Lowering Barriers to FAIR Data: A User-Friendly Tool for Semantic Data Transformation

Karolis Cremers[1], César Bernabé[1], Daphne Wijnbergen[1], Rosa Zwart[1], Anna Niehues[1] and Marco Roos[1]

[1]*Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, The Netherlands*

Abstract

The FAIR guiding principles are essential for improving data reusability but often require researchers to perform complex transformations, such as converting data into RDF and alignment with semantic models. These tasks demand significant technical expertise, posing barriers to adoption. We propose a user-friendly, minimalist tool that simplifies this process by providing an intuitive drag-and-drop interface for mapping data elements to metadata model classes, automatically generating YARRRML mappings, and producing FAIR-compliant RDF. Our proof of concept aims to lower the technical barriers to FAIR data transformation, empowering researchers to adopt FAIR practices more easily.

Keywords

User Interface, RDF, SHACL, data model, semantic webmk.m.p.cremers@lumc.nl

1. Introduction

The FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles are crucial for enhancing the value of data in life sciences and other research domains. Government agencies and funding bodies increasingly require researchers to integrate FAIR practices into their research [1, 2], which often requires transforming existing data into semantic web compliant,

FAIR enabling formats such as RDF. However, the technical requirements of making data FAIR—such as converting data into RDF formats and aligning with semantic metadata models—pose significant challenges [3]. This process often requires specialised technical skills, such as programming, computer logic, and applying semantic web technologies and standards. This may be daunting for researchers who have maintained their own domain-specific knowledge and best practices. The technical overhead may act as a significant barrier to FAIR adoption.

2. Proposed solution

We propose a tool that enables researchers to transform datasets into FAIR-compliant formats using a simple, user-friendly interface. Researchers using the tool would perform simple steps. First, the data source file in CSV or JSON, and the target metadata model in SHACL [4] are loaded into the tool. The tool then analyses the structure of both files to generate a drag-and-drop canvas that allows users to place source data elements next to target metadata elements visually. Secondly, with these mappings defined, the tool automatically generates a YARRRML file [5], which is used to transform the source data format into RDF conforming to the specified metadata model. This provides a permanent conversion tool for future data with the same structure, allowing reuse within the community. Consequently, by hiding the technical complexities behind abstract concepts, users can focus on conceptual mappings rather than technical details. This approach lowers the barrier for researchers to adopt FAIR practices. It facilitates the transformation of their data without requiring coding skills or extensive training in

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025
mk.m.p.cremers@lumc.nl (K. Cremers)
https://orcid.org/0000-0002-1756-3905 (K. Cremers); https://orcid.org/0000-0003-1795-5930 (C. Bernabé); https://orcid.org/0000-0002-7449-6657 (D. Wijnbergen); https://orcid.org/0000-0002-9839-5439 (A. Niehues); https://orcid.org/0000-0002-8691-772X (M. Roos)

Semantic Web technologies. Because we obtain the data model from a SHACL file, the mapping process gives immediate feedback if the mapped data follows the constraints in the model representation. This visual feedback and interactivity of the tool allows for interactive sessions between data owner and data steward where complexities of the model can be discussed.

3. Prototyping and Implementation

Our prototyping efforts focus on two main areas, which we aim to *accelerate by collaborating* in a hackathon-setting at SWAT4HCLS:

1. *User interface:* We aim to design an intuitive interface that requires minimal prior knowledge of semantic web technologies. It must consist of a familiar drag-and-drop canvas that supports placing source data elements next to metadata model components.

2. *Back-end functionalities:* We aim to develop a prototype backend to handle the data processing tasks, including generating YARRRML mappings and executing the transformation of source data into FAIR-compliant RDF using those mappings. This prototype will be based on the conversion workflow proof of concept available for the RDF Schemas for Phenopackets version 2 https://github.com/rosazwart/phenopackets-v2-rdf-schema/tree/main.

4. Impact of Work

The proposed tool has the potential to enhance researchers' ability to comply with FAIR principles by reducing the technical overhead associated with data transformation, this tool can increase the adoption of FAIR practices in life sciences and across disciplines. By streamlining the process of creating FAIR-compliant data, this tool can empower researchers to make their data more reusable, thus contributing to a more connected and efficient research ecosystem.

Acknowledgments

This work has been carried out for the LUMC Data Competency Centre and the Leiden Regional Health RI node with funding from the Dutch Research Council (grant N°ICT.001.DCC.015), the NextGenerationEU and the Dutch National Growth Fund (NGFHR2201). Additional contributions were funded by the European Joint Programme Rare Diseases (N°825575), the European Rare Disease Research Alliance (N°101156595), the Joint Action Rare Disease Integration of National health systems and ERNs (N°101129863), and the Patient Lifestyle and Disease Data Interactium (Innovative Health Initiative; N°101132943).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] M. Bloemers, A. Montesanti, The FAIR funding model: Providing a framework for research funders to drive the transition toward FAIR data management and stewardship practices, Data Intell. 2 (2020) 171–180.

[2] E. Commissie, D. generaal Onderzoek en Innovatie, Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data, Publications Office, 2018. doi:doi/10.2777/1524.

[3] B. Dos Santos Vieira, C. H. Bernabé, S. Zhang, H. Abaza, N. Benis, A. Cámara, R. Cornet, C. M. A. Le Cornec, P. A. C. 't Hoen, F. Schaefer, K. J. van der Velde, M. A. Swertz, M. D. Wilkinson, A. Jacobsen, M. Roos, Towards FAIRification of sensitive and fragmented rare disease patient data: challenges and solutions in european reference network registries, Orphanet J. Rare Dis. 17 (2022) 436.

[4] P. Pareti, G. Konstantinidis, A review of SHACL: From data validation to schema reasoning for RDF graphs, in: Reasoning Web. Declarative Artificial Intelligence, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2022, pp. 115–144.

[5] P. Heyvaert, B. De Meester, A. Dimou, R. Verborgh, Declarative rules for linked data generation at your fingertips!, in: Lecture Notes in Computer Science, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2018, pp. 213–217.