

How to achieve FAIRER research data by studying evaluation assessment protocols

Teresa Gomez-Diaz, teresa.gomez-diaz@univ-eiffel.fr
Université Gustave Eiffel-LIGM-CNRS, Est of Paris
<https://siteigm.univ-mlv.fr/>

Tomas Recio, trecio@nebrija.es
Universidad Antonio de Nebrija, Madrid
<https://www.nebrija.com/>

Goal: to study the relationships between the CDUR Protocols for *research data* evaluation proposed in the *Open Science* context and the *FAIR* data principles.

Keywords: *research data, evaluation protocols, FAIR principles, research outputs, research assessment.*

Definitions

Open Science (OS) [3] is the political and legal framework where *research outputs* are shared and disseminated in order to be rendered visible, accessible and reusable.

Research data (RD) [6,7] is a well identified set of *data* that has been produced (collected, processed, analyzed, shared & disseminated) by a research team. The *data* has been collected, processed and analyzed to produce a result published or disseminated in some article or scientific contribution. Each **RD** encloses a set (of files) that contains the dataset maybe organized as a database, and it can also include other elements as the docs, specifications, use cases, and other useful material as provenance information... It can include the **Research Software (RS)** [2,4] that has been developed to manipulate the dataset (from short scripts to **RS** of larger size) or give references to the software that is necessary to manipulate the data (**RS** or other).

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

The CDUR protocol was initially conceived for *Research Software (RS)* [2], it was extended to *RD* in 2022 [7]. It was designed to help evaluated researchers, evaluation committees and decision makers. It has four steps: (C) Citation, (D) Dissemination, (U) Use, (R) Research [2,7]:

- (C) Citation:** to measure if the **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:** “data” aspects of the **RD**: quality, documentation, tutorials, use examples etc., facilitate reuse, best data practices...
Reproducibility point: validation of scientific results
- (R) Research:** “research aspects”: quality of the scientific work, proposed and coded algorithms & data structures, related publications, collaborations...
Research point: impact

CDUR [2,7]	The FAIR Guiding Principles [1]
(C) Citation: The RD is well identified, involving issues concerning: <ul style="list-style-type: none">- 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: <ul style="list-style-type: none">- 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: <ul style="list-style-type: none">- 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

The FAIR Guiding Principles [1]
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

Table 1. This table illustrates the relationships between the FAIR principles [1] and the CDUR RD evaluation protocol [7].

References

[1] (2016) M. Wilkinson, M. Dumontier, I. Aalbersberg, et al. The FAIR Guiding Principles for scientific data management and stewardship, Sci Data, 3:160018

[2] (2019) T. Gomez-Diaz, T. Recio. On the evaluation of research software: the CDUR procedure, F1000Research 2019, 8:1353

[3] (2020-21) T. Gomez-Diaz, T. Recio. Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs, POLIS N.19 V3 dated 28 february 2021 available on Zenodo

[4] (2021) T. Gomez-Diaz, T. Recio. Open comments on the Task Force SIRS report: Scholarly Infrastructures for Research Software (EOSC Executive Board, EOSCArchitecture), Research Ideas and Outcomes (RIO), 7: e63872

[5] (2022) European Commission. Directorate-General for Research and Innovation, European Research Data Landscape: final report, Publications Office of the European Union

[6] (2022) T. Gomez-Diaz, T. Recio. Research Software vs. Research Data I: Towards a Research Data definition in the Open Science context (Definition), F1000Research 2022, 11:118

[7] (2022) T. Gomez-Diaz, T. Recio. Research Software vs. Research Data II: Protocols for Research Data dissemination and evaluation in the Open Science context (Dissemination, Evaluation/CDUR, FAIR), F1000Research 2022, 11:117



► OPEN SCIENCE ► FAIR ►

CHARTING THE COURSE:
REIMAGINING OPEN SCIENCE FOR NEXT GENERATIONS

25-27 SEPTEMBER 2023, MADRID, SPAIN

25-26 September - Reina Sofia Museum
27 September - El Ateneo de Madrid