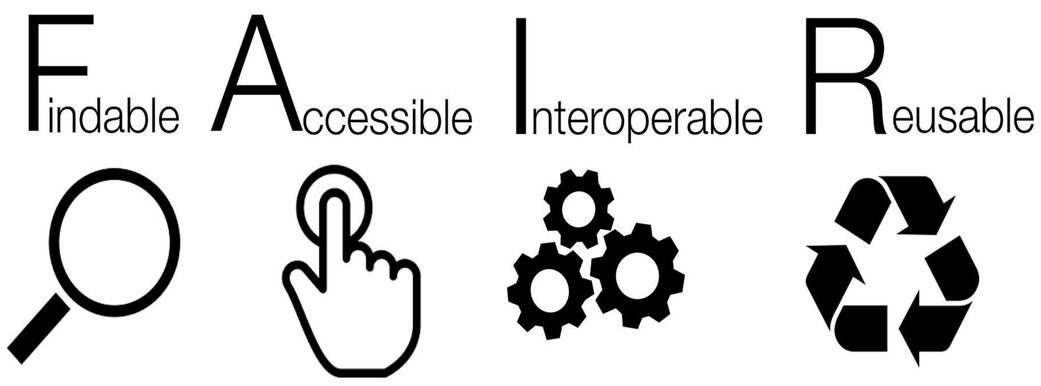
How to operationalize and to evaluate the FAIRness in the crediting and rewarding processes in data sharing: a first step towards a simplified assessment grid.

Romain David, Laurence Mabile, Mohamed Yahia, Anne Cambon-Thomsen, Anne-Sophie Archambeau, Louise Bezuidenhout, Sofie Bekaert, Gabrielle Bertier, Elena Bravo, Jane Carpenter, Anna Cohen-Nabeiro, Aurélie Delavaud, Michele De Rosa, Laurent Dollé, Florencia Grattarola, Fiona Murphy, Sophie Pamerlon, Alison Specht, Anne-Marie Tassé, Mogens Thomsen, Martina Zilioli, and the RDA-SHARC Interest Group.

Contacts: Romain David romain.david@imbe.fr Laurence Mabile: laurence.mabile@univ-tlse3.fr Anne Cambon-Thomsen: anne.cambon-thomsen@univ-tlse3.fr

DATA SHARING EVALUATION TO TRIGGER CREDITING/REWARDING PROCESSES

In order to foster data sharing, the RDA-SHARC (SHAring Rewards & Credit) interest group has been set up to unpack and improve crediting and rewarding mechanisms in the data/resources sharing process. As part of the objectives, **two assessment grids** are being developed using criteria to establish if data are compliant to the F.A.I.R principles (findable /accessible / interoperable / reusable). The criteria used are based on the work from **FORCE 11***, and on the basis of the Open Science Career Assessment Matrix designed by the **EC Working group on Rewards under Open science.**



Par SangyaPundir — Travail personnel, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=53414062

BUILDING FAIR- BASED ASSESSMENT GRIDS

To be generic and trans-disciplinary, assessment grids should be understandable by all scientist including the ones who are not expert in data science.

The two grids displayed as a tree-graph structure are based on previous works on FAIR data management (Reymonet *et al.*, 2018; Wilkinson *et al.*, 2016; Wilkinson *et al.*, 2018; and E.U.Guidelines about FAIRness DMPs):

1/ the self-assessment grid is conceived as a checklist for scientists to identify if her/his own activities are compliant to FAIR principles and to pinpoint the hurdles that hinder efficient sharing and reuse of data

2/ the two-level grid (simplified / extensive) is conceived as a chart for the evaluator to assess the quality of the researcher/scientist sharing practice, over a given period, taking into account the means & support available over that period. Assessment criteria are classified according to their level of stringency for FAIRness (essential / recommended / desirable).

First draft of the simplified FAIR criteria assessment grid

The aim of the simplified assessment grid is to focus on essential criteria only and to be completed by scientists who produce data. It is the summary of a more extensive grid designed for assessing optimal sharing of data (not yet possible at the moment for most scientists worldwide). The assessment is based on FAIR criteria compliance.

This grid can be used to get a first appreciation of the researcher's practice but cannot be used alone for a comprehensive assessment of the FAIRness of data sharing. Motivations related-criteria help to interpret further the results highlighted as good practices.

Motivations for Sharing (4 essential criteria)

Mandatory criteria - ☐ If non restricted access, are all datasets shared?
- ☐ Has any long term preservation strategy planned (e.g. in a long term archive)?
- ☐ Which motivations are declared by the researcher ?

☐ **To this aim, has established D.M.P.s been used?** If so, what tools/templates has been used?

Optional criteria - ☐ Any specific training followed? If so, what is the name of the programme ?
- ☐ If relevant, any use of open community software platform? If so, name of the platform?
- ☐ If relevant, any software management plans (S.M.P.s)? If so, any tool/template used?

INPUT NEEDED FROM RESEARCH COMMUNITIES

To implement a highly fair appraisal of the sharing process, appropriate criteria must be selected in order to design optimal generic assessment grids. This process **requires participation, time and input from volunteer data producers/users scientists in various fields**. The aim is to get feedback from a larger community as to the validity of the criteria over different fields. The assessment grids **will circulate in the RDA community as an online questionnaire as soon as possible**.

Are you producing or using data? **Please participate** in the development of the FAIRness assessment grids by completing the questionnaire when available.

It will help you get credit back for your efforts!

HOW?

Join the SHARC RDA community (free) at https://www.rd-alliance.org/get-involved.html and there join the SHARC interest group at https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig

You will then be informed in real time.

Reymonet N et al. Réaliser un plan de gestion de données « FAIR » : modèle, 2018. ⟨sic_01690547v2⟩
Wilkinson MD et al. (2018). A design framework and exemplar metrics for FAIRness. Scientific data, 5, 180118. doi:10.1038/sdata.2018.118
Wilkinson MD, The FAIR Guiding Principles for scientific data management and stewardship.Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18.
E.U. European Commission Directorate-General for Research and Innovation report: Evaluation of Research Careers fully acknowledging Open Science Practices; Rewards, incentives and/or recognition for researchers practicing Open Science. 2017

1) FINDABLE (8 essential criteria)

Indexed identifier ?
Identification
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Are each data/dataset identified by an indexed and independant 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
Has any identifying schema been used for data (e.g. DOI)?

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

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

Datasets linked to authority ?
Metadata traceability
☐ 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)?

Data description standards?
Metadata description and searchability
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
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?

Data format/type description?
Metadata description and searchability
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Are the types and formats of data generated / collected well described?

Result for Findable: .../8 Never/NA .../8 If Mandatory .../8 Sometimes .../8 Always

2) ACCESSIBLE (3 essential criteria)

Data repositories?
Repository
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Does the researcher use data 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 (various formats, visualisations, practical 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: .../3 Never/NA .../3 If Mandatory .../3 Sometimes .../3 Always

3) INTEROPERABLE (2 essential criteria)

Standard vocabularies, thesaurus, ontologies or data dictionary?
Identification
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
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?

Interoperability criteria explained?
Identification
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Are the interoperability criteria explained?

Result for Interoperability: .../2 Never/NA .../2 If Mandatory .../2 Sometimes .../2 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 row and transformed data?
Data traceability
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Are the provenance and type of all data properly specified (origin of raw, primary, transformed, secondary..)?

Information on methods and tools that permit the understanding, integrity of data?
Reusability tools
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
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)

Data sharing arrangements meet data ethics and protection?
Reusability right
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Do the data reuse control and data sharing arrangements meet the data protection and "local/national ethics requirements?

Legal reuse restriction properly justified?
Reusability right
☐ Never/NA ☐ If Mandatory ☐ Sometimes ☐ Always
Do the data reuse control and data sharing arrangements meet the data protection and "local/national ethics requirements?

Result for Reusable: .../5 Never/NA .../5 If Mandatory .../5 Sometimes .../5 Always

TOTAL FAIR simple criteria evaluation results:
.../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'

*advices will be provided according to the criteria predominantly obtained

Author affiliations : Romain DAVID (Aix-Marseille Université, CNRS, IRD, UAPV, Institut Méditerranéen de Biodiversité et d'Ecologie Marine et Continentale), romain.david@imbe.fr; Laurence Mabile, INSERM-Université Paul Sabatier Toulouse III, FR; Mohamed Yahia, INIST-CNRS, Nancy, FR; Anne Cambon-Thomsen, INSERM-Université Paul Sabatier Toulouse III, FR; Anne-Sophie Archambeau, GBIF-UMS PatriNat, Paris, FR ; Louise Bezuidenhout, University of Oxford, UK; Sofie Bekaert, Gent University, BE; Gabrielle Bertier, McGill University, Montréal, CA - Université Paul Sabatier Toulouse III, FR; Elena Bravo, Istituto Superiore di Sanità, Roma, IT; Jane Carpenter, University of Sydney, AUS ; Anna Cohen-Nabeiro, Independant, Paris, FR ; Aurélie Delavaud, Fondation pour la Recherche sur la Biodiversité, Paris FR ; Michele De Rosa, BONSAI, DEN; Laurent Dollé, Biothèque Wallonie-Bruxelles, BE; Florencia Grattarola, University of Lincoln, UK;Fiona Murphy, Murphy Mitchell Consulting Ltd, UK; Sophie Pamerlon, GBIF-UMS PatriNat, Paris, FR ; Alison Specht, Fondation pour la Recherche sur la Biodiversité, Paris, FR; Anne-Marie Tassé, P3G, Montreal, CA; Mogens Thomsen, INSERM-Université Paul Sabatier Toulouse III, FR; Martina Zilioli, CNR-IREA, Milan, IT.

JNSO 2018 Premières Journées Nationales de la Science Ouverte — De la stratégie à l'action — 4-6 décembre 2018 Paris, France — OUVRIR LA SCIENCE !