1st Conference on Research Data Infrastructure

Spreading RDM

https://doi.org/10.52825/CoRDI.v1i.415



Published: 07 Sept. 2023

# Experiences from FAIRifying community data and FAIR infrastructure in biomedical research domains

Dagmar Waltemath[1][https://orcid.org/0000-0002-5886-5563], Esther Thea Inau[1][https://orcid.org/0000-0002-8950-2239], Lea Michaelis[1][https://orcid.org/0000-0001-9691-2677], Venkata Satagopam[2][https://orcid.org/0000-0002-6532-5880], and Irina Balaur[2][https://orcid.org/0000-0002-3671-895X ]

[1] Medical Informatics Laboratory and Core Unit Data Integration Center, University Medicine Greifswald, Greifswald, Germany

[2] Luxembourg Centre For Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg

**Abstract.** FAIR data is considered good data. However, it can be difficult to quantify data FAIRness objectively, without appropriate tooling. To address this issue, FAIR metrics were developed in the early days of the FAIR era. However, to be truly informative, these metrics must be carefully interpreted in the context of a specific domain, and sometimes even of a project. Here, we share our experience with FAIR assessments and FAIRification processes in the biomedical domain. We aim to raise the awareness that "being FAIR" is not an easy goal, neither the principles are easily implemented. FAIR goes far beyond technical implementations: it requires time, expertise, communication and a shift in mindset.

**Keywords:** FAIR assessment, FAIRification, biomedicine, EOSC, COMBINE

## 1. Introduction

The Findability, Accessibility, Interoperability and Reusability (FAIR) guiding principles for scientific data management and stewardship [1] provide guidance for sustainable deposition and sharing of scientific outcomes. Adherence to these principles leads to a more systematic approach for machine-actionable work with scientific data. FAIR data has a better chance to be reproducible, is more trustworthy and is cited more often. Large research networks foster the exchange and discussion about FAIR data, leading to new concepts, methods, scientific infrastructures and tools.

FAIR-related actions, however, need to be consolidated to understand what communities mean by "FAIR", to ensure communication about reusable Research Data Management tools, to foster cross-community developments and to build common metadata standards. These needs have already been identified. For example, the Research Data Alliance [2] developed the FAIR Data Maturity Model [3], a standard reference system for FAIR assessment tools and FAIRification workflows.

Our experiences with FAIR evaluations and FAIRification tasks show that scientists overestimate the FAIRness of their data rather than being too critical. Oftentimes the FAIR scores are a surprise and even a disappointment. We also observe that while FAIR principles are generally known, only a few scientists can explain their meaning or interpretation in detail. Thus, it is imperative to educate scientists on what it actually means to strive for a "FAIRness" and to support them in the FAIRification process.

## 2. Results

We assessed five infrastructures and performed subsequent FAIRification.

1. The **Computational Modeling in Biology Network** [4] coordinates the development of community standards and formats for computational models in biomedicine. Building on the experience of mature projects, COMBINE continuously integrates new requirements for model sharing and reuse, e.g. harmonised metadata [5], model-specific FAIR metrics (https://github.com/FAIR-CA-indicators), and guidelines for FAIR data sharing [6]. Funded by the EOSC Future, we develop a domain-specific FAIR assessment tool. This community-oriented process allows us to work with cross-domain experts.
2. The **Disease Maps Project** (https://disease-maps.org) develops disease-specific comprehensive knowledge representations. The development of a disease map is a complex process requesting participants with interdisciplinary expertise over a considerable time period [7]. Therefore, it is essential that the maps are available to a broad scientific community in order to benefit from the invested efforts. We assessed the FAIRness of the COVID-19 Disease Map in the Molecular Interaction NEtwoRk Visualization (MINERVA) platform [8] following the template provided in the IMI FAIRplus Project (https://fairplus-project.eu/). MINERVA is a FAIR infrastructure, facilitating the discovery and the accessibility of the integrated biological content, supporting the authorisation/ authentication features, providing a licensing system at diagram/ project level and integrating a converter for systems biology standards, thus supporting interoperability.
3. The **Study of Health in Pomerania** is a population-based cohort study designed to investigate the long-term progression of subclinical findings, their determinants and prognostic values [9]. We specifically explored the FAIRification of the medical laboratory metadata, leading to the indication that successful FAIRification requires interdisciplinary collaboration between data stewards and domain experts.
4. The **German Center for Diabetes Research** has established a Core Dataset for diabetes research (https://medical-data-models.org/45430). Its FAIRification improves reusability and study comparability. Hence, we structured the data, mapped the codes to terminologies, and implemented a formalised, provenance-enabled and semantically enriched representation of (meta)data. A baseline FAIRness evaluation helped us plan the FAIRification and establish a fruitful collaboration between the data owners, clinicians and data curators.
5. The **German Network University Medicine** supports the COVID-19 data collection from German University Hospitals. Adherence to the FAIR principles has been discussed from the start of the project. In 2023, we set out to assess the FAIRness level of the overall project, its sub-projects and domain-specific datasets using a manual evaluation system [10]. Interestingly, the participants reported only little knowledge of the FAIR principles, and questions addressing the uptake of FAIR recommendations showed that knowledge about the actual data management processes had been missing.

## 3. Discussion and lessons learnt

Working with the FAIR principles is challenging and the FAIR journey of a research institution requires actions, change of workflows and mindsets and financial support. Data owners, data managers, scientists, stakeholders and funding agencies need to actively contribute at each step of the data lifecycle, from design/ collection to sustainability, to FAIRify data, to minimise errors, and ultimately, to save time and to reduce effort [11]. Openness and the willingness to accept that there is (always) space for improvement of the data management processes is a prerequisite. Having worked with different scientific communities, we remarked that most re-sources are on average FAIR (about 50% in manual assessment tools), but getting beyond

the 70% threshold involves extensive work. However, efforts towards data FAIRification are a worthwhile investment as the FAIRification is a gradual process towards improving the data quality and a FAIR data set positively affects research outcomes.

## Author contributions

DW, ETI, VS and IB contributed to all sections and revised and approved the final version.

## Competing interests

The authors declare that they have no competing interests.

## Funding

DW and IB acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017536 (EOSC Future) for the project on FAIR COMBINE archives. DW acknowledges funding from the German Ministry of Health (BMBF), FKZ 01KX2121.

## Acknowledgement

We thank all our collaborators from our initiatives on FAIR-related infrastructures and communities including the Disease Maps, COMBINE, EOSC Future/ RDA communities.

## References

1. M.D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi: https://doi.org/10.1038/sdata.2016.18
2. RDA COVID-19 Working Group, RDA COVID 19 Case Statement. Research Data Alliance. 2020. URL: https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-social [accessed 2023-04-27]
3. FAIR Data Maturity Model Working Group, FAIR Data Maturity Model. Specification and Guidelines, Zenodo (2020), doi: https://doi.org/10.15497/rda00050
4. Waltemath et al., The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE), Journal of Integrative Bioinformatics (2020) doi: https://doi.org/10.1515/jib-2020-0005
5. Neal et al., Harmonizing semantic annotations for computational models in biology, Briefings in Bioinformatics 20 (2019), doi: https://doi.org/10.1093/bib/bby087
6. Ramachandran et al., FAIR Sharing of Reproducible Models of Epidemic and Pandemic Forecast, Preprints (2022). doi: https://doi.org/10.20944/preprints202206.0137.v1
7. Mazein et al., Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms, NPJ Systems Biology and Applications (2018), doi: https://doi.org/10.1038/s41540-018-0059-y
8. Gawron et al., MINERVA—a platform for visualization and curation of molecular interaction networks, NPJ Systems Biology and Applications (2016), doi: https://doi.org/10.1038/npjsba.2016.20

9. Völzke et al., Cohort Profile Update: The Study of Health in Pomerania (SHIP), International Journal of Epidemiology 6 (2022). Doi: https://doi.org/10.1093/ije/dyac034
10. Michaelis et al., How FAIR is NUM? - Lessons learnt from a FAIR survey within the German Network University Medicine (NUM), Proceedings of the 2023 SWAT conference, Basel, Feb. 2023.
11. Bruno et al., FAIR and Open Data in Science: The Opportunity for IUPAC, Chemistry International (2021), doi: https://doi.org/10.1515/ci-2021-0304