

OWG_RFC_2: Use of DOID for “disease” field in the C2M2 model

CFDE-Draft-# (RFC Admin Only):	CFDE-DRAFT-2		
CFDE-Draft-Title:	Use of Human Disease Ontology for description of health conditions in C2M2 model		
CFDE-Draft-type: (Right click the box and select checkmark ✓ CTRL - Z to undo)	Standard Implementation ✓	Design Principle <input type="checkbox"/>	Policy <input type="checkbox"/>
Point of Contact Name:	Philippe Rocca-Serra		
Point of Contact Email Address:	philippe.rocca-serra@oerc.ox.ac.uk		
End of Date Comments (MM/DD/YY):	1/18/22		
Submitting Team/DCC Name:	Ontology Working Group		
CFDE-Draft-Status:	Pending	Active and Open for Comments <input checked="" type="checkbox"/>	
URL Link to the document:	https://docs.google.com/document/d/1OS_69jvdexMvH9SptGDAKAXCzyDtca2MoVjsGXImEE/edit		
License (Optional):			

Introduction

The C2M2 model, used to collect data from DCCs, defines entities such as ‘Subject’ and ‘BioSample’. Fields for essential qualifiers such as ‘taxonomic information’ or ‘anatomical part’ are currently available as part of the model. Fields for additional qualifiers to capture conditions, diseases, and health status will be part of the expanded C2M2. Thus, also required is the selection of a terminology resource against which to regularize annotation. The Ontology Working Group proposes the use of the Human Disease Ontology (DOID) terms for the values of a field capturing disease state.

The Ontology Working Group (OWG) has the following criteria for selection of an ontology for use with CFDE/C2M2. Ideally, the ontology/vocabulary should:

- be stable, but not static
- be under active development
- have a mechanism for requesting new terms and ontology changes (e.g. GitHub)
- be responsive to requests and questions
- have some level of community buy-in
- conform to community conventions on ontology/vocabulary development

An additional criteria, specific to this metadata area, is the availability of cross-references (mappings) to other well-established, disease-centric resources.

Rationale for this proposal:

The purpose of this proposal is to formalize the decision to use DOID to capture disease-related descriptors thus enhancing search capabilities of the CFDE Portal. Owing to extensive cross-references provided by DOID (e.g. mappings available to SNOMED-CT), the selection of DOID offers bridges to other resources, which also contribute to improving search functions. The selection process has been documented and notes are available [here](#). With regard to the role of the DOID for capture of disease information, a pragmatic approach has been taken. As the CFDE portal's main task is dataset discovery, the OWG decided that the role of DOID would be to capture information at the overall disease level. More granular descriptions will be handled via a more extended data model involving clinical and phenotypic terms from additional vocabularies. Work with the Clinical WG and C2M2 group will be carried out to understand and delineate the use/query cases and establish a model for capturing this information..

Proposed Actions

1. Use the Human Disease Ontology for describing health conditions as found in DCC datasets per the C2M2 model (once it is expanded to include capture of disease information.)
2. Handle new term requests as they emerge by following the DOID term request protocol which is clearly documented. The [DOID github issue tracker](#) will be used.
3. If there is a significant delay in getting new terms requested as described above in #2, or, if there is an urgent need for a new term, we can create temporary terms for use until the new terms are entered into DOID. Format for temporary ids: DOID:T###. Once the needed term is in DOID, we will replace the temporary local ids with official DOID term ids. Temporary DOID terms will be vetted by the Ontology Working Group. Temporary ids will be minted by the Ontology Working Group.
4. Collaborate with DOID and DCCs to address significant coverage gaps where and when they are identified (e.g. a DCC observed a gap in the area of pediatric oncology).
5. The Ontology Working Group is aware that there will be ongoing discussions around phenotypic features indicative of disease as well as clinical features of disease. As work in this area progresses, a new version of this RFC will be developed to reflect the C2M2 modeling plan that is adopted.