

FAIRification and semantic modelling for Duchenne and Becker Muscular Dystrophy rare diseases

Pablo Perdomo-Quinteiro^{1,*}, Sergiu Siminiuc², Paraskevi Sakellariou², Marco Roos¹, Pietro Spitali¹ and N ria Queralt-Rosinach^{1,*}

¹Leiden University Medical Center, Leiden, the Netherlands

²Duchenne Data Foundation, Veenendaal, the Netherlands

Abstract

The BIND project is a EU-funded project that attempts to improve the characterisation of brain involvement in Duchenne and Becker Muscular Dystrophies (DMD and BMD). Here, we present our ongoing work on making multimodal data FAIR, the semantic models, and discuss challenges and opportunities based on our FAIRification experience.

Keywords

FAIR data, semantic modelling, BYOD, Duchenne Muscular Dystrophy, Becker Muscular Dystrophy, Rare Diseases

Motivation

The BIND project is a EU-funded project that attempts to improve the characterisation of brain involvement in Duchenne and Becker Muscular Dystrophies (DMD and BMD). To achieve this, 19 organizations across Europe and Japan are collaborating and obtaining phenotypic and molecular data such as transcriptomic, proteomic, clinical, behavioral, and MRI brain images. It is crucial to make this information Findable, Accessible, Interoperable and Reusable (FAIR), as making data FAIR will allow machines to exchange, integrate and analyze these data within the consortium and with other external FAIR data sources. Additionally, making data FAIR improves transparency, reproducibility, and maximizes the use of scientific output. Ultimately, FAIR data will advance our understanding of DMD and BMD and drive innovation on new therapeutic strategies. In order to achieve FAIRification, i.e., the process of making data FAIR, data must be properly described and annotated using standard metadata models and ontologies used by the DMD and BMD community for research, and must be stored in a persistent and

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences

*Corresponding author.

✉ P.Perdomo_Quinteiro@lumc.nl (P. Perdomo-Quinteiro); sergiu@duchennedatafoundation.org (S. Siminiuc); sakellariou.elvina@duchennedatafoundation.org (P. Sakellariou); M.Roos@lumc.nl (M. Roos); P.Spitali@lumc.nl (P. Spitali); N.Queralt_Rosinach@lumc.nl (N. Queralt-Rosinach)

ORCID 0000-0001-8784-0907 (P. Perdomo-Quinteiro); 0000-0002-1531-1560 (S. Siminiuc); 0000-0002-9091-0053 (P. Sakellariou); 0000-0002-8691-772X (M. Roos); 0000-0003-2783-688X (P. Spitali); 0000-0003-0169-8159 (N. Queralt-Rosinach)



  2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

accessible repository. Additionally, data must be available in machine-readable formats and should follow best practices for data management and sharing.

The FAIRification process has been developed together with EBRAINS, a digital research infrastructure created by the EU-funded Human Brain Project with the objective of understanding human brain function and disease. By making use of EBRAINS standards, the project can benefit from all the information gathered in EBRAINS, and at the same time, other researchers can make use of the data provided by BIND. The FAIRification of the data has been done using Semantic Web technologies (RDF, OWL and SHEX) and reusing the semantic models provided by the European Joint Programme on Rare Diseases (EJP RD). Our approach is to use FAIR implementations selected by the rare disease community. To this end, we worked in collaboration with the EJP RD, as a key driver project on FAIRification of rare disease resources for research and a large representation of the rare disease community. Our method is based on the FAIRification workflow [1], where the first step is to assess the FAIR status of the data.

Several Bring Your Own Data (BYOD) workshops [2] have been organized throughout 2022, with the two-fold objective of assessing the FAIR status of the data and explaining the meaning of the data to FAIR experts and bringing the FAIRification methodology closer to the data owners. In these workshops, we as FAIR experts needed to fully understand the data to correctly perform the FAIRification process. One of the challenges is that data owners are often unfamiliar with the FAIRification process, and viceversa, FAIR experts are often not familiar with domain specific knowledge. Another challenge is to reach a consensus of what the data mean and how best to represent it. The modelling has been developed following the semantic core data model of the set of common data elements for rare disease patient registries provided by the EJP RD [3]. Here, we present our ongoing work on making multimodal data FAIR, the semantic models we are developing to represent data such as patient phenotypic behavioral information compiled in the form of questionnaires or single cell RNA-seq data obtained in preclinical mouse models; and discuss challenges and opportunities based on our FAIRification experience so far.

Acknowledgments

We thank all BIND team, the data owners and domain experts. We also thank the EJP RD for providing the semantic models used in this project. This project was supported by a grant from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847826 (Brain Involvement in Dystrophinopathies (BIND)).

References

- [1] A. Jacobsen, et al., A Generic Workflow for the Data FAIRification Process, *Data Intelligence* 2 (2020) 56–65. URL: <https://direct.mit.edu/dint/article/2/1-2/56/9988/A-Generic-Workflow-for-the-Data-FAIRification>. doi:10.1162/DINT_A_00028.
- [2] R. Hooft, et al., ELIXIR-EXCELERATE D5.3: Bring Your Own Data (BYOD) (2019). URL: <https://zenodo.org/record/3207809>. doi:10.5281/ZENODO.3207809.
- [3] R. Kaliyaperumal, et al., Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data, *Journal of*

Biomedical Semantics 13 (2022) 1–16. URL: <https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-022-00264-6>. doi:10.1186/S13326-022-00264-6/FIGURES/6.