

Why mappings matter and how to make them FAIR

A Biomedical Perspective



Henriette Harmse

Ontology Tools Project Lead, EMBL-EBI

FAIR Impact

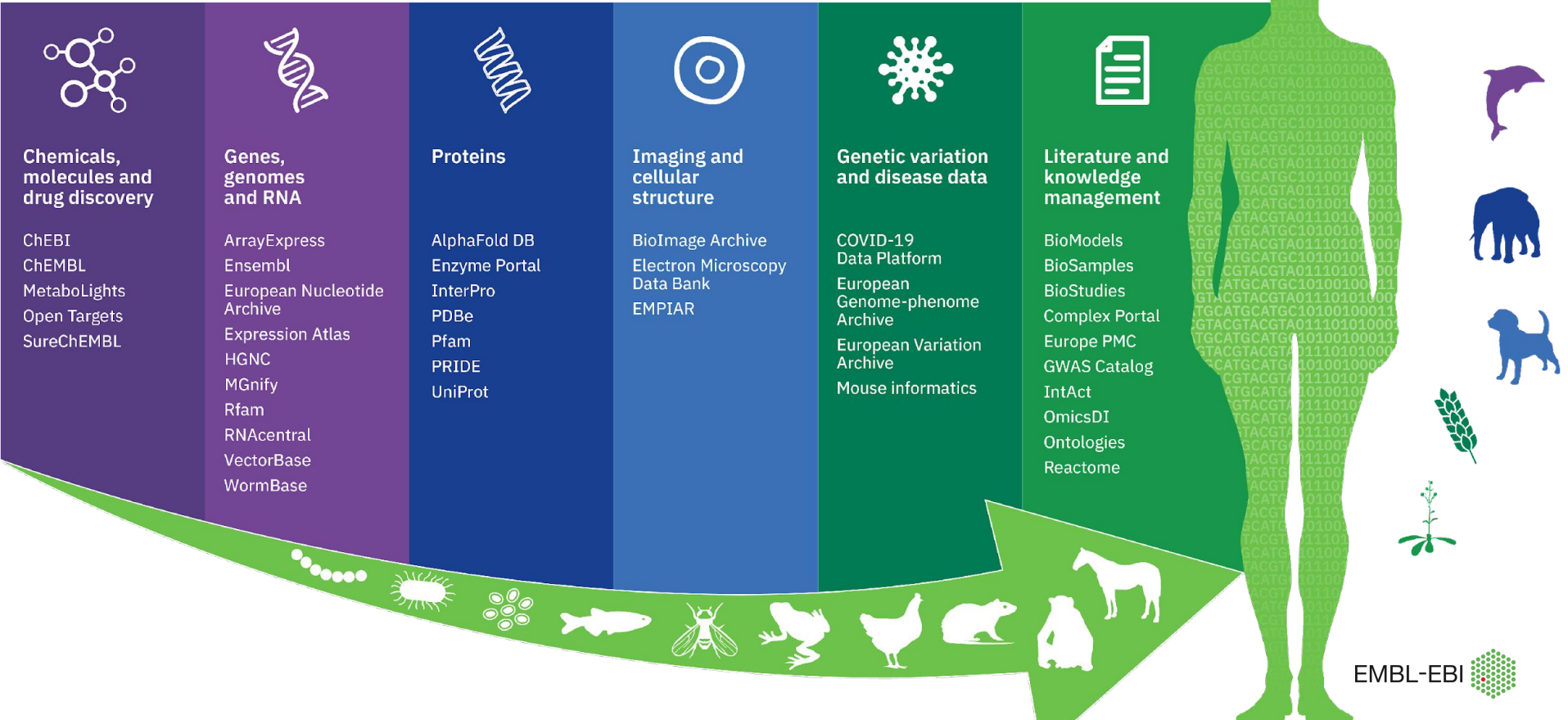
April 2023



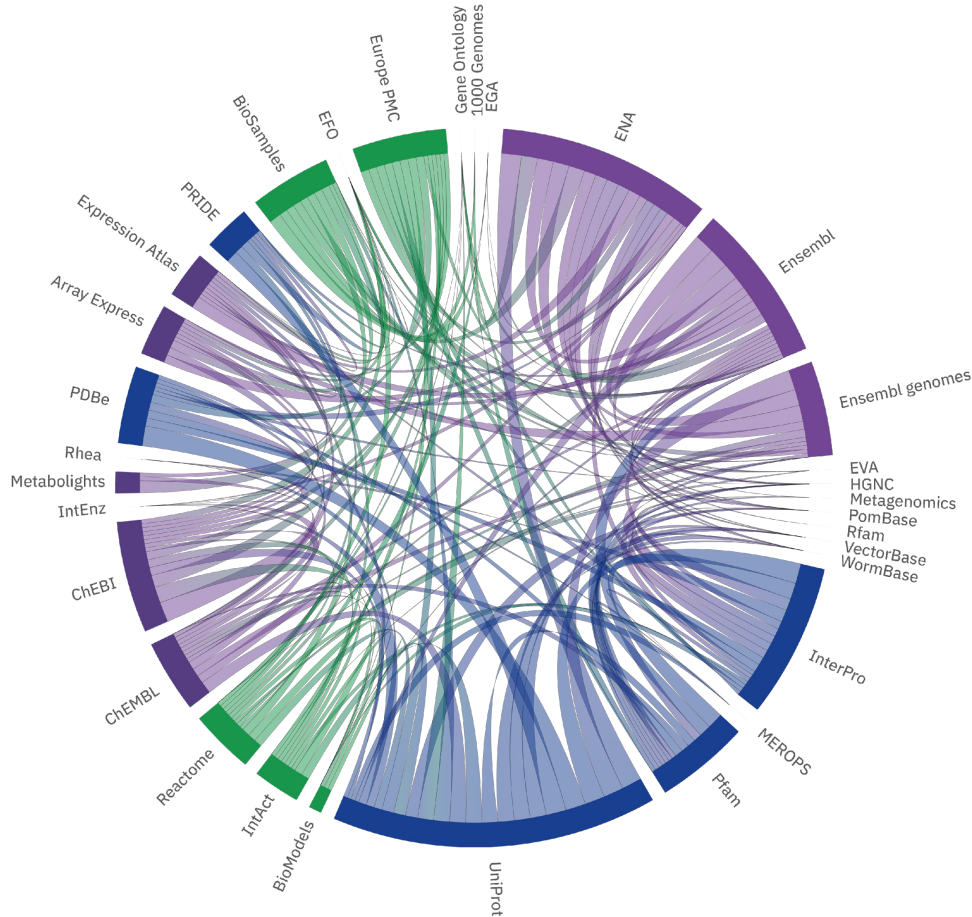
EMBL-EBI: Who are we?

- European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL), an international organization funded by over 20 member states.
- We are located at Wellcome Genome Campus 10 miles south of Cambridge, UK.
- We are a trusted source for biological and biomolecular data. Our core mission is to enable life science research and its translation to medicine, agriculture, industry and society.
- We have ±780 staff members from 66 nations.
- In 2021 we had 390 PB of raw data with 107 million requests to our websites and services coming from 41 million unique IP addresses.

The Data to Knowledge Spectrum at EMBL-EBI



Data integration across data resources



- Complex data flow between resources
- Interoperability between resources
- Aiming to minimize duplication of data and curation
- Direction of flow from archival data resources to knowledge bases
 - selection
 - adding value
 - QC
 - Update

Internal interactions between EBI data resources showing exchange of data. The width of each internal arc is weighted according to the number of different data types exchanged.

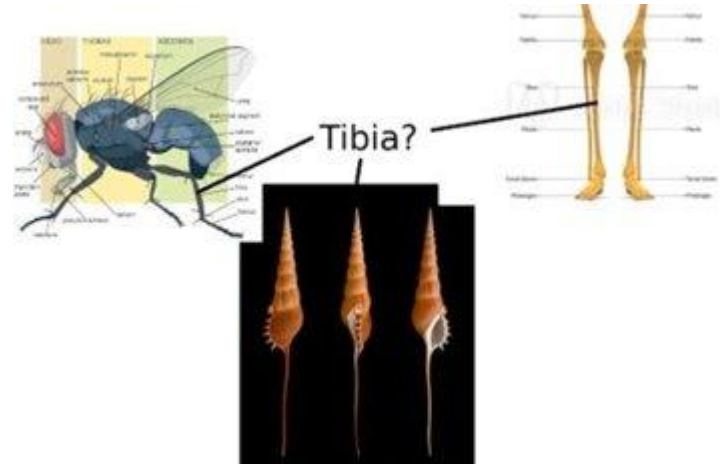
Two data related problems

Different words refer to the same concept

The same word refer to different concepts

18-day pregnant females	female (lactating)	individual female	worker caste (female)
2 yr old female	female (pregnant)	lgb*cc females	sex: female
400 yr. old female	female (outbred)	mare	female, other
adult female	female parent	female (worker)	female child
asexual female	female plant	monosex female	femal
castrate female	female with eggs	ovigerous female	3 female
cf. female	female worker	oviparous sexual females	female (phenotype)
cystocarpic female	female, 6-8 weeks old	worker bee	female mice
dikaryon	female, virgin	female enriched	female, sprayed
dioecious female	female, worker	pseudohermaphroditic female	femiale
diploid female	female(gynocious)	remale	metafemale
f	femele	semi-engorged female	sterile female
female	female, pooled	sexual oviparous female	normal female
female il	femalen	sterile female worker	sf
female	females	strictly female	vitellogenic replete female
female - worker	females only	tetraploid female	worker
female (afate sexual)	gynocious	thelytoky	hexaploid female
female (call)	healthy female	female (gynocious)	female (f-o)
hen	probably female (based on morphology)		

female [note: this sample was originally provided as a \"male\" sample to us and therefore labeled this way in the brawand et al. paper and original geo-submission; however, detailed data analyses carried out in the meantime clearly show that this sample stems from a female individual].



Courtesy of N. Silvester, European Nucleotide Archive, EMBL-EBI

Ontologies - Experimental Factor Ontology (EFO)

- EFO is designed to meet cross-domain EBI needs - application focused.
- Globally unique identifiers for concepts and relations, e.g. URI, IRI, PURL
- Machine readable syntax, e.g. XML, JSON-LD
- Defines a classification hierarchy from the most general to the most specific concepts.
- Depending on the formality of the ontology the hierarchy can be inferred based on reasoning.
- JSON-LD, RDFS and OWL are W3C standards.

What can possibly go wrong?

- Well... we are using ontologies ... at least >35 ontologies...
- This problem was already understood at EBI for some time and we created some tools
 - Ontology Lookup Service 2005
 - Zooma 2013
 - OxO 2016
 - OntoString 2021
- All are open source, web applications, REST API with JSON responses.



A case study - GWAS

Trait: myocardial infarction

GWAS Traits EFO_0000612

Trait information

Trait label myocardial infarction [Trait in OLS](#)

EFO ID EFO_0000612 [Trait in OXO](#)

Synonyms 28 synonyms [Trait in Open Targets](#)

Mapped terms - [Trait in PGS Catalog](#)

Description Gross necrosis of the myocardium, as a result of interruption of the blood supply to the area, as in coronary thrombosis. [+](#)

Reported Traits 24 reported traits [+](#)

Child traits 4 child traits [+](#)

[Highlighted Study](#)

Available data: [Associations 415](#) [Studies 62](#) [LocusZoom](#) [Download Catalog data](#)

Include background traits data

Include child trait data

Associations [415](#)

Variant and risk allele	P-value	P-value annotation	RAF	OR	Beta	CI	Mapped gene	Reported trait	Trait(s)	Background trait
rs544366796-G	4 x 10 ⁻¹⁰		0.009746	-	0.52 unit increase	[0.36-0.68]	SLC22A3	Myocardial infarction (PheCode 411.2)	acute myocardial infarction	-

GWAS Genes / SLC22A3

Gene information

Gene name SLC22A3

Description solute carrier family 22 member 3

Location 6:160348378-160452577

Cytogenetic region 6q25.3

Biotype protein_coding

[Gene in NCBI](#)

[Gene in Ensembl](#)

- [Phenotype information](#)
- [Pathway information](#)
- [Regulation](#)
- [Expression](#)

[Gene in Open Targets](#)

[Gene in OMM](#)

[Mouse ortholog in IMPC](#)

Available data: [Associations 208](#) [Studies 152](#) [Traits 208](#)

[Download Catalog data](#)

Why then support SSSOM?

myocardial infarction

http://www.ebi.ac.uk/efo/EFO_0000612 Copy

Gross necrosis of the myocardium, as a result of interruption of the blood supply to the area, as in coronary thrombosis. ⓘ

NECROSIS of the MYOCARDIUM caused by an obstruction of the blood supply to the heart (CORONARY CIRCULATION).

Synonym Attack - heart | Cardiac infarction, NOS | Cardiac infarction | Heart attack, NOS | heart attack ⓘ | Infarct, Myocardial | infarction (MI), myocardial ⓘ | INFARCTION (MI), MYOCARDIAL | Infarction of heart, NOS | Infarction of heart | Infarction, Myocardial | Infarctions, Myocardial | Infarcts, Myocardial | MI - Myocardial infarction | MI, MYOCARDIAL INFARCTION | MI, myocardial | infarction ⓘ | MI ⓘ | Myocardial Infarct | myocardial infarct ⓘ | myocardial infarction (disease) ⓘ | Myocardial infarction (disorder) | Myocardial infarction NOS | MYOCARDIAL INFARCTION, (MI) | myocardial infarction, (MI) ⓘ | Myocardial infarction, NOS | myocardial infarction ⓘ | Myocardial Infarctions | Myocardial Infarcts

Tree Graph

experimental factor (34,597)
└ material property BFO (16,973)
 └ disposition BFO (14,836)
 └ disease (14,785)
 └ cardiovascular disease (934)
 └ heart disease (488)
 └ myocardial disorder MONDO (19)
 └ myocardial infarction (6)

Show counts

Show obsolete terms

Show all siblings

Class Information

database_cross_reference

- MESH:D009203 ⓘ
- MedDRA:10028596
- OMIM:608446
- ICD10CM:I21 ⓘ
- NCIt:C27996
- myocardial infarction MONDO
- MeSH:D009203
- SCTID:22298006 ⓘ
- DOID:5844 ⓘ
- Myocardial Infarction NCIT ⓘ
- HP:0001658 🔍 2 ontologies ⓘ
- SNOMEDCT:22298006
- UMLS:C0027051 ⓘ

exactMatch

- <http://identifiers.org/snomedct/22298006>

gwas_trait

true

OBO foundry unique label

myocardial infarction (disease)

term editor

James Malone

- Semantics of mappings are not clear:
 - *dbXRef, narrower, broader, related.*
 - *A related B narrower C or A narrower B DbXref C or A narrower B broader C*
- Mappings often stored in ontology. In our example EFO:0000612 maps to SNOMEDCT:22298006. If SNOMEDCT:22298006 gets obsoleted & replaced, EFO has to change, but **nothing in EFO changed.**
- Provenance - i.e. author and date. As mapping information can be changed independently from ontology, term author and ontology version is a poor representation of mapping provenance.

OxO 2 - SSSOM

The screenshot displays the OxO 2 SSSOM web application interface. At the top, there is a navigation bar with the OxO logo and links for Home, Documentation, and About. A search bar contains the query "MP:0001262" and a "Search" button. Below the search bar, the results are displayed in a table format. The table has columns for Subject, Predicate, and Object. The Subject column shows "MP:0001262 (decreased body weight)". The Predicate column shows "http://www.w3.org/2004/02/skos/core#closeMatch", "http://www.w3.org/2004/02/skos/core#exactMatch", "http://www.w3.org/2002/07/owl#equivalentClass", and "http://www.w3.org/2004/02/skos/core#exactMatch". The Object column shows "HP:0004325 (Decreased body weight)". There are also filters for Mapping Justification and Predicate Id on the left side. The bottom of the page has a footer with "ABOUT SSSOM" and "PROJECT Issues / Feedback" links, and "EMBL-EBI 2023 Licensing" text.

Showing 4 from a total of 4

Confidence Download as... Show 10

Previous 1 Next

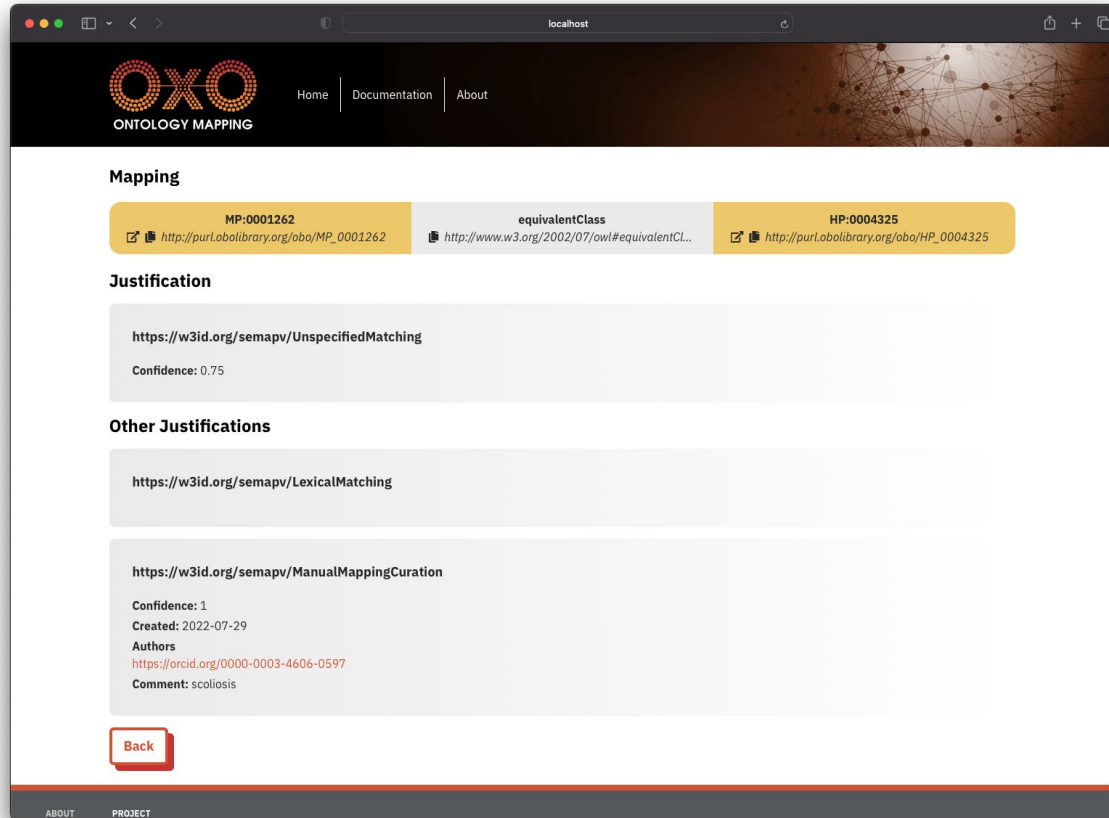
Mapping Justification	Subject	Predicate	Object
<input type="checkbox"/> LexicalMatching (2)	MP:0001262 (decreased body weight)	http://www.w3.org/2004/02/skos/core#closeMatch	HP:0004325 (Decreased body weight)
<input type="checkbox"/> UnspecifiedMatching (1)	MP:0001262 (decreased body weight)	http://www.w3.org/2004/02/skos/core#exactMatch	HP:0004325 (Decreased body weight)
<input type="checkbox"/> ManualMappingCuration (1)	MP:0001262	http://www.w3.org/2002/07/owl#equivalentClass	HP:0004325
<input type="checkbox"/> core#closeMatch (1)	MP:0001262 (decreased body weight)	http://www.w3.org/2004/02/skos/core#exactMatch	HP:0004325 (Decreased body weight)
<input type="checkbox"/> core#exactMatch (2)			
<input type="checkbox"/> owl#equivalentClass (1)			

Previous 1 Next

ABOUT PROJECT
SSSOM Issues / Feedback

EMBL-EBI 2023 Licensing

OxO 2 - SSSOM (Cont)



The screenshot displays the OxO 2 SSSOM web interface. At the top, there is a navigation bar with the OxO logo and the text "ONTOLOGY MAPPING". The main content area is divided into sections: "Mapping", "Justification", and "Other Justifications".

Mapping

MP:0001262 http://purl.obolibrary.org/obo/MP_0001262	equivalentClass http://www.w3.org/2002/07/owl#equivalentClass	HP:0004325 http://purl.obolibrary.org/obo/HP_0004325
--	---	--

Justification

<https://w3id.org/semapv/UnspecifiedMatching>
Confidence: 0.75

Other Justifications

<https://w3id.org/semapv/LexicalMatching>

<https://w3id.org/semapv/ManualMappingCuration>
Confidence: 1
Created: 2022-07-29
Authors
<https://orcid.org/0000-0003-4606-0597>
Comment: scoliosis

[Back](#)

ABOUT PROJECT

Acknowledgements – Funding and Collaborators



Open Targets

Collaborators



University of Colorado
Anschutz Medical Campus

NIH Common Fund Award to EMBL-EBI, QMUL, MRC Harwell **3UM1HG006370-10S2**

NIH NCI ITCR Award **5U24CA253539-02**

NHGRI GWAS Catalog - **5U41HG007823-08**

NHGRI Monarch Initiative - **2R24OD011883-10A1**

NHGRI Phenomics First - **7RM1HG010860-02**

MorPHiC - **1U24HG012674-01**

Health Data Research UK - PGS Catalog

EOSCLife has received funding from the European Union's Horizon 2020 programme under grant agreement number 824087

Intervene has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016775

OpenTargets funding for GWAS Catalog, eQTL Catalog and Ontology development

CZI for Human Cell Atlas

References

- [EMBL-EBI Highlights 2021](#)
- [The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries](#)
- [The Ontology Lookup Service: bigger and better](#)
- [OxO – a gravy of ontology mapping extracts](#)
- [The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource](#)

Applications

Tool	Address	Source
OLS	https://www.ebi.ac.uk/ols4	https://github.com/EBISPOT/ols4
OntoString	https://www.ebi.ac.uk/spot/ontostring	https://github.com/EBISPOT/OntoString
OxO	https://www.ebi.ac.uk/spot/oxo/	https://github.com/EBISPOT/OXO/
Zooma	https://www.ebi.ac.uk/spot/zooma/	https://github.com/EBISPOT/Zooma