

The VIVO Ontology: Enabling Networking of Scientists

Stella Mitchell¹, Shanshan Chen², Mansoor Ahmed²

¹Cornell University Library

sjm222@cornell.edu,
chenshan@umail.iu.edu,
mansoor@indiana.edu

Brian Lowe¹, Paula Markes³,
Nick Rejack³

²Indiana University Bloomington

bjl23@cornell.edu,
pamarkes@ufl.edu,
nrejack@ufl.edu

Jon Corson-Rikert¹, Bing He²,
Ying Ding², and the VIVO
collaboration

³University of Florida

jc55@cornell.edu,
binghe@indiana.edu,
dingying@indiana.edu

ABSTRACT

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. The application is in use at the seven institutions participating in the National Institutes of Health (NIH) grant to establish a national network of scientists, and has also been adopted by several universities in Australia and China as well as by the United States Department of Agriculture.

Categories and Subject Descriptors

I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods –*Predicate logic, Relation systems, Representation languages, Representations (procedural and rule-based), Semantic networks.*

General Terms

Management

Keywords

Ontology, VIVO

1. INTRODUCTION

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. The application is in use at the seven institutions participating in the National Institutes of Health (NIH) grant to establish a national network of scientists, and has also been adopted by several universities in Australia and China as well as by the United States Department of Agriculture (Gewin, 2009). A primary motivation behind VIVO is that modern science is increasingly team-based, interdisciplinary and cross-institutional, but that discovery across these boundaries is difficult (Hendler, 2003). VIVO aims to integrate researcher information from disparate, largely authoritative, sources into a common format establishing interrelationships and to make it publicly available.

In recognition of the need for standard representations that can provide meaningful linkage across different sets of data, the research community has begun to embrace the semantic web technologies as a powerful and effective way to represent and relate data. The Linked Open Data 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.gov¹ makes around 400 of its datasets, summing to 6.4 billion triples, available as Linked Data. VIVO produces Resource Description Framework (RDF) triples formatted according to a published ontology so that information can be exchanged, aggregated and searched by others on the web through standard protocols.

An ontology communicates the meaning of these RDF triples by defining types (classes) and the relationships between them (properties); the VIVO ontology is a unified, formal, and explicit specification of information about researchers, organizations, and the activities and relationships that link them together. An ontology can be constructed to align under broader upper level ontologies for more consistent interoperability across domains, and can also extend general concepts to the more specific needs of the target application. Moreover, a well-designed ontology can support reasoning to derive additional knowledge by inference using the logic relationships specified by the ontology's structure.

People are a uniquely pervasive nexus linking the inputs, outputs and outcomes of research over time and acting as a connector between different domains. The VIVO ontology concentrates on modeling scientists in the rich context of their activities, organizations and the products of their research rather than the knowledge in any single domain of science. Additionally, we are beginning to address the topic of how the ontology may contribute to more complex questions such as modeling a person's expertise. Another key issue is how to link to external controlled vocabularies, such as Library of Congress Subject Headings, MeSH, and SNOMED, in a reusable way that is compatible with semantic web principles. The VIVO project has awarded funding to Stony Brook University to use the National Library of Medicine's Unified Medical Language System to develop services to support linking to external vocabularies in a standard way.

VIVO is innovative in that its ontology is used not only to provide a logical structure for data representation, but also to drive the data entry and search interfaces of the application. In addition the ontology has served as a common language between domain experts, end users and information technology professionals during requirements gathering. The domain and information modeling needs were discussed by the different stakeholders in terms of ontology classes (types) and properties (relationships), recorded in a precisely defined language as an engineering artifact, and then used to drive the application directly. The focus on a high-quality, standardized ontology with the application built on top as a population and delivery

¹ <http://www.data.gov/semantic/index>

mechanism for the information protects long-term investment in the data. It also enables independent producers, such as Harvard Profiles and Collexis who have decided to offer RDF data formatted according to the VIVO ontology, to contribute to a national network of researcher information.

A key issue in building the VIVO ontology is enabling interoperability among different ontologies. Since heterogeneous ontologies have been developed in different contexts, ontology interoperability is a key factor essential for aligning and integrating distributed ontological resources over the Internet. Ontology interoperability is achieved by identifying or establishing semantic correspondence between entities (i.e., classes and properties) among multiple ontologies. The development of the VIVO ontology reuses several commonly used ontologies, including Dublin core, Event ontology, FOAF, Geopolitical ontology, SKOS, and BIBO. For the various ontologies mentioned here, please refer to <http://swl.slis.indiana.edu/repository/index.html> for a critical review.

Sharing a common representation format allows data from different institutions to be searched and aggregated in a meaningful way. A significant challenge in developing the core ontology has been to include enough detail to allow for meaningful cross-site interoperability and discovery, while keeping it simple enough to apply to diverse institutions. By using an iterative, consensus-based design process to develop an ontology that works for the diverse core partner institutions our hope is that it will have broad applicability across many research institutions. Individual institutions can then localize, or extend, the ontology to support local requirements.

Publishing large amounts of data about people, gathered from institutional databases and from self-assertions, on the web raise issues of trust (provenance) and privacy. We have begun to lay the groundwork in the ontology for addressing these issues by explicitly representing provenance information and public vs. private designations but this is a significant area of future work for the project as these considerations will only become more

important as awareness and usage of the linked open data commons increases. Another important area for future work will be ontology versioning and change management: that is, determining how to keep track of changes in the data definition, communicate them, and relate published data to the corresponding version of the ontology.

The VIVO ontology demonstrates the power of semantically annotated data by enabling a versatile set of applications. A SPARQL query builder facilitates the exploration of the ontological structure and provides a framework for developing search and reporting tools that go beyond VIVO's current text-based search capability to fully leverage the semantic structure in the ontology. These structures are effectively conveyed through the VIVO interface by visualizations leveraging co-author and co-investigator relationships to present a researcher in the context of collaboration over a ten-year time period. Amid calls at Universities to demonstrate the effectiveness of education, research and collaboration, and calls at the federal level for consistent data that can be used to evaluate the effectiveness of research funding and its impact on our economy, VIVO provides a practical ontology that represents institutional needs enough to provide 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.

2. ACKNOWLEDGEMENT

This work is supported by NIH grant U24 RR029822.

3. REFERENCES

- [1] Gewin, V. (2009). Networking in VIVO: An interdisciplinary networking site for scientists. *Nature* 462, 123.
- [2] Hendler, J. (2003). Science and the Semantic Web. *Science* 24: 520-521