datasets linked to authority ?
 Metadata availability
 Never/NA If Mandatory Sometimes Always
Are all datasets linked to an authority (legal entity) through a unique and persistent identifier over time (e.g. institution, association or established body) ?

Standards/dictionary for data description?
 Metadata description and accessibility
 Never/NA If Mandatory Sometimes Always
If relevant, are the metadata used and/or additional standards for data storage ? If so, are the data standards and particularly concerning data links, implemented in a machine-readable or machine-actionable standard ? If so, do they work, how the researcher could read/develop data links ?

Data format/type description?
 Metadata description and accessibility
 Never/NA If Mandatory Sometimes Always
Are the types and format of data provided (collected and described) ?

Result for Findable: Never/NA If Mandatory Sometimes Always

2) ACCESSIBLE (3 essential criteria)

Data repositories?
 Repository
 Never/NA If Mandatory Sometimes Always
Does the researcher use repositories for the storage of data ?

Efficient and rich services for various uses & users?
 Data security and services
 Never/NA If Mandatory Sometimes Always
Does the researcher use efficient and rich services to access data (service formats, visualizations, analytical tools and systems adapted to different types of use and users) ?

Data access restriction justification?
 Access restriction
 Never/NA If Mandatory Sometimes Always
In case of a non legal restricted access, is the restriction properly justified by the researcher ?

Result for Accessible: Never/NA If Mandatory Sometimes Always

3) INTEROPERABLE (2 essential criteria)

Standard vocabularies, thesaurus, ontologies or data dictionary?
 Identification
 Never/NA If Mandatory Sometimes Always
Are the researcher vocabularies or ontologies used for data type, provenance, datasets, or in data interoperability between well defined datasets ? If so, is such standard used data intercomparability ?

Interoperability criteria explained?
 Identification
 Never/NA If Mandatory Sometimes Always
Are the researcher fully criteria explained ?

Result for Interoperability: Never/NA If Mandatory Sometimes Always

4) REUSABLE (5 essential criteria)

Relevant actions for data reuse potential?
 Data potential
 Never/NA If Mandatory Sometimes Always
Which relevant actions have been undertaken by the researcher to enhance the data reuse potential ?

Provenance for raw and transformed data?
 Data availability
 Never/NA If Mandatory Sometimes Always
Are the researcher use type of all data provenance specified (type of raw primary, secondary, tertiary) ?

Information on methods and tools that permit the understanding, integrity of data?
 Availability tools
 Never/NA If Mandatory Sometimes Always
Does the researcher provide information on methods and tools that permit the understandability, integrity, trace and availability of data transferred to the user ? (e.g. recording, archival and long term reuse tools for protocols, software, required metadata and content to create, read and understand data)

Data sharing arrangements meet data ethics and protection?
 Availability rights
 Never/NA If Mandatory Sometimes Always
Do the data reuse control and data sharing arrangements meet the data producer and /or traditional ethics requirements ?

Legal reuse restriction properly justified?
 Availability rights
 Never/NA If Mandatory Sometimes Always
In case of a legal reuse restriction such as contractual data use and/or privacy limits, national or international data availability, information systems security, security in national and commercial context, is the restriction properly justified ?

Result for Reusable: Never/NA If Mandatory Sometimes Always

TOTAL FAIR simple criteria evaluation results:
 .../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'
.../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'
 *Indices will be provided according to the criteria predominantly obtained

Fairness assessment grids

Objectives : credit & reward for FAIRness in researchers sharing behaviors
 -> necessity to improve FAIRness (understandable and step by step processes)

Main properties:

- As simple as possible (understandable by non IT people)
- Easy to complete (due to FAIR skills availability in evaluation processes)
- Based on informations given by researchers in careers doc / activity reports
- CC author license (can be reused by anyone at the end of the implementation)

Assessment process: leading to recommendations to improve fairness

- Designed as a decision tree in each FAIR Principle
- 3 Level of criterion importance : **essential / recommended / desirable**
- 4 possible answers/criteria:
 - Never/NA If Mandatory Sometimes Always
- Evaluation based on scoring each answer for each F.A.I.R. principle
 ex: **Findable 2/8** Never/NA; **3/8** If Mandatory; **1/8** Sometimes; **2/8** Always
- Recommendations based on this scoring

Fairness assessment grids

1) FINDABLE (3 essential criteria)	
Indexed identifier ? Identification Are you assigned a identifier for an individual user/technical identifier ?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Unique, global, persistent ID? Identification Are the data identifier unique, global and persistent ?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
ID scheme? Identification Has any identifier scheme been used for data (e.g. DOI)?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Persistent metadata / data link ? Metadata traceability Are the metadata linked to the dataset through a persistent identifier ?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Metadata & authority linked ? Metadata traceability Are the metadata of each dataset linked to a person/authority responsible for the dataset at a given point?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Datasets linked to authority ? Metadata traceability Are datasets linked to an authority (legal entity) through a unique persistent identifier (e.g. metadata, accession or institutional code)?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Standard/dictionary for data description? Metadata description and searchability If present, has the researcher used and applied standard(s) for data description? If not, are the data described in a way that is understandable by users who are not an expert in the discipline? If so, are the standard(s) used, has the researcher created a self-describing data dictionary?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Data format/type description? Metadata description and searchability Are the type and format of data provided? collected well described?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Result for Findable: <input type="checkbox"/> Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	

2) ACCESSIBLE (3 essential criteria)	
Data repositories? Repository Does the researcher use data repositories for the storage of data?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Efficient and rich services for various uses & users? Data security and services Does the researcher use efficient and rich services to access data (online format, visualization, analytical tools and systems related to different types of use and users)?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Data access restriction justification? Access restriction In case of a non-legal metadata access, is the restriction properly justified by the researcher?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Result for Accessible: <input type="checkbox"/> Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	

3) INTEROPERABLE (2 essential criteria)	
Standard vocabularies, thesaurus, ontologies or data dictionary? Identification Are standard vocabularies, thesaurus or ontologies used for all data types present? If not, is a well-defined open data dictionary provided?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Interoperability criteria explained? Identification Are the interoperability criteria explained?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Result for Interoperability: <input type="checkbox"/> Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	

4) REUSABLE (5 essential criteria)	
Relevant actions for data reuse potential? Data potential What relevant actions have been undertaken by the researcher to enhance the data reuse potential?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Provenance for raw and transformed data? Data traceability Are the provenance and type of all data properly specified (stage of use, primary, transformed, secondary, etc.)	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Information on methods and tools that permit the understanding, integrity of data? Reusability tools Does the researcher provide information on methods and tools that permit the understanding, integrity, value and reusability of data intended to be kept in the long-term? (e.g., metadata, workflow and/or long-term data care for protocols, software, research methods and concepts to code, metadata, experimental data)	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Data sharing arrangements meet data ethics and protection? Reusability type Do the data reuse control and data sharing arrangements meet the data protection and confidentiality data requirements?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Legal reuse restriction properly justified? Reusability right In case of a legal reuse restriction (such as personal data, data and/or health records, national defence secret, confidentiality of administrative information, intellectual property, security or non-legal and commercial reasons), is the restriction properly justified?	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
Result for Reusable: <input type="checkbox"/> Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	

TOTAL FAIR simple criteria evaluation results:
.../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'
*figures will be provided according to the criteria predominantly obtained

Lessons learnt from the first tests:

- Essential criteria not always understandable without training
- Implementation of some criteria can be time consuming / need technical advisor / operator

Possible open issues:

- Develop gradual assessment of the *researcher FAIRness Literacy*
- Help identify needs to build FAIRness guidelines for a better researcher sharing capacity
(based on rewards and credits / *How to do and step by step tools*)

Next steps:

- **Upcoming SHARC-survey launch to evaluate the external assessment extensive grid usability:** please participate!
- RDA P13 Sharc's session: please attend!
- Tools experimentation in specific networks (IMI FAIRplus; BiodiFAIRse; Citizen science networks...)