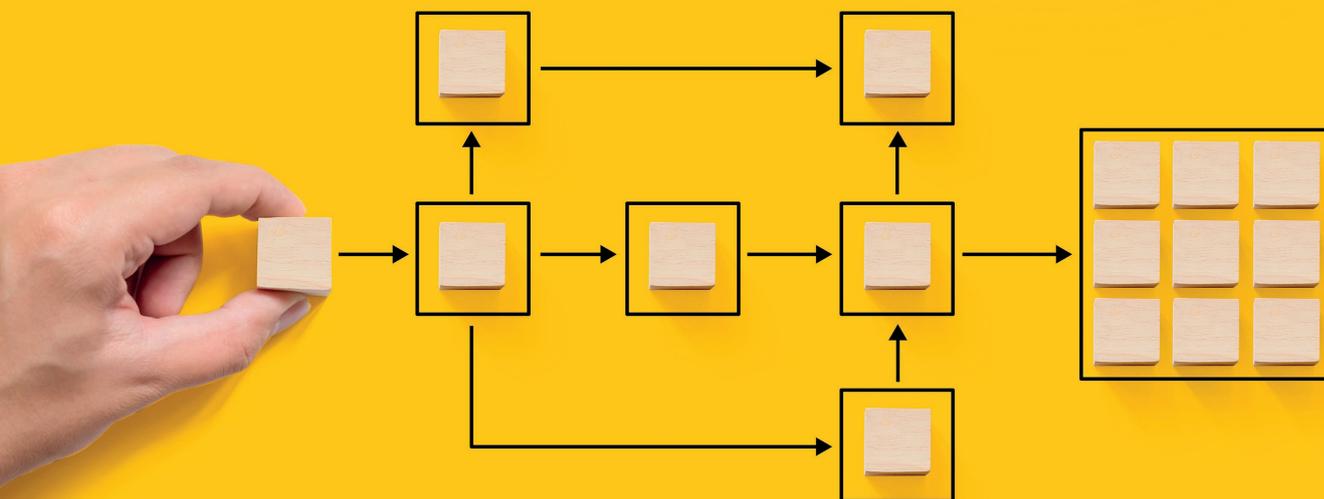# Making the metadata machine-readable to increase FAIRness of medical data in Poland

Authors names, affiliations and ORCIDs:

- ☐ Jakub Rusakow, Medical University of Gdansk Main Library, https://orcid.org/0000-0003-3233-7815

- ☐ Joanna Osika, Medical University of Gdansk Main Library, https://orcid.org/0000-0001-6401-5840

- ☐ Paulina Biczkowska, Medical University of Gdansk Main Library, https://orcid.org/0000-0001-7636-4773

- ☐ Agnes Jasinska, Digital Curation Centre (DCC), https://orcid.org/0009-0000-7934-3679

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

Support action:

FAIRness Assessment Challenge. During the 3-month challenge participants took part in three virtual workshops to self-assess and incrementally improve the FAIRness of their selected outputs. During the support action, participants benefited from interacting with a group of mentors representing the various FAIRness assessment tools and methods.

Keywords:

F-UJI, institutional repository, medical data, metadata, machine-readable

Summary:

A team from the Medical University of Gdansk (MUG) Main Library increased the FAIRness and machine readability of their repository datasets with the help of the F-UJI tool.

Introduction

We are a team of librarians at the Medical University of Gdansk (MUG) Main Library, and together with our counterparts at other medical universities in Poland, we manage the Polish Platform of Medical Research. The platform consists of several institutional repositories, each maintained by the home institution, but together forming a larger, interconnected system.

About three years ago, a major national funder in Poland, the National Science Centre, started requiring that all funded research projects must share their research data by depositing it in a trustworthy repository. The goal of our institutional repository, and the Polish Platform of Medical Research more broadly, is to support Polish medical researchers by giving them a place to deposit their data in compliance with the FAIR principles. An additional benefit of using our repository is that the researchers could contact us for advice and guidance with respect to FAIR principles.

However, we wanted to ensure that if a researcher deposited their data with us, they would be compliant with the funder's requirements, and we were not sure if we were there yet. We joined the support action to check the FAIRness of our datasets and improve our practices to best support our researchers going forward.

Approach taken:

We assessed the FAIRness of two different datasets in our repository using F-UJI[1]. Both datasets were described with the RDF metadata, and one had a PID and one did not, for comparison. However, both initial assessment scores were very low (35% and 4%).

After consulting with our software developer, we realised that the tool was not seeing some information, such as the licence that was included with the dataset but was not reflected in the results. The bug was eventually fixed, but we learned a valuable lesson: even when following established metadata standards like the RDF, the metadata can be expressed in multiple ways, and some of these ways will be less machine-readable than others.

We made two other changes that considerably improved the FAIRness score of our datasets: we added a controlled vocabulary for the medical subject headings and we provided open-format files for the datasets, where they were missing. Taken together, these changes made the data in our repository more machine-readable and easier to find and reuse.

Challenges encountered and addressed:

The main challenge was to figure out why our initial FAIRness scores from F-UJI were so low despite our datasets meeting the requirements of the F-UJI's template. Consulting with the support action mentor, Robert Huber, and with our software developer was very helpful, and once we realised that the issue was the automated tool not reading our RDF metadata properly, the solution was relatively simple. But it was still an important lesson for us to learn: that metadata that looks perfectly fine to the human eye may still be unreadable to automated tools, hindering machine-assisted discoverability. The subsequent changes we made to further improve the FAIRness of our datasets – stating the licence in a machine-friendly way, adding a controlled vocabulary of medical subject headings[2], and providing open-format files – were much easier.

---

1   https://www.f-uji.net/
2   https://www.nlm.nih.gov/mesh/meshhome.html

Impact:

Participating in the support action was a very valuable experience for our team. We gained a deeper understanding of the FAIR principles and how to assess and improve the FAIRness of datasets in practice, including through the use of external assessment tools. Now we can share these recommendations with the researchers, and PhD students at our institution and with the broader data management community in Poland. For example, we have presented our work with F-UJI at our working group for research data management on medical universities in Poland and it has created a lot of interest in the tool.

Our software developer also found F-UJI very useful to show the services they provide are in line with machine readability, and organised a webinar in his company on how he was using F-UJI to assess datasets hosted with their software.

In addition, we expect that higher levels of FAIRness will result in indexing our repository in OpenAIRE and will eventually lead to certification. This added recognition should encourage more researchers at our institution to deposit their data in our repository, helping us to advance our mission of facilitating the sharing, preservation, and reuse of medical data.

Key messages:

Human-readable metadata is not sufficient to comply with the FAIR principles. Machine-readable metadata is equally crucial.

Even when following established metadata standards, the metadata can be expressed in multiple ways, and some of these ways will be less machine-readable than others.

F-UJI and similar automated FAIRness assessment tools may yield different results depending on the PIDs and URLs of the dataset and on the chosen metric scheme.

@fairimpact_eu

in

company/fair-impact-eu-project/

fair-impact.eu