OLOBO: A new ontology for linking and integrating open biological and biomedical ontologies

Edison Ong[1], Yongqun He[2,*]

[1] Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

[2] Unit of Laboratory Animal Medicine, Department of Microbiology and Immunology, and Center of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

ABSTRACT

To link and integrate different Open Biological and Biomedical Ontologies (OBO) library ontologies, we generated an OLOBO ontology that uses the Basic Formal Ontology (BFO) as the upper level ontology to align, link, and integrate all existing 9 formal OBO Foundry ontologies. Our study identified 524 ontology terms shared by at least two ontologies, which form the core part of OLOBO. Current OLOBO includes 978 ontology classes, 13 object properties, and 90 annotation properties. OLOBO was queried, analyzed, and applied to identify potential deprecated terms and discrepancy among different OBO library ontologies.

1 INTRODUCTION

The OBO Foundry aims at establishing a set of ontology development principles (e.g., openness, collaboration) and incorporating ontologies following these principles in an evolving non-redundant suite (Smith et al., 2007). Currently, there are >150 ontologies in the OBO ontology library. These ontologies include 9 ontologies formally investigated and found to have satisfied the OBO principles and recommended as official foundry ontologies. The other ontologies have not been officially investigated and thus been labeled as candidate members of the OBO ontology library.

Ontology integration is a major issue. The bottleneck of ontology disintegration may also exist among OBO library ontologies (Ghazvinian et al., 2011). To support better ontology interoperation, we have developed a new ontology named “OLOBO” – the Ontology for Linking and Integrating Open Biological and Biomedical Ontologies.

2 METHODS

2.1 Collection of OBO ontologies

Instead of using all >150 ontologies, we chose the official 9 OBO Foundry ontologies (Table 1). OBO Foundry website includes their default ontology download websites. We downloaded the OWL source code of these ontologies.

2.2 Development of OLOBO

We developed an in-house Java program that uses OWLAPI to process ontology OWL files. Many ontologies also reuse other ontology terms. To integrate all the ontologies, we first identified and collected all ontology terms shared by at least two ontologies. This step identified the minimal sets of ontology terms shared by different ontologies. To notify from which ontology a term is imported, we generated a new annotation property called ‘*term from ontology*’ (OLOBO_0000001). Next, we used Ontofox (Xiang et al., 2010) to identify terms and axioms related to extracted ontology terms. To align with BFO, we manually examined ontologies and ensure appropriate upper level terms added for the purpose of ontology linkage and integration.

2.3 OLOBO source code, deposition, and queries

The OLOBO source code is openly available at GitHub: https://github.com/biomedontology/olobo. OLOBO is deposited in Ontobee: http://www.ontobee.org/ontology/OLOBO. Ontobee SPARQL queries was used to query OLOBO (Ong et al., 2017).

2.4 OLOBO applications

OLOBO was used to identify deprecated terms, and we used the Ontobeep tool (http://www.ontobee.org/ontobeep) to compare with OLOBO and other OBO ontologies to identify potential discrepancy among ontologies.

#	Prefix	Ontology full name	# of terms	# of shared terms
1	BFO	Basic Formal Ontology	55	55
2	ChEBI	Chemical Entities of Biological Interest	107,537	91
3	DOID	Human Disease Ontology	38,300	31
4	GO	Gene Ontology	48,745	10
5	OBI	O. for Biomedical Investigations	3,417	368
6	PATO	Phenotypic quality O.	2,664	79
7	PO	Plant Ontology	2,035	38
8	PRO	Protein Ontology	311,325	201
9	XAO	Xenopus anatomy and development	1,661	30
10	ZFA	Zebrafish anatomy and development	3,206	31

Table 1. Ontologies used as input for OLOBO development

* To whom correspondence should be addressed: yongqunh@med.umich.edu

3 RESULTS

3.1 Alignment of OBO ontologies in OLOBO

Our study found 524 classes, object properties, and annotation properties shared by at least two ontologies. In total, our pipeline removed 178,625 terms from the original 9 OBO ontologies. The OLOBO statistics is available on Ontobee: http://www.ontobee.org/ontostat/OLOBO. OLOBO includes 978 classes, and 103 object or annotation properties.

The overall OLOBO structure (Fig. 1) is aligned with BFO (http://ifomis.uni-saarland.de/bfo/), the recommended upper-level ontology in the OBO Foundry, and >100 ontologies are aligned to BFO. The chosen 9 ontologies are all aligned with BFO as defined by the ontologies themselves or manually by us.

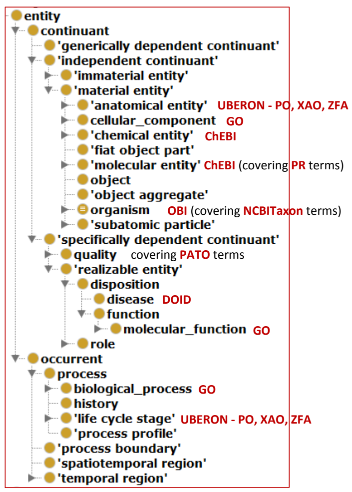

Fig. 1. Top level hierarchy of OLOBO.

3.2 OLOBO applications

Our study identified three deprecated PATO terms, PATO_0001237 (quality of a single physical entity), PATO_0001238 (quality of related physical entities), and PATO_0001631 (relational spatial quality), which are still used in OBI. PATO has recommended new PATO terms to replace these obsolete terms, e.g., 'physical object quality' (PATO_0001241) to replace PATO_0001238. We thus submitted a track issue to OBI (https://github.com/obi-ontology/obi/issues/826) to indicate such an issue.

Another application was to use OLOBO for ontology comparison and discrepancy detection. Ontobeep is an Ontobee-based ontology alignment and comparison tool. Our usage of Ontobeep identified that the UBERON term 'anatomical entity', which are included in OLOBO and OBI, also exists in CLO but under a different parent term. Such an ontology alignment and surveillance study is expected to identify achievements as well problematic issues in the top-down alignment study. These will significantly support collaborative ontology development and reuse.

4 DISCUSSION

The basic idea of our project is to generate an integrated upper level application ontology that links OBO ontologies in an integrated format. This study only aligned existing 9 official OBO ontologies. We plan to extend OLOBO to include the other candidate OBO ontologies.

The challenge of ontology integration and orthogonality is real. Using ontology term labels to analyze OBO ontology orthogonality (Ghazvinian et al., 2011), a previous study found that although progress made, a large amount of overlap remained among ontologies and achieving orthogonality would be difficult. Our study proposed and provided a first solution to address this challenge.

In addition to the demonstrated OLOBO usages, OLOBO may be used in other cases. For example, to develop a new OBO-oriented ontology, we may use OLOBO as the starting point to facilitate the ontology development process and lead to more efficient alignment. The usage of OLOBO may also provide an integrated ontology system for efficient software development for features such as consistency checking and standard integration. Given the large number of terms among different ontologies, it is practically infeasible to develop a single ontology that covers all ontology terms from all possible ontologies. However, OLOBO offers an integrative framework to link and integrate different ontologies in a semantically consistent strategy.

ACKNOWLEDGEMENTS

This work was partially supported by a BD2K-LINCS DCIC external data science research award, which is part of the grant U54HL127624 (BD2K LINCS Data Coordination and Integration Center, DCIC) awarded by the National Heart, Lung, and Blood Institute through funds provided by the trans-NIH LINCS Program and the trans-NIH Big Data to Knowledge (BD2K) initiative (http://www.bd2k.nih.gov). LINCS is an NIH Common Fund projects.

REFERENCES

Ghazvinian, A., Noy, N.F., and Musen, M.A. (2011). How orthogonal are the OBO Foundry ontologies? *J Biomed Semantics* **2 Suppl 2,** S2.

Ong, E., Xiang, Z., Zhao, B., Liu, Y., Lin, Y., Zheng, J., Mungall, C., Courtot, M., Ruttenberg, A., and He, Y. (2017). Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Res* **45,** D347-D352.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., The OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., and Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* **25**(11**),** 1251-1255.

Xiang, Z., Courtot, M., Brinkman, R.R., Ruttenberg, A., and He, Y. (2010). OntoFox: web-based support for ontology reuse. *BMC Res Notes* **3**(175)**,** 1-12.