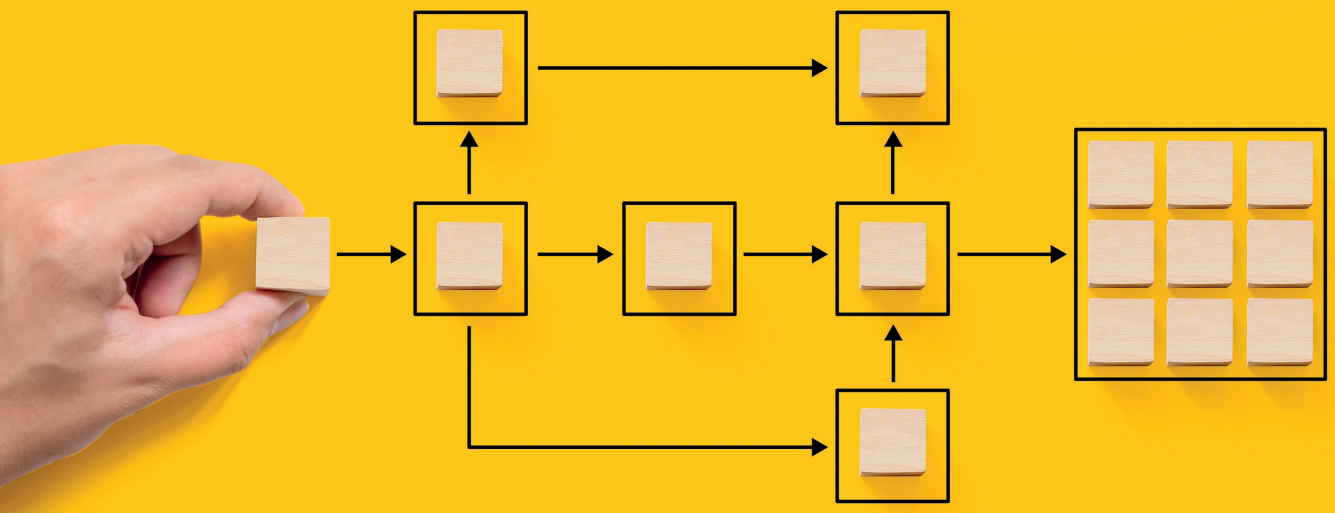
eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

FAIRSFAIR
Fostering Fair Data Practices in Europe

Evaluating the FAIRness of Health Research Software at Leipzig University Using an Automated FAIR Assessment Tool



IMPLEMENTATION STORY

Authors names, affiliations and ORCIDs:

- Mona Perbix, Institute of Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig University, https://orcid.org/0009-0006-3030-8606
- Matthias Löbe, Institute of Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig University, https://orcid.org/0000-0002-2344-0426

Summary

Software plays a crucial role in academic research, not only as a tool for data analysis but also as a research outcome or result, or even the object of research itself. FAIR (Findable, Accessible, Interoperable, Reusable) research software can increase the transparency, reproducibility, and reusability of research. For this to happen, software needs to be well-described (by metadata), inspectable, documented and appropriately structured so that it can be executed, replicated, built-upon, combined, reinterpreted, reimplemented, and/or used in different settings. The FAIR4RS Principles aim to guide software creators and owners on how to make their software FAIR. FAIR-IMPACT offered two support actions designed to enhance the FAIRness and impact of research software:

- Assessing and improving existing research software using a new extension of F-UJI which implements some of the metrics for automated FAIR research software assessment.
- Implementing the Research Software MetaData (RSMD) guidelines for better archiving, referencing, describing, and citing research software artefacts.

This FAIR Implementation Story outlines the specific aims and actions of the University of Glasgow in relation to their participation in one or both support action

Introduction

As part of Germany's Medical Informatics Initiative (MII) and the National Research Data Infrastructure for Personal Health Data (NFDI4Health), we aim to support the secondary use of health data from hospitals and research projects while ensuring compliance with strict legal and ethical standards. The development of new, specific software components is an essential part of the work in both initiatives. Our focus is on making complex research software more accessible and reusable, thereby promoting the effective use of health data in research.

Our goal was to assess selected repositories that already demonstrate advanced FAIRness using the F-UJI tool. We aim to identify areas for improvement and develop guidelines that can enhance the FAIRness of health data research software more broadly. This approach will help advance the responsible use and sharing of health data for research, driving better outcomes and innovation in the field. A particular focus is on the verifiability of the analysis of medical research studies by third parties, for which the software is being used and therefore its deployment must be precisely described.

Approach taken

One key activity involved assessing selected repositories using the F-UJI tool, which highlighted initial areas for improvement. The repositories lacked comprehensive metadata, which limited the machine readability and discoverability of the resources. Essential metadata components were either missing or not well-structured, reducing the ability for automated systems to find and interpret the data accurately. Although a basic CITATION BibTeX file existed in our GitHub repositories, it did not provide the necessary structure for optimal automated citation practices, affecting the ability to appropriately credit the original creators.

In response, we implemented changes to boost the repositories' FAIRness scores, including adding a CodeMeta file to enhance machine readability and discoverability. Additionally, we introduced a CITATION.cff file alongside an existing BibTeX file in our GitHub repositories, demonstrating the value of this approach in making automated citations more FAIR and readable.

During the support action, we also introduced a new automated test to F-UJI for a metric assessing whether software metadata includes an identifier for the software. The test works by checking for a persistent identifier, such as a Digital Object Identifiers (DOI) in the README or CITATION.cff file within a given GitHub repository. After that it is essential to confirm that the identifier resolves back to the software to ensure that the identifier is functional and accurately links to the intended software resource. This helps enhance citation practices, improves the software's traceability, and boosts its visibility and accessibility within research environments.

Challenges encountered and addressed

The main challenge that emerged was the utilisation of the automated FAIR assessment tool, F-UJI. Its incomplete state, as it is still undergoing full implementation and is currently only capable of testing a limited number of metrics, was an initial source of confusion. This presented a significant difficulty in conducting a comprehensive review of our repositories in collaboration with F-UJI. Since the final FAIR metrics percentage that a repository can achieve has a low threshold, it is not possible to increase the score even if the developer adds more metadata. Therefore, it is perceived that the repository has not yet met a substantial number of metrics, despite the repository actually meeting these which gives a lack of transparency regarding the fulfilment of the metrics.

Due to the limited implementation of metrics in F-UJI, we have introduced a new automated test. But the implementation of an additional metric into the current version of F-UJI has proven to be difficult due to the sparse documentation. Moreover, the specification of existing metrics complicated the process to define automated test requirements and methods. For instance, the implemented metrics second step "Does the identifier resolves back to the software" lacks clarity in terms of how it should be implemented programmatically and what defines it as a successful resolution.

Impact

Our work has begun to influence how we approach FAIRness within our team and organisations. We plan to develop comprehensive guidelines for implementing FAIR metrics in future projects. Once the F-UJI tool is fully released, we intend to incorporate it into some of our projects for FAIR assessments. Additionally, we aim to automate the process of adding the necessary metadata to new repositories, streamlining our workflows and ensuring that all projects meet the required FAIR standards from the outset.

Looking ahead, our action plan will continue beyond the support programme's lifecycle, focusing on ongoing training for team members on FAIR principles and integrating these practices into our project management processes. We will conduct workshops within our organisations to share our findings with our target audience. These sessions will aim to raise awareness about the importance of FAIR principles and the enhancements we've implemented. By engaging our stakeholders, we hope to foster a deeper understanding of these practices and encourage broader adoption throughout our organisations.

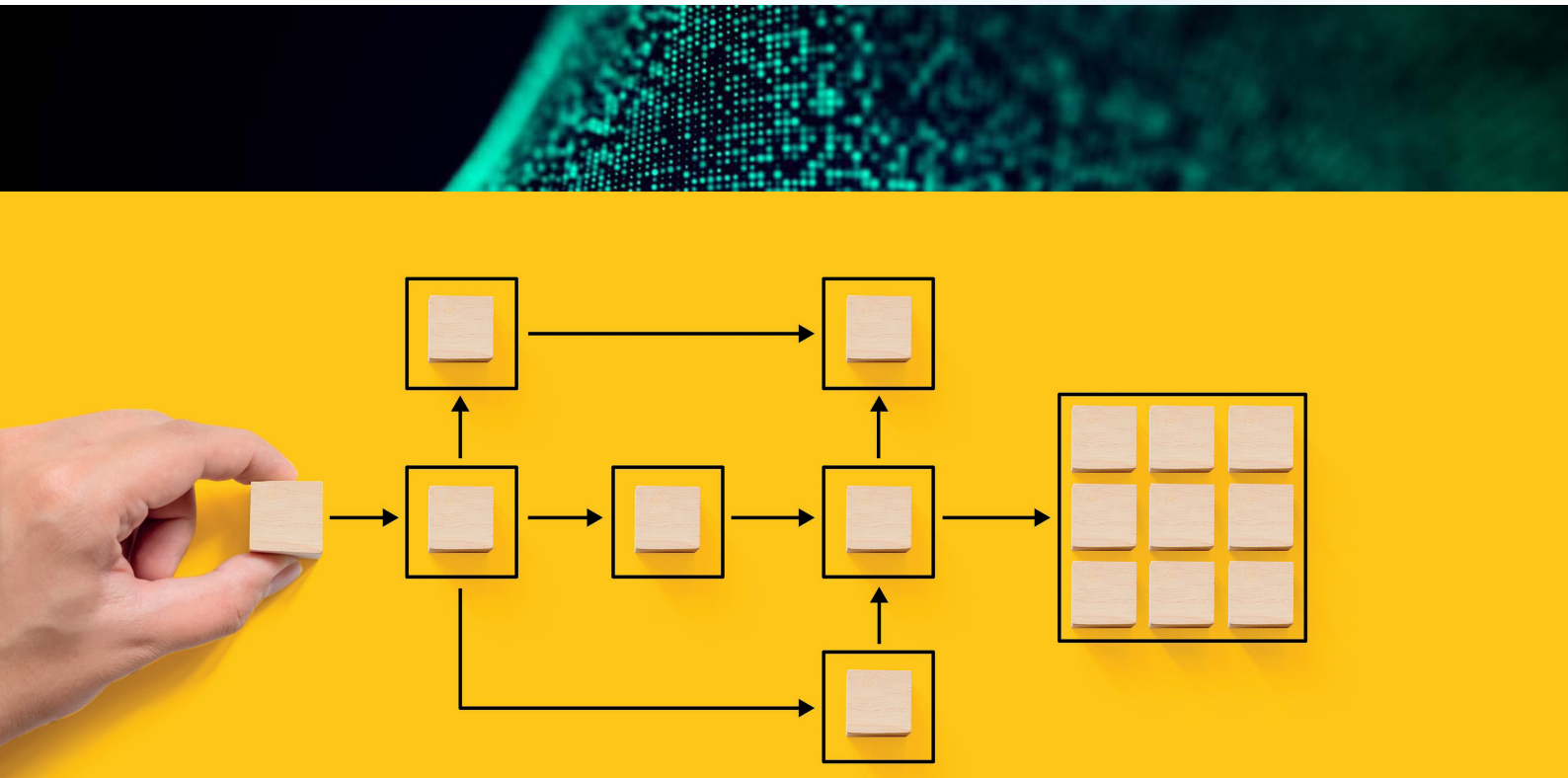
Key messages

1. FAIR metrics should be prioritised in health research software development to enhance accessibility, reusability and users' trust in further development.
2. Using the F-UJI tool aids in assessing the current state of FAIRness and identifying areas for improvement.
3. Descriptive metadata is essential; incorporating a CodeMeta file significantly improves machine readability and discoverability.
4. Continuous assessment and improvement of FAIR practices ensure that software remains accessible and reusable over time.

@fairimpact_eu

company/fair-impact-eu-project/

fair-impact.eu



Funded by the European Union