Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230155

Harvesting the Low Hanging Fruits From the FAIRtree

Esther Thea INAU[1], Atinkut ZELEKE and Dagmar WALTEMATH
Medical Informatics Laboratory, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

Abstract. Extensive workflows have been designed to FAIRify data from various domains. These tend to be cumbersome and overwhelming. This work summarises our own experiences with FAIRification in health data management and provides simple steps that can be implemented to achieve a relatively low but improved level of FAIRness. The steps lead the data steward to register the data in a repository and then annotate it with the metadata recommended by that repository. It further leads the data steward to provide the data in a machine-readable format using an established and accessible language, establish a well-defined framework to describe and structure the (meta)data as well as publish the (meta)data. We hope that following the simple roadmap described in this work helps to demystify the FAIR data principles in the health domain.

Keywords. FAIR, Simple Steps

1. Introduction

Comprehensive and transparent data stewardship involves making data findable, accessible, interoperable and reusable (FAIR) [1]. Canonical workflows have been developed to FAIRify data from specific domains such as geoscience and systems biology [2–4]. This work describes a few simple and helpful steps in the start of a FAIRification journey to achieve a relatively low but improved level of FAIRness in the health domain.

2. Simple Steps in a FAIRification Journey

To improve the findability, we share the dataset via a searchable data repository which will provide a persistent identifier, as well as versioning and tagging with keywords. We also generously enrich the dataset with metadata and register the metadata in preprint servers such as Zenodo which allows for DOI versioning [5]. We make efforts to provide related readme and provenance information alongside the dataset. Registering the data and the associated metadata in searchable data repositories and preprint servers also allows the metadata to remain accessible even after the data is no longer available. We encourage the data owners to decide on the license to choose for the dataset. We then publish the metadata in data repositories such as the Medical Data Models portal to allow

[1] Corresponding Author: Esther Inau, E-mail: inaue@uni-greifswald.de.

downloading and exporting the file in common technical formats and in a standardized manner [6]. We then add contextual knowledge to the dataset in the form of meaningful links to related datasets. Our final inclusion of qualified references increases the interoperability. We then attach rich metadata and standardized vocabularies to the data in a manner that allows to decipher the origin, lineage, usefulness, relevance and how to cite the data in the said context. We also employ domain-specific standards, i.e. (meta)data has the same type, standardized, follows a community accepted template, contains the same type of data organized in a standardized way and uses a common vocabulary in well-established and sustainable file formats.

3. Discussion and Conclusion

Indulging in the FAIRification journey typically leads the data steward to annotate the said data with codes from domain-specific repositories. Through this journey the (meta)data is provided in a machine-readable format using an established and accessible language. This further leads the data steward to establish a well-defined framework to describe and structure the (meta)data in order to facilitate its findability and interoperability. The annotated dataset is then published which increases its reusability the annotated dataset. This journey also leads data stewards to register the (meta)data in the data repositories and preprint servers to enable version control and for researchers to have access to it both now and in the future. This work serves as a basic roadmap that contains small but useful FAIRification steps. The process of FAIRification is far more intensive but for this work but we have chosen to describe only the simpler steps that can be taken to improve data FAIRness. We hope this work helps to demystify the FAIR data principles in the health domain and motivates various stakeholders to start the FAIRification journey by taking the introductory steps described in this work.

References

[1] Suver C, Thorogood A, Doerr M, Wilbanks J, Knoppers B. Bringing Code to Data: Do Not Forget Governance. J Med Internet Res [Internet]. 2020 Jul 28 [cited 2023 Jan 10];22(7):e18087. Available from: http://www.jmir.org/2020/7/e18087/

[2] Wittenburg P, Hardisty A, Le Franc Y, Mozaffari A, Peer L, Skvortsov NA, et al. Canonical Workflows to Make Data FAIR. Data Intell [Internet]. 2022 Apr 1 [cited 2023 Jan 10];4(2):286–305. Available from: https://direct.mit.edu/dint/article/4/2/286/109842/Canonical-Workflows-to-Make-Data-FAIR

[3] Koymans MR, Hinsbergen DJJ, Pastor‐Galán D, Vaes B, Langereis CG. Towards FAIR Paleomagnetic Data Management Through Paleomagnetism.org 2.0. Geochem Geophys Geosystems [Internet]. 2020 Feb [cited 2023 Jan 10];21(2). Available from: https://onlinelibrary.wiley.com/doi/10.1029/2019GC008838

[4] Reiter T, Brooks† PT, Irber† L, Joslin† SEK, Reid† CM, Scott† C, et al. Streamlining data-intensive biology with workflow systems. GigaScience [Internet]. 2021 Jan 13 [cited 2023 Jan 10];10(1):giaa140. Available from: https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giaa140/6092773

[5] European Organization For Nuclear Research, OpenAIRE. Zenodo: Research. Shared. 2013 [cited 2023 Jan 10]; Available from: https://www.zenodo.org/

[6] Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P, et al. Portal of medical data models: information infrastructure for medical research and healthcare. Database [Internet]. 2016 [cited 2023 Jan 10];2016:bav121. Available from: https://academic.oup.com/database/article-lookup/doi/10.1093/database/bav121