

▶ OPEN SCIENCE ▶ FAIR ▶

Lightning talk Submission

Janete Saldanha Bachⁱ, Brigitte Mathiakⁱ, Claus-Peter Klasⁱ, Yudong Zhangⁱ and Peter Mutschkeⁱ.

Enhancing FAIR Compliance in Research Data Infrastructures: Insights from Applications of the RDA FAIR Data Maturity Model and the F-UJI Automated FAIR Data Assessment Tool

Abstract

1. Introduction

The FAIR Data Principles [1] are widely applied to research data. These principles have been broadly adopted by scientific and scholarly institutions to guide research data infrastructures and services, ensuring data is findable, accessible, interoperable, and reusable. However, due to their interpretation scope, it is still challenging to assess the extent to which a data infrastructure addresses the FAIR principles. The Research Data Alliance (RDA) set up the FAIR Data Maturity Model Working Group to specify the required indicators for institutions to assess their levels of FAIR compliance, producing the FAIR Data Maturity Model (RDA-FDMM) [2]. The RDA-FDMM defines 41 FAIR indicators, organized into three classes (*Essential*, *Important*, and *Useful*), and five levels. We applied the RDA-FDMM to the PID service [3] of KonsortSWDⁱⁱ, which aims to assign PIDs to data elements below study level (such as survey variables). Furthermore, we discuss automatic assessment using the F-UJI Toolⁱⁱⁱ, which employs RDA-FDMM and FAIRsFAIR Metrics [4] in a machine-readable fashion.

2. Methodology: manual and automatic assessments

At the PID service, which is based on the data registration agency da|ra [5], we manually assessed some elements at the PID service level and others at the da|ra level, using the pass-or-fail method ('yes' or 'no') questions. For the automatic assessment we adopted the F-UJI tool, a web service to automatically assess FAIRness of research data objects based on FAIR object assessment metrics [6]. We applied the F-UJI tool to GESIS Search^{iv} in the context of KonsortSWD, motivated by the European landscape study [7], which also relies on F-UJI tool and led us to improve our metadata [8].

3. Results

3.1 The manual assessment

The manual assessment results show that the PID service meets all the indicators classified as *essential* [9] and most of the indicators from the classes *important* and *useful* (see Table 1).

ⁱ GESIS - Leibniz-Institute for the Social Sciences.

ⁱⁱ KonsortSWD (Consortium for the Social, Behavioural, Educational and Economic Sciences) is funded by the National Research Data Infrastructure Germany (NFDI) <https://www.konsortswd.de/>

ⁱⁱⁱ <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

^{iv} <https://search.gesis.org/>

Table 1. PID and da|ra service assessment results: levels' distribution.

| Framework | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---------------------|-------------|-------------|------------|------------|------------|
| <i>Essential</i> | 20 / 20 | 20 / 20 | 20 / 20 | 20 / 20 | 20 / 20 |
| <i>Important</i> | | 7 / 7 | 10 / 14 | 10 / 14 | 10 / 14 |
| <i>Useful</i> | | | | 3 / 3 | 3 / 7 |
| Achieved indicators | 20/20 | 27 / 27 | 30 / 34 | 33 / 37 | 33 / 41 |
| Scored | 20 | 27 | 30 | 33 | 33 |
| Results | 100% | 100% | 88% | 89% | 80% |

The results demonstrate outstanding achievements:

- fully comply at levels 1 and 2;
- at level 3 the service achieved 88% compliance;
- at level 4 the service achieved 89% compliance;
- at level 5, the service achieved 80% compliance.

The PID service meets all indicators classified as essential. The failed indicators are concerned with automatic features, including references and/or qualified references to other data, and data accessed automatically (i.e., by a computer program). The assessment checklist is available [10].

3.2 The automatic assessment

The automated assessment results allowed us to identify actions to improve our metadata and metadata representation by automated means. Implementing these measures led to a noticeable enhancement of our research data FAIRness, which increased from 47% to 74%. Based on this experience, we provide the following set of recommendations to improve FAIRness scores:

- ensure that the landing page is machine-readable, avoiding JavaScript generated contents;
- define available metadata in JSON-LD, both on the landing page and in the used PID registration system, e.g., DataCite;
- provide links to the content resources (e.g., the PD article, CSV datasets, etc.) on the landing page. Linked content resources of long-term readability such as plain text are preferred;
- ensure metadata for linked data is correct and complete;
- use the standards suggested by F-UJI to complement free-form descriptions;
- keep your re3data record up to date and define an OAI-PMI endpoint for it.

However, it should be noted that automatic tools only partially support FAIRness evaluation, as some aspects require human mediation and interpretation [11]. Furthermore, although the complete code of the F-UJI tool is accessible, the requirements to receive the highest score are not evident. On the other hand, tools like F-UJI are valuable for identifying weaknesses in metadata and metadata presentation that can be improved by automatic means. We propose a “FAIR by Design” approach which, following Privacy by Design (PbD) [12] [13] where privacy measures are embedded directly into technology and business practices from their inception. Accordingly, “FAIR by Design” aims to align research data infrastructures with FAIR principles through their entire lifecycle.

4. Conclusion

The RDA-FDMM is a comprehensive standard for manual FAIR assessment broadly recognized by the FAIR community. The in-depth FAIR analysis using RDA-FDMM helped us better understand where our services stand with regards to FAIR. Our main findings are:

- our experience highlights the importance of evaluating both machine-readable as well as non-machine-readable elements. Thus, we considered both cases in our study;
- Our experience showed that both broader standards such as RDA-FDMM as well as automated FAIR assessment using tools like F-UJI are needed to get a comprehensive picture regarding the FAIR compliance of research data infrastructures.
- Despite the fact that automated tools always have limitations and technical challenges, the F-UJI tool gave us valuable hints on how to improve our metadata, although the criteria how to receive a high score are not fully evident.

As the research ecosystem evolves, providing easily machine-readable metadata becomes increasingly important. To better achieve this, we recommend adopting a "FAIR by design" approach early in product or service development to ensure FAIR principles are embedded in project outcomes. This includes regular FAIR assessments throughout the project lifetime to monitor and evaluate how the ongoing improvement of research data infrastructures affects the FAIR maturity score.

Keywords

FAIR principles, FAIR assessment, RDA FAIR Data Maturity Model, Automated FAIR assessment tool.

References

-
- [1] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
 - [2] Research Data Alliance FAIR Data Maturity Model Working Group. (2020). *FAIR Data Maturity Model: Specification and guidelines*. <https://doi.org/10.15497/RDA00050>
 - [3] Saldanha Bach, Janete, Klas, Claus-Peter, & Mutschke, Peter. (2022). *Application of "RDA FAIR Data Maturity Model" to assess the PID registration service in terms of FAIRness*. <https://doi.org/10.5281/ZENODO.7409651>
 - [4] Devaraju, Anusuriya, Huber, Robert, Mokrane, Mustapha, Herterich, Patricia, Cepinskas, Linas, de Vries, Jerry, L'Hours, Herve, Davidson, Joy, & Angus White. (2022). FAIRsFAIR Data Object Metrics (0.5). Zenodo. <https://doi.org/10.5281/zenodo.6461229>
 - [5] Klas, Claus-Peter, Zloch, Matthäus, Bach, Janete Saldanha, Baran, Erdal, & Mutschke, Peter. (2022). *KonsortSWD Measure 5.1: PID Service for variables report (1.0)*. Zenodo. <https://doi.org/10.5281/ZENODO.6397367>
 - [6] Devaraju, Anusuriya, & Herterich, Patricia. (2020). *D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories*. <https://doi.org/10.5281/ZENODO.3678715>.
 - [7] European Commission. Directorate General for Research and Innovation., Visionary Analytics., DANS., DCC., and EFIS., *European Research Data Landscape: final report*. LU: Publications Office, 2022. Accessed: Aug. 18, 2023. [Online]. Available: <https://data.europa.eu/doi/10.2777/3648>.
 - [8] J. S. Bach, C.-P. Klas, B. Mathiak, Y. Zhang, and P. Mutschke. "FAIRness assessment: a comparison of the RDA model and the F-UJI Automated tool report," Zenodo, 2023, doi: 10.5281/zenodo.8308902.
 - [9] Saldanha Bach, Janete, Klas, Claus-Peter, & Mutschke, Peter. (2022). *Application of "RDA FAIR Data Maturity Model" to assess the PID registration service in terms of FAIRness*. <https://doi.org/10.5281/ZENODO.7409651>
 - [10] Saldanha Bach, Janete. (2023). FAIR assessment practices: Experiences from KonsortSWD and BERD@NFDI [Dataset 2 KonsortSWD] (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8339809>
 - [11] Anusuriya Devaraju, & Robert Huber. (2020, September 30). F-UJI : An Automated Assessment Tool for Improving the FAIRness of Research Data. Zenodo. <https://doi.org/10.5281/zenodo.4068347>.
 - [12] A. Cavoukian, "Privacy by Design (PDF)", Information and Privacy Commissioner, https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf.
 - [13] P. Hustinx, "Privacy by design: delivering the promises," IDIS, vol. 3, no. 2, pp. 253–255, Aug. 2010, doi: 10.1007/s12394-010-0061-z.