Extending the VIVO Ontology to iSchools: Enabling Networking of Information Scientists[1]

Ying Ding
Indiana University
dingying@indiana.edu

Erjia Yan
Indiana University
eyan@indiana.edu

Ali Ghazinejad, Han Jia, & the VIVO ontology team
Indiana University

Abstract

VIVO, funded by NIH, utilizes Semantic Web technologies to model scientists and provides federated search to enhance the discovery of researchers and collaborators across disciplines and organizations. VIVO ontology is designed with the focus on modeling scientists, publications, resources, grants, locations, and services. VIVO data is annotated based on the VIVO ontology to semantically represent and integrate information about faculty research, teaching, and service. This paper introduces the birth and development of the VIVO ontology, and discusses the potential of expending it to the information science community to facilitate networking of information scientists.

Keywords: VIVO, ontology, Semantic Web

Introduction

The exponential growth in complexity and scope of modern science has dramatically increased the demand for more collaboration among scientists in different fields and at different levels. Modern science is team-based, interdisciplinary and cross-institutional, but discovery across these boundaries is difficult. Researchers seeking answers to one research questions may have to consult scientists or studies from other scientific domains. Scientists have been heavily relying on the World Wide Web for supporting their research endeavors, especially for interdisciplinary and international collaboration (Hendler, 2003). However, currently popular Web technology is not satisfying for the needs of the collaborative and interdisciplinary "e-Science". For example, boundaries of institutions, distributed data in different formats, and specialized terminology still impede the communication of scientific information between scientists. Therefore, new models of communication need to be forged so as to establish the next paradigm of tools of scientific collaboration on the Web.

The Semantic Web is designed to build a standard representation that can provide meaningful linkages across different sets of data to promote integration and communication. Many communities have embraced the Semantic Web technologies as a powerful and effective way to represent and relate data. The Linked Open Data (LOD) initiative currently contains 203 linked datasets which together serve 25 billion RDF triples to the Web and are interconnected by 395 million RDF links. The US government portal Data.gov[2] makes around 400 of its datasets, summing to 6.4 billion triples, available as Linked Data. The Semantic Web technology may serve as an effective solution to the increasing and urging demand of broader and more in-depth communication between scientists in the academic community all over the world. However, there are currently no Semantic Web applications authorized to integrate official information of academic communities. VIVO can fill this gap.

VIVO is an open source Semantic Web application that, when populated with researcher interests, activities, and accomplishments, enables discovery of research and scholarship across disciplines and organizations. The VIVO core ontology models the academic community in order to

[1] This poster is developed based on a poster presentation at the Conference on Semantics in Healthcare and Life Sciences (CSHALS2012).
[2] http://www.data.gov/semantic/index

Ding, Y., Yan, E., Ghazinejad, A., & Jia, H. (2013). Extending the VIVO ontology to iSchools: Enabling networking of information scientists. *iConference 2013 Proceedings* (p. 905-908). doi:10.9776/13460

provide an consistent and connected perspective on the research community to various shareholders, including students, administrative and service officials, prospective faculty, donors, funding agencies, and the public (Karfft, Cappadona, Devare, et al., 2010). The major impetus for NIH to fund the VIVO effort 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"[3].

VIVO can support discovering potential collaborators with complementary expertise or skills, suggesting appropriate courses, programs, and faculty members according to students' interests, and facilitate research currency, maintenance and communication. The implementation of the VIVO ontology to iSchools allows one to search across experts among varied sub-domains in the field of information studies. In this paper, we present a relatively comprehensive discussion of the development of the VIVO core ontology.

VIVO Ontology Development

According to Gruber (1993), an ontology is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. Simply put, it contains a hierarchical taxonomy or controlled vocabulary and secondary associations between terms. Complex real world entities can be expressed by assigning properties (i.e, relationships/associations) to classes/subclasses. Subclasses usually can inherit properties from their upper classes. For example, *faculty member* have subclasses *assistant professor* and *associate professor*. If *faculty member* has the property *teach class*, then *assistant professor* and *associate professor* will also have the property *teach class.* Furthermore, ontologies also contain inference rules to enable machine-processable computing and reasoning (Berners-Lee, Hendler, & Lassila, 2001). An ontology may express the rule that if woman A is the mother of woman B, and woman B is the mother of man C, then woman A is the grandmother of man C. With the reasoning power, machines can manipulate the concepts in a well-defined logic way that can be easily understood by human beings (Guarino, 1998; Guarino & Giaretta, 1995). An ontology models the semantics of components by defining concepts (classes) and the relationships (properties) between them. The VIVO ontology is a unified, formal, and explicit specification of information about researchers, organizations, activities and relationships that link them together.

Modeling Principles

A significant challenge in developing the VIVO core ontology is to include enough detail to allow for meaningful cross-site discovery of data, while keeping it simple enough to apply to diverse academic and clinical institutions (Karfft, Cappadona, Devare, et al., 2010). The development of the VIVO ontology follows four principles:

- Core and localization: The design of VIVO core and localization combines the advantages of standardization with those of customization.
- Modularized design: It subdivides a system into independent but connected modules that can be used in different contexts to enable multiple functionalities.
- Aligning with other ontologies: VIVO ontology is iteratively refined to align with other ontologies.
- Modeling complex relationship as classes: In VIVO ontology, complex relationships are modeled as entities in order to add properties to these relationships.

Conceptualization

The major classes covers the three important areas of the faculty research activities are: research (bibo:Document, vivo:Grant, vivo:Project, vivo:Software, vivo:Dataset, vivo:ResearchLaboratory), teaching (vivo:TeacherRole, vivo:AdvisingRelatioship), and services (vivo:Service, vivo:CoreLaboratory, vivo:MemberRole). For each of the three parts, there are existing ontologies that conceptualize them.

The VIVO core ontology, on one hand, reuses the useful classes and properties of FOAF and, on another hand, develops customized properties and classes targeting at modeling social network of scientists. For information resources, BIBO ontology models publications; DOAP models projects; and

[3] http://grants.nih.gov/grants/guide/rfa-files/RFA-RR-09-009.html

Doblin core models documents in general. Other resources, including courses and grants, are usually organized and stored in databases. The VIVO core ontology reuses useful classes and properties from those existing ontologies, and further connects, creates, or modifies those distributed classes in a uniform framework of modeling social network of scientists.

Figure 1 shows concept map of VIVO's (version 1.3) main classes and objective properties. Bubbles in different colors represent different sources of the classes. Lines between bubbles indicate their relationships, which are object properties in VIVO ontology.

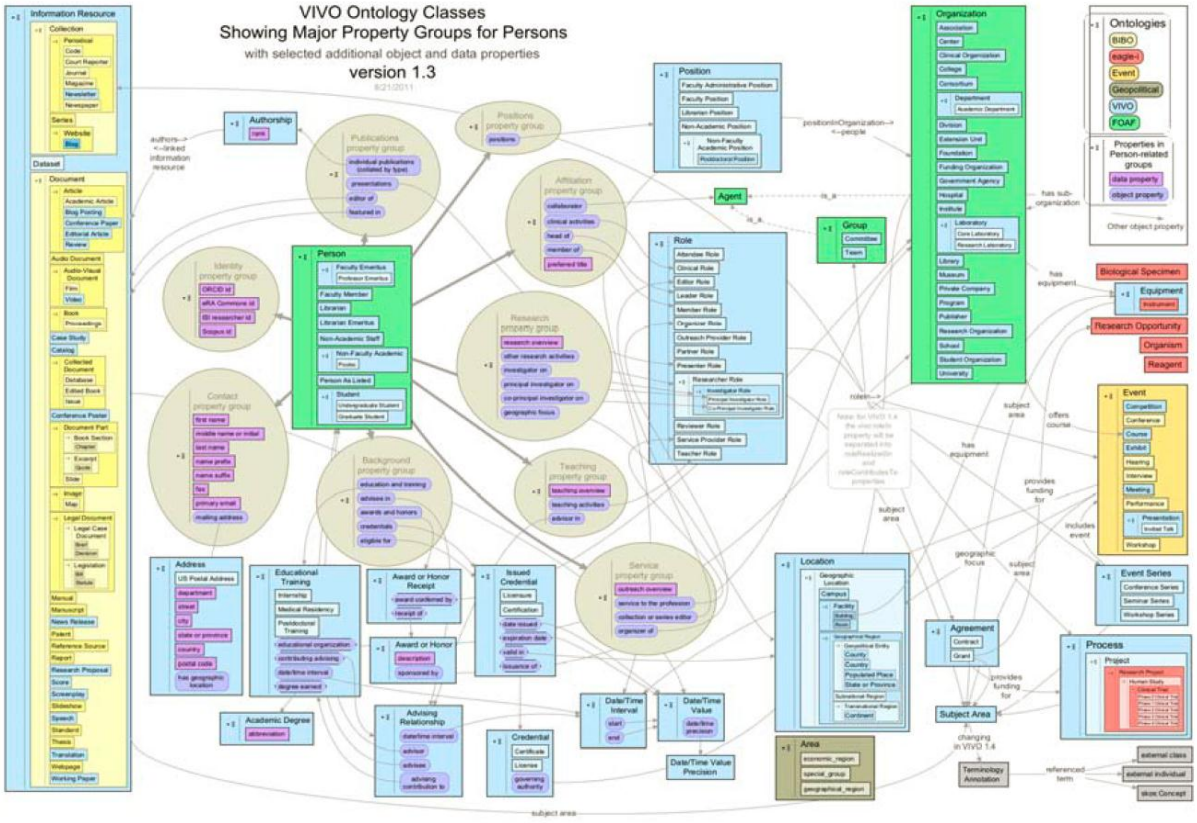

Figure 1 VIVO's main classes and objective properties[4]

Interoperability

A key issue in building the VIVO ontology is enabling interoperability among different ontologies. Since ontologies have been developed in different contexts, ontology interoperability is a key factor essential for aligning and integrating distributed ontological resources over the Internet. Mappings to these ontologies will enable VIVO data to be shared among a variety of systems. The development of the VIVO ontology reuses several commonly used ontologies, including upper level ontologies Event ontology, Basic Formal Ontology-OBO Foundry), as well as domain ontologies (i.e., FOAF, Geopolitical ontology, SKOS, and BIBO).

Localization

Individual institutions can localize, or extend, the ontology to support local requirements. Different namespaces are adopted for different localized versions in specific institutions. Individual installations may extend the core with ontologies that reflect available data sources according to their contextual needs. While all institutional installations of VIVO share the core ontology, each institution is free to

[4] High definition figure can be found at:
http://sourceforge.net/apps/mediawiki/vivo/nfs/project/v/vi/vivo/8/83/VIVOCoreOverview.1.3.byPropertyGroup.2011-08.png

extend this ontology or add additional ontologies as desired. The VIVO core ontology plays the role as an integration layer that permits data from different institutions to be queried in a consistent way.

Conclusion and future development

The *iSchools* Project is a consortium of 36 institutions located at 11 countries[5]; meanwhile, the studies of information is interdisciplinary. VIVO provides a practical ontology that represents *iSchools*' needs by providing local value for sustainability, while putting those institutions who adopt it at a competitive advantage for discovery through linked data and for compliance with federal data initiatives. In the future, we intend to expend the VIVO ontology to all *iSchools* and expect to see an enriched knowledge base of academic communities among *iSchools* and even beyond. There are at least three ways to participate in VIVO: 1) Download, Adopt, and Implement: The open source VIVO software and ontology are now available for download. 2) Provide Data: *iSchools* can participate by providing machine readable data for research discovery. Bibliometric and funding data are of great interest to the research community. 3) Develop Applications: Many software applications can benefit from using information that will be provided by the national network of VIVO. New applications can use information from the national network to provide enhanced search, new collaboration capabilities, grouping, finding and mapping scientists and their work[6].

References

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American Magazine*. Retrieved March 26, 2011 from http://www.sciam.com/article.cfm?id=the-semantic-web&print=true

Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.

Guarino, N. (1998). Formal ontology and information systems. In N. Guarino (Ed.), *Formal ontology in information systems: Proceedings of FOIS '98* (pp. 3-15). Amsterdam, The Netherlands: IOS Press.

Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars (Ed.), *Towards very large knowledge bases: Knowledge building and knowledge sharing* (pp. 25-32). Amsterdam, The Netherlands: IOS Press.

Hendler, J. (2003). Science and the semantic web. *Science, 299*(5606), 520-521

Krafft, D. B., Cappadona, N. A., Devare, M., & Lowe, B. J. (2010). VIVO: Enabling national networking of scientists. Web Science Conference, April 26-27 2010, Raleigh, NC, USA. Retrieved February 11, 2011 from http://vivoweb.org/files/websci10_submission_82.pdf

[5] http://www.ischools.org/site/descriptions/
[6] http://vivoweb.org/participate