The VIVO Ontology: Enabling Networking of Scientists

Ying Ding, Stella Mitchell, Jon Corson-Rikert, Brian Lowe, Bing He, and the VIVO ontology team

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. It incorporates classes from popular ontologies, such as BIBO, Dublin Core, Event, FOAF, geopolitical, and SKOS. VIVO data is annotated based on the VIVO ontology to semantically represent and integrate information about faculty research (i.e., educational background, publications, expertise, grants), teaching (i.e., courses, seminars, training), and service (i.e., organizing conferences, editorial boards, other community services). The VIVO ontology has been adopted nationally and internationally, and enables the national and international federated search for finding experts.

1 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 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[1] makes around 400 of its datasets, summing to 6.4 billion triples, available as Linked Data. The Semantic Web technology may

[1] http://www.data.gov/semantic/index

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 fills 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 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"[2].

The application is in use at the seven institutions of the NIH VIVO project and has been adopted or to be adopted by several other universities (e.g., Harvard University) and organizations in the USA (e.g., the United States Department of Agriculture), and several universities or institutions in Australia and China (e.g., Queensland University of Technology, Chinese National Academy of Sciences) (Gewin, 2009). More specifically, 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. For example, a Computer-Aided Drug Discovery (CADD) group may want to find and team up with a computer specialist and a group using *in vivo* experiments in drug discovery. If the VIVO core ontology is implemented in the hypothetical situation, the group leader can search across experts in computer science and molecular biology. In this paper, we present a relatively comprehensive discussion of the development of the VIVO core ontology, including the latest updates.

2 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).

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

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. An ontology can be constructed to align under broader upper level ontologies for interoperability across domains, and can also extend general concepts to the more specific needs of target applications. Moreover, a well-designed ontology can support reasoning to derive additional knowledge by inferring the logic relationships specified by the ontology.

The external ontologies to which the VIVO core ontology refers can be found in the following table.

Table 1 External ontologies to which the VIVO core ontology refers

External ontologies	namespace	Descriptions	Use in VIVO
Bibliographic ontology	bibo	The Bibliographic Ontology Specification provides main concepts and properties for describing citations and bibliographic references (i.e. quotes, books, articles, etc) on the Semantic Web.[3]	Used to model scientific products, mainly publications.
Dublin Core elements & Dublin Core terms	dc	The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. "core" because its elements are broad and generic, usable for describing a wide range of resources.[4]	Used to model generic documents
Event Ontology	event	This ontology is centered around the notion of event, seen here as the way by which cognitive agents classify arbitrary time/space regions.[5]	Used to be the upper class of other specific events
FOAF	foaf	FOAF defines an open, decentralized technology for connecting social Web sites, and the people.[6]	Used to model scientists, students, staff, etc.
Geopolitical ontology	geopolitical	The geopolitical ontology is a mechanism to describe, manage and exchange data related to geopolitical entities such as countries, territories, regions and other similar areas.[7]	Used to describe the geopolitical aspects of any entities
Simple Knowledge Organization System (SKOS)	skos	SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists and	Used to represent knowledge organization systems

[3] http://bibliontology.com/
[4] http://dublincore.org/documents/dces/
[5] http://motools.sf.net/event/event.html
[6] http://www.foaf-project.org/
[7] http://aims.fao.org/website/Geopolitical-Ontology/sub2

		taxonomies within the framework of the Semantic Web. [8]	

There are existing ontologies that have similar purposes to the VIVO core ontology, for example, Semantic Web for Research Communities (SWRC)[9], Advanced Knowledge Technologies (AKT)[10], etc. The SWRC (Semantic Web for Research Communities) is an ontology for modeling entities of research communities such as persons, organizations, publications (bibliographic metadata) and their relationships (Sure et al., 2005).

2.1 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. For example, scientific resource section is modeled as a separate module.
- Aligning with other ontologies: VIVO ontology is iteratively refined to align with other ontologies. More specifically, VIVO ontology develops interoperability through reusing popular upper level ontologies and domain-specific ontologies. For example, by sharing a number of classes and properties with the Eagle-I ontology, the VIVO core ontology builds direct semantic correspondence with Eagle-I. For example, by the equivalent property *has_role* with domain *Agent* and range *Role* in both the Eagle-I and VIVO ontology. Another example is that the VIVO ontology reuses several existing domain ontologies (e.g., FOFA, BIBO, etc.) and is also based on an upper-level ontology (e.g. BFO).
- Modeling complex relationship as classes: In VIVO ontology, complex relationships are modeled as entities in order to add properties to these relationships. For example, researchers' affiliation and positions change over time. Instead of using a simple property relating researchers to organizations, the VIVO core ontology, however, models employment as a resource that is related not only to a person and an organization but which can have additional properties describing the date that relationship began and the date when it ended. (W3C Working Grou, 2006).

2.2 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. FOAF is a very popular ontology for modeling person and organization. The inverse functional email property in FOAF has been the biggest

[8] http://www.w3.org/2004/02/skos/
[9] http://ontoware.org/swrc/
[10] http://www.aktors.org/

contributor to the adoption of that ontology, as it's the best practical way to link two URIs with some assurance that they refer to the same person (Williams, Weaver, Atre, & Heldler, 2010). However, FOAF assumes that its users and audience is people in general, so it only models universal features of human being and organizations. By contrast, scientists, as a specific group of persons, present a special set of characteristics, the same with academic institutions. Thus 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 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. We have investigated different ontologies and make decisions on whether to adopt them as they are (BIBO), versus whether to extend for more details (FOAF), versus whether to have minimal overlap since the other ontology is focused on a different domain. 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.

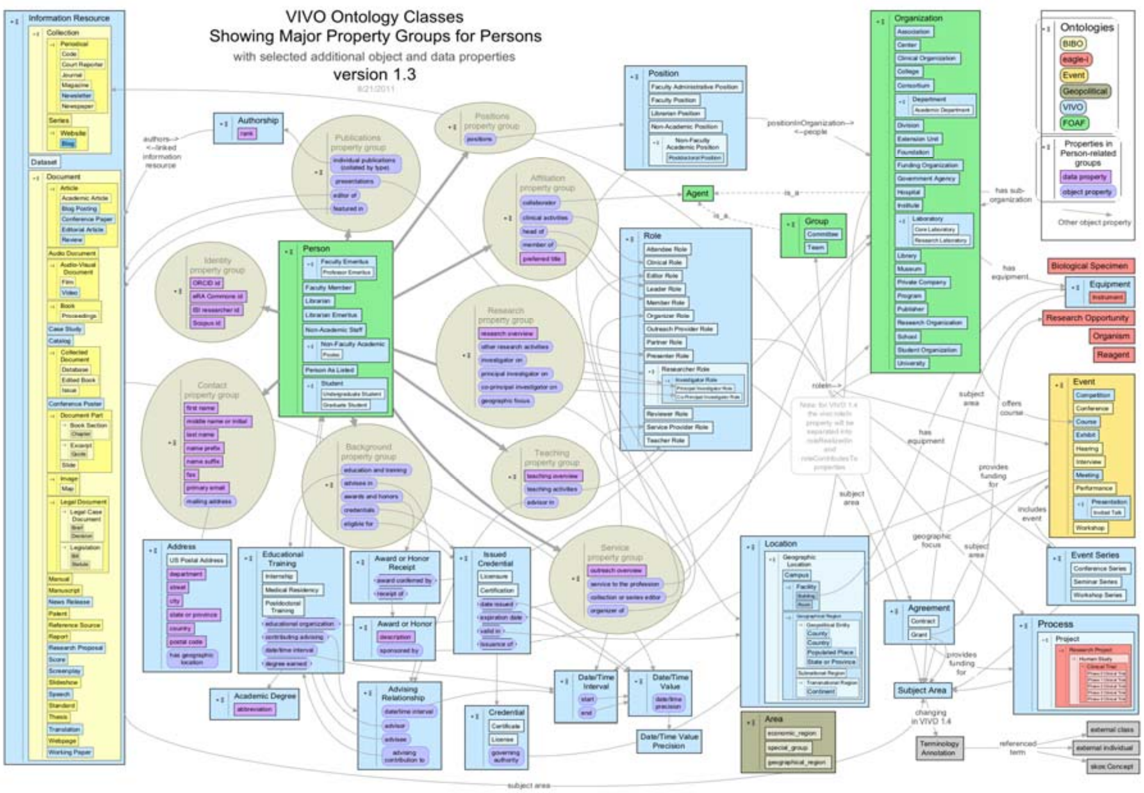

Figure 1 VIVO's main classes and objective properties

2.3 Modeling person

The VIVO ontology concentrates on modeling scientists in the rich context of their activities, organizations and products of their research. Figure 2 shows Katy Borner's profile.

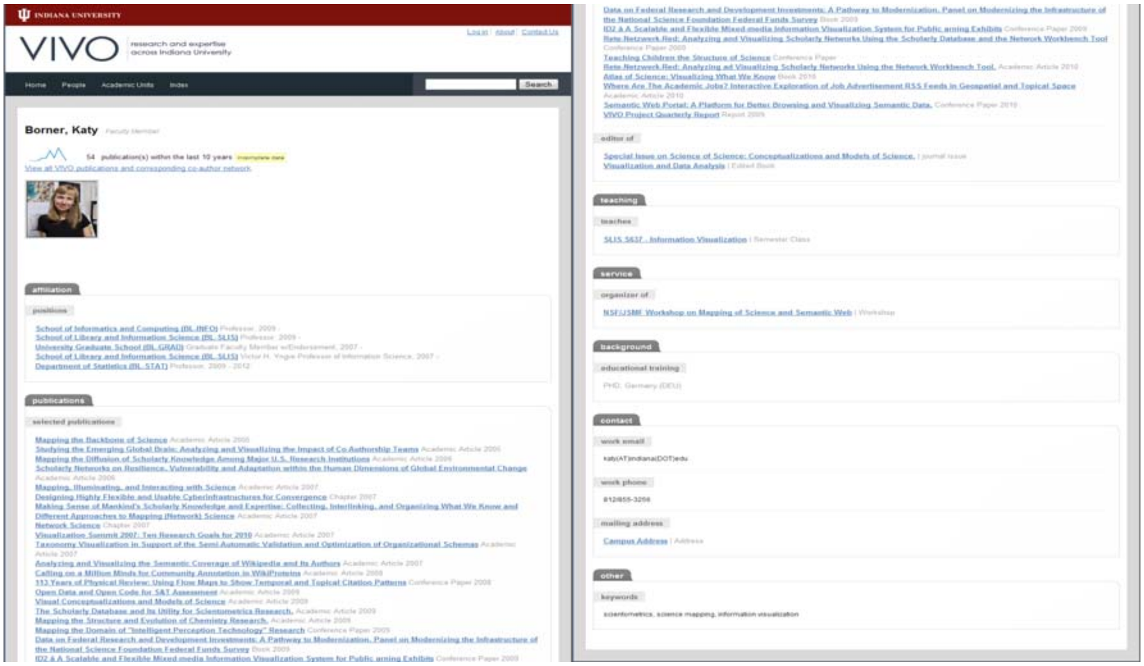

Figure 2 Katy Borner's VIVO profile

Those profiles are filled with information collected from official, verifiable sources, including authoritative information about researchers' positions provided by the organizations, publications pulled directly from scientific journals, and researchers themselves.

```
<http://vivo.iu.edu/individual/person25557> rdf:type <http://vivoweb.org/ontology/core#FacultyMember > .

<http://vivo.iu.edu/individual/person25557>
<http://vivoweb.org/ontology/core#authorInAuthorship>   <http://vivo.iu.edu/individual/n74>  .

<http://vivo.iu.edu/individual/n74 > rdf:type <http://vivoweb.org/ontology/core#Authorship>  .

<http://vivo.iu.edu/individual/n74>
<http://vivoweb.org/ontology/core#linkedInformationResource>
<http://vivo.iu.edu/individual/n7109>  .

<http://vivo.iu.edu/individual/n7109> rdf:type <http://purl.org/ontology/bibo/Article>  .

<http://vivo.iu.edu/individual/person714388> rdf:type <http://vivoweb.org/ontology/core#NonAcademic> .

<http://vivo.iu.edu/individual/person714388>
<http://vivoweb.org/ontology/core#authorInAuthorship>   <http://vivo.iu.edu/individual/n2881>  .

<http://vivo.iu.edu/individual/n2881>  rdf:type <http://vivoweb.org/ontology/core#Authorship>  .

<http://vivo.iu.edu/individual/n2881>  <http://vivoweb.org/ontology/core#authorRank>   2  .

<http://vivo.iu.edu/individual/n2881>
<http://vivoweb.org/ontology/core#linkedInformationResource>
<http://vivo.iu.edu/individual/n7109>  .
```

Figure 3 RDF triples ,'Professor Katy Borner's coauthorship with other colleagues

In RDF triples shown in Figure 3, Professor Katy Börner is represented by a uniform resource identifier (URI) http://vivo.iu.edu/individual/person25557. Katy Börner and Nianli Ma are coauthors of one academic paper. The reason to use one bridging class to model this authorship relationship is that it enriches the context of this relationship, for example, there are data properties core:authorRank to describe the rank of a specific author in one publication's author-list.

2.4 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). For example, the VIVO core ontology extends FOAF to provide the basis for describing persons and organizations, and BIBO to model documents such as books and journal articles. Figure 4 shows a list of external ontologies that VIVO is linked to.

Ontologies

Add new ontology

#	Ontology	Namespace	Prefix
1	Bibontology	http://purl.org/ontology/bibo/	bibo
2	Dublin Core elements	http://purl.org/dc/elements/1.1/	dcelem
3	Dublin Core terms	http://purl.org/dc/terms/	dcterms
4	Event Ontology	http://purl.org/NET/c4dm/event.owl#	event
5	FOAF	http://xmlns.com/foaf/0.1/	foaf
6	geopolitical.owl	http://aims.fao.org/aos/geopolitical.owl#	geo
7	SKOS (Simple Knowledge Organization System)	http://www.w3.org/2004/02/skos/core#	skos
8	UF Vivo Extension	http://vivo.ufl.edu/ontology/vivo-ufl/	ufVivo
9	VIVO core	http://vivoweb.org/ontology/core#	core

Figure 4 external ontologies that VIVO is linked to

The VIVO ontology team has collaborated extensively with ontologists of the Eagle-I project (http://www.eagle-i.org/home/), a companion NIH-funded project addressing the challenge of finding and representing research resources such as laboratory equipment, cell lines, protocols, reagents, and other tools and products of science that too often remain invisible in the larger research community (Segerdell et al., 2010). The VIVO ontology shares a number of classes and properties with the eagle-I ontology, for example, by the equivalent property *has_role* with domain *Agent* and range *Role* in both the eagle-I and VIVO ontology. In a nutshell, by cross-importing of classes and properties between VIVO ontology and Eagle-I, we achieve a targeted overlap between them.

2.5 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. For example, librarians at one institution may be members of the faculty while librarians at a second are not: the core Librarian class may be extended in local ontologies to represent this. While all institutional installations of VIVO share the core ontology, each institution is free to 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. For example, UF has extended their local ontology in their namespace of http://vivo.ufl.edu/ontology/vivo-ufl/. New localized classes are added to expand VIVO core ontology on faculty member and organizations, such as ufVivo:CourtesyFaculty, ufVivo:CourtesyFacultyPosition, ufVivo:UFAcademicDepartment and ufVivo: UFResearchLaboratory.

3 Conclusion and future development

Universities are called 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 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 expect to see a dramatically growing in the number participating institutions in the VIVO ontology implementation and an enriched knowledge base of academic communities in the United States and all over the world. 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: institutions 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. New applications can use information from the national network to provide enhanced search, new collaboration capabilities, grouping, finding and mapping scientists and their work. [11]

4 Reference

[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

[3] 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

[4] Devare M., Corson-Rikert, J., Caruso, B., Lowe, B., Chiang, K., & McCue, J. (2007). VIVO: Connecting People, Creating a Virtual Life Sciences Community. *D-Lib Magazine, 13*. Retrieved February 11, 2011 from http://www.dlib.org/dlib/july07/devare/07devare.html

[5] Gewin, V. (2010). Collaboration: Social networking seeks critical mass. *Nature* 468, 993-994

[6] Gruber, T.R. (1993). "A translation approach to portable ontology specifications" (PDF). Knowledge Acquisition 5 (2): 199–220. Retrieved April 26th, 2011 from http://tomgruber.org/writing/ontolingua-kaj-1993.pdf

[11] http://vivoweb.org/participate

[7] Berners-Lee, Tim; James Hendler & Ora Lassila (2001). "The Semantic Web". Scientific American Magazine. Retrieved March 26, 2011 from . http://www.sciam.com/article.cfm?id=the-semantic-web&print=true
[8] 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: IOS Press. Available at www.ladseb.pd.cnr.it/infor/ontology/PUBL15.html
[9] 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: IOS Press. Available at www.ladseb.pd.cnr.it/infor/Ontology/Papers/KBKS95.pdf
[10] York Sure, Stephan Bloehdorn, Peter Haase, Jens Hartmann, Daniel Oberle. The SWRC Ontology - Semantic Web for Research Communities. In Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA 2005). Springer, Covilha, Portugal, December 2005.
[11] W3C Working Group. (2006). Defining N-ary Relations on the Semantic Web. Retrieved April 26, 2011 from http://www.w3.org/TR/swbp-n-aryRelations/
[12] Williams, G. T., Weaver, J., Atre, M., & Hendler, J. A. (2010). Scalable reduction of large datasets to interesting subsets. [doi: DOI: 10.1016/j.websem.2010.08.002]. *Web Semantics: Science, Services and Agents on the World Wide Web, 8*(4), 365-373.
[13] VIVO Cornell. http://vivo.cornell.edu/
[14] VIVO University of Florida. http://vivoweb.org/
[15] VIVO Indiana University. http://vivo.iu.edu/
[16] Southwest Biodiversity Knowledge Environment in China. http://ske.las.ac.cn/
[17] Plant Evolutionary Biology Knowledge Environment in China. http://botany.las.ac.cn/
[18] Biomedical and Health Knowledge Environment. http://health.las.ac.cn/
[19] NCBO Bioportal. http://bioportal.bioontology.org/
[20] Loki. http://www.icts.uiowa.edu/Loki/
[21] Harvard Catalyst. http://catalyst.harvard.edu/
[22] RDF standard. http://www.w3.org/RDF/
[23] SPARQL standard. http://www.w3.org/TR/rdf-sparql-query/
[24] Linked Open Data. http://linkeddata.org
[25] Bibontology. http://bibliontology.org/
[26] Dublin Core elements. http://purl.org/dc/elements/1.1/
[27] Dublin Core terms. http://purl.org/dc/terms/
[28] Event Ontology http://purl.org/NET/c4dm/event.owl#
[29] FOAF. http://xmlns.com/foaf/0.1/
[30] Geopolitical. http://aims.fao.org/aso/geopolitical.owl#
[31] SKOS. http://www.w3.org/2004/02/skos/core#
[32] UF VIVO Extension. http://vivo.ufl.edu/ontology/vivo-ufl/
[33] VIVO Core. http://vivoweb.org/ontology/core#