FAIRness assessment of datasets and ontologies at Eurac Research
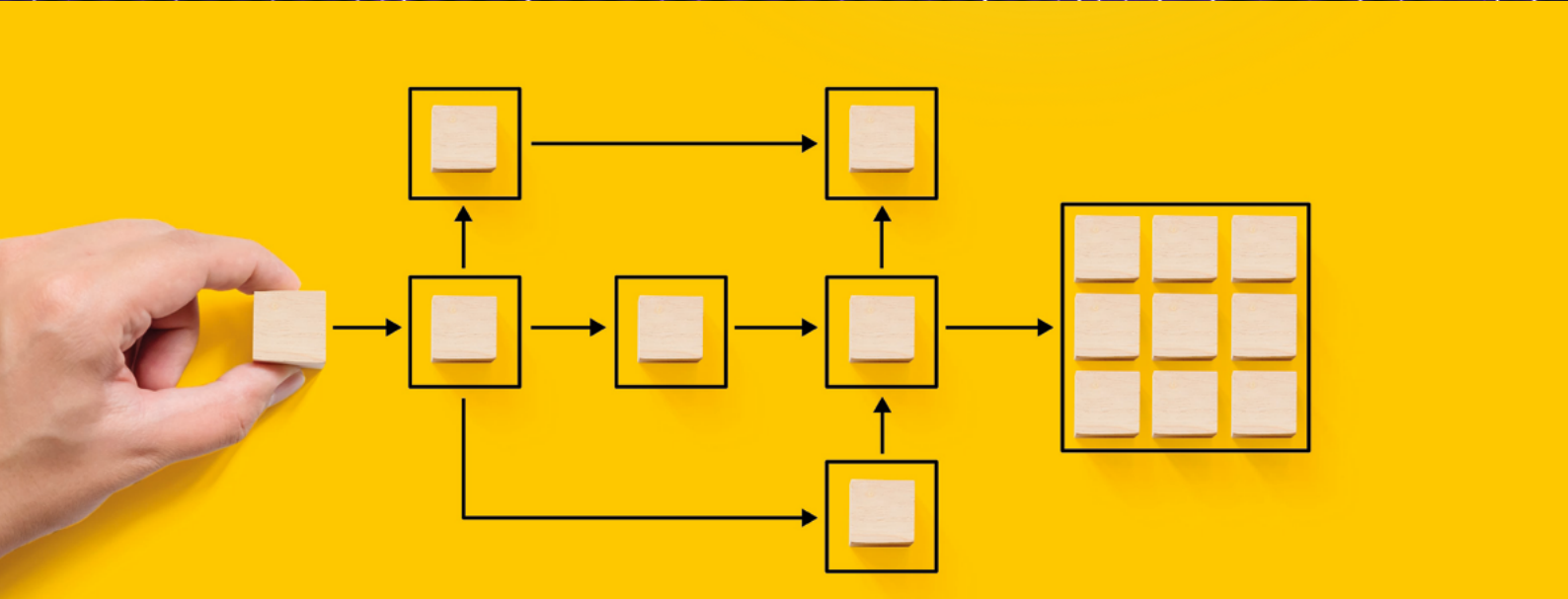
Authors names, affiliations and ORCIDs:

☐ Andrea Vianello, Eurac Research,
https://orcid.org/0000-0002-3017-6045

☐ Ekaterina Chuprikova, Eurac Research,
https://orcid.org/0000-0002-8981-0442

☐ Clara Linés, Digital Curation Centre (DCC),
https://orcid.org/0000-0003-3437-5145

Summary:

A team from Eurac Research assessed the FAIRness of a dataset hosted in their Environmental Data Platform using F-UJI and an ontology using O'FAIRe. After an initial low score in F-UJI, they implemented signposting to improve the machine interoperability of their metadata catalogue.

Introduction

RADAR4Chem is a free publishing service for chemistry research data based on the established data repository RADAR[1] by FIZ Karlsruhe - Leibniz Institute for Information Infrastructure. The service is suitable for research data from all areas of chemistry. It can also be used by researchers, regardless of their institutional affiliation, for datasets that do not easily fit into an existing dedicated chemistry repository, for instance because they contain a variety of data types or are interdisciplinary in nature. The repository uses persistent identifiers (DOIs) and guarantees a retention period of at least 25 years to ensure the research data remains available, citable, and discoverable in the long term.

While RADAR4Chem already incorporates several service features and measures aligned with the FAIR principles, we recognized that there were areas in need of improvement. This was also shown when RADAR was evaluated as a part of the European Research Data Landscape study[2], which examined the research data repository landscape using the F-UJI FAIR assessment tool[3]. Although the F-UJI tool focuses primarily on analysing the FAIRness of individual datasets, we believe that the results can be extrapolated to the entire repository.

Approach taken:

In this support action, we decided to use F-UJI again to assess the FAIRness of selected datasets published in RADAR4Chem as representative of the FAIRness of the repository infrastructure, for continuity with the European Research Data Landscape study. We also selected F-UJI because it is well known in the community, easy to implement for data repositories, data providers and data users, and the source code and documentation are available on GitHub, which proved helpful when interpreting the messages from F-UJI.

Challenges encountered and addressed:

F-UJI is quick and easy to use and produces immediate results: the overall FAIRness percentage score and the results for each individual criterion. However, the error messages and warnings are not always self-explanatory, and the interpretation of what needs to be improved is sometimes difficult. We gained helpful information from the F-UJI source code itself (available via GitHub) and by consulting with our mentor, Robert Hubert.

One additional disadvantage of the F-UJI tool is that there are only limited options for development and test instances, where no real resolvable DOIs are created and which are not registered in re3data[4]. The scores achieved for datasets on development and test instances are not comparable with the scores for actually published datasets (with DOIs). This makes software optimization more difficult after error analysis. Further, we would welcome the opportunity to assess the FAIRness of datasets before they are published in the repository, so that any potential improvements could be flagged and implemented before publication.

We also quickly realized that there are two partners who contribute to the FAIRness of datasets in a repository. The first partner is the data repository itself, which has to ensure that their services are FAIR-enabling. But the second partner are the data providers. If, for example, a data provider does not provide any information in the "Related Identifier" metadata field (which is optional in our metadata schema), this omission will lower the F-UJI assessment score by about 5%. Therefore, it is essential to advise data providers on how to improve the FAIRness of their datasets as well.

---

1   https://www.radar-service.eu/radar/de/home
2   https://data.europa.eu/doi/10.2777/3648
3   https://www.f-uji.net/
4   Registry of research data repositories: https://www.re3data.org

Impact:

Our score was already high, but with help from the mentor, we were able to identify some improvements to increase the accessibility level. We included some additional signposting links to the header and made some changes to our OAI-DC[5] and DataCite[6] formats and JSON-LD[7].

One immediate action we have already discussed with the RADAR User Advisory Council is to add a "FAIRness Check" to our repository. For instance, if a data provider does not include PIDs when using our form-based metadata editor (since these fields are optional in the RADAR MD schema), a recommendation could be displayed to include them. Over time, the built-in FAIRness Check should change the habits of our data providers, including more frequent use of PIDs, better quality of metadata annotations, and better machine readability.

Taken together, the enhanced FAIRness of RADAR datasets should result in better visibility, interoperability, and reuse of the data. Furthermore, these improvements should well position our repository to be integrated in (future) knowledge graphs, e.g. in the context of the German National Research Data Infrastructure (NFDI)[8], which should further increase the reach and impact of our services. We also plan to act as an advisor to other repositories with which we are in contact within the NFDI consortia (e.g. NFDI4Chem and NFDI4Culture[9]).

Finally, while our focus in this support action was on assessing and improving the FAIRness of datasets in our repository, we also learned about the FOOPS! tool[10] to assess the FAIRness of ontologies. This tool might be valuable to us and to other departments in our institution in the future, when evaluating any internally developed ontology.

Key message:

If the results of an FAIRness assessment tool are difficult to interpret or implement, consult the source code and/or human experts for guidance.

Data repositories must partner with data providers (and occasionally nudge them to change their habits) in order to improve the FAIRness of the datasets.

---

5   https://guidelines.openaire.eu/en/latest/literature/use_of_oai_dc.html
6   https://schema.datacite.org/
7   https://json-ld.org/
8   https://www.nfdi.de/?lang=en
9   https://radar.products.fiz-karlsruhe.de/en/radarabout/ueber-radar#radar+and+nfdi
10   https://catalogue.fair-impact.eu/resources/foops