

IMI2 101034344 – EPND

European Platform for Neurodegenerative Diseases

WP2 – Establishment of the platform

**D1.2 – Online publication EPND glossary, core data models and metadata
specifications as a FAIR implementation profile**

Lead contributor (partner organisation)	1 – UM
Other contributors	4 – DZNE
	6 - UNILU
	7 - BBMRI
	12 - EATRIS
	13 - UMCG

WP1	
Due date	30/04/2023
Delivery date	Month 18
Delivery type¹	DEM
Dissemination level²	PU

Description of Action	
V2.0	27/10/2022

Document History

Version	Date	Status
V1.0	15/11/2023	Draft
V2.0	11/04/2024	Comments
V3.0	29/04/2024	Final

Abstract

This deliverable involves standardisation of language and data. The first in the form of a Glossary, which can be helpful when working with a multidisciplinary team with a wide variety of backgrounds. The second as a definition of a core data model and metadata specifications for case study 1, which is essential when working with data from multiple sources. The Glossary is a living document that uses a Wiki-based approach to define terms. A public view of its content is published and we expect the terms to undergo constant revisions and additions from EPND members. The core data model and metadata specifications for case study 1 were published in the form of a FAIR implementation profile and it is also publicly available and enables versioning.

Methods

Glossary

Innovation requires effective communication enabling transdisciplinary collaboration. Transdisciplinary glossaries are lacking in the field of neurodegenerative disorders; we developed a Glossary for EPND aiming in standardising communication across different disciplines and backgrounds.

We used the open-source tool [NocoDB](#) to create a Glossary with Wiki-based editing and sharing principles. The procedure of how the Glossary was deployed and how it is managed is available on [GitHub](#). We identified relevant modules, drafted a starting list of terms, and asked specific experts from EPND¹ to complete the modules. The experts edited and integrated definitions and references and provided additional terms. A short tutorial also gives user guidance for the platform. This initial draft is being reviewed by EPND members. A questionnaire is being developed to collect feedback on utility and possible improvements for use within and outside EPND. The plan is to distribute it during Q2-2024 and process the results during Q3-2024. We are also recording metrics of access and contribution.

FAIR implementation profile

Data harmonisation is critical for research that involves data from multiple sources, but it is timely and costly. Data FAIRification has overlapping goals with data harmonisation and combining these efforts can be beneficial in various ways; there are benefits for the research consortia and for the wider research community. Documenting the choices made by a research community during the FAIRification process in FAIR implementation profiles (FIPs) is a way for improving interoperability, in particular by repurposing community metadata and data models.

A common data model (CDM) for case study 1 (CS1) was defined using Shapes Constraint Language ([SHACL](#)), following W3C's standard recommendation. Different members of the consortium had the clinical expertise to define the common data model and to write the SHACL file that specifies this model. In order to facilitate the joint effort of formalising the CS1 CDM as a SHACL, Python scripts were created that takes as input an Excel file that defines the core data elements and their standard

¹ Davit Chokoshvili, Alex Gardiol, Nektarios Liaskos, Marjon Pasmooij, Maike Tauchert, Francesco Vigna

terms and outputs a SHACL file. The [Python scripts](#) were created by technical experts and the Excel file by clinical experts.

The bronze level cohort metadata specifications defined by task 1.3 from work package 1 were defined as a [CEDAR template](#). The [CEDAR initiative](#) aims in facilitating creation of metadata templates that uses controlled terms, which later can be used by a wider community.

We chose to use SHACL and CEDAR as both are FAIR formats and thus interoperable. An online tool called [FIP Wizard](#) was used to declare the FAIR implementation profile for case study 1. This tool aims in facilitating the capturing of FAIR implementation choices as a FIP.

Results

Glossary

Currently, the Glossary includes 524 entries across eight Modules: Acronyms, Hub Framework, Data Terms, Legal, Regulatory, Biomarker, Business, and Clinical. A [public view](#) is open for consultation by EPND members and the general public. We enable usage attributing a CC-BY access, and the use of our [GitHub repository](#) instructions to build analogous platforms.

FAIR implementation profile

The common data model includes items about identification, demographics, diagnosis, global cognitive measures, genetics, samples and Amyloid PET.

The CEDAR template defining the bronze level metadata items includes 17 cohort properties related mostly to contact details and basic information about the available data.

The SHACL file and CEDAR template are ways of tackling the FAIR principles I2 and I3. These solutions were published as a FIP using the FIP Wizard and are publicly available for consultation in [machine-readable](#) and [human-readable](#) formats.

Discussion

Glossary

Based on transdisciplinary collaboration and Wiki-like verification standards, the EPND Glossary will facilitate interactions between industry, academia and other partners within EPND and can be used or adapted to other initiatives. The ongoing evaluation of its efficacy will allow to derive metrics to evaluate transdisciplinary collaboration.

FAIR implementation profile

The FIP for CS1 describes possible solutions for tackling the FAIR principles I2 and I3 and thus generating FAIRer data. The FIP was published as a nanopublication, which enables versioning and automatic retrieval of the latest versions of both the core data model and cohort metadata model. We want to stress that we chose SHACL and CEDAR templates to represent these models as an initial proof-of-concept, but other technical solutions are available and possible. As the project evolves, the FIP can be updated to accommodate changes.

Conclusion

We published an online Glossary for EPND that is meant to be a living document that can be constantly updated to reflect EPND's multidisciplinary and standardise the understanding of terms amongst the different participants, but also the wider public. Besides this, we published a FIP for CS1 that defines possible technical implementations for generating FAIR data. FIPs published as a nanopublication are documents where versioning can be applied and thus also be a living document that can be updated to reflect technical implementation changes.