Aligning Research Resource and Researcher Representation: The eagle-i and VIVO Use Case

Stella Mitchell[1], Carlo Torniai[2], Brian Lowe[1], Jon Corson-Rikert[1], Melanie Wilson[2], Mansoor Ahmed[3], Shanshan Chen[3], Ying Ding[3], Nicholas Rejack[4] , Melissa Haendel[2], the eagle-i Consortium, and the VIVO Collaboration

[1]Albert R. Mann Library, Cornell University, Ithaca, NY, USA
[2]Oregon Health and Science University, Portland, OR, USA
[3]Indiana University, Bloomington, IN, USA,
[4]University of Florida, Gainesville, FL, USA,

Keywords: expertise, research resources, linked open data, application ontology, researcher profile

1 Introduction

People are a uniquely pervasive nexus, linking the inputs, activities, and outputs of research over time and thus enabling their discovery through contextual relationships with other people, organizations, and events. However, an artificial separation between information about people and information about the research resources they use has evolved in institutional information systems. This lack of interoperability has recently been recognized as impairing the efficiency and effectiveness of research. The impairment stems not only from the overall lack of discoverability, but also from the lack of connectivity between systems for managing information about people and those, if any, that gather and maintain information about research resources. The lack of data integration is compounded when working across institutions, where local needs may dictate inconsistent approaches to data collection, management, and display.

Many institutions are adopting or building researcher profiling systems to encourage collaboration, and these typically encompass publications and grants as the most visible and widely accepted evidence of scientific productivity and expertise. However, few if any, of these systems encompass other products of research or research resources. Research resources can be, however, as compelling evidence of expertise and experience as publications, presentations, and other more commonly tracked outcomes. They provide another vehicle for discovery as well as rich connection opportunities across the often-hidden research infrastructure of a university or medical school.

The National Center for Research Resources (NCRR) 2009 U24 request for proposals, which funded the eagle-i and VIVO projects, aimed to "develop, enhance, or extend infrastructure for connecting people and resources to facilitate national discovery of individuals and of scientific resources by scientists and students to encourage interdisciplinary collaboration and scientific exchange" [1]. The U24 RFA distinguished projects about research resources from those about the researchers themselves, but eagle-i [2] and VIVO [3] have been strongly encouraged by the NIH to work together to address information technology needs of the research community.

2 VIVO and eagle-i are Ontology-Driven Applications

VIVO and eagle-i store data natively in RDF and use OWL ontologies as primary data models for their respective applications. Both projects have been successful in building flexible and extensible applications leveraging the relationships as much as the type differentiation inherent in ontology class hierarchies. Both offer web-based editing of content in addition to display and search functionalities. As ontology-driven applications that also want to interoperate with the Linked Open Data (LOD) cloud, eagle-i and VIVO bridge some of the gaps dividing the LOD, RDF-centric world from the more constrained domain of OWL. Both ontologies are designed to support extensive instance data, on the order of tens of millions of statements per university.

The ontologies are also influenced by the high value placed on having eagle-i and VIVO data available as LOD. Our applications support a number of sub-properties to help consumers of linked data to decide what path to follow; we can also facilitate linked data crawling by returning the labels as well as URIs and rdf:type of the object individuals of object property statements. The ontologies developed by VIVO and eagle-i aim to facilitate data publishing of. To this end, VIVO uses a local extension process to support individual site needs while still enabling class subsumption to support multi-site search at the VIVO core level. For example, VIVO at Cornell has included research resource types from eagle-i and populated them as a way to have facilities and related services more broadly visible in the Cornell research community. This local extension methodology will likely be applied to eagle-i in the future.

3 Ontology Interoperability

The eagle-i ontology [4] has where possible adopted classes and properties from OBO ontologies that represent entities common in the biomedical context, while the VIVO ontology [5] covers the full range of disciplines represented in major research universities. OBO Foundry principles for ontology development [6] are intended to be orthogonal (non-overlapping) to permit any domain to develop its necessary classes and properties and encourage re-use of classes and properties from other ontologies at points of intersection. The goals are to avoid the ambiguity of closely aligned but distinct class and property definitions, and to avoid the need for mapping by importing and reusing actual URIs from other ontologies. To this end, both projects have aligned their ontologies under the Basic Formal Ontology (BFO) [7] to provide a consistent upper ontology framework guiding the class and property structure. Both use the MIREOT approach [8] to bring together the required ontologies and vocabularies, which reduces complexity and allows interoperability with other systems.

The alignment between VIVO and eagle-i has focused on areas of natural overlap between our respective domains as well as points of intersection with external ontologies of common interest. Some of these classes in common exist in the eagle-i namespace, some in the VIVO namespace, and some are imported and defined in external ontologies such as the Ontology of Clinical Research (OCRe) [9], Ontology for Biomedical Investigations (OBI) [10], Bibliographic Ontology (BIBO) [11] and Friend of a Friend (FOAF) [12]. Since not all of these are within the OBO library of orthogonal ontologies, this has sometimes proven a challenge, as there can be significant overlap and therefore the need to choose and/or map between them. The eagle-i ontology encompasses many more types (classes) than VIVO in order to allow resource contributors to classify data with important nuances. The VIVO ontology has many fewer classes but a relatively larger number of properties to distinguish the many relationships among people and organizations.

4 Challenges and Future Directions

There are areas of overlap between the ontologies where eagle-i uses research terminology while VIVO represents similar concepts in general terms applicable across academia, as illustrated by distinctions between foaf:Person and eagle-i:HomoSapiens. This raises the question of how using foaf:Person to represent a researcher should be reconciled with the notion of Homo sapiens as a research subject, e.g. as a source of a biospecimen. In other cases, both ontologies use the same term classified in different ways under the BFO hierarchy based on the anticipated context of use. For example, one ontology represents a service as the offering of a service at a particular point in time, while the other represents ongoing services offered by a core facility. Definitions must also be clarified for properties as well as types: the hasLocation property, for example, may relate not only to geographic locations but also to areas of the brain. This is the difference between the use of an ontology in combination with LOD versus simply linking to the same URI based on the class or property label alone -context is important. Consistent application of ontology terms and properties will be aided by developing a common approach to ontology metadata.

We recognize that controlled vocabularies of terminology (such as MeSH and other UMLS vocabularies) must remain distinct from the eagle-i and VIVO ontologies. However, it is clear that there should be a mechanism for interoperability due to the large amount of research outputs indexed with these vocabularies. Certain subsets, such as the MeSH disease tree, may be imported to promote consistency in annotation of resources. The strategy we chose to follow is to reference external URIs, leveraging the work of the bioinformatics research team under Dr. Moisés Eisenberg at Stony Brook University.

Teams of scientists are now much more common than single scientists in the production of biologically meaningful and clinically consequential breakthroughs [13]. It is common for new grant programs to require multi-institutional and multi-disciplinary participation, as well as to prioritize the participation of underserved populations. Nevertheless, it is still often difficult for investigators to find active researchers and research resources in their field, especially when required to move beyond their own professional circles. We believe that increasing the visibility of the full range of researcher activities and outputs will greatly enhance the discovery potential and promote collaboration. Encoding this knowledge based on the semantic relationships between them lays the groundwork for new methods of inferring expertise based on a wider range of data representing research resources and not just grants and publications as outputs. Of particular interest is mapping clinical experience to medical vocabularies in order to bridge the gap between clinical and basic research domains.

Acknowledgements

VIVO is supported by U24RR029822 and eagle-i by 1U24RR029825 from NIH/NCRR.

References

1. Recovery Act 2009 Limited Competition: Enabling National Networking of Scientists and Resource Discovery (U24), http:// grants.nih.gov/grants/guide/rfafiles/RFA-RR-09-009.html
2. eagle-i project, http://www.eagle-i.org
3. VIVO project, http://www.vivoweb.org
4. Torniai, C., Bashor, T., Bourges-Waldegg, D., Corday, K., Frost, H.R., Johnson, T., Segerdell, E., Shaffer, C.J., Stone, L., Wilson, M.L., Haendel, M.A.: eagle-i: an ontology-driven framework for biomedical resource annotation and discovery. Bio-Ontologies 2010: Semantic Applications in Life Sciences, ISMB (2010).
5. Krafft, D.B., Cappadona, N.A., Caruso, B., Corson-Rikert, J., Devare, M., Lowe, B.J., VIVO Collaboration VIVO: Enabling National Networking of Scientists. Proc WebSci10: Extending the Frontiers of Society OnLine, April 26-27, 2010, Raleigh, NC: US (2010).
6. Smith, B., et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. In: Nat Biotechnol, 25 (11), 1251-5. PMCID: PMC2814061. (2007)
7. Basic Formal Ontology (BFO), http://www.ifomis.org/bfo/home
8. Courtot, M., Gibson, F., and Lister, A. MIREOT: the Minimum Information to Reference an External Ontology Term. ICBO (Buffalo, NY) (2009).
9. Ontology of Clinical Research (OCRe), http: //bioportal.bioontology.org/ontologies/39956
10. Ontology for Biomedical Investigations, http: //obi-ontology.org/
11. Bibliographic Ontology (BIBO), http://bibliontology.com/
12. Friend of a Friend (FOAF), http://www.foaf-project.org/
13. Börner K, Contractor N, Falk-Krzesinski HJ, Fiore SM, Hall KL, Keyton J, Spring B, Stokols D, Trochim W, Uzzi B. A multi-level systems perspective for the science of team science. Science Translational Medicine, 2(49), 49cm24 (2010).