The ImmPort Antibody Ontology

William Duncan[1], Travis Allen[1,2], Jonathan Bona[3], Olivia Helfer[1], Barry Smith[1,2,3], Alan Ruttenberg[4], Alexander D. Diehl[1,3,5]
[1]NYS Center of Excellence in Bioinformatics and Life Sciences, [2]Department of Philosophy, [3]Department of Biomedical Informatics, [4]Oral Diagnostics Sciences, [5]Department of Neurology
University at Buffalo
Buffalo, NY, USA
addiehl@buffalo.edu

I. Introduction

Monoclonal antibodies are essential biomedical research and clinical reagents that are produced by companies and research laboratories. The NIAID ImmPort (Immunology Database and Analysis Portal) resource provides a long-term, sustainable data warehouse for immunological data generated by NIAID, DAIT and DMID funded investigators for data archiving and re-use [1]. A variety of immunological data is generated using techniques that rely upon monoclonal antibody reagents, including flow cytometry, immunofluorescence, and ELISA. In order to facilitate querying, integration, and reuse of data, standardized terminology for describing monoclonal antibody reagents and their targets needs to be used for annotating data submitted to ImmPort.

A major problem with monoclonal antibody-associated data is that data producers typically report antibody clones or target markers using non-standardized terminology:

- CD3 vs. CD3e (protein names)
- HIT3e vs. UCHT1 (antibody clones for CD3e)
- 550367 vs. 300401 (catalog numbers for anti-CD3e antibody reagents)

In order to address this problem, we have created the ImmPort Antibody Ontology (AntiO) to provide a source of standardized names for monoclonal antibodies and their protein targets for use by ImmPort investigators and the scientific community in general, and to provide robust querying for monoclonal antibody reagents via a variety of criteria.

II. Methods

We curated monoclonal antibody-protein target relationships by identifying names and information about monoclonal antibodies based on published papers, data submissions to ImmPort, and commercial monoclonal products for immunology research such as the BD Lyoplate products. We selected standardized monoclonal antibody names (clone names) and curated information about the protein targets of the antibodies using Protein Ontology and UniProt identifiers [2]. For both the monoclonal antibody clone names, and the protein targets of the monoclonal antibodies, we have included many additional synonyms to facilitate querying.

The resulting antibody registry was transformed into the AnitO ontology using the Reagent Ontology (ReO) as a paradigm for the representation of monoclonal antibody reagents [3]. Monoclonal antibodies are classified via isotype and species of origin and are formally related to their protein targets via the *recognizes* relation. For example, monoclonal antibody clone HI100 *recognizes* **some** 'receptor-type tyrosine-protein phosphatase C isoform CD45RA'. We supplemented the information in AntiO by creating classes for entries in the NIF Antibody Registry [4] that represent products that contain particular monoclonal antibody clones. These classes are types of 'monoclonal antibody offering' in our ontology and are linked to clone name classes via *has_part* relations. We have also mined and standardized additional information from the NIF Antibody Registry that is associated with particular monoclonal antibody offering classes, including information about product vendors, catalog numbers, conjugations (fluorchromes, biotin, etc.) of antibody products, antibody species specificity, and experimental usage.

AntiO is built in an automated fashion using scripts that combine information about monoclonal antibodies and their targets found in curated spreadsheets with information text-mined from relevant NIF Antibody Registry entries to create a base set of OWL2 modular ontologies that are imported into the AntiO ontology (see Figure 1) along with import files for ReO and Protein Ontology terms. Additional terms from the Ontology for Biomedical Investigations [5], the BioAssay Ontology [6], the Molecular Interactions Ontology [7], and the NCBI Taxonomy [8] are included as MIREOT'ed terms as well [9]. The resulting combined ontology is viewable and queryable in Protégé 5 [10], and is loaded into a publicly available RDF triple store for SPARQL queries.

III. Results

AntiO contains 941 monoclonal antibodies of common use in immunology experiments, and represents about 30,000 monoclonal antibody products from 80 vendors based on information derived from the NIF Antibody Registry. We have included the NIF 'AB_XXXXXX' identifiers as part of our monoclonal antibody offering labels

The AntiO triple store is based on OWLIM [11], is pre-reasoned, and contains over a million RDF triples. A variety of queries using AntiO are possible. One can for instance search for all monoclonal antibodies that have a particular protein target (Figure 2). Or, similarly, all monoclonal antibody offerings (products) from a given vendor that have a particular target. More indirect querying is possible; for instance, one can

Supported by NIGMS 2R01GM080646 (Protein Ontology), NIAID HHSN272201200028C (ImmPort), NIAID HHSN272201200028C (HIPC).

search for the protein targets of monoclonal antibodies using only the catalog number of the products used. There are additional ways to search as well; one can limit searches to antibodies that work only in particular types of experiments, for instance. We have created a Bitbucket repository and wiki to provide information about the ontology, as well as example SPARQL queries (see Table 1 for URLs).

TABLE I

Important URLs	
AntiO Triple Store	http://protein.ctde.net:8080/openrdf-workbench/repositories/antio/query
AntiO Wiki	https://bitbucket.org/wdduncan/antio/wiki/Home

IV. Discussion

Through careful curation and data extraction using computer programs, we have developed an ontology of monoclonal antibodies used in immunological research with a focus on ImmPort clinical studies and other recently published papers in immunology. Our effort developing AntiO is complementary to existing antibody registries. While such resources let researchers find useful antibodies and the companies that produce them, they do not provide standardized terms for clone names, targets of the antibodies, conjugations, etc. and so are difficult to use computationally. In collaboration with the NIH-funded NIF Antibody Registry, we have developed a framework that will allow researchers to more easily query for monoclonal antibodies, the vendors that sell them, and their protein targets and experimental usage, and provides standardized terminology for all these data types and more. Our long-term goal is to develop web interfaces that will enable submitters of data not only to query for monoclonal antibodies and their targets, but also facilitate the finding of experimental results, such as clinical studies within the ImmPort system, in which particular monoclonal antibodies were used.

Of further note is our reuse within AntiO of the compiled NIF Antibody Registry data on antibody products, which is part of the Research Resource Identification Project [4]. By associating the monoclonal antibody offerings in AntiO with the RRIDs provided by NIF Antibody Registry, we ensure AntiO contributes to the goals of the Research Resource Identification Project by linking to this common resource to enable better reuse and integration of scientific data while adding value to the NIF Antibody Registry data through our careful curation and standardization steps.

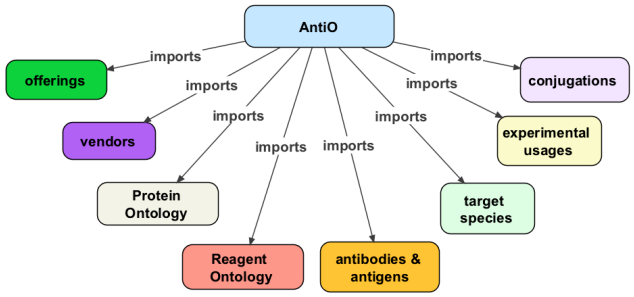

Fig. 1. AntiO Ontology ImmPort Schema

Acknowledgment

We thank Sanchita Bhattacharya, Patrick Dunn, Atul Butte, Matthew Brush, Melissa Haendel, and Anita Bandrowski for helpful comments and support.

References

[1] Bhattacharya S, et al., "ImmPort: disseminating data to the public for the future of immunology," Immunol Res. 2014, 58:234-9.

[2] Natale DA, et al., "Protein Ontology: a controlled structured network of protein entities," Nucleic Acids Res. 2014, 42:D415-21.

[3] Brush MH, et al., "Developing a Reagent Application Ontology within the OBO Foundry," 2011, http://ceur-ws.org/Vol-833/paper32.pdf.

[4] Bandrowski A, et al., "The Resource Identification Initiative: A cultural shift in publishing," F1000Res. 2015, 4:134.

[5] Bandrowski A, et al., "The Ontology for Biomedical Investigations," PLoS One. 2016, 11:e0154556.

[6] Visser U, et al., "BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results," BMC Bioinformatics. 2011, 12:257.

[7] Orchard S, Kerrien S, "Molecular interactions and data standardisation," Methods Mol Biol. 2010, 604:309-18.

[8] Sayers EW, et al., "Database resources of the National Center for Biotechnology Information," Nucleic Acids Res. 2009, 37:D5-15.

[9] Courtot M, et al. "MIREOT: The minimum information to reference an external ontology term," Applied Ontology. 2011, 6:23-33.

[10] http://protege.stanford.edu

[11] Kiryakov A, Ognyanov D, Manov D, "OWLIM–a pragmatic semantic repository for OWL." 2005, In Web Information Systems Engineering–WISE 2005 Workshops, Springer Berlin Heidelberg.


```
SELECT distinct ?offering ?vendor ?clone ?target
WHERE {
  ?offeringt rdfs:subClassOf offering: . ?vendori rdf:type vendor: .
  ?clonet rdfs:subClassOf mAB: . ?targett rdfs:subClassOf protein: .

  ?r1 owl:onProperty has_part: . ?r1 owl:someValuesFrom ?clonet .
  ?offeringt rdfs:subClassOf ?r1 .

  ?r2 owl:onProperty is_sold_by: . ?r2 owl:hasValue ?vendori .
  ?offeringt rdfs:subClassOf ?r2 .

  ?r3 owl:onProperty recognizes: . ?r3 owl:someValuesFrom ?targett .
  ?clonet rdfs:subClassOf ?r3 .

  ?offeringt rdfs:label ?offering . ?clonet rdfs:label ?clone .
  ?vendori rdfs:label ?vendor . ?targett rdfs:label ?target .

  filter(?clonet != mAB:)
  filter(?vendor = "Abcam")
  filter (?target = "E-selectin") }
```

Fig. 2. Example SPARQL query and Results (see wiki for complete query listing).

Query Result (13)

Limit results: 100

Offering	Vendor	Clone	Target
"AB_1658362 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_868979 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_305609 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_1566023 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_868981 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_2040871 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_726629 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_868980 monoclonal antibody offering"	"Abcam"	"1.2B6"	"E-selectin"^^xsd:string
"AB_726624 monoclonal antibody offering"	"Abcam"	"B-P7"	"E-selectin"^^xsd:string
"AB_726625 monoclonal antibody offering"	"Abcam"	"B-P7"	"E-selectin"^^xsd:string
"AB_726628 monoclonal antibody offering"	"Abcam"	"ENA1"	"E-selectin"^^xsd:string
"AB_306322 monoclonal antibody offering"	"Abcam"	"ENA2"	"E-selectin"^^xsd:string
"AB_448259 monoclonal antibody offering"	"Abcam"	"P2H3"	"E-selectin"^^xsd:string