ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

Developing a Reagent Application Ontology within the OBO Foundry Framework

Matthew H. Brush[1], Nicole Vasilevsky[1], Carlo Torniai[1],
Tenille Johnson[2], Chris Shaffer[1], Melissa A. Haendel[1]

[1]Oregon Health and Science University, Portland, OR, USA; [2]Harvard University, Cambridge, MA, USA

1 Introduction

In light of the increasing complexity and cost of biomedical research, the ability to find and re-use research resources has emerged as an important challenge for information technologies to address. Among the most expensive resources to produce, and challenging to discover, are the material reagents employed in performing biomedical experiments. The eagle-i Consortium (www.eagle-i.org/home) represents an open and collaborative effort to develop ontology-driven applications aimed at cataloging biomedical research resources, including reagents, and making them available through a semantic search portal. Here, we describe the development of a reagent ontology module within our larger eagle-i ontology, that uses Open Biomedical Ontologies (OBO) Foundry [1] best practices to facilitate interoperability and extend external ontology efforts.

As part of an 'application ontology', key goals for the eagle-i reagent module were to represent real data collected from research laboratories, and to drive the user-interface (UI) and application logic of data collection and search tools. Reagent modeling was informed by analysis of data collected from participating laboratories, whereby seven top-level reagent categories were identified. These included 'Antibody,' 'Cell Line,' 'Chemical Reagent,' 'Construct,' 'Nucleic Acid Reagent,' 'Protein Reagent,' and 'Reagent Library' classes. An equally important goal in developing the reagent module was to facilitate interoperability and linking of data across systems, such that reagents described in eagle-i could be linked to experiments, protocols, publications, or data sets cataloged in external sources. This Linked Open Data (LOD) [2] approach can add significant value to data repositories, particularly when their underlying ontology models are logically integrated to allow computational reasoning across linked bodies of data. This approach serves not only to enhance the performance of the eagle-i search portal by exploding the information linked to cataloged resources, but can also support a range of external applications that can be fed by the data housed in the eagle-i repository. To promote interoperability and publication of LOD, we looked to OBO Foundry Library as set of orthogonal ontologies with which to align our efforts. The OBO Foundry provides an evolving set of shared principles for ontology development, and houses a collection of 'reference ontologies' designed according to these standards. In developing our reagent ontology module, OBO principles and reference ontologies were consulted and reused as described below.

2 The Ontological Nature of Reagents

The term 'reagent' means different things to different people. For example, in chemistry a reagent is considered an inert substance that catalyzes a reaction, while in biomedical research, reagents span a broader range of granularities, and might be any material entity input into an experiment (drugs, antibodies, cell lines, etc). In this sense, reagents are defined by their playing a role a scientific experiment. For instance, antibodies exist naturally in organisms where they serve to fight infection, but those applied to detect some analyte in an experimental setting are considered reagents. Similarly, many chemical compounds are produced naturally through biological processes (antibiotics, toxins, etc), but qualify as reagents only if used in a research investigation. To capture this idea, we have defined reagents as "material entities used in an experimental process to detect, measure,

examine, or produce other substances". Accordingly, classification of a material entity as a reagent depends not on some shared physical attribute, but rather on its use in a particular context (biomedical experimentation) and for a specific purpose (to generate data or other materials for experimentation). Therefore, rather than assert that material entities are members of a reagent class in our ontology, they are classified with other entities that share common inherent physical features, and inferred to be reagents based on an axiom asserting their role in scientific experimentation. For example, a 'plasmid' is classified as a subtype of 'double-stranded DNA', and described by an axiom that indicates it to play a 'reagent role'. As described below, the OWL-DL language in which our ontology is written offers mechanisms for using such axioms to generate a unified hierarchy of reagents that meets our application needs.

3 Preliminary Landscape Analysis

Two key principles advocated by the OBO Foundry are *interoperability*, whereby existing ontologies and classes are re-used whenever possible, and *orthogonality*, whereby modeling of a particular domain converges upon a single reference ontology. In keeping with these principles, we performed a preliminary analysis of existing ontological representations of reagents to determine whether any might offer a single, comprehensive model that could be adapted to meet our application-specific needs. An examination of the 266 ontologies cataloged by the National Center for Biomedical Ontology (NCBO) Bioportal [3] revealed very limited modeling of reagents. A 'reagent' class appeared in fewer than ten ontologies, and none of these offered sufficient logical descriptions or classification of reagent subtypes for our application needs. What sparse modeling had been done was inconsistent, with 'reagent' classes being modeled alternately as 'material entities', 'roles', and 'features' in different ontologies. This modeling was also restricted in scope, often describing only those reagents specific for a particular domain, application, or granularity. Therefore, convinced that no existing resource offered a suitable, comprehensive, and broadly applicable model of reagents for our needs or that of the community at large, we set out to construct one.

4 Modeling Approach

Having found no suitable representation of reagents among existing ontologies, our landscape analysis was extended with two goals in mind: (1) to identify "source" ontologies from which to re-use individual classes representing reagent types (ensuring interoperability); and (2) to identify an appropriate "home" ontology in which to re-use these classes to construct our reagent model (ensuring orthogonality). Below, we describe our plans for identifying source and home ontologies, and their roles in developing a stand-alone reagent ontology module that meets eagle-i application needs.

4.1 Source Ontologies for Reagents

Reagents are represented across many levels of granularity – from small molecules such as drugs and chemicals, to biological macromolecules such as proteins or DNA constructs, to cells and cell lines, to libraries comprised of large collections of peptides or genomic clones. Many of these reagent types are represented across the OBO library of ontologies, but are not defined therein as reagents. For example, proteins are modeled in the Protein Ontology (PRO) [4], chemicals are represented in the Chemical Entities of Biological Interest Ontology (ChEBI) [5], and many types of nucleic acids are modeled in both ChEBI and the Sequence Ontology (SO) [6]. When materials representing reagents in our seven top-level categories fall within the scope of existing ontologies, relevant classes from these sources will be imported into our home ontology, where they will be extended to model reagents. For example, because all instances of eagle-i 'Construct' reagents are types of 'ChEBI: nucleic acid > DNA > double-stranded DNA', these classes will be imported from ChEBI into our home ontology. Here, construct reagents and their more specific subtypes (e.g. 'plasmid', 'viral plasmid', 'retroviral plasmid'), which lie outside the scope of ChEBI, will be modeled beneath the ChEBI 'double-stranded DNA' class.

4.2 A Home Ontology for Modeling Reagents

As reagent classes are defined by their participation in experimental processes, we determined that the process-oriented Ontology for Biomedical Investigations (OBI) [7] would serve as a suitable "home" for our modeling efforts. OBI is an actively developed OBO foundry candidate ontology driven by representatives from over 20 research communities, which describes all phases of biomedical investigations. OBI also models the material entities that participate in these processes, including reagents, instruments, specimens, and agents. Furthermore, OBI has defined a mechanism for importing (re-using) terms from external sources, called MIREOT (Minimum Information to Reference an External Ontology Term) [8]. Thus, classes representing reagents that are implemented in various "source" ontologies can be imported into a single "home" ontology (OBI) using the MIREOT principle.

Once all relevant classes are assembled in OBI, we will be able to infer a stand-alone hierarchy of reagents sufficient to drive our suite of eagle-i applications. This will be achieved by attaching logical axioms to all classes representing reagents asserting that they *necessarily* bear a reagent role. A 'Reagent' equivalent class can then be created, and defined with a *necessary and sufficient* axiom equating it to any 'material entity' that has a 'reagent role'. An OWL-DL reasoner can then generate an inferred 'Reagent' hierarchy that unites all reagents in OBI. This will provide a single reagent module, encoded in OWL/RDF, that can be imported into our eagle-i ontology to drive application functionality, while also providing interoperability with external efforts that point to shared OBO classes we used in our model.

5 Conclusions

Our approach for application ontology development aims to balance practical project requirements with OBO best-practices that support interoperability and orthogonality. While this approach creates some initial overhead in the form of a landscape analysis and technical hurdles to re-using of existing terminologies, it offers many long-term benefits for both application developers and the community at large. For eagle-i, our reagent modeling efforts have been enhanced in many ways by aligning with OBO principles. First, we have benefited from the collaborative nature of ontology development, applying feedback and resources from Foundry members toward improving our reagent model. Second, we are contributing back to the community by extending several Foundry ontologies to include classes and properties relevant to our domain of interest. Third, we are enabling our application to access and use data from external systems built on models that comply with OBO foundry principles. Finally, we hope to provide a reagent ontology module suitable for re-use in the community, which will continue to add value to eagle-i by expanding data available to the system.

References

1. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 2007;25(11):1251-5.
2. Bizer C HT, Berners-Lee, T. Linked data-the story so far. Int. J. Semantic Web Inf. Syst. 2009;5:1-22.
3. Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, et al. National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. Omics 2006;10(2):185-98.
4. Natale DA, Arighi CN, Barker WC, Blake JA, Bult CJ, Caudy M, et al. The Protein Ontology: a structured representation of protein forms and complexes. Nucleic Acids Res;39(Database issue):D539-45.
5. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 2008;36(Database issue):D344-50.
6. Eilbeck K, Lewis SE, Mungall CJ, et al. The Sequence Ontology: a tool for the unification of genome annotations. Genome Biol 2005;6(5):R44.
7. Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, et al. Modeling biomedical experimental processes with OBI. J Biomed Semantics;1 Suppl 1:S7.
8. Courtot M GF, Lister AL, Malone J, Schober D, Brinkman RR, Ruttenberg A. MIREOT: the minimum information to reference an external ontology term. Nature Proceedings 2009.