# **Mapping Project Data To Ontologies**

### C. Purpose

This document describes the process and instructions for mapping H3Africa project data and metadata to existing ontologies, enabling harmonisation and standardisation of data.

## D. Scope

This document is applicable to members of the H3ABioNet Ontology Mapping team or any member of aiming to map project data and metadata to existing ontologies. These include clinical and experimental data and metadata, as well as secondary data use.

### E. Abbreviations

CRF -	Case Report Form
DOID -	Human Disease Ontology
DUO -	Data Use Ontology
EFO -	Experimental Factor Ontology
HP -	Human Phenotype Ontology
ID -	Identifier
OLS -	Ontology Lookup Service

#### F. Procedure



1. Process Map

#### 2. Mapping Project Metadata to Ontologies

A. Identify the relevant experimental and(or) clinical/phenotype data and(or) metadata labels for the prospective/running project.

Note #1: Phenotype metadata can be identified by using a project's Case Report Form (CRF), Data Dictionary, or a survey, which allows collection of data labels.

Note #2: H3Africa CRFs are grouped by project on Google Drive: <u>https://drive.google.com/drive/folders/1c3wQs0xu7qTbjoQzjyClpvLjDKv</u> <u>2bdfs</u>

Note #3: Examples of experimental metadata include Technology Used, Protocol Used, Study Type, etc.

B. Once data labels are identified, input these values into a mapping mastersheet. Example Below:

Project	Demographics							
Project	Date of Birth	Age	Sex	Country of Birth (COB)	Language	Ethnicity	Parents' COB	Parents' Lan
Project 1								
Project 2								

Note #1: The H3Africa mapping mastersheet can be found on Google Drive:

https://drive.google.com/file/d/1E2w6Qvc9VRP5Xj0RiQcKcntTBLQLwjh H/view?usp=sharing

#### Mapping Clinical or Phenotype Data

- C. Navigate to Zooma, a online ontology mapping tool: <u>https://www.ebi.ac.uk/spot/zooma/</u>
- D. Input data labels (identified in Step A) as free text in the "What's this" text space. Thereafter, click"Annotate". See below figure:

ZOOMA					
Home Explore Help About ZOOMA					
What's this? •	What's this?  Show me some examples		Zooma is a tool for mapping free text annotations to ontology term based on a cura repository of annotation knowledge. Where mappings are not found in the curated respository one or more ontologies or selected from the <u>Ontology Lookup Service</u> to increase coverage. For example if y want to map GWAS annotations select the GWAS datasource and a common dise ontology such as EFO or DOID to maximise coverage when terms have no curated mappings. Use the text box to find possible ontology mappings for free text terms in the ZOOI repository of curated annotation knowledge. You can add one term (e.g. <i>Homo</i> sag per line. If you also have a type for your term (e.g. <i>torganism</i> ), put this after the ter separated by a tab.		
Configure Datasources	Annotate				
Results The table below shows a report describing h	ow ZOOMA annotates text terms supplied above.			Hide results that did not map? 🕑	
Term Type • Term Value	Ontology Class Label	Mapping Confidence	Ontology Class ID	Source	

Note #1: The tool maps free text to ontology terms based on pattern matching.

- E. Determine whether the matching is appropriate as follows:
  - a. Check if the Term Value matches the Ontology Class Label.
  - b. Identify the Mapping Confidence Low, Medium, or High.
  - c. Check the definition by clicking on the Ontology Class ID result.

Note #1: Example of Ontology Class Label - Diabetes Mellitus

Note #2 Example of Ontology Class ID - NCIT:C2985.

NB! Frequently, IDs from more than one ontology can be matched to the data label. Therefore, depending on use, the user may either map to all matching terms OR select one ontology ID to map to. The following guidelines can be used to prioritise ontologies

- Determine the ontology ID with the definition best suited to your data label.
- Where possible, use phenotype-specific ontologies.

- If no specific phenotypes are available, the disease ontologies DOID of HP can be used for disease label mappings.
- Use ontologies that are more frequently updated (well maintained) or ones that have been updated most recently.
- Some ontologies absorb structure from others, in these cases it's more appropriate to use the more specific ontology (former).
- F. If the mapping is suitable, match the data label (identified in Step A) to the Ontology Class Label and ID (identified in Step E).
- G. If the mapping is not suitable (based on the matching criteria in Step 5). Navigate to the Ontology
   Lookup Service (OLS): <u>https://www.ebi.ac.uk/ols/index</u>

EMBL-EBI			Services Research Training About us
Contology L	ookup Service	9	
Home Ontologies Documentation About			ය Contact Us
Welcome to the EMBL-	EBI Ontology Lookup Service.		Data Content Updated 14 Feb 2020 03:31
Search OLS Examples: diabetes, GO:0098743		Q Looking for a particular ontology?	<ul> <li>244 ontologies</li> <li>5,750,672 terms</li> <li>27,266 properties</li> <li>484,813 individuals</li> </ul>
			Tweets by @EBIOLS
About OLS The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the <u>Samples</u> , <u>Phenotypes and Ontologies Team</u> (SPOT) at EMBL-EBI.	Activity of the service of the servi	For feedback, enquiries or suggestion about OLS or to request a new ontology please contact ois-support @ ebi.ac.uk. For bugs or problems with the code or API please report on <u>GitHub issue</u> For announcements relating to OLS, such as new releases and new features sign up to the <u>OLS announce mailing list</u>	EBISPOT OLS       Image: Construction of the second s

Note #1: OLS is a searchable repository of biomedical ontologies.

- H. Type in the data label in the search box and click the magnifying glass.
- I. The result provides a list of matching Ontology Class Label, their Class ID and definition. Select the best suited definition, and match the data label to the associated Ontology Class Label and ID.

# NB! If mapping for more than one project at a time, an additional step is required! In cases such as this, proceed to Step J.

J. Check which projects are collecting which data labels. This can be marked with 1's if the projects are collecting the data labels, and 0's if not.

#### Mapping Experimental Data & Metadata to Ontologies

G. Navigate to the Experimental Factor Ontology (EFO) in OLS:

#### https://www.ebi.ac.uk/ols/ontologies/efo

	Expe	erimenta	al Factor On	tology			
	The Experimental Factor Ontology (EFO) provides a systematic description of many experimental variables available in EBI databases, and for external projects such as the NHGRI GWAS catalogue. It combines parts of several biological ontologies, such as anatomy, disease and chemical compounds. The scope of EFO is to support the annotation, analysis and visualization of data handled by many groups at the EBI and as the core ontology for OpenTargets.org						
	Search EFO				٩		
	Terms	🗄 Download	# Ontology Homepage	Contact			
Bro	wse Terms	Browse Properties	dil Ontology history		Ontology info		
exper	imental factor				Ontology IRI: <i>Phtp://www.ebi.ac.uk/efo/efo.owl</i> Version IRI: <i>Phtp://www.ebi.ac.uk/efo/releases/3.14.0/efo.owl</i>		

- H. Search the ontology for the data labels (identified in Step A).
- I. If a suitable match is found, match the data label (identified in Step A) to the Ontology Class Label and ID (identified in Step D).
- J. If a suitable match is not found, navigate to OLS: <u>https://www.ebi.ac.uk/ols/index</u>
- K. Type in the data label in the search box and click the magnifying glass.
- L. The result provides a list of matching Ontology Class Label, their Class ID and definition. Select the best suited definition, and match the data label to the associated Ontology Class Label and ID.

# NB! If mapping for more than one project at a time, an additional step is required! In cases such as this, proceed to Step 9.

M. Check which data labels apply to which projects. This can be marked with 1's if the data labels apply, and 0's if not.

#### 1. Mapping Secondary Data Use to Ontologies

A. Navigate to the Data Use Ontology (DUO) in OLS: <u>https://www.ebi.ac.uk/ols/ontologies/duo</u>

DLS > The Data Use Ontology DUO			JSON
The Data Use O	ntology		
DUO is an ontology which represent data us	e conditions.	Q	
Terms ± Download # Onto	logy Homepage		
Browse Properties all Ontology     data use limitation     data use requirements     investigation	Nistory  Preferred root terms  All terms	Ontology info Ontology IRI: A http://purl.obolibrary.org/obo/duo.ow/ Version IRI: A http://purl.obolibrary.org/obo/duo/releases/2020-02- 03/duo.ow/ Ontology id: duo Version: 2020-02-03 Number of terms: 282 Last loaded: Thu Feb 06 17:34:10 GMT 2020	

- B. In the "Browse Terms" section, expand "data use limitations", "data use requirements" and "investigations. The Ontology Class Terms found in these sections will also serve as the data labels for mapping of Secondary Data Use.
- C. Once data labels are identified, input these values into a mapping mastersheet.

Note #1: The H3Africa mapping mastersheet can be found on Google Drive:

https://drive.google.com/file/d/1E2w6Qvc9VRP5Xj0RiQcKcntTBLQLwjh H/view?usp=sharing

D. Review the prospective/running project's consent form.

Note #1: H3Africa consent forms are grouped by project on Google Drive: <u>https://drive.google.com/drive/folders/1ITCm2UDbzp3RzGML3G8vRcvb</u> muh4qt3c

E. Based on the definitions of the Ontology Class Terms (identified in Step A), check which data Secondary Data Use considerations apply to the prospective/running project. This can be marked with 1's if the data labels apply, and 0's if not.