Letting the Good Times Roll through Alignment: Meeting Institutional Missions and Goals with VIVO, a Web-based Research Discovery Tool

Sara Russell Gonzalez, Ph.D., M.L.I.S.
UF VIVO Implementation Lead
Marston Science Library, University of Florida

Valrie Davis, M.L.S.
VIVO National Implementation Coordinator
Marston Science Library, University of Florida

Michele R. Tennant, Ph.D., M.L.S.
UF VIVO Outreach Lead
Health Science Center Libraries and UF Genetics Institute, University of Florida

Kristi L. Holmes, Ph.D.
WUSTL VIVO Outreach Lead
Becker Medical Library, Washington University School of Medicine

Mike Conlon, Ph.D.
PI, VIVO
Associate Director, Clinical and Translational Science Institute
University of Florida

VIVO Collaboration

Abstract

SLA's Alignment Project seeks to align library and information services and resources with the vision, mission and goals of the parent institution. For institutions with research missions, collaboration and discovery are key. Twenty-first century science requires a team-based multi-disciplinary approach. Researchers need to discover experts with similar and complementary experience in order to create research teams. Institutional administrators require tools to identify research trends at their home institutions as well as identify and track competitors. Librarians need to identify institutional research and educational priorities in order to create appropriate information services and build relevant collections. One tool –

VIVO – can lead researchers, administrators, and librarians to the information required to meet the institution's goals.

In 2003 the Cornell University Library staff aligned their expertise with the research and educational missions of their institution and created VIVO – a web-based researcher directory and discovery tool. In 2009, Cornell and partner institutions the University of Florida, Washington University School of Medicine in St. Louis, Indiana University, Ponce School of Medicine, Scripps Research Institute, and Weill Cornell Medical College were awarded $12.2 million by the National Institutes of Health to enhance and expand VIVO into a national network of researchers. Most of the partner institutions have followed the lead of Cornell and have placed the library at the center of user instruction, adoption facilitation, and other key activities, thus aligning clearly the libraries' services with institutional needs.

1. Introduction

After a century of significant change in librarianship and technology, the 2009 SLA Alignment project is an opportunity for the association to re-evaluate its core values and mission and ensure that its vision and identity communicate the value of the profession and set an ambitious course for the future. Similarly, the nature of research is quickly changing, becoming increasingly multi-disciplinary as technological advances facilitate new methods of collaboration across institutions and into new media. Identifying potential collaborators is challenging within complex institutions such as universities or academic medical centers; it is even more difficult when the required expertise resides outside the home institution. Libraries have traditionally enjoyed strong liaison relationships with their communities through offering assistance on projects and providing customized information and educational opportunities. These strong liaison relationships position the library to play a key role as connector on campus.

This transition to interdisciplinary science is an exciting opportunity for academic libraries to further strengthen their liaison relationships to their research community and re-align the library as a full partner in scholarly research. The interconnectedness of the academic library makes it ideal to assist researchers in forging these collaborations and information discovery. To facilitate the discovery process, a collaboration of seven academic institutions is developing an open-source Semantic Web application, VIVO, that contains publicly viewable researcher profiles, grants, publications, and seminars and course listings. Librarians are integral team members on the VIVO project and are assisting in developing it into a national network for researchers.

This paper will introduce VIVO and describe how implementing this web-based discovery tool allows libraries to significantly raise their profile at an institution and align their services with the research and collaboration needs of the parent institution. We will discuss ways in which VIVO can be used to meet the goals and information needs of the typical research institution and the major stakeholders in the research process. Finally, the future of VIVO will be outlined and opportunities for interested organizations and individuals to become participants in the VIVO national network will be presented.

2. Description of VIVO

VIVO is a Semantic Web database that displays and makes fully searchable the research activities of an institution. Researcher profiles contain information such as publications, grants, educational background, position history, courses, research keywords, contact information, and more. The VIVO database is installed and managed by an institution and data are added using a variety of methods. The first is through the ingest of bulk data from local authoritative sources, such as human resources, grants, and faculty activity sources. Second, through ingest from external data sources, such as publication (PubMed) and grant (NIH Explorer) databases. Third, data can also be manually inputted by librarians, departmental assistants, or self-editors. As all data is considered linked open data, VIVO provides a way of exposing, sharing, and connecting pieces of dataon the Web using URIs and RDF.

In 2003, Cornell University Library began development of VIVO as a way to align their mission with the research and education mission of their institution. Two campus initiatives were timely to the development of VIVO, the first designed to increase interdisciplinary collaborations within the Life Sciences and the second to increase graduate student recruitment. VIVO contains profiles for all Cornell researchers, not just those in the life sciences. A researcher searching in VIVO on a topic, such as "microarray", provides information on all researchers, publications, facilities, departments, grants, courses, and more, related to that topic. VIVO is a resource open to anyone to search, whether employed by Cornell or a community member, and allows researchers from other institutions to locate Cornell collaborators. Cornell then created a secondary portal for graduate students which utilized VIVO data to highlight graduate programs and advising faculty. In 2007, University of Florida librarians were consulted by faculty interested in tracking publications authored by the UF community. VIVO was identified as a possible solution, as well as a method for supporting a growing need for collaboration within the UF community. By the end of 2007, the UF Libraries began implementation of their own local VIVO that

would first cover the science and engineering faculty and then later expand to other disciplines.

In 2009, the National Institutes of Health (NIH) announced a American Reinvestment and Recovery Act grant solicitation, seeking proposals to develop a national network to facilitate discovery of scientists and resources and encourage interdisciplinary collaboration. University of Florida and Cornell University, and partner institutions Washington University School of Medicine in St. Louis, Indiana University, Ponce School of Medicine, Scripps Research Institute, and Weill Cornell Medical College proposed and were subsequently awarded $12.2 million over two years to enhance and expand VIVO into a national network of researchers. VIVO would be the platform and the seven institutions would be the first participants in a growing national network.

The VIVO project teams are Development, Outreach, Implementation, Ontology, Evaluation, and Publications. Development, led by Cornell, is separated into application development (Cornell); development and expansion of the data ingest structure (UF); and the creation of the social networking and visualization features (IU). Outreach has representatives, primarily librarians, from the seven institutions and was responsible for interfacing VIVO with potential participants, marketing and branding efforts, and website creation (vivoweb.org). Implementation also included representatives, both librarians and IT support, responsible for the implementation of local VIVO systems. Ontology focuses on the development of the VIVO ontology and also includes representatives from all institutions. Evaluation, led by Washington University School of Medicine, works closely with sites to evaluate the usefulness of VIVO. Lastly, Publications works with database providers and vendors to create agreements for data sharing to the VIVO participants.

At the six month mark, all seven institutions have implemented a functioning version of the VIVO platform and have begun inputting authoritative data and marketing VIVO to the research community. Furthermore, the application is now open source and RDF compliant, providing opportunities for additional institutions and individuals to explore implementing their own instance of VIVO or collaborate by developing external applications that consume and repurpose VIVO data.

3. Needs of stakeholders

Researchers & Faculty

Science is rapidly moving away from strictly defined fields into a new era of interdisciplinary research and collaboration. Identifying potential

collaborators is challenging within complex institutions such as universities or academic medical centers; it is even more difficult when the required expertise resides outside the home institution. VIVO gives researchers the ability to locate new collaborators and identify competitors along with their research activity. The social networking features will allow a researcher to contact potential collaborators or monitor and stay current about new developments in their field.

VIVO also provides researchers an outlet to showcase their own research and attract collaborators. Profiles are primarily created and maintained from authoritative internal and external data sources rather than relying on manual input. These profiles can be individually customized to include information that researchers wish to highlight including descriptive overviews of their research programs and notable accomplishments. Faculty profiles can also be used to attract potential students who are seeking advisors to work with in a graduate program or undergraduate project.

Because VIVO is capable of exposing all data as RDF, external applications are being written to capture VIVO data and utilize and transform it for a myriad of purposes. One such application currently being integrated with VIVO consumes the profile data and outputs it into a CV or NIH biosketch format. This meets a need for many researchers who are required to submit curriculum vitas or biosketches for annual reporting purposes or as part of a grant proposal. Another output option currently being developed is the ability to create a webpage that displays a researcher's or organization's profile data in VIVO. Currently, researchers and faculty maintain descriptions and listings of their research and teaching activity in multiple sites including departmental webpages, tenure or promotion packets, annual reports, and vitas and biosketches. A similar idea being developed is that external sites, such as grant and conference submission forms, could interface with VIVO and fill in automatically the necessary professional data. The possibility that researchers can update and approve data in a single, already pre-populated site provides a strong draw to utilize VIVO.

Administration

A major challenge for an institution's administration is to review and understand the productivity and accomplishments of the research faculty and staff. Administrators often also need to identify areas of instructional strength, especially when in the process of restructuring the organization or realigning the mission of the institution. Data is often scattered throughout multiple institutional sources and may even be in paper, rather than digital, format. These factors make institutional review prohibitively difficult for administrators and can result in lost opportunities for showcasing the institutional's research strength and forming collaborations and partnerships with other entities.

By enabling and encouraging research staff, departments, and laboratories to maintain data within VIVO, administrations will acquire a powerful tool to view and search research data within one location. An example of such a situation is an administrator hosting an international delegation who would like to identify all research projects that involve that particular country. Typically, the provost would rely upon institutional knowledge, the internet, or word of mouth to locate the projects and affiliated researchers. However, this can be daunting in a large, scattered institution (such as many universities or multi-campus corporations) and there is a distinct possibility an important project will be overlooked. With VIVO, the administrator can simply browse by geographic location, identifying researchers and projects through pertinent grants, publications, research activities, and descriptive overviews. Another scenario is an administrator who uses information contained in VIVO to assist with promotion and pay raise decisions rather than require lengthy paper reports from staff and faculty. The flexibility and strength of the VIVO ontological model holds promise for enhancing and widening the communication channels between an organization's research staff and administration

Students

Students need faculty and department information when they are investigating majors and classes and selecting a school or advisor. It can be challenging to find appropriate mentors and programs, especially when seeking information about an unfamiliar school. Undergraduates can struggle locating graduate advisors in the suitable department and this problem is compounded in interdisciplinary research areas where there is not an obvious graduate department within which to search. VIVO offers students an alternative to Googling or browsing websites by providing a method to efficiently locate a potential advisor, dissertation committee, or department through keyword and controlled vocabulary search results. Cornell University's Life Sciences graduate school realized very early the potential of VIVO for reaching potential graduate students and developed a portal that filters Cornell's VIVO instance to display and highlight relevant graduate advisor research activity and availability (http://gradeducation.lifesciences.cornell.edu/).

Donors and Funding Agencies

The richness of the VIVO ontological model allows researchers and departmental units the ability to showcase the entire breadth of their research activity. These profiles can be searched and browsed by donors and funding agencies to discover potential projects and researchers to support. Fundraising arms of institutions can use VIVO to highlight areas of current research to better demonstrate funding merit and outcomes of

currently funded projects. Funding agencies also regularly need panel and proposal reviewers with specialized expertise. VIVO provides these agencies with a database to search and retrieve experts that might not have been normally identified, especially in complex multi-disciplinary areas of research.

Public and Media

Many institutions and states have expertise databases that list affiliated researchers and topics about which they are willing to speak to the media. These databases are traditionally manually derived, relying on the researchers and institutional personnel to voluntarily self-identify areas of expertise, contact information, and educational background. VIVO meets the needs of journalists by providing a search interface to locate experts through keyword matching research descriptions, grants, and publications. Additionally, the media and general public can easily search or browse VIVO to learn about new research advances and find public events occurring at local institutions. VIVO aggregates events such as seminars, conferences, and lecture series - information that is often distributed across various departmental websites - and future functionality will provide users to create notification feeds to alert them of relevant happenings.

Libraries

Libraries need information about their dependent departments and researchers to adequately provide services and focus collection development. Specific information that is valued includes recent publications, journals used, new areas of research, course topics, recent grants, and the arrival of new faculty. The growing involvement of libraries in the data curation effort means that librarians also need to identify datasets that need to be curated. Commonly, there is no single strategy for librarians to acquire this data. Librarians often monitor faculty and lab webpages or attend department seminars to try to keep abreast of developments but this can be difficult, especially if librarians have multiple subject responsibilities or work with large research-intensive departments.

VIVO provides the ideal data source for libraries to learn more about their research community. It delivers easily searchable and browsable department, faculty, and research staff profiles, enabling libraries to closely monitor departmental activities and align services, collections, and budget to meet each department's unique needs. VIVO also allows libraries and staff to create their own profiles, showcasing library resources and services in a research context. Finally, the inclusion of publications within VIVO gives libraries an excellent opportunity to link an institutional repository to VIVO and provides another avenue to serving material to patrons.

4. Role of Library Staff in VIVO

While VIVO provides an opportunity to realign the library with the mission and goals of the institution, the participation of libraries and information specialists contributes greatly to the success of VIVO both as a national network and as a local institutional network. Building a research community requires an understanding of the information needs of a community, the utilization of established institutional channels for dissemination and support, and a strong understanding of the structure of information-all knowledge librarians and other library staff possess. Such staff have experience organizing information and have traditionally enjoyed strong liaison relationships with their communities through offering assistance on projects and providing customized information and educational opportunities. This transition to interdisciplinary science is an exciting opportunity for academic libraries to further strengthen their liaison relationships to their research community and re-position the library as a full partner in scholarly research. The interconnectedness of the academic library makes it ideal to assist researchers in forging these collaborations and information discovery. The liaison role enables libraries at the VIVO partner institutions to contribute in a productive way to all aspects of the VIVO project, including: development, dissemination, education and support, ontology development, data curation, usability, and more.

Within the VIVO national project, library staff are responsible for facilitating the development of the core and local ontologies. The ontology is the underlying structure of how the data is modeled in VIVO and this work is similar to traditional "cataloging" work. Ensuring that the ontology meets the data needs of an institution is essential to ensuring the value of VIVO to the research community. Along with the ontology management, librarians are also involved in locating and selecting subject vocabularies to incorporate into VIVO. These are necessary for improving the search and browse capabilities of VIVO so that users retrieve accurate and relevant results.

With their familiarity with patron information search behavior, library staff are also assisting with the development of user-centered design by identifying and prioritizing many of the possible technical features. This prioritization is based upon the research and networking needs of the research community, identified by focus group and usability testing. Library staff have drawn upon their liaison relationships with departments and researchers to identify faculty who are open to new technology and willing to offer opinions and user testing to improve the application.

The true worth of the VIVO project is determined by the quality and quantity

of researcher information authoritatively ingested into the database. A primary role of the library staff implementing VIVO at each institution is identifying local data sources and negotiating with data stewards both internal and external for rights to obtain and display the data. This non-traditional task has placed the library staff in contact with other institutional groups that manage information, leading to new relationships with the library across campus.

A core responsibility of library staff on the VIVO project is to participate in and lead marketing and outreach to the academic community. The majority of the outreach team members across the seven partner schools are librarians or informaticsts and this role acknowledges the close relationship librarians have with the institution's research community. Such experts are familiar with the needs, style and goals of their liaison departments and how best to market a service such as VIVO to best appeal to that subset of researchers. Although much marketing is being done at an institutional level, library staff are also presenting at library and scientific conferences to reach out to potential users and collaborators. Along with marketing, outreach also is responsible for providing instruction to new users both in the library and in departments.

The successful marketing by the outreach team and multiple grant announcements has raised VIVO, although still early in development, to a high enough level that campus faculty are becoming aware of the service and associating VIVO with the library. This association has prompted some curious queries from faculty about why library staff are involved in developing a database that is not clearly a traditional type of library service. These questions present an outstanding opportunity to inform faculty and administrators about the technical and knowledge management skills librarians can provide in the research process. VIVO provides libraries a new platform to demonstrate their value and skills to their research community and facilitate relationships and communication that can improve alignment of the library with the research mission of the institution.

5. Future steps

Throughout the next year and a half, the project collaboration will continue to develop VIVO into a robust, full-featured Semantic Web application. The ontology is being finalized and local instances will continue to identify and harvest internal and external datasets. Social networking and data visualization features are under development and researcher recognization is growing steadily. There is considerable opportunity for interested institutions and individuals to become involved in the VIVO collaboration as

adopters, data providers or data consumers. Now that the application is open source and publicly available at www.vivoweb.org, potential users are encouraged to download and test the software and consider becoming part of the national network. External applications that utilize the VIVO data are being sought and full details can be found at http://www.vivoweb.org/participate.

Acknowledgements

VIVO Collaboration: Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, Elizabeth Hines, Huda Khan, Brian Lowe, Joseph McEnerney, Holly Mistlebauer, Stella Mitchell, Anup Sawant, Christopher Westling, Rebecca Younes. University of Florida: Mike Conlon (VIVO and UF PI), Chris Barnes, Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Linda Butson, Chris Case, Christine Cogar, Valrie Davis, Mary Edwards, Nita Ferree, George Hack, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Meghan Latorre, Yang Li, Paula Markes, Hannah Norton, Narayan Raum, Alexander Rockwell, Sara Russell Gonzalez, Nancy Schaefer, Dale Scheppler, Nicholas Skaggs, Matthew Tedder, Michele R. Tennant, Alicia Turner, Stephen Williams. Indiana University: Katy Borner (IU PI), Kavitha Chandrasekar, Bin Chen, Shanshan Chen, Jeni Coffey, Suresh Deivasigamani, Ying Ding, Russell Duhon, Jon Dunn, Poornima Gopinath, Julie Hardesty, Brian Keese, Namrata Lele, Micah Linnemeier, Nianli Ma, Robert H. McDonald, Asik Pradhan Gongaju, Mark Price, Yuyin Sun, Chintan Tank, Alan Walsh, Brian Wheeler, Feng Wu, Angela Zoss. Ponce School of Medicine: Richard J. Noel, Jr. (Ponce PI), Ricardo Espada Colon, Damaris Torres Cruz, Michael Vega Negrón. The Scripps Research Institute: Gerald Joyce (Scripps PI), Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaeleen Trimarchi. Washington University School of Medicine in St. Louis: Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Caerie Houchins, George Joseph, Sunita B. Koul, Leslie D. McIntosh. Weill Cornell Medical College: Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Mark Bronnimann, Adam Cheriff, Oscar Cruz, Dan Dickinson, Richard Hu, Chris Huang, Itay Klaz, Kenneth Lee, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Vinay Varughese, Virgil Wong.
This project is funded by the National Institutes of Health, U24 RR029822, "VIVO: Enabling National Networking of Scientists".