European Rare Blood Disorders Platform

ENROL recommendations for interoperability:

Codification schemes for Rare Hematological Diseases

Co-funded by the European Union

Disclaimer

European Rare Blood Disorders Platform (ENROL)

Call: HP-PJ-2019

Type of action: HP-PJ

GA number: 947670

Authors:

Stella Tamana - The Cyprus Foundation for muscular dystrophy research (CING)
Froso Sophocleous - The Cyprus Institute of Neurology and Genetics (CING)
Dimitris Theodorides - The Cyprus Institute of Neurology and Genetics (CING)
Sara Reidel - Vall d'Hebron University Hospital (HUVH) - Vall d'Hebron, University Hospital Foundation - Research Institute (VHIR)
Victoria Gutiérrez Valle - Vall d'Hebron University Hospital (HUVH) - Vall d'Hebron, University Hospital Foundation - Research Institute (VHIR)
Claire Diot - Vall d'Hebron University Hospital (HUVH) - Vall d'Hebron, University Hospital Foundation - Research Institute (VHIR)
Ines Labidi - Erasme University Hospital (ERASME) / LHUB-ULB
Petros Kountouris - The Cyprus Institute of Neurology and Genetics (CING)
Béatrice Gulbis - Erasme University Hospital (ERASME) / LHUB-ULB
María del Mar Mañú Pereira - Vall d'Hebron University Hospital (HUVH) - Vall d'Hebron, University Hospital Foundation - Research Institute (VHIR)

Short Description

Recommendation on codification schemes for RHD to ensure interoperability with ENROL.

Introduction

The European Rare Blood Disorders Platform (ENROL) has been conceived in the core of European Reference Network on Rare Haematological Diseases (ERN-EuroBloodNet) and with the endorsement of the European Haematology Association (EHA), as an umbrella for both new and already existing registries on Rare Hematological Disorders (RHDs). ENROL aims at avoiding fragmentation of data by promoting the standards for patient registries' interoperability released by the EU RD platform, with the principle to maximise public benefit from data on RHDs opened-up through the platform with the only restriction needed to guarantee patient rights and confidentiality, in agreement with EU regulations for cross-border sharing of personal data.

One of the first important building block for the EU RD platform is the "set of common data elements" (EU RD-CDS), which was released in December 2017 as a result of a dedicated working group facilitated by the Joint Research Centre (JRC) and composed of experts from projects related to common data sets: EUCERD joint action, EPIRARE and RD-Connect, as well as the JRC EU RD platform team. By defining the minimum data elements to be registered by all RDs registries across Europe, this set represented the first practical instrument released by the EU RD platform to increase interoperability of the data registries, while providing instructions on how and in which format each data element should be registered. Since in the light of the data set, all existing and new data registries across Europe are recommended to use this standard as the basis for their data collection activities, ENROL has included the data elements defined as part of the mandatory elements of the data set.

ENROL's strategy for data gathering addresses two ENROL objectives through the combination of the exhaustiveness of data coverage at the EU level, for health planning and epidemiological porpoises, with the RHD granularity, for promoting research. Therefore, the design of the ENROL platform allows the integration of data from any EU source and offering the possibility of available types of sources contribution through different levels of data granularity. Considering the different types of data sources, one of the pillars of the ENROL strategy for data gathering was to build a platform that will be opened to any EU data provider willing to contribute to ENROL (ERN-EuroBloodNet members / Other EU healthcare providers / European/national/regional existing registries). Thus, collecting data in a standardised manner will increase the power of that data in several ways while also, allow ENROL to become syntactically and semantically interoperable. Furthermore, new and existing registries of RHDs will greatly benefit when data is collected in a certain way, using recommended tools and standards, and thus, achieve a lifetime beyond the initial purpose of direct care delivery. Once pseudonymised, data can be pooled to advance diagnostics, knowledge, and understanding of the disease and of its accompanying symptoms. Moreover, standardisation allows information held inside multiple locations to be interrogated and produce aggregate results without requiring sensitive data to travel.

Despite the high number of RHD entities, as for other rare diseases (RDs), it is often challenging to bring together sufficient patient data, due to their low prevalence compounded by the fact that problems often arise in achieving an accurate diagnosis, especially for less prevalent diseases. Moreover, standards governing collection, organisation, or availability of RD patient data have been lacking until recently, leading to large number of unstructured and non-interoperable sources of information at national, regional, and local levels that have been running for last decades.

The following suggested recommendations and practices were developed as an initial set of data standardisation principles to ensure interoperability with ENROL.

Recommendations on Codification schemes for RHD

Making the existing registries interoperable is a major challenge directly linked to the exploitation of available data for research and epidemiological surveillance. To achieve the FAIRification (make data Findable, Accessible, Interoperable and Reusable) of data and promote interoperability with the European platform for RDs registries fragmentation of data, it is essential to use international codifications for the definition of a European registry dataset as ENROL.

Nowadays, multiple codification schemes are available for the international classification and categorisation of diseases and health conditions. Importantly, most of these codification schemes provide a definition along with the unique code to enable common understanding/interpretation of the different terms.

The EU RD-CDS codification schemes have been used for the ENROL Rare Haematological Diseases Common Data Set (RHD-CDS) and recommended.

1.1. Disease definition

- Orphanet Disease Ontology (ORDO; promoted by Orphanet) represents the only structured nomenclature specifically created for Rare Diseases (RDs) and the one with wider coverage of RDs. In addition, ORDO provides interoperability between international codifications, aligned with ICD10 and ICD11, and other nomenclatures including SNOMED-CT, MeSH, MedDRA and OMIM.

- New and already existing registries – and their 'affiliated' HCPs– during their broader, daily activities- should promote and utilise the OrphaCode as the preferred nomenclature for capturing the suspected or confirmed diagnosis of a rare disease. HCPs seeking hands-on guidance and support in adopting the ORPHA classification might consider consulting Ontology experts or FAIR data experts.

1.2. Observed phenotypes

In complex rare diseases, a patient may have the same (apparent) genetic variation but exhibit very different clinical presentations, with varying severity and prognosis. To capture and understand these variations and translate this knowledge a) to better diagnostics and care for the patient under review, and b) to drive forwards the pace of knowledge and understanding for the field at large, it is often necessary to capture detailed phenotypic descriptions.

Given the scarcity and thus value of data in RHDs, it is important to optimise the utility of this clinical information, in terms of immediate, one- to-one patient benefit but also re-use, for instance by searching databases and computing similarity: the best way to do this is to use an agreed ontology for capturing phenotypes.

The **Human Phenotype Ontology** (HPO; promoted by the EU RD platform and NIH as part of the Global Alliance for Genomics and Health (GA4GH)) is considered the most appropriate ontology for capturing the clinical presentation of rare diseases.

1.3. GENETIC DIAGNOSIS

International codification schemes for reporting patient genotype have been defined by the:

- **HUGO Gene Nomenclature Committee (HGNC):** is a worldwide authority responsible to assign standardised nomenclature to human genes (including protein coding genes, ncRNA genes and pseudogenes), to allow unambiguous scientific communication.

- **Human Genome Variome Society (HGVS):** reporting of genetic variants (i.e., changes in DNA, RNA and protein sequences) is described using a specific standard that defines the position, type and change of each variant. The standard is well-established and is widely used world-wide for molecular diagnosis. To ensure interoperability with ENROL, the use of HGVS for reporting of the patient genotype by registries and their affiliated HCPs is strongly recommended. When possible, ENROL provides a predefined list of genes and their corresponding variants in HGVS format, to ensure the quality of the provided genetic data.

1.4. LABORATORY PARAMETERS

LOINC is the most widely used international standard for classification of health-associated biological measurements, observations, and documents. It provides semantic interoperability (i.e., is the ability to exchange data with unambiguous, shared meaning) by representing the result of a test or observation in a standardised manner and when needed aligned with other codes such as SNOMED-CT. In ENROL, it is highly recommended when reporting the results of laboratory tests.

1.5. BIOMEDICAL CONCEPTS

NCI thesaurus (NCIT; promoted by National Cancer Institute) provides a large and wide range of standardised terms (integrated by NCI and other partners) for biomedical concepts and reference. It represents biomedical and clinical symptoms classifications where other standardised terms may lack a proper definition (e.g., blood count). In ENROL, NCIT is used to complement the codification of different types of elements, specifically in cases when a code is not found in other ontologies shown above.

Furthermore, as for other RDs, some disparities and gaps remain uncovered for some RHDs definition. Thus, for those terms remained uncovered, an internal unique code is given, and ENROL will liaise with the different ontology developers to propose additional terms relevant to RHDs, when needed.

http://www.eurobloodnet.eu/e

https://ec.europa.eu/health/ern_en

Co-funded by the European Union