

# OxO – a gravy of ontology mapping extracts

Simon Jupp, Thomas Liener, Sirarat Sarntivijai,

Olga Vrousitou, Tony Burdett and Helen Parkinson

<sup>1</sup> Samples, Phenotypes and Ontologies Team,  
European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory,  
Cambridge, United Kingdom

---

## 1 INTRODUCTION

Data is increasingly being annotated and described using controlled terminology or ontology standards. There are often multiple ontologies for any given domain, so the ability to map related or similar concepts is a necessary tool for data integration. Several techniques and tools have emerged that support ontology mapping, but finding and harmonising mappings from multiple sources remains a challenge for users. To address this we have developed OxO, a repository of known ontology mappings and cross-references extracted from multiple datasources. OxO provides a Web interface and API to access mappings and the functionality for users to upload their own sets of mappings.

Getting access to mappings is extremely valuable for large data integration efforts, such as the OpenTargets (Koscielny, An et al. 2017) platform, that pools data supporting gene-disease/phenotype associations from a wide variety of resources. OpenTargets requires that disease and phenotype information be normalised to terms in the Experimental Factor Ontology (Malone, Holloway et al. 2010). However, data coming into the platform may have been pre-annotated with a different ontology. For example, OpenTargets subsumes information from UniProt where disease terms are annotated with OMIM labels, while it identifies Orphanet Rare Disease Ontology (ORDO), which is imported by EFO, as the preferred standard for rare diseases in the platform. The OMIM accessions associated with UniProt diseases need to be consolidated to ORDO terms to unify the representation of disease in the platform. The Monarch Initiative (Mungall, Koehler et al. 2016) makes extensive use of cross-reference mappings to build a unified representation of disease and the Pistoia alliance is currently running the Ontologies Mapping Project to evaluate tools to support Ontology Mapping (<http://www.pistoiaalliance.org/ontologies-mapping-rfi-guidelines/>).

A good source of ontology mappings is within the ontologies themselves. Many of the Open Biomedical Ontologies use the xref property to indicate a cross-reference between entities. Whilst the semantics of xrefs are not explicitly defined by OBO, it has predominantly been used to indicate equivalence between terms. Mappings can be found in a number of other places, including some dedicated resources like the UMLS that provide a large number of mappings for medical terminologies. Although challenges remain when integrating term identifiers from multiple sources due to a lack of consistency in how identifiers are reported (McMurry et al. 2017). For example both MSH:D009202 and MESH:D009202 are both compact URIs (CURIEs) in common use for MeSH terms.

OxO extracts ontology mappings from multiple sources and harmonises the identifiers to provide an integrated resource of mappings. OxO constructs a graph of the mappings that allows users to explore how mappings from different resources intersect. The OxO mapping graph can be used to explore both direct and indirect mapping between terms, thus can be used to help traverse gaps where no direct mappings exist between two ontologies.

## 2 METHODS

OxO crawls the Ontology Lookup Service API (Jupp et al. 2015) to discover mappings based on annotation properties such as the OBO xref property. OxO uses identifiers.org, the OBO library, and prefixcommons.org to identify datasources using the identifier prefix. This is also used to assign identifiers to either an ontology/terminology category or a database category. For example, the Gene Ontology provides cross-reference mapping to databases like Reactome, whereas the Disease Ontology provides cross-references as mapping to other disease terminologies, such as MeSH. Additional mappings of interest to OpenTargets, such as disease terminology mappings for SNOMED-CT, ICD, Meddra, OMIM, NCIT and MeSH are also integrated

