1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.344

Published: 07 Sept. 2023

FAIR assessment practices

Experiences from KonsortSWD and BERD@NFDI.

Janete Saldanha Bach[1] [https://orcid.org/0000-0001-9011-5837], Fidan Limani[2] [https://orcid.org/0000-0002-5835-2784], Yudong Zhang[1] [https://orcid.org/0009-0006-7108-475X], Atif Latif[2] [https://orcid.org/0000-0003-3085-3031], Brigitte Mathiak[1] [https://orcid.org/0000-0003-1793-9615], and Peter Mutschke[1] [https://orcid.org/0000-0003-3517-8071]

[1] GESIS - Leibniz-Institute for the Social Sciences, Germany

[2] ZBW – Leibniz Information Centre for Economics, Germany

Abstract. The poster presents FAIR assessment experiences in the context of the two NFDI consortia KonsortSWD and BERD@NFDI, employing the established Research Data Alliance's FAIR Data Maturity Model (RDA-FDMM) and the F-UJI Tool, an automated solution. RDA-FDMM, a manual technique, is more comprehensive, while the automated F-UJI tool effectively detects areas of improvement in metadata presentation that automated means can address. Our experiences highlight the need to examine both machine-readable as well as non-machine-readable elements and acknowledge automated tools' limitations, while valuing their insights. As the research ecosystem advances, metadata representation should be made increasingly machine-readable. We recommend a "FAIR by design" approach from the beginning to ensure alignment with FAIR principles in project outcomes. Continuous assessments during a project's lifetime promote ongoing research data infrastructure improvements within the NFDI consortia context, contributing to NFDI infrastructure innovation and optimization.

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

1. Introduction

The FAIR Data Principles [1] are widely applied to research data infrastructures. 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 has proposed a FAIR Data Maturity Model (RDA-FDMM) [2], based on which we share FAIR assessment experiences from KonsortSWD's[i] PID service [3] and BERD@NFDI's[ii] metadata schema. Moreover, we studied a FAIR automatic assessment using the F-UJI tool [4], which employs the RDA-FDMM and the FAIRsFAIR metrics [5] in a machine-readable fashion. We applied the F-UJI tool to GESIS Search in the context of KonsortSWD, motivated by the European landscape study [6] which also relies on F-UJI tool and led us to improve our metadata [7]. Our findings

[i] KonsortSWD (Consortium for the Social, Behavioural, Educational and Economic Sciences) is funded by the National Research Data Infrastructure (NFDI). KonsortSWD Homepage: https://www.konsortswd.de/

[ii] BERD@NFDI is an initiative to build a powerful platform for collecting, processing, analysing, and preserving Business, Economic, and Related Data. It is funded by the National Research Data Infrastructure (NFDI). BERD@NFDI Homepage: https://www.berd-nfdi.de/

represent a great opportunity to support convergence towards FAIR adoption and implementation at the NFDI level, while overcoming challenges related to different projects.

2. FAIRness Assessment Methodologies

FAIR assessments based on automatic tools [8] [9] [10], online self-assessment surveys [11], [12], [13], [14], manual [15] [16] [17] and hybrid methods [18] exist. The RDA-FDMM stands out as the most prominent manual method due to its completeness and comprehensive acknowledgement from the broader FAIR community. Most assessment tools rely partially on the RDA-FDMM model's indicators and measures. The RDA-FDMM defines 41 FAIR indicators, organized into three classes (*Essential*, *Important*, and *Useful*), and five levels (see Table 1).

Table 1. RDA-FDMM: indicators classes in five levels [19].

Classes	**Indicators Quantity**	Level 1	Level 2	Level 3	Level 4	Level 5
Essential	20	20	20	20	20	20
Important	14		7	14	14	14
Useful	7				3	7
Total according to the sum indicators		**20**	**27**	**34**	**37**	**41**

The RDA-FDMM measures indicators based on binary questions, by a "progress" evaluation, or by a hybrid mode. The KonsortSWD applied the binary method on each indicator, while BERD relied on a hybrid approach [2].

An example of an automated FAIR assessment is provided by the F-UJI tool, which, as a consequence, considers only indicators that can be assessed automatically, leading to a subset of just 16 [5] out of the 41 indicators proposed by RDA-FDMM.

3. FAIR Assessment using the RDA FAIR Data Maturity Model

3.1 KonsortSWD PID Service

In KonsortSWD we applied the RDA-FDMM to its PID service aiming to assign PIDs to data elements below study level (such as for survey variables). The PID service is based on the data registration agency da|ra [3], and the indicators were manually assessed at the PID service or at the da|ra level, using the pass-or-fail method. This approach is focused on determining how a resource under evaluation performs on meeting the indicators across the FAIR areas. In that sense, it gives a binary answer to each indicator. The results show that the PID service meets all the indicators classified as *essential* and most of the indicators from the classes *important* and *useful* (see Table 2).

Table 2. PID and da|ra service assessment results: level distribution.

Classes	**Indicators Quantity**	Level 1	Level 2	Level 3	Level 4	Level 5
Essential	20	20 / 20	20 / 20	20 / 20	20 / 20	20 / 20
Important	14		7 / 7	10 / 14	10 / 14	10 /14
Useful	7				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%**

We fully comply with levels 1 and 2, achieve 88% compliance at level 3, 89% at level 4, and 80% at level 5. 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).

3.2 BERD Metadata schema

In BERD@NFDI, we assessed the core part of the project's metadata schema, represented by DataCite Schema [20]. We wanted to assess the extent to which the elements of this schema can support the FAIR principles as identified in the RDA-FDMM, and not the metadata values of an (digital) object. Thus, our evaluation scope included only the FAIR principles that relate to the metadata aspects, resulting with 26 indicators in total, of which 14 *essential*, 9 *important*, and 3 *useful*. We applied the "0 - not applicable" score to the data-related indicators; Table 3 shows the indicators that pertain to both data and metadata.

Table 3. RDA-FDMM: Indicators/FAIR category for the data and metadata.

RDA Maturity Model Per Category	Indicators Per Category	Metadata Related	Data Related
F	7	5	2
A	12	6	6
I	12	7	5
R	10	8	2
Total	**41**	**26**	**15**

The assessment results show a relatively low FAIRness progress for the A and I categories of FAIR, whereas F and R perform better. The FAIR principles encompass data, metadata, and infrastructure [21] and the lack of any entity during evaluation normally affects the assessment score. Thus, since the data-related indicators are not being considered for this case, this disfavours the final assessment score.

4. Lessons learned from using an automated FAIR Assessment Tool

As an example of an automated FAIR assessment, we used the F-UJI tool to assess the GESIS Search as a relevant repository in the context of KonsortSWD. The automated assessment allowed us to identify actions to improve our metadata and metadata representation by automated means. Implementing the measures led to a noticeable enhancement of our research data FAIRness, and to a set of recommendations to improve FAIRness scores. The recommendations include:

- 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.

It is essential to highlight that automatic tools can support FAIRness evaluation only partially. Although automation saves effort, not all components of the research data ecosystem are machine-readable. Some FAIR principles' aspects still require human mediation and interpretation [22]. On the other hand, using a tool like F-UJI is valuable in identifying weaknesses in metadata presentation that can be improved by automatic means, e.g., when the required metadata do not exist in machine-readable way, such as metadata generated by JavaScript.

We propose a "FAIR by Design" approach which, following Privacy by Design (PbD) [23] [24] where privacy measures are embedded directly into technology and business practices

from their inception. “FAIR by Design” aims to align research data infrastructures with FAIR principles through their entire lifecycle. Regular FAIR assessments and continuous improvements in FAIR scores should become an integral part of any data infrastructure development.

5. Conclusion and recommendations

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, whereas the F-UJI tool gave us valuable hints on how to improve our metadata, despite the fact that automated tools always have limitations and technical challenges. Our experience with RDA-FDMM and the F-UJI tool highlights the importance of evaluating both machine-readable as well as non-machine-readable elements. Thus, we considered both cases in our study.

There are two major take-aways from our studies: (1) Apply both broader standards for manual FAIR assessment like RDA-FDMM, as well as automated tools like F-UJI to finally get a comprehensive picture regarding the FAIR compliance of research data infrastructures. (2) Adopt a “FAIR by design” approach early in product or service development to ensure that the FAIR principles are embedded in the development of research data infrastructures from the beginning, including regular FAIR assessments throughout the project lifetime to evaluate how the ongoing improvement of research data infrastructures affects the FAIR maturity score. This approach should be applied to NFDI as well, to finally innovate NFDI infrastructures.

Author contributions

The listed authors have prepared and written this extended abstract (role: Writing – original draft according to CreDIT guidelines, Contributor Roles Taxonomy).

Competing interests

The authors declare that there are no competing interests.

Funding

KonsortSWD is funded by the German Research Foundation (DFG) within the framework of the NFDI – project number: 442494171.

BERD@NFDI is funded by the German Research Foundation (DFG) within the framework of the NFDI – project number: 460037581.

References

1. M. D. Wilkinson *et al*., “The FAIR Guiding Principles for scientific data management and stewardship,” Sci Data, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

2. Research Data Alliance FAIR Data Maturity Model Working Group, “FAIR Data Maturity Model: specification and guidelines,” 2020, doi: 10.15497/RDA00050.

3. C.-P. Klas, M. Zloch, J. S. Bach, E. Baran, and P. Mutschke, “KonsortSWD Measure 5.1: PID Service for variables report,” Zenodo, Mar. 2022. doi: 10.5281/ZENODO.6397367.

4. A. Devaraju *et al*., “From Conceptualization to Implementation: FAIR Assessment of Research Data Objects,” Data Science Journal, vol. 20, p. 4, Feb. 2021, doi: 10.5334/dsj-2021-004.

5. A. Devaraju and P. Herterich, “D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories,” Feb. 2020, doi: 10.5281/ZENODO.3678715.

6. 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.

7. 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: 1*0.5281/zenodo.8308902.*

8. M. D. Wilkinson *et al*., “Evaluating FAIR maturity through a scalable, automated, community-governed framework,” Sci Data, vol. 6, no. 1, p. 174, Sep. 2019, doi: 10.1038/s41597-019-0184-5.

9. T. Rosnet, V. Lefort, M.-D. Devignes, and A. Gaignard, “FAIR-Checker, a web tool to support the findability and reusability of digital life science resources,” Jul. 2021, doi: 10.5281/ZENODO.5914307.

10. A. Ganske *et al*., “ATMODAT Standard (v3.0),” 2021, doi: 10.35095/WDCC/ATMODAT_STAN
DARD_EN_V3_0.

11. Universidade Federal da Paraíba (PPGCI/MPGOA - UFPB), “FairDataBR: a tool for datasets evaluation”. https://wrco.ufpb.br/fair/.

12. Australian Research Data Commons (ARDC), “ARDC FAIR data self-assessment tool,” 2022, https://ardc.edu.au/resource/fair-data-self-assessment-tool.

13. Data Archiving and Networked Services, “*SATIFYD*: self-assessment tool to improve the fairness of your dataset,” https://satifyd.dans.knaw.nl.

14. Data Archiving and Networked Services, “Fairaware: your first step towards your FAIR data(set),” https://fairaware.dans.knaw.nl/.

15. WDS/RDA Assessment of Data Fitness for Use WG, “WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations”, doi: 10.15497/RDA00034.

16. C. Bahim et al., “The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments,” Data Science Journal, vol. 19, p. 41, Oct. 2020, doi: 10.5334/dsj-2020-041.

17. J. Yu and S. Cox, “5-Star Data Rating Tool.” CSIRO, 2017. doi: 10.4225/08/5A12348F8567B.

18. D. J. B. Clarke *et al.*, “FAIRshake: Toolkit to Evaluate the FAIRness of Research Digital Resources,” Cell Systems, vol. 9, no. 5, pp. 417–421, Nov. 2019, doi: 10.1016/j.cels.2019.09.011.

19. J. Saldanha Bach, C.-P. Klas, and P. Mutschke, “Application of ‘RDA FAIR Data Maturity Model’ to assess the PID registration service in terms of FAIRness,” Oct. 2022, doi: 10.5281/ZENODO.7409651.

20. DataCite Metadata Working Group, “DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4,” p. 82 pages, 2021, doi: 10.14454/3W3Z-SA82.

21. E. Schultes, Official GO FAIR Foundation icons for the Three-Point FAIRification Framework. Zenodo, 2021. doi: 10.5281/ZENODO.4678333.

22. A. Devaraju and R. Huber, “F-UJI : An Automated Assessment Tool for Improving the FAIRness of Research Data,” Sep. 2020, doi: 10.5281/ZENODO.4068347.

23. A. Cavoukian, “Privacy by Design (PDF)”, Information and Privacy Commissioner, https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf.

24. 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.