VIVO: Enabling National Networking of Scientists

Dean B. Krafft
Cornell University
201 Olin Library
Ithaca, NY 14853
607-255-9214
dean.krafft@cornell.edu

Nicholas A. Cappadona, Brian Caruso, Jon Corson-Rikert
Cornell University
Albert R. Mann Library
Ithaca, NY 14853
nac26,bdc34,jc55@cornell.edu

Medha Devare, Brian J. Lowe, and VIVO Collaboration
Cornell University
Albert R. Mann Library
Ithaca, NY 14853
mhd6,bjl23@cornell.edu

ABSTRACT

The VIVO project is creating an open, Semantic Web-based network of institutional ontology-driven databases to enable national discovery, networking, and collaboration via information sharing about researchers and their activities. The project has been funded by NIH to implement VIVO at the University of Florida, Cornell University, and Indiana University Bloomington together with four other partner institutions. Working with the Semantic Web/Linked Open Data community, the project will pilot the development of common ontologies, integration with institutional information sources and authentication, and national discovery and exploration of networks of researchers.

Building on technology developed over the last five years at Cornell University, VIVO supports the flexible description and interrelation of people, organizations, activities, projects, publications, affiliations, and other entities and properties. VIVO itself is an open source Java application built on W3C Semantic Web standards, including RDF, OWL, and SPARQL. To create researcher profiles, VIVO draws on authoritative information from institutional databases, external data sources such as PubMed, and information provided directly by researchers themselves. While the NIH-funded project focuses on biomedical research, the current Cornell implementation of VIVO supports the full range of disciplines across the university, from music to mechanical engineering to management.

There are many ways a person's expertise may be discovered, through grants, presentations, courses and news releases, as well as through research statements or publications listed on their profile—resulting in the creation of implicit groups or networks of people based on a number of pre-identified, shared characteristics. In addition to formal authoritative information and relationships, VIVO can also support the creation of personal work groups and associated properties to represent the informal relationships evolving around collaboration.

Keywords

VIVO, semantic web, linked open data, ontologies, profiles.

1. INTRODUCTION

VIVO touches on a number of the key issues of Web Science [5]. By describing the interrelated web of people, organizations, interests, projects, grants, publications, and more, it can serve the needs of research, scholarship, and higher education as an enabling technology for a wide range of applications. VIVO at Cornell already supports the discovery of potential collaborators; connects students with the right courses, projects, and faculty; and enables the outside world to discover and explore the full range of research and scholarship at Cornell. The VIVO project plans to build applications over participating institutions to recommend collaborators, visualize the network of scientific relationships, and support communication among researchers with common interests.

VIVO also raises some interesting issues in the areas of trust, privacy, and information reliability. By drawing most of its information from institutional databases, the semantic data provided by VIVO is potentially much more authoritative than either self-reported or third-party data. At the same time, exposing institutional data brings potential privacy risks, both in terms of exposing confidential data and in terms of providing information on grants, publications, or research before their official public release. Since VIVO also supports the self-editing of selected information, there are further issues of both information reliability and control. Finally, VIVO can be independently installed at a potentially very wide range of institutions and professional organizations. Choosing which institutions to include in a particular search or analysis system raises issues of both selection criteria and trust.

In this paper we describe the history of the VIVO effort, touch on some of the technology underlying VIVO, explain how VIVO is implemented and sustained, talk about its relationship to both librarians and institutions, and explore some of its opportunities and challenges. Given the broad interest in VIVO extending far beyond the participating institutions, VIVO has the potential to become a key element in the infrastructure of the Semantic Web.

2. A BRIEF HISTORY OF VIVO

The Cornell University Library developed VIVO to support two multi-million dollar initiatives in the Life Sciences (1997) and Social Sciences (2004), which were created to promote inter-disciplinary collaboration and to help recruit competitive faculty and students in targeted focus areas spanning traditional department boundaries. However, finding collaborators across Cornell's many departments, colleges, and four geographically distant campuses, and representing its rich research landscape effectively to candidates using existing tools proved challenging. VIVO was developed to provide an integrated view of the life sciences across disciplinary and administrative boundaries, and has provided a solution to these research discovery issues.

Through early engagement with key administrative and faculty stakeholders, VIVO's initial launch in 2004 soon generated requests that coverage be expanded to other disciplines, and for filtered views by major academic priority areas for the College of

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.

Agriculture and Life Sciences, and this was accomplished in 2005 [4]. In 2007 the Office of the Provost committed funding to the Library to expand the scope of VIVO to the entire university. This funding has supported the development of an interactive editing tool and web services to facilitate data sharing from VIVO to other websites and information consumers at Cornell. Rapid expansion of content scope made it essential to go beyond manual content acquisition and to tap other sources of information including external publications databases and Cornell administrative systems of record. By working closely with managers and IT staff, VIVO is now able to integrate regular feeds from the human resources system (Peoplesoft), the sponsored research (grants) database, the course database, publications databases, and faculty reporting systems for several Cornell colleges. Thus, appropriate publicly-visible data is mapped into VIVO from a variety of authoritative sources as updates that require no additional effort by faculty.

An informal network of department, research center, and core facilities staff has also volunteered to review and update content, and individuals wishing to update information themselves can log in via Cornell's single sign-on to supplement the information from systems of record. VIVO not only supports a public search and browse interface, but it provides an integrated source of rich information to supplement Cornell websites focused on international activities, graduate student recruitment, and campus-wide initiatives in entrepreneurship.

VIVO was originally designed as an ontology implemented in a relational database, but one which emulated Semantic Web principles and presented the relationships among the people, departments, grants, publications, and events at Cornell as a primary feature of the interface. By 2007 improvements to Web Ontology Language [1] (OWL) and Resource Description Framework[2] (RDF) tools, including Java libraries to manage large RDF models, the SPARQL query language [3], and improved reasoners, made it clear that VIVO would be both more flexible and more sustainable if converted to read and write OWL ontologies and RDF data. This conversion was accomplished [3], and it has enabled VIVO to fit into a much larger ecosystem of tools, as well as helping to ensure that data will not be locked in as standards continue to evolve. The emergence of the Linked Open Data[4] movement provides an opportunity for aggregating data from VIVO installations at multiple institutions without requiring a centralized database or support structure.

In September 2009 the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH) provided a grant of $12.2 million to develop and implement a new version of VIVO to enable national networking of scientists. Led by the University of Florida, the VIVO team consists of Cornell University and Indiana University Bloomington as lead development partners, with Weill Cornell Medical College, the Washington University at St. Louis School of Medicine, The Scripps Research Institute, and Ponce School of Medicine as implementation partners. The remainder of this paper describes how we propose to accomplish this goal, and some of the challenges and opportunities involved.

3. THE VIVO TECHNOLOGY

VIVO is an application built on Semantic Web technologies that combines a web-based ontology and instance editor, content integration tools, and a lightweight content management system. The VIVO core ontology provides the framework to model scientists and their activities in a consistent manner. The VIVO data ingest tools enable populating the ontology from a range of local and national data sources, while the self-editing interface allows end users to supplement standard sources. VIVO's content management system provides site curators additional control over the information architecture and display features as well as branding. Via the Linked Data standard, VIVO can be connected to the rest of the Semantic Web and the range of services interoperating on it.

VIVO is implemented as a Java web application, and the current release of VIVO is designed to run in a Tomcat[5] servlet container. VIVO supports OWL ontologies, which are represented in the system at runtime as Jena[6] object models. These ontologies can either be developed using an external editor, such as Protégé[7], or they can be developed within VIVO itself. Section 4 describes the specifics of the VIVO ontology.

Once the standard VIVO ontology is loaded, with possible local extensions, it can be populated either by direct entry and editing through an administrative interface, or information can be loaded in batch from sources of record within the institution. In addition to directly entered instances, the system uses a Pellet[8] reasoner to create selected inferences from the available data. The level of inferencing can be selected in the VIVO configuration.

Once the data is available, VIVO creates a public web site, with pages showing profile information and affiliations for individuals, and membership, affiliations, and other information for organizational units. Information display is theme-driven and fully customizable. Site information is indexed using Lucene[9], and the site is fully searchable.

VIVO supports authenticated self-editing of a subset of the information about an individual. This allows, for example, faculty to upload images, refer to their personal or project web sites, and provide a research statement or a list of research interests. Authentication can be through Shibboleth [10] or some other institutional single sign-on system.

Starting in mid-April 2010, the latest version of the VIVO software, ontology, and documentation will be available for download at http://www.vivoweb.org. The VIVO software is licensed under a standard open-source BSD license.

[1] http://www.w3.org/TR/owl-features/

[2] http://www.w3.org/RDF/

[3] http://www.w3.org/TR/rdf-sparql-query/

[4] http://linkeddata.org

[5] http://tomcat.apache.org/

[6] http://jena.sourceforge.net/

[7] http://protege.stanford.edu/

[8] http://clarkparsia.com/pellet/

[9] http://lucene.apache.org/java/docs/index.html

[10] http://shibboleth.internet2.edu/

4. THE VIVO ONTOLOGY

All data in VIVO are represented as RDF statements using classes and properties from OWL ontologies. These ontologies specify the types of resources described in VIVO and their relationships. The project is developing a VIVO core ontology to model the people, organizations, and activities involved in scientific research as a collaborative endeavor of the seven partner institutions. In keeping with the spirit of linked data, the VIVO core ontology extends existing ontologies such as the Friend-of-a-Friend (FOAF) ontology [11], which provides the basis for describing persons and organizations, and the Bibliographic Ontology (BIBO) [12], which models documents such as books and journal articles. The VIVO project will identify additional ontologies not directly extended by the core ontology that are likely candidates for interoperability. Mappings to these ontologies will enable VIVO data to be shared among a variety of systems.

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 core ontology is not a constraining schema that prescribes the data that may be entered into VIVO. Instead, it serves as an integration layer that permits data from different institutions to be queried in a consistent way. Individual installations may extend the core with ontologies that reflect available data sources or which model things that may be important to display locally but are not necessary for aggregation at the national level. Differences in institutional structure may also be modeled. 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. These local ontology additions can be created in the VIVO software's built-in ontology editor as well as by loading OWL ontologies made by other tools or found on the Web.

A significant challenge in developing the core ontology is to include enough detail to allow for meaningful cross-site discovery of data, while keeping it simple enough to apply to very diverse academic and clinical institutions. There is a limit to what can be anticipated in advance, and it is expected that the core ontology will evolve in response to the RDF data that is ingested into local installations. Another factor that complicates the VIVO core ontology is the need to track how a researcher's affiliations and activities change throughout his or her career. Employment relationships, for example, may be represented by a simple property relating persons 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 employment relationship began and the date when it ended. While this is a well-understood technique for modeling complex relationships[13], it is not especially convenient for applications interested only in current relationships. We plan to add a simple property linking persons directly to those organizations in which they are currently employed; the values of this property would be populated automatically by a query of the complex relationships.

While the VIVO ontology was based on the ontology used in Cornell University's existing implementation, the VIVO project has been working hard to generalize it to meet the diverse needs of the VIVO partners. With the public release of the ontology in mid-April 2010, we will be seeking extensive outside review and comment. Our goal is to create a core VIVO ontology that can serve as the base for describing individuals and their affiliations at any higher-education or research-focused institution.

5. LIBRARIAN DRIVEN

The academic library is a strong resource for information technology expertise and information management and dissemination at any institution. Further, it tends to be a neutral and trusted entity, with employees who regularly engage with researchers and have a good understanding of the academic landscape and the needs of the research community. These attributes make the institutional library an ideal driver for the VIVO project. As has already been noted, VIVO was developed within the Cornell University Library, and librarians have played a critical role in its development, growth, and adoption. This successful model is being broadened to other institutions within the VIVO project consortium, and libraries will play a central role in a variety of areas including content development, support provision, and outreach to foster adoption.

Librarians are increasingly being hired with MS and PhD credentials as subject experts with intimate knowledge of how research is conducted and the ways in which the business of research might be better supported through information-related solutions. Medical librarians in particular are involved in translational research efforts as well as in instruction in efforts to bridge the "bench to bedside" gap.

Librarian expertise and knowledge is being brought to bear in guiding the development of features, content types and entry, and the appropriate context for institutional implementations of VIVO. The ontology team includes librarians to ensure that the core ontology appropriately models content types of importance in the research and clinical setting. Implementation teams at each institution include librarians in lead or key roles, to inform customization that will more accurately reflect the local scenario, and to provide installation and use feedback to developers. Although much of the data in VIVO profiles is populated via automated feeds, initial data entry and testing to refine content categories in the VIVO interface tends to be manual, and managed by librarians.

Work with VIVO at Cornell has demonstrated that delivering technical capability alone is not sufficient to ensure adoption and usage by either the individual or the institution. For successful adoption, technical innovation must be backstopped by human facilitation—in this instance, by information specialists from institutional libraries. Our assumption is that researcher engagement and outreach by information specialists will promote adoption, usage, and maintenance of VIVO by the research communities in their institutions. The liaison programs already in place at many academic libraries mean that librarians already have the proven capability of engaging with their users, with medical librarians involved in translational efforts often being at the forefront of such engagement.

Good knowledge of user communities and their information needs means that librarians recognize that the motivations for diverse user groups in utilizing VIVO are often different. For instance, while the possibility of finding potential collaborators via VIVO may in itself be a motivator for scientists, clinicians who do not

[11] http://www.foaf-project.org/

[12] http://bibliontology.com/

[13] http://www.w3.org/TR/swbp-n-aryRelations

depend on such tools to "draw business" may be less enthusiastic. Administrators are motivated by yet other functionalities, including the need to quickly find experts in a given area. Similarly, Rogers [6] typified technology adopters by their openness to new technology as innovators (who tend to be interested in the technology itself and quick to test and adopt new tools), the early adopters (who are willing to experiment and enthusiastic users of technology), the early majority (who wait and evaluate a technology before adopting it), the late majority (who may use a tool because others are), and the laggards (who tend to be reluctant to embrace a new tool). Ensuring that VIVO is widely adopted and of use to this variety of user types necessitates knowledge of their motivations and fears—which librarians are very well placed to provide at any academic institution.

6. FREEING INSTITUTIONAL DATA

The goal of the Linking Open Data movement is to "extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources."[14] Typically, data about researchers, projects, and affiliations at institutions of higher education is at best scattered in text form across a variety of web pages, and at worst simply locked up in institutional databases. An explicit goal of the VIVO project is to unlock this institutional data and to make it part of the Open Data universe. This allows researcher information to be combined with other open sources of publication, project, grant, and disciplinary information to support discovery and analysis beyond the VIVO network itself. In addition, as more research data becomes available in linkable forms through sites like Data.gov, VIVO can provide the "researcher context" to find, understand, and explore that data universe.

Every higher education institution maintains a number of information systems containing human resources, grant, faculty reporting and other types of institutional data. Frequently, the data in these systems are not available simply because the system provides no human or machine accessible feeds of that data, or has no mechanism to screen out private data and enable re-use for broad institutional and public purposes. Data from different silos may also be missing common identifiers or be normalized to incompatible units (e.g., the definition of a department in a financial system may differ from the human resources system). Providing a mechanism to integrate publicly visible data elements from multiple sources and present these data in a useful format is a key component of the VIVO project.

There are two elements of opening up this institutional data. The first is to present human users with an integrated, web accessible view in the form of a web site, with the institutional data optionally supplemented by direct entry and by external sources such as NIH Reporter[15], NSF award search[16], PubMed[17], or licensed publications databases. From such a web site, VIVO provides a single point for discovery of research and scholarship across an entire institution, where, based on Cornell's experience, as much as 90% of traffic will come through external search engines. The second is to leverage the integrated database to create a source of appropriately filtered data in standard formats (RDF, XML, or JSON) for use by automated systems via web services or as linked data. The effort to establish and maintain a VIVO installation can provide a return on investment all across the institution via search interfaces and/or live feeds of selected content to additional websites.

The data found in institutional systems come in a variety of formats. Working with our initial partners, the VIVO implementation team at the University of Florida is expanding on the initial set of VIVO tools to develop new extract, transform and load (ETL) processes to get institutional data into local VIVO installations. The goal of this process is to transform the data into RDF statements that match the VIVO core ontology, with potential local extensions. The process must also support incremental update of the existing information in VIVO.

Once the data has been loaded into a VIVO installation it will also be exposed as linked data in various RDF formats. A linked data request comes into VIVO as a standard HTTP request modified only to specify RDF/XML (or another format of RDF) rather than HTML as the response format. The request can be interpreted much as a web crawler indexes an HTML page, with the added advantage that RDF provides machine-readable structure expressed in the defined namespaces of the VIVO ontology—linking directly to a term in a controlled vocabulary or to the related resources referenced with the page requested. VIVO will also incorporate RDFa markup within pages to make data visible to search engines not yet ready to consume linked data directly.

In addition to exposing researcher information through accessible RDF, the VIVO project is also exploring making SPARQL endpoints available to support arbitrary queries on a per-institution or network-wide basis. We are interested in working with the Linking Open Data community to ensure that information from VIVO is as usable, useful, and used as possible.

7. ENABLING NATIONAL NETWORKING

The primary goal of our approach to enabling national networking of scientists is to offer an open and evolving perspective on research across institutions not just to scientists, but also to students, administrative and service officials, prospective faculty and students, donors, funding agencies, and the public, empowering them both to contribute to the network and to make it their own. The focus at the individual, institutional, and national level is on illuminating previously "invisible" human assets and ensuring that this network of assets within and between academic institutions will be adopted, used and maintained as a result of its value to constituents.

Apart from driving use by providing pre-populated profiles that are easily modifiable, VIVO provides value to the individual by enabling the ability to advance research and academic standing; discover potential collaborators, find "people like me" or people with complementary skills; and facilitate research currency, maintenance and communication. Individuals are more likely to use a tool like VIVO if it minimizes the need for duplication of effort, such as that involved in generating a CV or biosketch for a grant application, for instance. VIVO already has much, if not all of the content necessary for such documents; we envision a feature in the near-future that will enable users to easily publish a

[14] http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description

[15] http://projectreporter.nih.gov/reporter.cfm

[16] http://www.nsf.gov/awardsearch/

[17] http://www.ncbi.nlm.nih.gov/pubmed

biosketch from their profiles. Similarly, the ability to visualize co-authorship, collaboration, or co-PI networks is very valuable to researchers, and this planned feature is another driver for use and sustainability.

To drive institutional use of VIVO, institutions must perceive value in it. VIVO provides value to the institution by increasing efficiencies associated with the management and dissemination of information about people and resources through data interoperability and sharing. Further, it represents a powerful recruitment and retention tool for both faculty and students, and has the potential to advance the standing of the institution by seamlessly representing and showcasing research and scholarship across the entire academy.

At Cornell, VIVO's campus-wide cross-referencing search capability and large index of researchers, resources, and facilities make it a core service whose timeliness and need are becoming clear to faculty and administrators alike, as evidenced by some of the comments we received from the life sciences community and beyond:

"VIVO helps build a "virtual" research community [which can] have several advantages – it is fluid, can change relatively rapidly and can present itself in more interesting ways to the rest of the world than more traditional structures."

"As an administrator, it is often difficult to access information in a manner that is quick and in a form that is useful. VIVO and the CALS portal could be a major step forward in reducing the amount of time we spend collecting information and in better utilizing [it]."

"VIVO provides unparalleled access to information about the life sciences at Cornell in a user-friendly way. This will be of particular benefit not only to those researchers and students already at Cornell, but potential faculty and students as well, by offering a much-needed, integrated view of the life sciences community at Cornell."

"VIVO saved my life as a new faculty member at Cornell; I used it all the time to find facilities and people I might work with."

It is very likely that the continued use and success of VIVO will also depend on an institution's central IT administration finding value in it. For this entity, VIVO's worth might be that it seeks to reach the widest possible audience by leveraging – not supplanting – existing systems and data warehouses to allow the materials prepared by information providers to be discovered via as many paths as possible. Additional value may be provided by the fact that VIVO affirms the value of a service-oriented architecture (SOA) model by enhancing access to cross-disciplinary research and collaboration across the institution—which is invaluable in this era of ever-increasing inter-disciplinary work among scholars and scientists.

Enabling national networking does not entail a central portal; institutional VIVO installations can provide access to both local and national-level research information. Institutions will be able to participate in national-level networks either by implementing VIVO, or by providing RDF data that is mappable into the core ontology.

The VIVO project will build an exemplar network search capability to support discovery and visualization across all the participating institutions. This code will be released, and can serve as the basis for national, regional, or other specialized discovery networks over the universe of participating institutions. In addition, we expect outside groups, both commercial and non-commercial, to harvest the open VIVO data and to use their own tools to support discovery, analysis, visualization, and browsing across it.

Ultimately, the success of VIVO as a part of the research infrastructure will depend on the institutions who adopt it, the researchers who populate it with information and use it to discover new opportunities, and the tool and infrastructure operators who make it widely available. We believe that VIVO provides compelling value for all of these users.

8. FUTURE PLANS

The major impetus for the National Center for Research Resources of the NIH to fund the VIVO effort was 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" [18]. By creating a common system for uniformly describing the entire context surrounding a researcher and his or her projects, expertise, and interests, VIVO provides the basic information and infrastructure to enable the discovery of potential collaborators, the creation of research teams, and the visualization and analysis of the scientific social networks in which these researchers live and work.

Over the next 18 months the VIVO project will use the researcher-related data from each of our partner institutions to create open-source exemplar systems to:

- Aggregate RDF data from the institutions to support search and discovery over the national network of VIVO institutions and compatible linked data providers;
- Build on the work of Katy Börner and her colleagues [1, 2] to analyze and visualize collaborators and collaborations across the network, drawing on common publication, grant, and project participation; and
- Support discovery of collaborators through analysis of affiliations (network distance metrics) and selection based on specific disciplinary vocabularies exposed as RDF, such as MeSH[19] and the Library of Congress Subject Headings[20].

More broadly, VIVO is committed to the open publication, use and reuse of this researcher related data. The project plans to make data in VIVO available through embedded RDFa, through web APIs supporting the bulk download of RDF, and potentially through SPARQL endpoints. One of our biggest opportunities is in working with existing Semantic Web efforts to ensure that they can fully leverage VIVO data and integrate it into larger scientific and social systems.

To assist in our engagement with the broader technical community, the project has created a Technical Advisory Board to provide input and guidance on our efforts. Consisting of fourteen

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

[19] http://www.nlm.nih.gov/mesh/

[20] http://id.loc.gov/authorities/

experts from the semantic web and web science communities, this group is working to help us identify specific groups and opportunities for potential collaboration.

VIVO itself is a relatively small, focused effort. Its success will be measured by the extent to which it becomes "infrastructure", a seamless piece of the larger fabric of discovery, collaboration, networking, analysis, and understanding. Our most important challenge and opportunity is to fully engage with the research, educational, and technical communities to make this happen.

9. COLLABORATORS

The VIVO Collaboration consists of the following researchers and contributors: *Cornell University:* Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Huda Khan, Brian Lowe, Holly Mistlebauer, Stella Mitchell, Anup Sawant, Christopher Westling, Rebecca Younes. *University of Florida:* Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Yang Li, Paula Markes, Sara Russell Gonzalez, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. *Indiana University:* Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. *Ponce School of Medicine:* Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. *Scripps Research Institute:* Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaeleen Trimarchi. *Washington University, St. Louis:* Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. *Weill Cornell Medical College:* Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klaz, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.

10. ACKNOWLEDGMENTS

The project described was supported by Grant Number U24 RR029822 from the National Center for Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health. The authors would like to thank both Mann Library and Cornell University for their support of and commitment to the VIVO project over many years.

11. REFERENCES

[1] Börner, K., Dall'Asta, L., Ke, W. and Vespignani, A. Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams, 2005, http://arxiv.org/abs/cond-mat/0502147v1.

[2] Boyack, K.W., Klavans, R. and Börner, K. Mapping the backbone of science. Scientometrics, 64 (3). 351-374, http://dx.doi.org/10.1007/s11192-005-0255-6

[3] Corson-Rikert, J., Caruso, B. and Lowe, B. Vitro - Integrated Ontology Editor and Semantic Web Application. Available: http://vitro.mannlib.cornell.edu/. (2009) Accessed 2009 June 11, 2009.

[4] Devare, M., Corson-Rikert, J., Caruso, B., Lowe, B., Chiang, K. and McCue, J. VIVO: Connecting People, Creating a Virtual Life Sciences Community. D-Lib Magazine, 13 (7/8), http://www.dlib.org/dlib/july07/devare/07devare.html.

[5] Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T. and Weitzner, D. Web science: an interdisciplinary approach to understanding the web. 51 (7). 60-69, http://doi.acm.org/10.1145/1364782.1364798.

[6] Rogers, E.M. Diffusion of innovations. Simon and Schuster, New York, NY, 1995.