How to achieve FAIRER research data by studying evaluation & assessment protocols

Teresa Gomez-Diaz [*], Tomas Recio [**]

[*] CNRS - Laboratoire d'informatique Gaspard-Monge (Est of Paris)

[**] Universidad Antonio de Nebrija (Madrid)

Poster Session

Open Science FAIR, Madrid, 25-27 september 2023

The FAIR Guiding Principles

You know them by heart...

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

To be Accessible

- A1 (meta)data are retrievable by their identifier using a standardized 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 are accessible, even when the data are no longer available

To be Interoperable

- I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2 (meta)data use vocabularies that follow FAIR principles
- I3 (meta)data include qualified references to other (meta)data

To be Reusable

- R1 meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1 (meta)data are released with a clear and accessible data usage license
- R1.2 (meta)data are associated with detailed provenance
- R1.3 (meta)data meet domain-relevant community standards

(2016) M. Wilkinson, et al. The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data

CDUR Protocol(s) for RD (and RS) evaluation

CDUR has four steps:

You *will* know them by heart...

(C) Citation
to measure if the Research Data (RD) is well identified as a research output : good citation form, but also metadata, best citation practices...

Legal point. Authors (if $\exists$ copyright), producers, affiliations, participation %

(D) Dissemination
best dissemination practices, in agreement with the scientific policy of the evaluation context

Policy point. Open Science. ★ **Legal point.** other legal issues, licenses.

★ **(U) Use**
to evaluate the "data" aspects of the RD: quality, documentation, tutorials, use examples etc., facilitate reuse, best data practices...

Reproducibility point. validation of scientific results

(R) Research
to evaluate the "research aspects" of the RD: quality of the scientific work, proposed and coded algorithms & data structures, related publications, collaborations...

Research point. impact of the RD research related work

(2019) TGD, T. Recio. On the evaluation of research software: the CDUR procedure

(2022) TGD, T. Recio. Research Software vs Research Data II: Protocols for Research Data dissemination and evaluation...

How to achieve FAIRER research data...

Put them together....

CDUR [2,7]	**The FAIR Guiding Principles** [1]
(C) Citation: The **RD** is well identified, involving issues concerning: - citation form or reference - metadata (including PIDs)	**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 **To be Interoperable:** I3. (meta)data include qualified references to other (meta)data **To be Reusable:** R1.2. (meta)data are associated with detailed provenance
(D) Dissemination: The **RD** is well disseminated, involving issues concerning: - list of included components - RD licence - RD deposit	**To be Findable:** F4. (meta)data are registered or indexed in a searchable resource **To be Accessible:** A1. (meta)data are retrievable by their identifier using a standardized 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 are accessible, even when the data are no longer available **To be Reusable:** R1.1. (meta)data are released with a clear and accessible data usage license
(U) Use: The **RD** facilitates its reuse, involving issues like: - documentation, tutorials, examples... - reproducibility and replicability issues	**To be Interoperable:** I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation I2. (meta)data use vocabularies that follow FAIR principles **To be Reusable:** R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.3. (meta)data meet domain-relevant community standards
(R) Research: Measures the impact of the **RD** related scientific work	Not applicable

Table 1. Relationships between the FAIR principles and the CDUR RD evaluation protocols.

(2022) TGD, T. Recio. Research Software vs Research Data II: Protocols for Research Data dissemination and evaluation...

(2023) TGD, T. Recio. How to achieve FAIRER research data by studying evaluation assessment protocols, Open Science FAIR

CDUR Protocol(s) for RD (and RS) evaluation

The CDUR protocols (2019) were initially conceived for Research Software (RS), they have been extended to Research Data (RD) in 2022.

There are designed to help evaluated researchers, evaluation committees and decision makers.

There are four steps :

(C) Citation

(D) Dissemination

(U) Use

(R) Research

They are flexible enough to be adapted to different evaluation situations.

Each evaluation committee sets its own protocol adapted to the evaluation context: recruitment, career evolution, publication...

If followed correctly, CDUR may clearly contribute towards FAIRER RD (and RS).

(2019) TGD, T. Recio. On the evaluation of research software: the CDUR procedure

(2022) TGD, T. Recio. Research Software vs Research Data II: Protocols for Research Data dissemination and evaluation...

A related protocol in the context of FAIR/Nanomaterials, see TRAAC 2021, TRAAC 2023 (2).

RD and RS definitions

CDUR translation from RS to RD is possible because we have similar formulations for the RS and RD definitions:

Research software (2019) *is a well identified set of code*
that has been written by a (again, well identified) research team.
It is software that has been built and used to produce a result
published or disseminated in some article or scientific contribution...

Research data (2022) *is a well identified set of data that has been produced*
(collected, processed, analyzed, shared & disseminated)
by a (again, well identified) research team.
The data has been collected, processed and analyzed to produce a result
published or disseminated in some article or scientific contribution...

(2019) TGD, T. Recio. On the evaluation of research software: the CDUR procedure

(2022) TGD, T. Recio. Research Software vs Research Data I: Towards a Research Data definition in the Open Science context

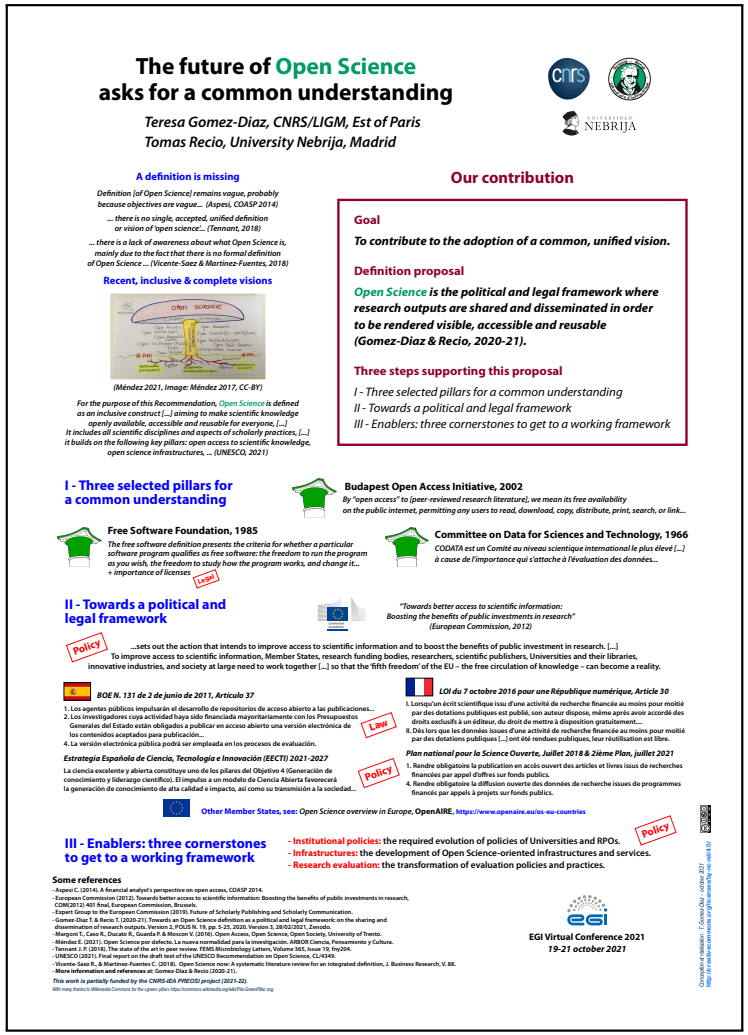
(2022) TGD, T. Recio. Research Software vs Research Data II: Protocols for Research Data dissemination and evaluation...

The Open Science context

The future of Open Science asks for a common understanding

Open Science definition

Open Science is
the political and legal framework
where research outputs are shared
and disseminated in order to be rendered
visible, accessible and reusable.



EGI Virtual Conference 2021, Lisbon, 19-21 october 2021
https://padlet.com/gwenfranck/EGI2021Posters

(2020-21) TGD, T. Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs

References

2016 M. Wilkinson, M. Dumontier, I. Aalbersberg, et al. The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3:160018, https://www.nature.com/articles/sdata201618

2019 TGD, T. Recio. On the evaluation of research software: the CDUR procedure, F1000Research, Research on Research, https://doi.org/10.12688/f1000research.19994.2

2020-1 TGD, T. Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs, **POLIS N. 19, 2020**, https://uet.edu.al/polis/wp-content/uploads/2022/01/polis-19.pdf, **V3 du 28/02/2021**, https://zenodo.org/record/4577066

2021 TGD, T. Recio. The future of Open Science asks for a common understanding, Poster EGI Virtual Conference 2021, https://indico.egi.eu/event/5464/contributions/15729/

2021 Shandilya, N., Barreau, M. S., Suarez-Merino B., et al. TRAAC framework for regulatory acceptance and wider usability of tools and methods for safe innovation and sustainability of manufactured nanomaterials, https://web.archive.org/web/20220201053706/https://assets.researchsquare.com/files/rs-1158958/v1/ebcd0186-d77f-4178-a2f9-ac283a25005e.pdf

2022 European Commission, Directorate-General for Research and Innovation. European Research Data Landscape - Final report, https://doi.org/10.12688/f1000research.78459.2

2022 TGD, T. Recio. Research Software vs Research Data I: Towards a Research Data definition in the Open Science context, F1000Research, Research on Research, https://f1000research.com/articles/11-118/v2

2022 TGD, T. Recio. Research Software vs Research Data II: Protocols for Research Data dissemination and evaluation in the Open Science context, F1000Research, Research on Research, https://f1000research.com/articles/11-117/v2

2023 P. Schmidt. Podcast Code for Thought: Open Data, Open Software - with Teresa Gomez-Diaz, https://codeforthought.buzzsprout.com/1326658/13216530-en-open-data-open-software-with-teresa-gomez-diaz

2023 TGD, T. Recio. Podcast Code for Thought: Research Software and Research Data in Open Science, Zenodo preprint, https://zenodo.org/record/8159906

2023 TGD, T. Recio. How to achieve FAIRER research data by studying evaluation assessment protocols, Conference Open Science FAIR 2023, http://opensciencefair.eu/2023/posters

2023 Shandilya, N., Barreau, M. S., Suarez-Merino B., et al. TRAAC framework to improve regulatory acceptance and wider usability of tools and methods for safe innovation and sustainability of manufactured nanomaterials, NanoImpact, Volume 30, April 2023, https://www.sciencedirect.com/science/article/pii/S2452074823000125

2023 Shandilya, N., Barreau, M. S., Suarez-Merino B., et al. 81 Traac Framework and tool for Regulatory Acceptance and Wider Usability of Tools/Methods for safe Innovation and Sustainability of Nanomaterials, Annals of Work Exposures and Health, Volume 67, May 2023, https://doi.org/10.1093/annweh/wxac087.201