

OxO – a gravy of ontology mapping extracts

Simon Jupp, Thomas Liener, Sirarat Sarntivijai,
Olga Vrousitou, Tony Burdett and Helen Parkinson

¹ 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

into OxO directly from the UMLS. Curators at the EBI are able to log into OxO and upload their own mappings and we encourage external users to submit any mappings they have generated to us for review and integration into OxO.

3 RESULTS

OxO identified 75 database sources that have records mapped to ontology terms and 104 ontologies that have mappings to another ontology or database. In total OxO identified over 1.4 millions xrefs for terms where we can automatically identify their source. There remain around 80 identifier prefixes that we are yet to map to an authoritative source¹. The OxO user interface (<http://www.ebi.ac.uk/spot/oxo>) allows users to explore mappings between terminologies and export the data in variety of formats.

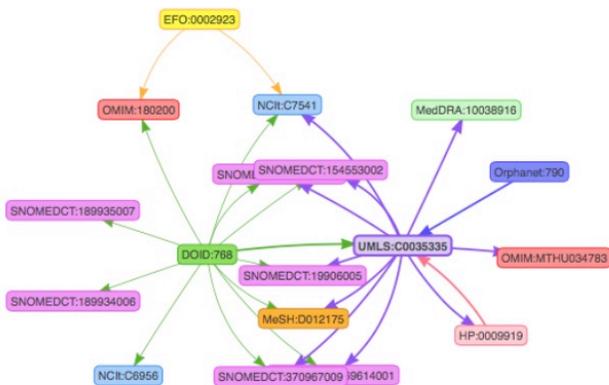


Figure 1. Mappings for NCI (Retinoblastoma) with evidence from over 6 sources (UMLS, HP, EFO, DOID, Orphanet and BAO)

Users can use the OxO graph to traverse potential gaps in coverage, for example, OxO found direct mappings for “Retinoblastoma” from NCI (NCI:C7541) to DOID:768 and UMLS:C0035335. By expanding the search space to 2 hops, we find additional mappings from NCI:C7541 to HP, SNOMEDCT, MEDDRA, ORPHANET, MeSH and OMIM that are derived from a number of different sources (See Figure 1.). The more sources that verify these mappings, the greater the confidence the user can have that these mappings are valid. OxO is also being integrated into the Ontology Lookup Service user interface, so users can view mappings for terms directly from OLS rather than the OxO interface.

4 CONCLUSION

Mappings between ontologies are a useful tool for data integration, but they are often poorly represented and semantically ambiguous. Cross-references are used to represent a range of mapping types including equivalence, subclass and related in some other way. There is still much work to do before we can infer equivalence through logical axioms within ontologies, so these kinds of mappings will remain a crucial tool, especially in the mapping of disease terminologies. By bringing these mapping together and harmonising the identifiers, OxO provides a platform for curators to collaborate on improving the mappings that exist. We plan to use OxO as background knowledge to feed into automated ontology matching algorithms such as LogMap (Jiménez-Ruiz and Cuenca Grau 2011) to compute a reference set of gold standard mapping for a number of disease ontologies.

OxO is open source and source code is available from <https://github.com/EBISPOT/OLS-mapping-service>

ACKNOWLEDGEMENTS

This resource is funded in part by EMBL-EBI core funds, CORBEL This project receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654248.

EXCELERATE ELIXIR-EXCELERATE is funded by the European Commission within the Research Infrastructures programme of Horizon 2020, grant agreement number 676559. The Open Targets project and the organisers and participants of the 2016 BioHackathon in Tsuruoka, Japan.

REFERENCES

- Jiménez-Ruiz, E. and B. Cuenca Grau (2011). LogMap: Logic-Based and Scalable Ontology Matching. The Semantic Web – ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23–27, 2011, Proceedings, Part I. L. Aroyo, C. Welty, H. Alani et al. Berlin, Heidelberg, Springer Berlin Heidelberg: 273–288.
- Koscielny, G., et al. (2017). "Open Targets: a platform for therapeutic target identification and validation." *Nucleic Acids Research* 45(D1): D985–D994.
- Malone, J., et al. (2010). "Modeling sample variables with an Experimental Factor Ontology." *Bioinformatics* 26(8): 1112–1118.
- Mungall, C. J., et al. (2016). "k-BOOM: A Bayesian approach to ontology structure inference, with applications in disease ontology construction." <http://biorxiv.org/content/biorxiv/early/2016/04/15/048843.full.pdf>
- Jupp S. et al. (2015) A new Ontology Lookup Service at EMBL-EBI. In: Malone, J. et al. (eds.) Proceedings of SWAT4LS International Conference 2015"
- Julie McMurry, & 40 additional authors (see file). (2016). Identifiers for the 21st century: How to design, provision, and reuse identifiers to maximize data utility and impact. <http://doi.org/10.5281/zenodo.163459>

¹ Unmapped prefixes list <https://goo.gl/xasfND>