

VIVO response to 2011 JISC name identifier questionnaire

2011 JISC name identifier questionnaire

The UK's Joint Information Systems Committee (JISC) which is looking at the way that other countries are building name identifier systems relating to researchers in higher education.

The questions and VIVO's responses are listed below.

1. What was the motivation for developing the identifier system?

VIVO is a semantic web platform that stores all data natively in RDF and uses uniform resource identifiers, or URIs, as internal as well as public identifiers. VIVO was developed at Cornell University from 2003-2009 to provide a single Web presence at Cornell for discovering information about researchers throughout the University. VIVO not only bridges administrative boundaries but integrates the publicly-visible data from many internal and external data sources of record. VIVO publishes this integrated information as linked open data and provides convenient feeds in other Web-friendly formats to facilitate re-use of data and a greater return on the University's overall IT investment.

As other institutions learned about VIVO, Cornell began to receive requests to offer the software for use elsewhere. As described in question 13 below, VIVO in 2009 became the central technology behind a major U.S. National Institutes of Health initiative to foster the development of a multi-institutional research networking platform predicated on the value of stable, institutionally-hosted URIs and linked open data (LOD). The VIVO: Enabling National Networking of Scientists collaboration (<http://vivoweb.org>) has recently completed in the second annual VIVO conference, in Washington, DC. A prototype cross-institutional search (<http://vivosearch.org>) leverages linked open data across the VIVO collaboration and Harvard Catalyst Profiles (<http://catalyst.harvard.edu/spotlights/profiles.html>).

VIVO has been adopted at institutions outside the United States and by several private and governmental organizations – notably the American Psychological Association, the United States Department of Agriculture, Griffith University (<http://www.dlib.org/dlib/may11/wolski/05wolski.html>) and the University of Melbourne (<http://blogs.unimelb.edu.au/vivoands/>) in Australia, and the Chinese Academy of Sciences (<http://ske.las.ac.cn/>).

VIVO's ontology has been designed to store any additional identifiers, typically about people, organizations, publications, or grants, and we have already added an ORCID (<http://orcid.org>) identifier to our ontology to encourage participation in that initiative. VIVO is well positioned to provide a platform for institutions to submit consistent, authoritative information about researchers in batch format to jumpstart ORCID adoption in the spring of 2012.

2. Which organization(s) are responsible for the name identifier system?

The institution hosting VIVO is responsible for creating and maintaining VIVO URIs. Institutions normally populate VIVO with additional local identifiers for employees, departments, and grants, allowing VIVO to serve as a crosswalk between local and external identifiers and systems.

VIVO is distributed under a the open-source BSD license (<http://www.opensource.org/licenses/bsd-license.php>) and maintained by a community coordinated by the development teams from the NIH-funded VIVO collaboration and additional contributors from several institutions beyond the grant.

Tools presently include VIVO itself (<http://sourceforge.net/p/vivo/home/VIVO/>), the VIVO ontology (<http://vivoweb.org/ontology/core>), the Harvester (<http://sourceforge.net/p/vivo/harvester/home/VIVO%20Harvester:%20Enabling%20ETL%20for%20the%20National%20Network%20of%20Scientists/>), the search tools (<http://vivosearch.org/about>), and the underlying Vitro ontology editor and semantic web platform (<http://vitro.mannlib.cornell.edu>). Additional related tools are available for download at <http://sourceforge.net/projects/vivo/files/Utilities/>.

3. What is the scope of your identifier system, in terms of the type of people it covers?

(For example, does it include: book authors, active current researchers, formerly active researchers, doctoral students, masters students, etc.)

The scope is a matter of local institutional decision, both in terms of inclusion and retention of information about people who have move elsewhere or become deceased. Cornell and several other institutions anticipate adding graduate students as a group in the future; until such time students are typically included only through co-authorship of publications.

VIVO encourages entering all book authors as individuals of `rdf:type foaf:Person`, using the Friend of a Friend ontology, even if no further information is known about the author at the time. By storing all authors as first-class entities all other available properties of a `foaf:Person` can be applied in the event that additional information such as name components, email address, or institutional affiliation becomes available.

We anticipate using `owl:sameAs` relationships to connect multiple references to the same person – experiments to this effect at the May, 2011 VIVO Hackathon event (<https://github.com/timrdf/csv2rdf4lod-automation/wiki/Example:-vivohack11>) were initially promising.

4. How is your system populated with data?

(by researchers themselves/their institutions/funding bodies)

VIVO is designed to be populated either by batch ingest or direct end user entry, with the latter managed via the institution's single sign-on system.

Batch ingest tools have been written for both institutional data and data from external sources including Pubmed, grant funding agency databases, and publishers. Data imported from any given source may optionally be stored in a separate RDF graph to facilitate tracking provenance, and this is a prime area identified for further development.

5. Who is authorised to make changes to the information in the system?

VIVO has a robust authorization system to associate defined roles and mine semantic relationships inherent in the data to evaluate whether the currently authenticated user may add or modify any specific data. The latter system has been implemented primarily to govern self-editing, and extensions are planned to support proxy editing, either by explicit permission or via semantic relationships.

6. How are identifiers assigned?

The VIVO application makes no assumption of any semantic content in the final "local name" component of the URI, following the institutional namespace. Some institutions assemble the local names for URIs in ways that allow ingest and updating tools to reliably predict what a URI will be given a data source and another institutional identifier, independently of VIVO itself.

7. What form does the identifier take?

A standard URI – examples include <http://vivo.cornell.edu/individual/individual5227>, <http://vivo.ufl.edu/individual/n25562>, <http://vivo.iu.edu/individual/person25557>

HTTP requests to these URIs return HTML in a normal browser but RDF through standard content negotiation (e.g., via Marbles, <http://marbles.sourceforge.net/>)

8. What information is maintained in the system?

(e.g., names, alternative forms of names, email addresses, dates of birth, institutional affiliation(s), details of publications, details of grants)

From the beginning, VIVO has maintained information about much more than people, in the belief that research and expertise can best be communicated in the full context of the academic and research activities as well as outcomes. VIVO represents and publishes as linked data information about publications, grants, roles, organizations, research facilities, and events together with bi-directional relationships to each other and to people.

9. With which other systems (if any) does your identifier system interact?

The VIVO project has worked closely with Drs. Melissa Haendel and Carlo Torniai from the companion NIH-funded eagle-i consortium for research resource discovery (<http://eagle-i.org>). This joint work has established high-level connections between the VIVO and eagle-i ontologies, and we see closer interoperability of VIVO and eagle-i through coordinated ontologies and linked open data as an important contribution to improving discovery and analysis of the many dimensions of expertise.
(http://icbo.buffalo.edu/ICBO-2011_Proceedings.pdf, page 267)

A forthcoming release of Harvard Profiles (<http://profiles.catalyst.harvard.edu/>) uses the VIVO ontology natively and exposes its content as linked open data, as demonstrated through the prototype multi-institutional search at <http://vivosearch.org>.

VIVO is an extensible platform that can represent other types of information and other relationships through its extensible ontology without modification to the software (except where customized composite editing or display functions are desired). Projects not using the VIVO ontology as a starting point may elect to use the underlying Vitro software (<http://vitro.mannlib.cornell.edu>); the developers of the Ontology for Clinical Research (OCRe, <http://rctbank.ucsf.edu/home/ocre>) are evaluating Vitro as a tool for populating and maintaining registries of clinical trials within one university or across many
(http://icbo.buffalo.edu/ICBO-2011_Proceedings.pdf, page 303).

10. Is the information in the system made available to other services?

If properly configured, any VIVO or Vitro instance publishes its data in conformance with linked open data standards.

Some VIVO instances opt to support SPARQL endpoints using the Jena Joseki libraries and a connection to the same database driving a production VIVO instance. This configuration avoids the need to synchronize data between VIVO and the triple store while buffering VIVO from the dual demands of HTML and SPARQL request traffic.

11. Is there a license on the data? If so, what is the license?

The following Terms of Use statement for data in VIVO is provided by default:

To the extent copyrightable, the information presented on the VIVO website and available as Resource Description Framework (RDF) data from VIVO at [institution name] is intended for public use and is freely distributed under the terms of the Creative Commons CC-BY 3.0 <http://creativecommons.org/licenses/by/3.0/> license which allows you to copy, distribute, display and make derivatives of this information provided you give credit to [institution name]. Any non-copyrightable information is available to you under a CC0 waiver <http://creativecommons.org/publicdomain/zero/1.0/>. However, source documents, images or web pages attached to or linked from VIVO may contain copyrighted information and should only be used or distributed under terms included with each source or in accordance with the principles of fair use.

12. If yes, how is this achieved (what interfaces/protocols are used) and is the system free to access?

A link to the above Terms of Use statement is included in the footer of every HTML page in VIVO.

Every VIVO installation we are aware of has remained freely accessible for viewing as HTML and offers its data via standard linked open data requests.

13. How is the system funded?

Cornell University funded the original development and deployment of VIVO and continues to fund operational costs at Cornell.

In 2009, the U.S. National Institutes of Health awarded a 2-year grant entitled "VIVO: Enabling National Networking of Scientists" to a consortium of 7 research universities and medical schools (including Cornell) lead by the University of Florida (NCRR U24 RR029822, principal investigator Dr. Michael Conlon). This grant has supported deployment at the 7 institutions, 5 successive releases the VIVO software, the development of the current VIVO ontology, creation of a prototype multi-institutional search using linked open data, and transition to an open source community. Although the grant officially ended on August 31, 2011, work on the project will continue at a somewhat lower intensity under a no-cost extension valid through the 2011-2012 academic year.

Members of the VIVO community are exploring a range of longer-term sustainability options. In addition, institutions within and beyond the original VIVO collaboration have received grants for additional VIVO-related work, and additional grant proposals are pending.

14. Is the system still under active development? If so, what are your priorities for future enhancements?

The system is still under development at Cornell, Indiana University, and the University of Florida using NIH grant funds. Priorities for enhancements will be determined by assessment of the potential benefits for the VIVO community or related communities such as ORCID, alignment with development road maps for the core infrastructure and ontology, feedback from adopting institutions, and the requirements of funded projects.

15. Do you have any plans for integrating your system with external initiatives such as ORCID, ISNI, Mendeley, Zotero, Academia.edu?

The VIVO Collaboration funded five mini-grants from January, 2011 through August, 2011, including an award to ORCID to "explore the interaction of VIVO and ORCID in the scholarly identity ecosystem" (<http://orcid.org/node/229>). The project has developed a connector from VIVO to CrossRef for searching and pulling in publications and has explored the potential for VIVO to provide batch submissions of authoritative institutional information on researchers. In the future, we hope it will be possible for individuals to edit or supplement information in their ORCID profiles from VIVO and vice-versa using open identity exchange mechanisms.

The VIVO collaboration is very interested in promoting exchanges and, where possible, direct interoperability with additional systems including Academia.edu, LinkedIn, Mendeley, and Zotero at the individual researcher and institutional levels, for both data population and data sharing. With version 1.3, VIVO has a rich export capability through which a single linked data request can optionally provide the results of an arbitrary SPARQL query. Connectors can in theory be written to any systems having public APIs available, subject to appropriate authentication and authorization provisions.