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.

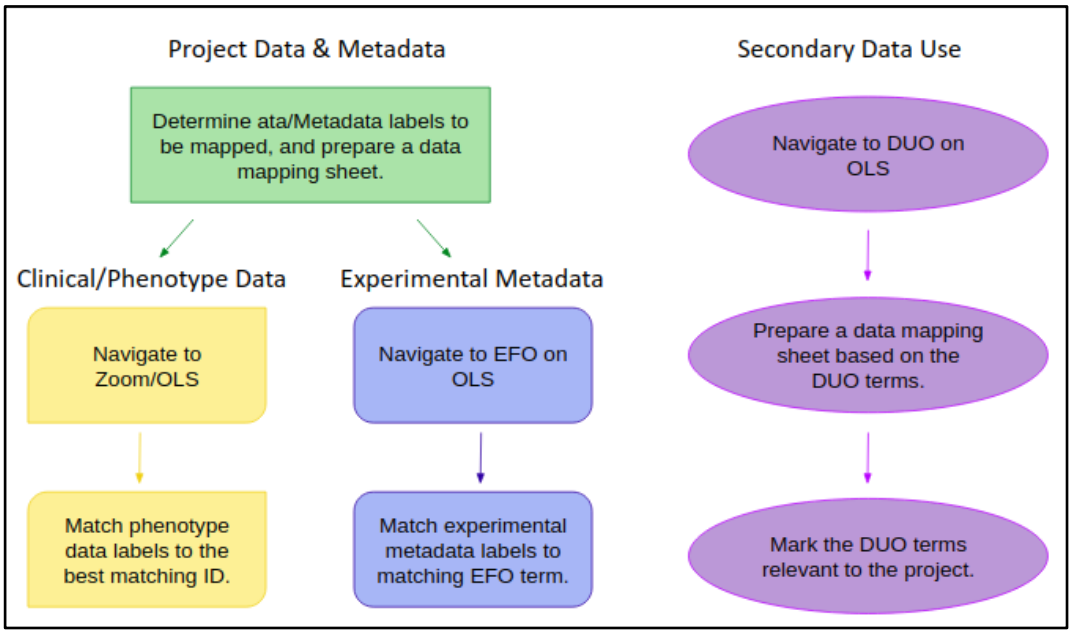
D. Scope

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: https://drive.google.com/drive/folders/1c3wQs0xu7qTbjoQzjyCIpvLjDKv2bdfs

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

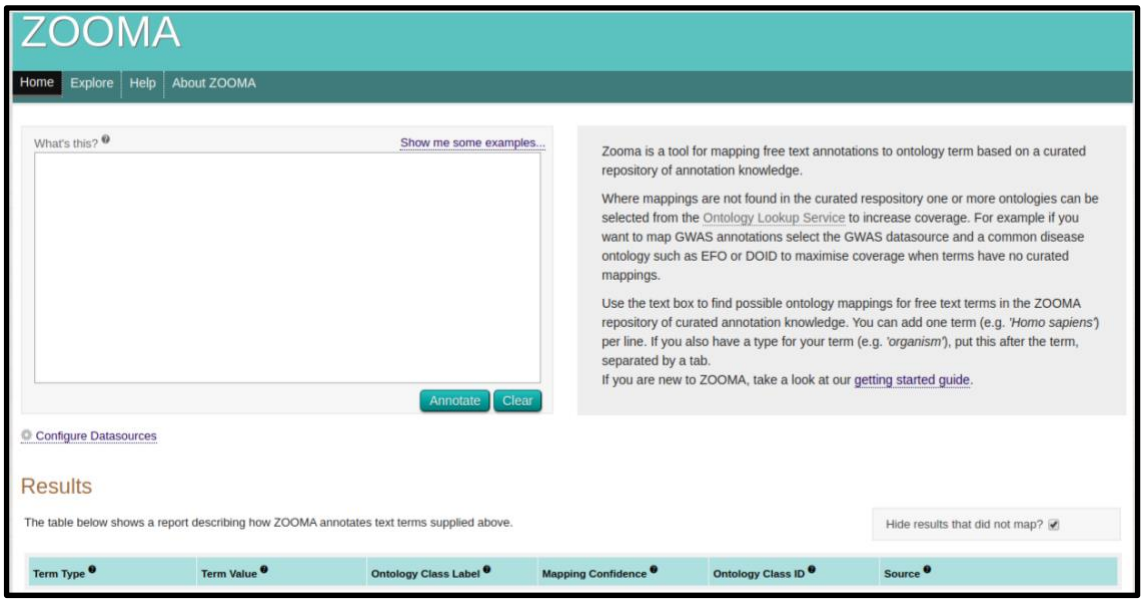
Note #1: The H3Africa mapping mastersheet can be found on Google Drive:

https://drive.google.com/file/d/1E2w6Qvc9VRP5Xj0RiQcKcntTBLQLwjhH/view?usp=sharing

Mapping Clinical or Phenotype Data

C. Navigate to Zooma, a online ontology mapping tool: https://www.ebi.ac.uk/spot/zooma/

D. Input data labels (identified in Step A) as free text in the "What's this" text space. Thereafter, click "Annotate". See below figure:



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.

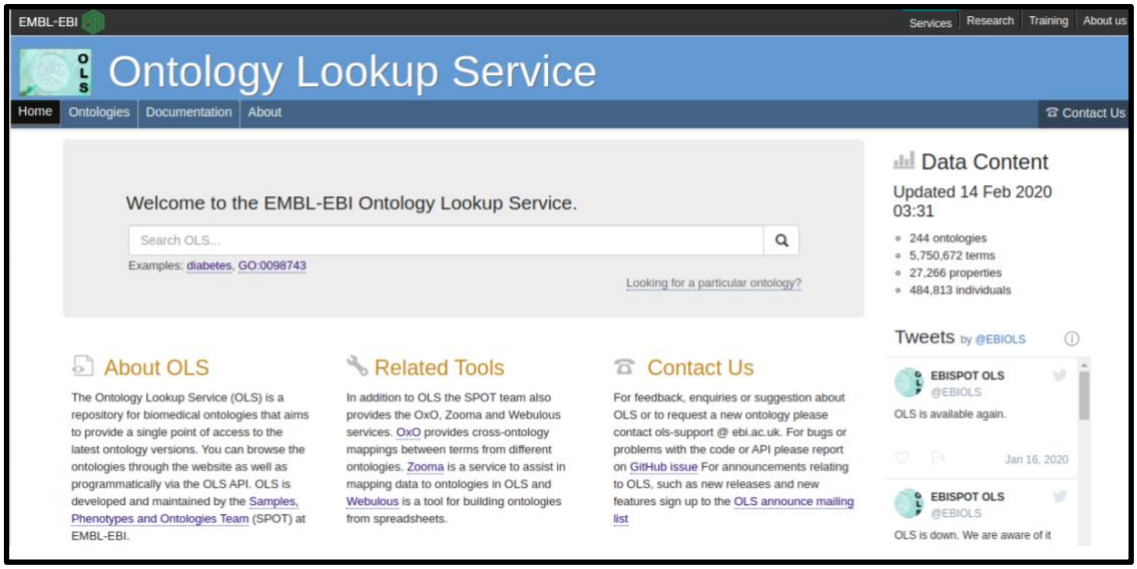
NB! Frequently, IDs from more than one ontology can be matched to the data label. Therefore, depending on use, the user may either map to all matching terms OR select one ontology ID to map to. The following guidelines can be used to prioritise ontologies

- 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): https://www.ebi.ac.uk/ols/index



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.

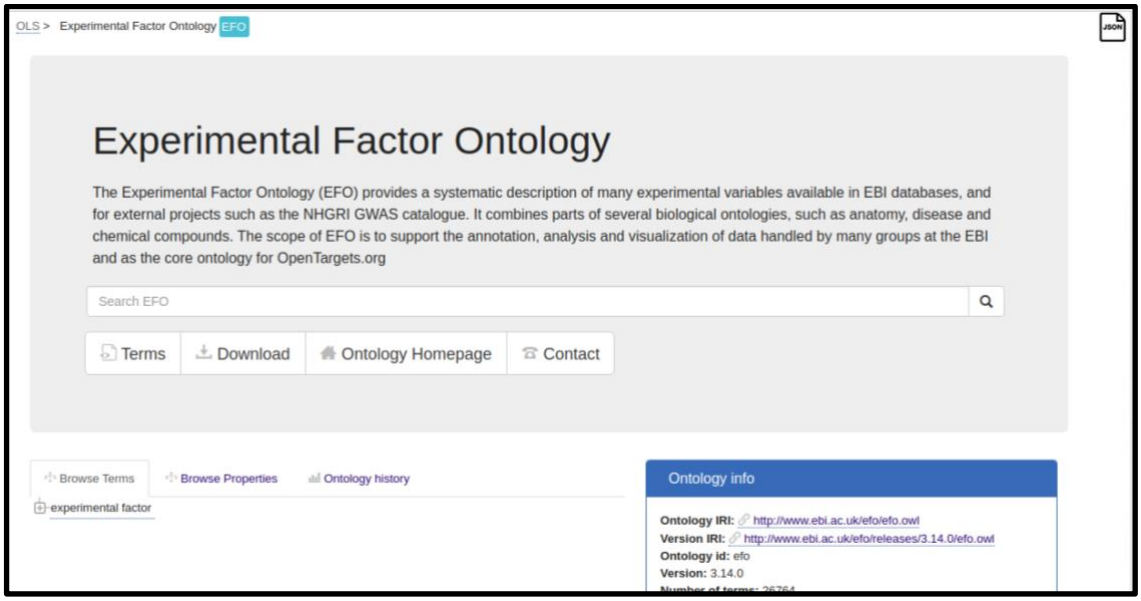
NB! If mapping for more than one project at a time, an additional step is required! In cases such as this, proceed to Step J.

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



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: https://www.ebi.ac.uk/ols/index

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.

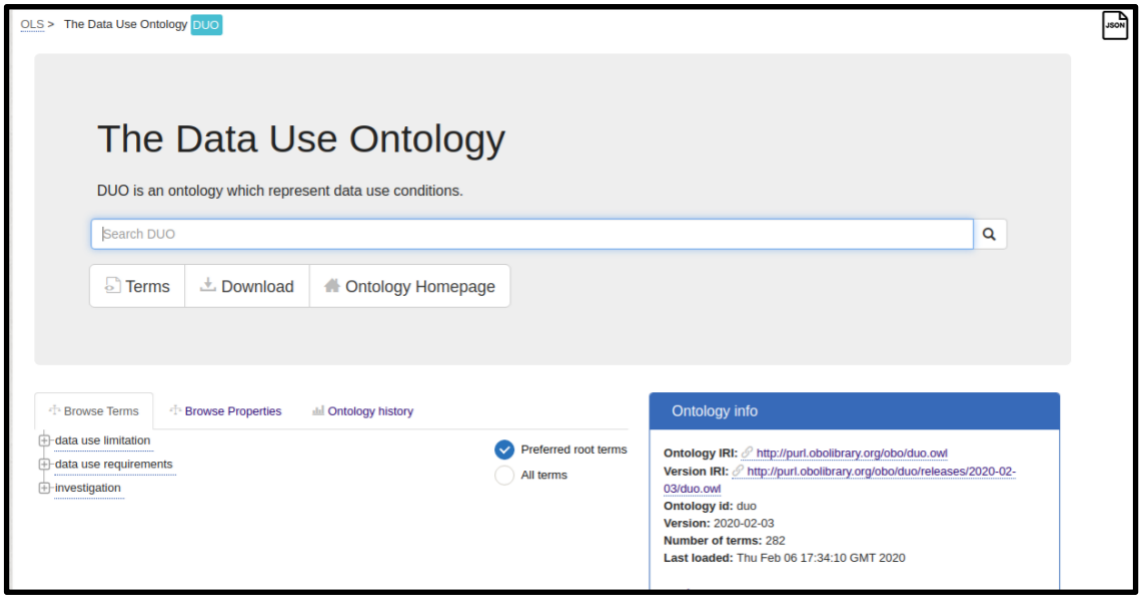
NB! If mapping for more than one project at a time, an additional step is required! In cases such as this, proceed to Step 9.

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:

https://www.ebi.ac.uk/ols/ontologies/duo



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/1E2w6Qvc9VRP5Xj0RiQcKcntTBLQLwjhH/view?usp=sharing*

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

> *Note #1: H3Africa consent forms are grouped by project on Google Drive: https://drive.google.com/drive/folders/1lTCm2UDbzp3RzGML3G8vRcvbmuh4qt3c*

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.