

Research Discovery through Linked Open Data

Paul Albert
Weill Cornell Medical Library
Weill Cornell Medical College
New York, NY
paa2013@med.cornell.edu

Kristi L. Holmes
Washington University in St. Louis
St. Louis, MO 63110
holmeskr@wustl.edu

Katy Börner
Indiana University
School of Library and Information
Science
Bloomington, IN
katy@indiana.edu

Mike Conlon
University of Florida
Gainesville, FL
mconlon@ufl.edu

VIVO Collaboration

ABSTRACT

VIVO is an open source semantic web platform that contains information about scholars and their interests and activities. This demonstration will highlight the platform and ontology, data sources, features of the software and the ways that VIVO data can be leveraged for a variety of purposes within and beyond an institution to facilitate collaboration and research discovery.

Categories and Subject Descriptors

D.0 [Software]: General

Keywords

VIVO, Semantic Web, Open Source, Ontology, Research Representation, Profiles

1. BACKGROUND

VIVO [1] is an open source semantic web platform that enables the discovery of research and scholarship across disciplinary and administrative boundaries through interlinked profiles of people and other data relevant to researchers (Figure 1). VIVO allows researchers to highlight areas of expertise, display academic credentials, visualize academic and social networks, and display information such as publications, grants, teaching, service, awards, and more. Profile information can be programmatically imported from authoritative data sources such as institutional sources as well as bibliographic and grants databases. By storing and exposing data as RDF and using standard ontologies, the information in VIVO can either be displayed in a human-readable web page or delivered to other systems as RDF, allowing researcher metadata in VIVO to be harvested, aggregated, and integrated into the Linked Open Data cloud. Data reuse and repurposing enhances the accuracy and currency of the data while increasing efficiencies across an organization.

VIVO was originally developed at Cornell University and is being expanded as a platform to support national and international research networking via semantic web standards with funding from the National Institutes of Health to the University of Florida, Cornell University, Indiana University and implemented at Ponce School of Medicine, Scripps Research Institute, Washington University in St. Louis, and Weill Cornell Medical College. Each site has a local installation that shares data in flexible and openly accessible ways to facilitate research discovery.

2. DATA

VIVO gathers data from a variety of sources: internal data sources such as human resources databases, course listings and university calendars can be programmatically harvested



Figure 1. A VIVO profile.

and leveraged to provide a good perspective of the local environment; award, funding, and publication data can be harvested from external systems; and individuals and their proxies can supplement the profile with additional information. A number of sites are working to highlight the efforts of their researchers contributions to Open Access journals, and VIVO offers institutions an elegant way for sites to highlight their institutional repository as a critical campus resource. By leveraging the data in an institutional repository and prioritizing resources for content, VIVO can serve as an ideal gateway for scholars to access the content

that they need throughout the course of their scholarly activity. Open Access articles can result in greater impact of the work than may otherwise occur in traditional journals. [2] A number of institutions successfully leverage their individual institutional repository platforms (e.g. DSpace, Berkeley Electronic Press) and integrate the data into their VIVO.

3. RESEARCH DISCOVERY

VIVO promotes discovery in several key ways. Because profiles are rich in content and the data are highly structured, VIVO profiles will rank higher in general internet searches than typical webpages or social networking sites. Research discovery is facilitated via content browsing and the structured data lends itself to a variety of applications that can use these data. Data are repurposed and reused in a variety of tools and settings. Across institutions, VIVO provides a uniform semantic structure to enable a new class of tools that can use the data to advance science. Among some of the most compelling applications are VIVO Search [3], a deep semantic search which facilitates discovery of people, papers, grants, and more across VIVO sites; VIVO Searchlight [4] finds people whose work is related to the material on any web page allowing readers to find experts for collaboration, media requests, foundation relations, and more; An inter-institutional collaboration explorer [5] depicts co-authored papers across selected schools and organizations; and VIVO widgets for Drupal, WordPress and OpenSocial provide easy reuse of VIVO data.

Research discovery is facilitated via content browsing and the structured data lends itself to a variety of applications that can make use of these data. A SPARQL query builder

provides a configurable way to access RDF data for many potential uses.

One of the most visible ways of leveraging VIVO data is through the network analysis and visualization tools available in the VIVO platform itself [7]. Visualizations of networks of co-authors and co-investigators on grants have recently been complimented by novel science map visualizations that support the comparison of publication profiles of up to three people, groups, or organizations (Figure 2). Visualizations are interactive and offer options for export as images or data. VIVO enables high quality data to be revealed about researchers, their collaborators, their funding sources, and more. These data can serve as the foundation for further network analyses and elegant visualizations of the research enterprise on the individual, local, and global levels.

4. THE VIVO COMMUNITY

VIVO enjoys a robust open source, open community space on SourceForge. The site offers a comprehensive project wiki, focused topic listservs, and ways of connecting with other VIVO project members such as event listings and collaborative phone calls. The VIVO software and ontology are publicly available at the SourceForge site, along with content that supports implementation, adoption, and development efforts at sites around the world. For more information, visit <http://vivo.sourceforge.net>.

VIVO and other compatible applications produce a rich network of information across institutions, agencies, and other organizations that can be searched to foster collaboration and enable open research discovery. The VIVO software and ontology are publicly available at <http://vivo.sourceforge.net> along with content that supports implementation, adoption, and development efforts around the world. To learn more about VIVO, please visit <http://vivoweb.org/>. VIVO is supported by NIH Award U24 RR029822.

5. ACKNOWLEDGMENTS

VIVO is supported by NIH Award U24 RR029822.

6. REFERENCES

- [1] <http://vivoweb.org/>
- [2] The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. <http://opcit.eprints.org/oacitation-biblio.html> (accessed 30 January 2012)
- [3] <http://vivosearch.org/>
- [4] <http://vivosearchlight.org/>
- [5] <http://xcite.hackerceo.org/VIVOviz/visualization.html>
- [6] <http://code.google.com/p/linked-data-api/>
- [7] <http://vivo.cns.iu.edu/gallery.html>
- [8] <http://vivo.sourceforge.net>.

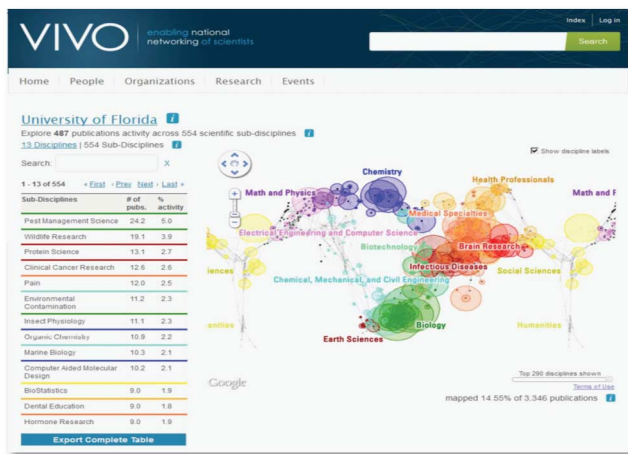


Figure 2. Map of Science visualization for the University of Florida.

is part of the standard VIVO distribution and allows for end-user data requests from a VIVO implementation, enabling access to data for subsequent analysis, evaluation, and visualization of the research enterprise. One new promising option is the Linked Data API [6], which