VIVO | connect • share • discover

2015 VIVO Conference

August 12-14, 2015 Hyatt Regency Cambridge

VIVO Conference Program

Short Papers

The Value of the Librarian in Implementing VIVO

Bart Ragon, Andrea Denton, Inhye Son, Megan Nunemaker,
Jeremy Bartczak and Michael Wilson

Abstract: In 2014, the University of Virginia began implementing VIVO using Symplectic Elements as a primary source of publication data. A team of librarians at the Claude Moore Health Sciences Library worked to pre-curate 1,150 profiles for School of Medicine faculty. This work enabled the project to commence at a faster rate than anticipated. Pre-curation also helped alleviate potential concerns faculty may have had over initial false positive citations and allowed librarians to create customized filters that ensured a high degree of accuracy moving forward. As the product was introduced to departments, the library adopted a curation support role, providing training through consultations, online tutorials, and drop-in clinics. The success of the pre-curation strategy established the library is a key project member, creating additional roles for the library including administrative reporting and data architecture. This paper will discuss the value of library participation in the implementation of VIVO and make an argument for its continued inclusion throughout other facets of the project. The team-based curation strategy, lessons learned, and perceived value-added will be discussed in detail.

Fulling VIVO with rich meta-data from a RIS system

Thomas Vestdam

Abstract: A modern Research Information System (RIS) aims to capture very complete, fine-grained and coherent meta-data about research at an institution - such as research outputs, projects, external income and expenditure, awards, activities, impact, and research data sets as produced by the academics/faculty at the institution. One of the main purposes of having a RIS is to perform internal or external research assessments and evaluation of the workfore, or just plain measurement in context of income and output. The over-all requirements for a RIS is a rich and fine-grained model, advanced tools for managing and visualising data within the RIS, supplying workflows, reporting on data, as well as means for public exhibition of data stored in the RIS, in effect showcasing the strength of the institution.

In contrast to a RIS like VIVO - most comercial RIS systems are build on top of a relational database - for many good and sound reasons. So the question is - what if you wan't to supply networking capabilities between traditional, enterprise level, RIS systems, similar to what you can achieve with a number of connected VIVO systems? You could 1) create you own networking tool that aggregates appropriate information from the individual systems as done before for VIVO systems, 2) you can setup a VIVO "index" for each system populated RIS with information from each individual system, or 3) build a SPARQL parser that translates SPARQL quires into you own internal data-model. Common for all these solutions is that they allow your RIS system to connect to a VIVO network. So far we have implemented two of these

options - and, would like to share the experiences we have made so far, both in terms of the usefulness of the tools, but also in terms of the experiences we made build the tools.

Option 1: The concept of a VIVO “index” is quite simple - just push whatever information you have in your master system to a tripple store, while adhering to the VIVO ontology. So you basically just need a triple store and a SPARQL server, like Apache Jena (Fuseki). The challenges here are the mapping of meta-data between the two diffrent meta-data models and implementing a mechanism for incremental updates of the triple store based on updates in the master system. Benefit of the solution is having a VIVO “compliant” exhibition of data via a SPARQL endpoint. Downside being that you now have yet another “server” to maintain. We will elaborate much more on these pros and cons during the presentation.

Option 2: Creating your own networking tool - or a Community Service as we call it - is fairly simple - especially if you are only aggregating data from systems that have a well-defined standard interface for harvesting. Our Community Service currently aggregates more than 90.000 researchers, 3.000.000 publications, 115.000 grants and a small number of equipment (new feature). The Community Services is fed with information from Pure systems (or RIS) via their individual web-services. This web-service provide a basic harvesting mechanism that allows clients (the aggregator) to harvest all information in the instance, or simply harvest changes in the RIS since last harvest. Benefit of such an approach is a lot of aggregated data, where the downside is a closed ecosystem. We will elaborate more on the details of this solution, as well as present pros and cons during the presentation.

Finally we will discuss how we could establish a standard for exhibition for networking tools that is platform agnostic.

Using openWordnet-PT to improve VIVO

Alexandre Rademaker, Daniela Brauner, Glauco Roberto Munsberg and André Guimarães Peil

Abstract: VIVO is an open source semantic web application for research discovery. The power of VIVO relies mainly on the VIVO-ISF ontology and its expressivity to represent all information about researchers and the research domain. VIVO-ISF makes all those types of information interconnected and browsable in the VIVO application. Nevertheless, although VIVO has good support for faceted search across disciplines, it is still not anything more than a keyword-based search engine. That is; it is still not using the power of semantics for information retrieving. OpenWordnet-PT is a lexical-semantic resource describing (Brazilian) Portuguese words and their relationships. It is modelled after and fully interoperable with the original Princeton WordNet for English, relying on the same identifiers as WordNet 3.0. We present ideas to use openWordnet-PT to empower VIVO providing two usable services: concepts disambiguation and query expansion.

Leverage Your VIVO Profiles with Modern Metrics to Tell the Stories of Your Research

Andrea Michalek, Bruce Herbert and Marianne Parkhill

Abstract: You have implemented VIVO – created the researcher profiles, associated research output with the researchers and have massaged the metadata to meet your standards. Now what? What value can you get for this work?

PlumX from Plum Analytics is an impact dashboard that utilizes modern metrics, aka altmetrics, to uncover and tell the hidden stories of your research output. By integrating your VIVO profiles with PlumX profiles you uncover what is

happening with research output. This is especially helpful for recent research and early-career researchers. This helps you answer questions you could not answer before. For example, who is engaging with my research? What societal impact is my research having? How do I showcase my research for grant opportunities?

Texas A&M has integrated VIVO and PlumX. We will showcase the TAMU VIVO-PlumX system and how the library is using the tools to advance faculty online identity, support departmental development of program narratives, and address TAMU's strategic plan, Vision 2020.

The Complexity of Scholar Affiliation in ISNI and VIVO

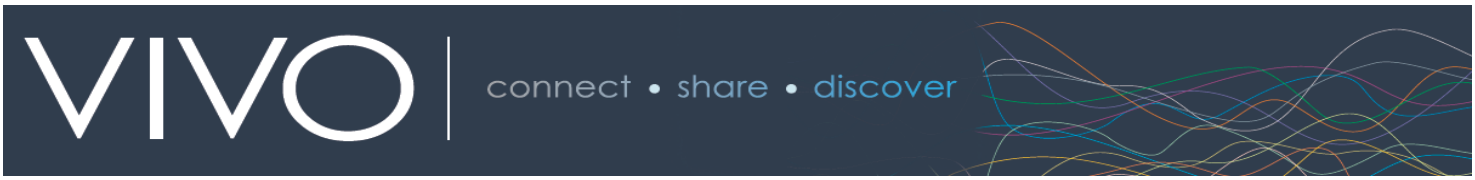
Jing Wang, Karen Smith-Yoshimura and Janifer Gatenby

Abstract: Scholars' affiliation provides the context for their work and is often represented on the publications and grant applications. However, universities and departments change their names, and scholars change their affiliations, which present challenges for matching and aggregating data among multiple systems. Institutional identifiers are crucial to accurately represent scholars' affiliations. Institutions may not realize they already have such an institutional identifier, ISNI, and that this identifier has already been disseminated, used by ORCID and included in VIAF and Wikidata. In this presentation we summarize the current work of an OCLC Research Task Force on Representing Organizations in ISNI, discuss the challenges of ISNI and VIVO with organizational data, and identify points of collaboration.

Leveraging Personalized Google Analytics Information for Greater RNS Engagement

Brian Turner, Anirvan Chatterjee and Eric Meeks

Abstract: Many installations of Research Networking Systems (RNSs) like VIVO and Profiles get significant web traffic. However, most visits are anonymous and relatively short, leading us to believe we are providing little more than a directory service. If RNSs are to evolve from a research directory into true networking systems, we must engage users much more significantly. We will do this by providing personalized, valuable content to researchers, the primary target audience of RNSs. UCSF Profiles gets over 100,000 visits per month. However, most visits are anonymous (only 4% sign in) and relatively short (89% are under two minutes). We plan to entice users to log in and spend more time within UCSF Profiles by offering them personalized content that is interesting to them. We will base the personalized content on our recent success with a 'vanity' email. In that email, we sent researchers easily-understandable statistics from Google Analytics, including the number of visitors and information about those visitors to their page. That email was very well-received, based on click-through data and survey feedback. Our plan is to build a dashboard page using the Open Research Networking Gadget platform to provide similar information profile holders can sign in any time to see. We will show them how often their page is viewed, and some information about those visitors such as geographic location and their domain. This information will be available under sign only, as that is needed to identify the user. We will promote the feature via the traditional means of email, newsletters and on the site itself. We expect this personalized content to drive engagement in a virtuous cycle. As we offer more reasons to sign in, more people will. That will lead more people will use more features of UCSF Profiles in quantifiable ways. That in turn will enforce the utility of providing users more personalized content. When researchers are signing more frequently, we can and will introduce more engaging features such as opportunities for collaboration. This will help fully realize the utility of research networking.



Delivering Trending Publications to Researchers

Brian Turner and Anirvan Chatterjee

Abstract: UCSF is piloting a way to deliver publications, which are matched to researchers' current interests and trending in the research community, to researchers in a timely way. UCSF searches PubMed for publications that match a researcher's stated interests on her or his profile page. We then assess those publications using Altmetric parameters to determine which are getting the most new-media buzz: tweets, shares, reads and posts. The most popular publications are ranked and the top five presented to the researcher via email. If this pilot is successful, we plan to build an Open Research Networking Gadget to display the content to the researcher when s/he signs in to UCSF Profiles.

Combining Flexible Faculty Activity Reporting with Public Research Profiles: Converis and VIVO

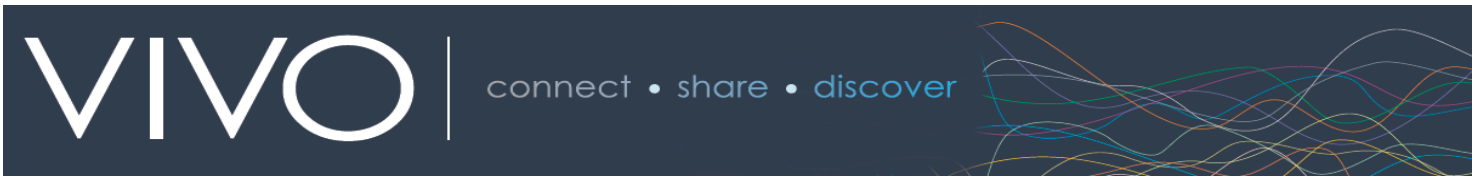
Ann Beynon, Thorsten Höllrigl, Julia Laurin and Brigitte Joerg

Abstract: Research organizations need to fulfill a variety of needs, from enhancing public visibility of scholarly activities to efficiently reporting on faculty activity across campus for accreditation, strategic planning, and faculty evaluation purposes. Converis, a comprehensive faculty activity reporting system, helps universities around the world to minimize the burden on faculty for maintaining a comprehensive record of their activities. Universities use Converis to manage their faculty review processes through customizable workflows and templated reports. Converis enables a seamless integration of institutional data with trusted publication sources such as Web of Science and PubMed, to easily meet grant submission and accreditation requirements. Converis modernizes the faculty activity reporting process through minimizing data entry, leading to efficient administrative reporting and insightful analytics. Curated faculty data from Converis can also populate a rich VIVO instance based on semantic mappings, whereby faculty benefit from enhanced visibility of their scholarly activities through the VIVO application software. Examples of this integration between Converis and VIVO will be shared.

CrossLinks – Towards a Single VIVO Profile Ecosystem

David Eichmann and Eric Meeks

Abstract: Profiling systems have achieved notable adoption by research institutions. Multi-site search of research profiling systems has substantially evolved since the first deployment of systems such as DIRECT2Experts. CTSAsearch is a federated search engine using VIVO-compliant Linked Open Data (LOD) published by members of the NIH-funded Clinical and Translational Science (CTSA) consortium and other interested parties. Sixty-four institutions are currently included, spanning six distinct platforms and three continents (North America, Europe and Australia). In aggregate, CTSAsearch has data on 150-300 thousand unique researchers and their 10 million publications. The public interface is available at http://research.icts.uiowa.edu/polyglot. We are now in the process of cross-linking co-author data from UCSF's ProfilesRNS to their respective home institution profiles through the CrossLinks project. CrossLinks interrogates the CTSAsearch SPARQL endpoint (http://marengo.info-science.uiowa.edu:2020), then provides real-time JSON-LD, supporting cross-site linking (with thumbnail images), and effectively creating a single inter-institutional information space.



The ISF Ontology: Lessons in Mapping

Graham Triggs

Abstract: Much has changed since Symplectic first introduced harvester extensions in 2012 to allow a VIVO instance to be populated with curated data from an Elements RIM system. Since then, Symplectic has used the CASRAI dictionary to extend and standardize the data stored in Elements, and the VIVO ontology has evolved into the Integrated Semantic Framework. In early 2015, Symplectic released an upgrade of the harvester, bringing the mappings up to date with the VIVO ISF, ensuring all of the CASRAI-based standard Elements data was included. But while both systems have ties with CASRAI and influences in their data, the structure of the data can be very different. In addition to handling these differences, we were also able to extract keywords into assertions against controlled vocabularies, infer institution types, and create people for co-authors at external institutions. This presentation discusses the lessons learnt from upgrading data translations to the VIVO ISF ontology, the challenges encountered in mapping differing data structures, and the techniques used to overcome them.

VIVO ISF: Investigating Speed Factors

Graham Triggs

Abstract: Maintaining and upgrading harvester extensions for populating VIVO from Elements requires testing. Through our development of Elements, we are able to generate large databases of test data - in excess of 25,000 “affiliated” users and 147,000 co-authors for more than 170,000 research outputs. In VIVO, these account for over 30 million triples. This presentation explores the limitations of populating and maintaining a dataset of such size, the effects of highly linked data (not just users and publications, but concepts, grants, even external co-authors), the hardware requirements, and investigates how scalability could be addressed.

Reflections of the Service Provider

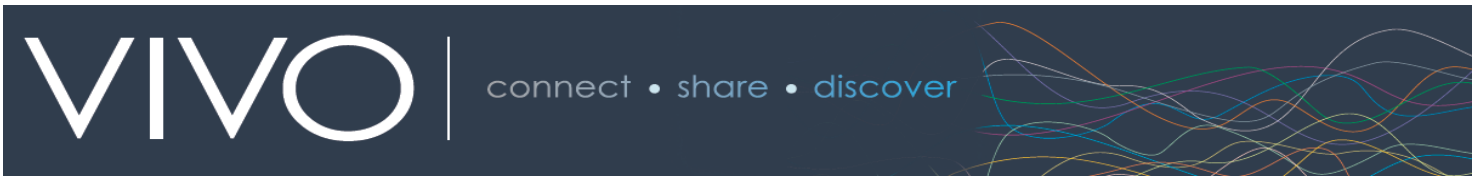
John Geischen, Alex Viggio, Jonathan Breeze and Sabih Ali

Abstract: As the first official DuraSpace Registered Service Provider for VIVO, Symplectic provides services that cover everything from installation to support of VIVO on client servers. Our continued involvement in user groups, the DuraSpace community and participation in VIVO working groups mean that we are well placed to support the needs of the VIVO community. Symplectic is also a registered service provider for Profiles RNS, the open source Semantic Web research networking platform based on the VIVO ontology. This presentation will reflect on Symplectic’s first year as a Registered Service Provider, showcasing the projects we have supported for both VIVO and Profiles RNS - covering installations, data analysis, data population and customisation. As well as our current engagements, we will introduce a number of new initiatives we plan to introduce in support of both open source communities.

The Electronic Notebook Ontology

Stuart Chalk

Abstract: Science is rapidly being brought into the electronic realm and electronic laboratory notebooks (ELN) are a big part of this activity. The representation of the scientific process in the context of an ELN is an important component to making the data recorded in ELNs semantically integrated. This presentation will outline initial developments of an Electronic Notebook Ontology (ENO) that will help tie together the ExptML ontology, HCLS Community Profile data descriptions, and the VIVO-ISF ontology.



VIVO: Data Ingest from Google Scholar, Microsoft Academic Search, ORCID and Enrich VIDWAN Profile Using Impactstory Altmetrics

Kannan P and Hetal Oza

Abstract: Identifying potential collaborator for a research project, expert for project proposal review and committee is a challenging task and time-consuming process. There is an increasing demand to develop a faculty profile management system in the institute and national level that gather research and scholarly information about the researcher and enable the discovery of experts to address new challenges in collaborative way. Information about research and research activities scattered in various places such as HR system, funding agencies and scholarly publishers. There is a conflict that most of the ranking system, research reporting, and impact factor analysis depends on scholarly article indexed in the commercial citation information provider such as SCOPUS, Web of Science. There is huge amount of digital resources such as scholarly article published in national level journal, pre-print articles and Ph.D. theses in the institutional repository, open source citation information provider such as Google Scholars and Microsoft Academic Search are available at the national level and emphasis to be given such resources for research reporting and impact factor analysis. The paper built on the work done on the project called VIDWAN: Expert Database and National Researcher's Network developed by the Information and Library Network Centre with the support of National Mission on Education through Information and Communication Technology (NMEICT) under Ministry of Human Resources Development, Govt. of India. We are in the process to move from the existing Expert Database system to VIVO based Expert Database and National Researcher's Network system. We have developed the API to import bibliographic information from the open sources citation information providers such as ORCID, Google Scholar, Microsoft Academic Search. The BibTex import process includes parse the BibTeX file using jbibtex 1.0 parser, cross check the article name and author name using SPARQL query (Ask?), the author or article not matched with the existing author or article, it will generate new URI, generate RDF using Apache Jenna and upload the RDF in the VIVO store. Also bibliographic information import directly from the information provider based on the ids such as author's Google Scholar ID, ORCID ID and Microsoft Academic Search ID using HTML parser and API provided by citation information providers. Alternative metric is another dominant area in the impact factor calculation apart from the traditional article citation. These metric includes the how many times the respective article cited, downloaded, talked and discussed in the social media. We enhanced the VIVO profile with the metrics from the open source altmetric tool called Impactstory. We will discuss the importance of the API to ingest publication data from various sources, Impactstory widget and its limitations.

Semantics and Design: Reflections on Beautiful Metadata

Richard Whaling and Andre Marques

Abstract: We describe UChicago Faculty Profiles, a semantic platform for exploring faculty research interests and output, based on VIVO-ISF. In the talk, we will describe how this project faced different constraints and requirements than most VIVO implementations; in particular, our directive was to work closely with our university's Office of Communications to provide carefully curated editorial content, and to maintain consistent design and branding with the University's other flagship public-facing sites. These aesthetic concerns, as well as other factors, led us to first explore existing software options, principally VIVO and Profiles RNS, before ultimately deciding to build a system ourselves.

We will discuss our experiments in building the features and workflows we needed with a variety of web application technologies, including Ruby on Rails, Node.js, Expression Engine, and Meteor. In particular, we will focus on the obstacles we faced in building proper handling of Linked Data with these tools, rather than a pure RDF approach, and in

which cases we were capable of overcoming them. Finally, we will reflect on the fundamental limitations of our approach compared to VIVO and related technologies, and contemplate what future work could allow us to combine a design-first sensibility with the power and rigor of a more purely semantic system.

Zero to VIVO in 7 Days

Jonathan Breeze, John Geischen and Sabih Ali

Abstract: Notwithstanding the collaborative efforts of the VIVO community, the technical and data related tasks associated with implementing and populating VIVO continue to act as significant barriers to wide adoption of the VIVO platform.

In direct response to these challenges, Symplectic has developed a quick and effective means of installing, implementing and populating a production ready VIVO instance for subscribers of Symplectic's Research Information Management System, Elements.

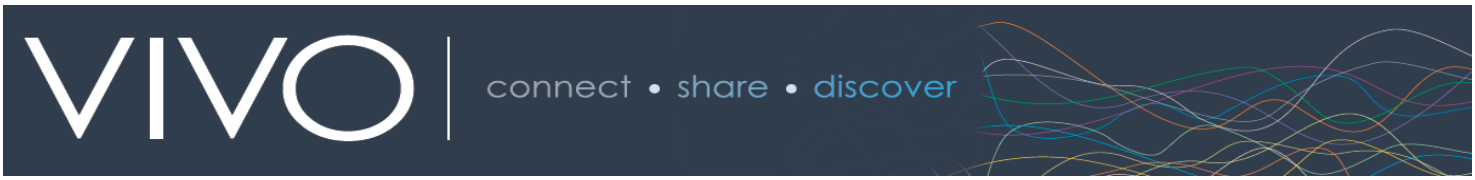
Titled 'Zero to VIVO in 7 Days', this session will introduce how this process works, from the acquisition of a new server through to the population of VIVO using linked data captured in Elements. This new service also includes an analysis of triples captured in VIVO and a report to help inform further data curation inside Elements.

This session will conclude with a short demonstration by Juergen Wastl, Head of Research Information at University of Cambridge.

Graphs without Ontologies: Using Neo4j to Explore Nascent Research Networks

Clifford Anderson, Suellen Stringer-Hye and Ed Warga

Abstract: Graphs are the fundamental concept behind both linked data and the semantic web. Research networks are, like all social networks, essentially graph structures. We argue that exploring research networks simply as graphs–without the complexities of the semantic web stack–provides an easy way to start visualizing and analyzing them. Neo4j is an open source graph database that makes it straightforward to represent data as vertices and edges. Neo4j provides a simple, declarative query language called Cypher, allowing users to readily explore and ask questions of large graph datasets. At the Vanderbilt University Library, we have initiated several network analysis projects using Neo4j, including investigation of the epistolary correspondence of Flannery O'Connor. Using a graph data model, we encoded key details for each of the letters and wrote Cypher queries to expose underlying connections between correspondents. We have also explored research networks among faculty authors by loading and querying faculty publication data in Neo4j. We extracted the data from a retrospective bibliography of faculty publications targeted for inclusion in our institutional repository. The resulting graph and preliminary analysis revealed relationships between authors based on co-authors and shared research topics. This preliminary work suggests that relationships stored in bibliographic data may fruitfully be visualized as graphs. Building on this work-in-progress, our next step will be to harvest a larger dataset from other university repositories to expose analogous relationships at other universities. Our talk will present graph theory as distinct from both linked data and the semantic web. We will also provide a quick introduction to Neo4j and the Cypher query language. Using those tools, we will visualize and analyze several scholarly networks. Our goal is to show how using graph databases as a kind of scratch pad assists with representing emergent research networks. By way of conclusion, we will point to complementarities between graph databases like Neo4j and semantic web projects like VIVO.



Tackling Silos and Cultures: VIVO Outreach and Engagement at Duke University

Lamont Cannon and Julia Trimmer

Abstract: What's the best way to reach faculty members? Most colleges and universities are commonly siloed across schools and departments, so it is a major challenge to effectively communicate with faculty to inform and educate them consistently on any topic. Cultures across schools can be very different -- what is important can vary greatly across disciplines. In addition, we found that by talking with Duke school communicators, administrators, and other teams who have implemented systems in the past, there is no single "best way" in which to consistently reach all faculty.

Given these issues and challenges, the Scholars@Duke team took a multi-pronged approach to engage our faculty concerning our VIVO implementation (Scholars@Duke). In this presentation, we will discuss our methods of outreach and how we interacted with faculty, power users, data consumers, and other Scholars@Duke stakeholders. We will discuss the successes, as well as, the "not so happy" reactions of faculty with the hope that our methods and strategies help other VIVO institutions add to their "outreach toolbox" at their own institutions.

Publishing Linked Open Data about University Scientific Outputs using the VIVO Ontology

Roberto Garcia, Jordi Virgili-Gomá and Rosa Gil

Abstract: Initiatives promoting that institutions open their data are starting to have impact and slowly but consistently, as reported for instance by the World Wide Web Foundation in its Open Data Barometer . This is also reaching universities, especially public universities, which under budgetary constraints must transparently show where resources are spent and what results are being obtained. In many cases, as in the one reported here, all the data is already available but scattered across different information systems and databases controlled by different institution units and using different vocabularies and custom terms. Therefore, the first step in order to provide an integrated view of all this data is to define a reference vocabulary. Universitat de Lleida, a Spanish university, is currently undergoing this Linked Open Data publishing process of all its the research outputs. This includes papers, research projects, patents, grants, PhD thesis, etc. Given the broad range of entities under consideration, many of the evaluated reference vocabularies failed short in their coverage and required the combination of many different vocabularies, with the consequent integration burdens at the conceptual level. However, the VIVO Ontology, part of the VIVO project, showed the right coverage as it included all the required entities and a wide range of properties that cover their interrelationships. Moreover, it is built on top of well know and already commonly used ontologies like the Bibliographic Ontology (BIBO), which facilitates its adoption. This document reports about the experience mapping existing institutional databases at Universitat de Lleida, containing information about scientific outputs, their impacts, involved researchers, their organization into research groups, etc. All these data is then published in an integrated and semantic form, as Linked Data, using a semantic data exploration tool called Rhizomer. The aim is to facilitate the exploration and visualization of all the available data about scientific production, also facilitating the automation of the generation of reports like annual research reports at the department, research center or at the level of the whole university.

A New Research Data Mechanics

Simon Porter

Abstract: Over the past 10 years, research systems have evolved from systems that focused on how to structure and record information on research, to systems capable of allowing significant insights to be derived based upon years of

high quality information. In 2015, the maturity of the information now collected within many Current Research Information Systems, and the insights that this can provide is of equal or greater value than the insights that could be gleaned from established externally provided research metrics platforms alone. The ability to intersect these external and internal worlds provides new levels of strategic insight not previously available.

With the addition of platforms that track altmetrics, and their ability to connect university publications data with a constant flow of real time attention level metrics, an image of a dynamic network of systems emerges, connected together by ever turning 'cogs' pushing and translating information. Add to this, the success of ORCID as researcher identifier infrastructure across systems and it becomes possible to extend this network back from the systems that track and record research information, right to the platforms through which research knowledge is created.

The 'Mechanics' of this network of systems is more than just getting the 'plumbing' right. As research information moves through the network, its audience and purpose changes, the requirements for contextual metadata can also change.

Reflecting on the connected implementations of Symplectic, Figshare, and Altmetric for Institutions at the University of Melbourne, this presentation till explore the concept of a New Research Data Mechanics, and VIVO's role within it.

Using Virtuoso as an alternate triple store for a VIVO instance

Paul Albert, Eliza Chan, Prakash Adekkanattu and Mohammad Mansour

Abstract: Background: For some time, the VIVO for Weill Cornell Medical College (WCMC) had struggled with both unacceptable page load times and unreliable uptime. With some individual profiles containing upwards of 800 publications, WCMC VIVO has relatively large profiles, but no profile was so large that it could account for this performance. The WCMC VIVO Implementation Team explored a number of options for improving performance including caching, better hardware, query optimization, limiting user access to large pages, using another instance of Tomcat, throttling bots, and blocking IP's issuing too many requests. But none of these avenues were fruitful.

Analysis of triple stores: With the 1.7 version, VIVO ships with the Jena SDB triple store, but the SDB version of Jena is no longer supported by its developers. In April, we reviewed various published analyses and benchmarks suggesting there were alternatives to Jena such as Virtuoso that perform better than even Jena's successor, TDB. In particular, the Berlin SPARQL Benchmark v. 3.1[1] showed that Virtuoso had the strongest performance compared to the other data stores measured including BigData, BigOwlim, and Jena TDB. In addition, Virtuoso is used on dbpedia.org which serves up 3 billion triples compared to the only 12 million with WCMC's VIVO site. Whereas Jena SDB stores its triples in a MySQL database, Virtuoso manages its in a binary file. The software is available in open source and commercial editions.

Configuration: In late 2014, we installed Virtuoso on a local machine and loaded data from our production VIVO. Some queries completed in about 10% of the time as compared to our production VIVO. However, we noticed that the listview queries invoked whenever profile pages were loaded were still slow. After soliciting feedback from members of both the Virtuoso and VIVO communities, we modified these queries to rely on the OPTIONAL instead of UNION construct. This modification, which wasn't possible in a Jena SDB environment, reduced by eight-fold the number of queries that the application makes of the triple store. About four or five additional steps were required for VIVO and Virtuoso to work optimally with one another; these are documented in the VIVO Duraspace wiki.

Results: On March 31, WCMC launched Virtuoso in its production environment. According to our instance of New Relic, VIVO has an average page load of about four seconds and 99% uptime, both of which are dramatic improvements. There are opportunities for further tuning: the four second average includes pages such as the visualizations as well as pages served up to logged in users, which are slower than other types of pages.
[1]http://wifo503.informatik.unimannheim.de/bizer/berlinsparqlbenchmark/results/V7/#comparison

Leveraging Institutional Data for Author Name Disambiguation

Michael Bales, Paul Albert, Jie Lin and Stephen Johnson

Abstract: Author name disambiguation is a challenging problem in computer science. The problem arises from the fact that many authors share similar or identical names. Although some scholarly databases assign unique author identifiers, levels of accuracy are often unacceptable—especially for authors with common names. Existing algorithms have largely not leveraged institutional data on individual researchers. We are extending ReCiter, an agglomerative clustering algorithm for author name disambiguation, for use in publication management at our institution. The system uses available institutional data on researchers, including primary and secondary departments, history of co-investigatorships on grants and co-authorships, favored journals, and years of authors' terminal academic degrees. We are investigating the use of machine learning approaches to optimize system performance, and are planning to make the system available as a suite of freely available, open-source tools.

A Team Assembly Tool Based on NU Scholars VIVO endpoint

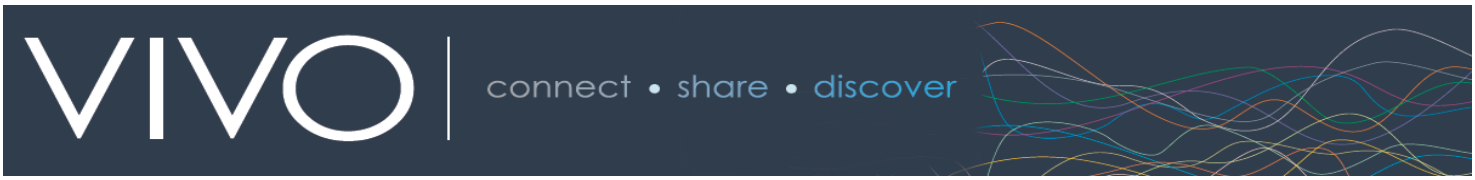
Raja Lalith Sabbisetti, Anup Sawant, Harshad Gado and Noshir Contractor

Abstract: In earlier work, we developed a suite of heuristics for building collaborations between researchers. These heuristics are informed by empirical studies that test theories from the social sciences regarding the formation of effective collaborations and teams. In the current work, the initial prototypes of these heuristics were ported to operate over data represented in the VIVO ontology. The C-IKNOW VIVO team recommender is a web application that allows users to assemble teams based on the NU Scholars VIVO endpoint.

Our efforts demonstrate that the architectures and programming techniques of the semantic web are well suited to the problem of building practical software tools that can be leveraged to apply to diverse sources of data.In general, much of the infrastructure required for developing useful researcher recommender systems is available today. In particular, we found that the interoperability between researcher networking systems (RNSs) from diverse institutions and vendorsoffered by the VIVO ontology is a sound basis on which to build researcher networking recommender tools.

The C-IKNOW VIVO Team Recommender aims to assemble teams based on the preferences of an individual. These preferences encompass team size, homophily and network properties of co-authorship networks. The tool also aims to support team assessment feature that would allow individuals to evaluate different potential teamsbased on their preferences.

The ‘VIVO Team Recommender’ utilizes the World Wide Web Consortium (W3C) standard SPARQL query language for real-time retrieval of semantic web data. We found the SPARQL implementation available in open source software to be robust. Further, because programmers can target only the particular data needed, performance is enhanced by reducing unnecessary network traffic. We found the learning curve and technical skills needed for SPARQL programming to be similar to that of more traditional, relational, SQL-based programming competencies which are more generally available.



Developing a Model for Expert Networking Across Federal Government: The HHS Profiles Pilot

Jessica Hernandez and Nichole Rosamilia

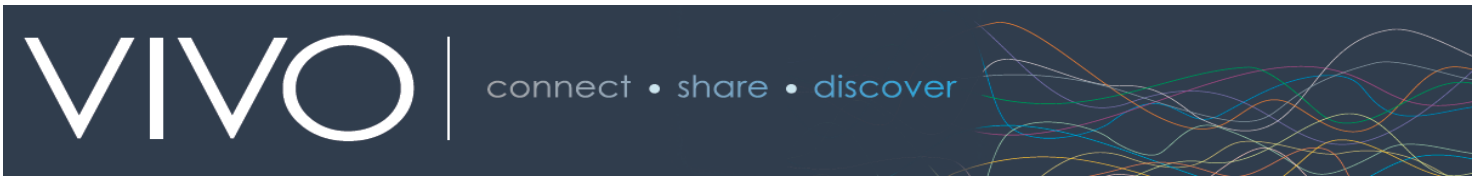
Abstract: Organizations within the United States government are increasingly beginning to explore research networking software as a way to locate and match federal expertise for collaboration and problem-solving. The U.S. Department of Agriculture (USDA) launched a public-facing VIVO instance in 2014, and internal research networking pilots are underway at the Environmental Protection Agency (EPA), Food and Drug Administration (FDA), and Smithsonian Institution. However, the proliferation of agency- and department-specific systems raises concerns about reproducing existing information silos. A driving force behind government adoption of expert networking is to more quickly mobilize federal expertise in crisis situations like the recent Ebola outbreak. Yet nearly every executive branch department was engaged in Ebola-related research, prevention, or remediation, so an effective response requires expertise identification and collaboration on a government-wide scale. This study therefore aims to develop (1) a model for a multi-organizational, federal government-wide expert networking system and (2) an evaluation framework to determine whether the system effectively increases the speed and agility of government operations. To test the viability of the model, we are launching a public-facing Department of Health and Human Services (HHS) pilot using Profiles Research Networking Software. This pilot will include experts from three HHS agencies: FDA, the National Institutes of Health (NIH), and the Centers for Disease Control and Prevention (CDC). We will present lessons learned and preliminary results from this effort to design a platform-agnostic, integrated framework to harness intra-government and cross-sector expertise.

Integrating VIVO and eagle-i to develop a Resource Recommender System

Suzanne Thompson and Amarendra Das

Abstract: Over thirty academic institutions, including Dartmouth, participate in the eagle-i research resource network, which provides information to researchers on both local and nationally networked resources, such as cores, labs, specimens, instruments, reagents, and software. Discovery of these eagle-i resources requires active user searching of the semantically structured information via its web interface. In this poster presentation, we discuss the design and implementation of a recommender system for eagle-i resources that is part of the Inspire research management tool we have built for the Dartmouth SYNERGY Clinical and Translational Science Institute. The recommender system automatically matches relevant resources to investigators based on information collected within their VIVO profile, a system we are currently in the process of launching institution wide.

Background: Since 2009, Dartmouth has participated in the eagle-i network, and continues to update eagle-i with curated clinical and translational resources. In 2013, the Dartmouth SYNERGY Clinical and Translational Science Institute was funded through a NIH Clinical Translational Science Award (CTSA), becoming the newest hub in the CTSA network. As part of these efforts, the SYNERGY-supported Informatics Collaboratory for Design, Development, and Dissemination (ic3d) has developed and released in 2014 a web-based, mobile friendly, open-source research management system, called Inspire, to allow investigators to access and manage a range of CTSA-supported resources, which are encoded in eagle-i. In addition, the ic3d team is also implementing the VIVO system to automatically profile researchers by extracting publication and grant information and listing investigator-provided data on research interest. Our team is currently building an integrated platform that connects data on eagle-i resources, VIVO profiles, and Inspire project information to provide a researcher-specific recommendation of which eagle-i resources are relevant.



Objective: To design and build a recommender system (called Inspiration) that uses investigator profile in VIVO and project activity information in Inspire to determine relevance of eagle-i resources to a researcher.

Methods: We chose to design a recommender system based on a vector space model that would represent eagle-i resources and faculty profiles as weighted terms vectors. The vector space model is widely used in web searching as a highly scalable information retrieval method. In our approach, we use the proximity of a vector representing an eagle-i resource to a vector representing an investigator's research portfolio in the vector space model as an indication of the relevance of that resource to the investigator. The vector for the investigator is based on concepts associated with their publication titles, grant titles, provided research interests, and research project descriptions.

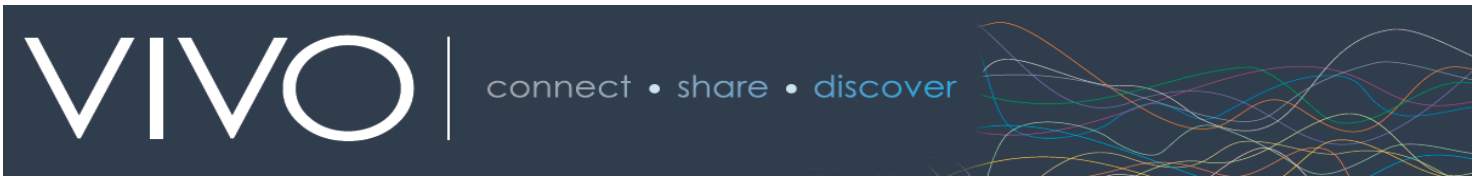
Results: To create the vector space model, we used the APIs of VIVO, eagle-i, and the Inspire systems to digest the data. Although VIVO and eagle-i store information using the semantic resource description framework (RDF), which have been merged together into the Integrated Semantic Framework, the descriptions of the research resources and information on publication and grant titles are in free text. Similarly, project description fields within Inspire used for request management are unstructured text. We used a dictionary-based approach to map these words into standardized terms, and then created term vectors for each resource and investigator. We implemented the vector space model within python using the gensim library. We are currently working with faculty investigators to evaluate the relevance of the matched eagle-i resources in Inspiration, and will present the results of this study in the poster presentation.

Acknowledgments: Research reported in this poster was supported by the Dartmouth SYNERGY Clinical and Translational Science Institute, under award number UL1TR001086 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Using VIVO-ISF Ontology in Open Repositories

Violeta Ilik, Piotr Hebal and Kristi Holmes

Abstract: The intent of the Galter Health Sciences Library team is to enable the digital repository, built on Fedora/Hydra stack, to work with numerous semantic web data standards to provide for interoperable data. We introduce the VIVO-ISF Ontology and local VIVO ontology extensions to represent National Library of Medicine (NLM) publication types, enabling the users of the repository to represent their scholarly outputs with deep granularity. The Galter Health Sciences Library team, as a member of the Northwestern University Clinical and Translational Sciences Institute (NUCATS), Feinberg School of Medicine (FSM), looks to establish a digital repository that will house traditional and non-traditional scholarly outputs. Non-traditional outputs include measurement devices, patient education materials, curriculum materials, conference materials, community engagement materials, etc. A strong effort will be made to support a variety of dissemination strategies to optimize visibility of outcomes by NUCATS and NUCATS Members, including an Open Access repository at Galter Library to enhance accessibility of traditional science products (e.g., peer-reviewed literature), as well as more non-traditional outputs which may be created as part of the research process by the many roles that exist in today's diverse translational workforce (e.g., recruitment protocols, survey instruments, manuals and brochures for research participants, NUCATS reports, and myriad other outputs). Every effort will be made to make the evaluation process and results available to members, NUCATS stakeholders and leadership, and the CTSA consortium to help NUCATS deliver accountability; to advocate for new methods or opportunities, and facilitate future analyses.



Laying the Foundation for a Robust Faculty Data Reporting Infrastructure at a Medical College

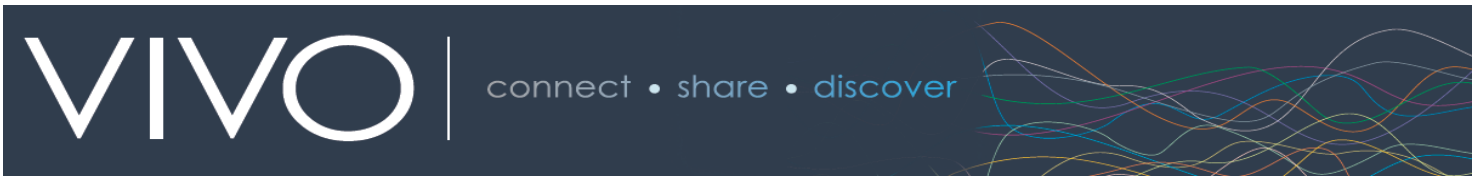
Paul Albert and Curtis Cole

Abstract: Background: In 2014, Weill Cornell Medical College's (WCMC) Information Technologies Services was charged with producing high quality and on demand reports that would empower administrators to make key decisions and fulfill external reporting requirements. One dean is particularly keen on being able to do analyses in a disintermediated way - that is, without depending on middle men. Information systems originally designed for a single purpose can be serviceable for the occasional ad hoc system-specific reports, but the need for reliable, on-demand, and sophisticated reporting across systems highlighted our key unmet needs. These include: business process improvements for maintaining the systems; building open channels for feedback from end users and other stakeholders; improved processes for supplying data to downstream systems; and documenting and communicating the meaning and context of data. Such considerations cannot be an afterthought.

Analysis: Weill Cornell maintains 10 systems especially relevant to faculty including those that capture faculty reviews, board certifications, hospital credentials, and appointments as well as VIVO. To lay the groundwork for a faculty data reporting infrastructure, we scored all 10 systems according to these custom set of criteria. (1) Authoritative data is accurate; (2) Secondary data is accurate; (3) Data is well-structured; (4) End user can view; (5) End user can update, or at least provide feedback; (6) Accurate assignment of institutional identifier; (7) Avoids duplicate records; (8) Well-connected to other systems; (9) Relevant information is collected; (10) Technically easy to output reports; and (11) Transparent reporting process.

In total, the 110 scores can be grouped as follows: 30 needs improvement; 42 okay; 28 good; and 10 unknown.

Conclusion: At first blush, producing reports on faculty seems like it should be a straightforward proposition. Similarly, some VIVO implementation sites might assume that the majority of work towards propping up a new VIVO will be devoted to figuring out the technical mechanics of moving data between systems and tools. But this analysis clarified for our key stakeholders and ourselves certain prerequisites for creating a nimble and reliable faculty reporting infrastructure. When they come to fruition, we expect our efforts will also improve the quality of data in our VIVO.



Panels/Long Papers

The Linked Data for Libraries Project: A Progress Report

Dean Krafft and Jon Corson-Rikert

Abstract: We will report on the first eighteen months of the Mellon-funded two-year Linked Data for Libraries (LD4L) project (http://ld4l.org). LD4L is a partnership of Cornell University Library, Stanford University Libraries, and the Harvard Library Innovation Lab. The goal of the project is to use Linked Open Data to leverage the intellectual value that librarians and other domain experts and scholars add to information resources when they describe, annotate, organize, select, and use those resources, together with the social value evident from patterns of usage. The project is producing an ontology, architecture, and set of tools that work both within and across individual institutions in an extensible network.

This progress report will describe the LD4L use cases, which focus on linking data about library bibliographic resources for well-described assets (the catalog) with other silos of information, including people's scholarly profiles (from VIVO, Harvard Faculty Finder, and Stanford CAP), curation and annotation data, and information about usage. We will describe the current state of the LD4L ontology, and how it addresses these use cases. The ontology includes elements of BIBFRAME, VIVO-ISF, OAI-ORE, PAV, and others, and it makes use of a number of standard global identifiers, including VIAF, ORCID, ISNI, and OCLC Works.

We will also report on the outcomes of the LD4L workshop, which brought together fifty linked data experts at Stanford in late February 2015 and provided extensive feedback on the use cases, ontology design, and engineering work to date. Finally, we will describe the engineering work both planned and completed on converting bibliographic, person, curation, and usage data to shareable linked data from our three institutions; making that data available as linked open data on the web; and creating a demonstration search across the scholarly resources at Cornell, Harvard, and Stanford.

Visualizing the VIVO Profile Ecosystem Using CTSAsearch

David Eichmann

Abstract: CTSAsearch (http://research.icts.uiowa.edu/polyglot) is a federated search engine using VIVO-compliant Linked Open Data published by 64 institutions using six distinct platforms. In aggregate, CTSAsearch has data on ~140,000 researchers and their ~8,300,000 publications. Since its introduction in 2013, the query and visualization mechanisms in CTSAsearch have proven to be the primary elements of user interest. I first present a number of comparative analytic visualizations, characterizing distinctions between classes in the set of institutions as a whole. The remainder of the presentation then covers my approaches to query formulation and visualization.

NYCCSC: Extending VIVO Ontology and Application to Facilitate Discovery through Blacklight and to Support Climate Change Science Content Curation

Huda Khan, Jon Corson-Rikert, Darcy Branchini and Ingrid Zabel

Currently, the policy makers and associated researchers who analyze the impacts of climate change and formulate response strategies must gather information and data on climate change from scattered and diverse sources. The New York Climate Change Science Clearinghouse (NYCCSC) project aims to provide a central source of climate change science data and information specific to New York State. The NYCCSC application consists of multiple components, including a

Blacklight front-end that provides search and discovery functions. This Blacklight instance shares a Solr search index with a VIVO instance used as the main curation interface for the project. This presentation will cover the ontology and architecture used for the project, including the following: - Ontology development: We have extended the VIVO-ISF ontology to represent climate change science concepts and linkages between those concepts and the data and information being represented in the Clearinghouse. These concepts include climate changes, effects of these changes on organisms and built and natural systems, response strategies, and actions to implement these strategies. This project's use of semantic relationships has helped to expose the connections within the content in a richer way. - Software extensions to support search and discovery using Blacklight and VIVO: As Blacklight expects certain fields and formats from the search index and we needed to display certain information in the user interface, we had to modify VIVO's search index to enable discovery and display through Blacklight. Furthermore, we had to modify Blacklight to enable querying RDF data directly from VIVO as not all content is included in the search index. In addition, we worked on boosting results for specific classgroups in the VIVO application to prioritize documents, data, and GIS over people, organizations, and events. - VIVO and PostGIS for custom boosting and geographical location representation: The NYCCSC application uses PostGIS for storing geographical information and associations between geographic locations and geotagged items represented in VIVO. We extended the Blacklight front-end to allow the user action of zooming in and out of the map displayed on the search page to update search results based on relevance to the resulting bounding box of the map. To support this geographical search, we extended VIVO's Solr search with a custom function that utilized PostGIS ranking of items based on a bounding box. In addition, we extended the NYCCSC VIVO instance and utilized VIVO's built-in N3 editing framework to allow a user to search the PostGIS database for geotagging a particular entity represented in VIVO. - Interface design: As part of this project, we have utilized usability testing and other user feedback to help inform the design of the front-end and the features included in the clearinghouse site. A demo of the site will demonstrate how the different components of the architecture work together to support search and discovery through semantic relationships.

Examining the Differences Between Scientific Team Types

Denis Agniel and Griffin Weber

Abstract: Introduction: There may be fundamental differences between the way that scientific research gets done by teams of different types. We seek to understand the differences between teams of three different types: biomedical teams, non-biomedical teams, and interdisciplinary teams. In particular, we are interested in examining the impact of team type on translation into clinical practice. We used the Scopus bibliographic database of over 54 million publications and over 28 million authors to explore these areas.

Defining team types: We defined a researcher as a biomedical researcher if over 50% of their publications were identified as having a PubMed ID in Scopus. Then biomedical teams were defined as teams with all biomedical researchers, and non-biomedical teams were defined as those with all non-biomedical researchers. Interdisciplinary teams were those that had a mix of biomedical and non-biomedical researchers.

Team type characteristics: We examined the behavior of each team type. Publications were classified by Scopus according to the subject matter of the journal in which they were published. We identified which subject areas were most likely to publish research by interdisciplinary teams. We also examined the breakdown of team types by number of authors, finding that the majority of publications with fewer authors were non-biomedical, but as the number of authors increases articles tend to be more biomedical and interdisciplinary. More than 540 million citations were available to be

analyzed, and we examined the citation rate for each article. Biomedical teams were more likely to be cited regardless of the subject matter or the number of authors.

Measuring translation impact by team type: We finally measure the impact of team type on translation according to the Triangle of Biomedicine, which maps the 22 million biomedical journal articles in PubMed to a triangle, whose corners represent research related to animals, cells and molecules, and humans. We examined where on the triangle articles of each team type fell on the Triangle according to the subject matter of the article (see figure). We also modeled the impact of team type on the time until translation.

The SEO State of the Union: How 35 VIVO and Profiles Sites Show up on Search Engines — and What We Can Do To Make Research Networking Platforms more Discoverable to Real Users

Anirvan Chatterjee and Brian Turner

Abstract: Research networking platforms vary substantially in terms of their discoverability via search engines like Google – the front door to the majority of online resources for many researchers. For example, UCSF Profiles receives about 82% of its traffic through search engines (about 90,000 visits per month); another comparable site receives 87% of its traffic from search engines. Other research networking platforms, however, appear to be largely invisible to search engines.

UCSF has been performing the first large scale study of search engine optimization (SEO) in the research networking space, assessing the effectiveness of 35 different public research networking implementations in making their people-related data findable to Internet searchers. These include systems built on VIVO, Profiles RNS, and SciVal Experts, as well as home-grown systems, leveraging the work of the R2R initiative.

We will present on our findings, including an overview of how different implementations and platforms rate in terms of search engine discoverability, stratifying by the size of deployment, platform.

These findings will help us offer an improved set of search engine optimization best practices, informed by real-world data on which approaches are most likely to succeed in the real world. We will leave the audience with an updated guide on search engine optimization for research networking.

Research Networking Systems: A Data Treasure Trove for Creating Strategic Research Initiatives and Analyses

Ruth Allee, Cynthia Cleto, Jan Fransen, Kate McCready, Sharlini Sankaran and Julia Trimmer

Abstract: Research Networking Systems (RNS) were primarily established to stimulate collaboration, which may cross disciplines, sectors, institutional or even national boundaries; and studies have shown that collaborations that cross such boundaries, may increase citation rates by as much as a two-fold factor. Within North America, a critical mass of institutions has been able to invigorate their research programs by implementing RNS. However, once implemented, how does the institution utilize the system and the data treasure trove it contains to facilitate impactful research initiatives? This session brings together the project leads from Duke University, University of Minnesota, Northwestern University, and REACH NC (a 20-institution, North Carolina state-wide RNS) who have implemented Pure Experts Portals, Elsevier's RNS. They will share their perspective on how their organizations have utilized RNS, and in certain cases repurposed data to pursue further analysis and visualizations that address strategic and diverse organizational needs.

What are their most impactful use cases, and how can data from RNS facilitate these use cases? We will present the use cases below and more:

1. Using RNS to enhance and expand upon contextual data of other Northwestern sites – ex. Global Research and Undergraduate Research Opportunities, and graduate program sites.

2. Visualizing collaborations between the U of Minnesota's College Biological Sciences (CBS) and College of Food Ag Natural Sciences (CFANS).

3. Identifying expertise and forming partnerships to advise on collaborations between diverse sectors including academic researchers, the economic development community, the military and defense community, and private businesses in North Carolina using REACH NC.

4. Provisioning publications for VIVO, for department websites, and for reporting and analytics at Duke University.

VIVO in a Networked Research Ecosystem

Jing Wang, Don Elsborg, Huda Khan, David Eichmann and Eric Meeks

Abstract: One of the goals stated in the VIVO Strategic Plan 2015-2016 is to promote a more open and networked research ecosystem (Goal 2). As more systems become interconnected, the demand for systems integration and interoperability increases but the problems become more complex.

System of systems (SoS) refers to a set of operationally and managerially independent systems interacting with each other to provide capabilities which cannot be accomplished by any single system. SoS has its own characteristics and challenges, such as emergent behavior and evolutionary development. SoS principles and methods can be applied to the networked research ecosystem and can be helpful in identifying the pain points and opportunities as well as the requirements for future systems integration across VIVO instances and between VIVO and complementary platforms.

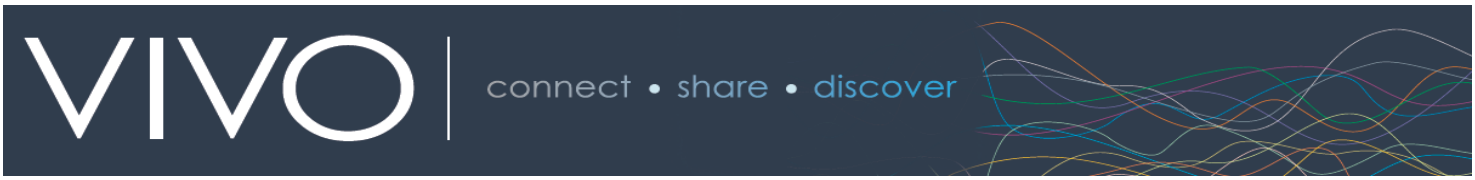
In this presentation we will summarize preliminary work that has been done in a System of Systems approach for a network of VIVOs; review the necessary architecture components; discuss the pros and cons of different integration styles and patterns; identify challenges and opportunities; and highlight some of the SoS level integration requirements for VIVO to function optimally in a networked research ecosystem.

EarthCollab: Implementing, Extending and Connecting VIVO Instances to Enable Discovery Across Scientific Research Networks

Huda Khan, Matthew Mayernik, Mike Daniels, Keith Maull, Steve Williams, Linda Rowan, M. Benjamin Gross, Jon Corson-Rikert, Erica Johns, Dean Krafft and Dave Eichmann

Abstract: Given the interdisciplinary and interorganizational nature of research conducted in the Geosciences, some of the work is conducted using distributed networks of researchers and resources including instruments and platforms. To better enable the discovery of the research output from the scientists and resources used within these organizations, UCAR , Cornell University, and UNAVCO are collaborating on the EarthCollab project which seeks to leverage semantic technologies to manage and link scientific data. The EarthCollab project is working within the following two geoscience-based communities to deploy VIVO instances: (1) the Bering Sea Project , an interdisciplinary field program whose data archive is hosted by NCAR's Earth Observing Laboratory (EOL), and (2) diverse research projects informed by geodetic

tools which are operated and/or maintained by UNAVCO. The lessons we learn about extending or integrating domain specific information into a VIVO instance and about searching across and linking information in multiple VIVO instances should prove relevant and interesting to the VIVO community as a whole. In this presentation, we will discuss our goals and work in the following areas: - Ontology design and implementation driven by Geoscience use cases: We are designing the EarthCollab ontology based on what the two use case communities have identified as information they would like to represent and the information they currently have about data and researchers. We are populating the two use case-specific VIVO instances using the VIVO ontology as the foundation. We have also reviewed multiple existing ontologies, such as GCIS and DCAT to explore how to extend VIVO and represent the relationships between resources, such as instruments and platforms, data, and researchers. We will discuss our ontology design approach and how the different needs of each use case have informed the ontology design. - Cross linking VIVOs and furthering linked data: To support searching across distributed information represented in separate VIVO instances, we are exploring mechanisms for linking information from multiple VIVOs without necessarily having to duplicate or import this information in each instance. Central to these mechanisms are the abilities to: (1) designate multiple URIs from separate VIVO name spaces as equivalent to each other or to an independent unique identifier (such as an ORCiD ID) using sameAs assertions, (2) retrieve the appropriate URIs that might designate the same person using a lookup service based on (1), and (3) display information in a VIVO instance for a URI from a different VIVO instance without having to copy or duplicate information.



Posters

Profiles Research Networking Software – An Open Source Project

Nick Brown and Griffin Weber

Abstract: Introduction: Profiles Research Networking Software (RNS) is a free semantic web application which uses the VIVO ontology to generate searchable online profiles of an organization's investigators (http://profiles.catalyst.harvard.edu). As an open source product, Profiles RNS benefits from a community of developers who contribute code to the software, customize the website in unique ways for their institutions, and provide helpful suggestions for future functionality. This poster describes how the Profiles RNS open source code is managed and how we have built a community around it.

Developers: Profiles RNS has a core development team at Harvard Medical School and also receives community submissions, with significant parts of the code base written and updated by University of California San Francisco (UCSF) and Boston University (BU). Additional institutions have contributed modular "gadgets" they built for Profiles RNS; and, some sites hired commercial vendors (e.g., Recombinant Data Corp) to build custom features for Profiles RNS, which they ultimately made available to others for free.

Release Process: Updates to the Profiles RNS open source code occur about 2-3 times per year. We use GitHub for source control. The distributed nature of Git is ideal for collaborative open source project. A continuous build system hooks into GitHub and deploys Profiles RNS to multiple environments each time code is contributed. We perform three types of automated testing: (1) Link Checking: This spiders a site looking for broken links and identifying 404 and 500 errors. This is easy to configure and provides broad coverage of the pages in a Profiles RNS installation. (2) API tests: These are custom tests that query the Profiles API and compare the results with the test data, covering database install scripts and the APIs. (3) Selenium UI testing: Selenium allows automated interaction with a site, allowing for testing of search and edit functionality. Selenium requires significant development effort but has deep coverage. Community: We use several approaches to building a Profiles RNS open source community and engaging sites that use the software: (1) a restricted mailing list for official Profiles RNS announcements; (2) an open Google group allowing discussion of Profiles RNS; (3) a monthly Developers webinar to discuss technical topics; (4) a monthly Users Group webinar to discuss long-term plans and for guest presentations; and (5) a partnership with Symplectic, which provides commercial support to institutions using Profiles RNS. Future Plans: In the future we would like to use issue tracking software to link every source code commit to a bug or enhancement. This would increase accessibility to the code base and help provide a bridge between users and developers. Additionally, we would like to create a community Wiki, which would provide easier management of the software documentation and enable other sites to contribute to it.

Profiles Research Networking Software – System Architecture

James Norman and Griffin Weber

Abstract: Introduction: Profiles Research Networking Software (RNS) is an open source semantic web application that generates searchable online profiles of an organization's investigators (http://profiles.catalyst.harvard.edu). It uses the VIVO ontology and generates RDF (Resource Description Framework) identical to the VIVO software, though it adds unique features such as automatic matching of PubMed articles to investigators and interactive network visualizations. This poster describes three components of the Profiles RNS system architecture. Presentation Layer: RDF provides a

standardized way of extending the data model of Profiles RNS. An institution can add new types of data by describing them in the ontology, without any changes to the database schema. However, in Profiles RNS we also sought to have a standardized way of configuring the presentation of these data. In other words, we wanted to be able to design the profile page of a person to look different than a profile of a department, without using hard-coded HTML. We therefore defined an XML document called the PresentationXML, which describes the content that should appear in different parts of the page and how that content should be rendered. Is it used in combination with the page's RDF-XML document (i.e., the data file). For example, the PresentationXML can indicate that the title for a page should come from the "rdfs:label" tag in the RDF-XML. Different PresentationXML documents can be defined for specific data types (e.g., person, publication, concept, etc.). This enables us to customize the layout of these pages with minimal .NET code changes. Security Model: When a user or external system initiates a request for data in Profiles RNS, a set of SecurityGroup IDs are assigned to the user's session. The SecurityGroups provide authorization to different types of data or functionality. Profiles RNS comes with several pre-defined SecurityGroups (e.g., public user, web bot/spider, authenticated user, admin), which can be extended by an institution. Each investigator who has a profile page also has a personal security group for private content. Each property in the VIVO ontology has a default SecurityGroup. Investigators can override these to control what content appears on their profiles. RDF Caching: Profiles RNS uses the PresentationXML and SecurityGroups to cache RDF data efficiently and greatly improve performance of the website. Generating the full RDF-XML for a person is slow (often 30+ seconds). However, the PresentationXML indicates which RDF properties are needed for a page, and only that subset of the RDF is requested from the database. We can also rely on the fact that all users with the same SecurityGroups will see the same content. We therefore store the RDF in a .NET memory cache using a key that combines the URI and the SecurityGroup. As a result, Profiles RNS does not have to query the database again if another user with the same SecurityGroup visits the website. The cache expires after a defined period of time or immediately if the underlying data changes (e.g., a person edits her profile). Conclusion: Profiles RNS uses the VIVO ontology but includes key software components that contribute to front-end configurability, security, and performance.

Cross-Linking DOI Author URIs Using Research Networking Systems

Nick Benik, Timothy Lebo and Griffin Weber

Abstract: A proof-of-concept application was created to automatically cross-link publications that were written by the same person through harvesting linked open data from institution-based research networking systems. This is important because it (1) helps people identify related articles when exploring the biomedical literature, (2) gives scientists appropriate credit for the work they have done, and (3) makes it easier to find experts in a subject area.

Four Use Cases for Research Networking: A Medical School, University, Federal Agency, and Physician Network

Griffin Weber

Introduction: Profiles Research Networking Software (RNS) is an open source semantic web application built using the VIVO ontology (http://profiles.catalyst.harvard.edu). We originally created Profiles RNS in 2008 to break down silos in biomedical research by helping investigators at a medical school find new collaborators across their institution. Since then, dozens of organizations have adopted Profiles RNS, and many have found novel use cases for research networking tools that we did not anticipate when we first started. This presentation compares four implementations of Profiles RNS,

which illustrate the broad range of ways that institutions benefit from research networking and the unique policies, data sources, and software customizations required for those websites.

Harvard Catalyst Profiles (HC Profiles): Our first implementation of Profiles RNS was for medical school faculty in Harvard's Clinical and Translational Science Award (CTSA) program, called Harvard Catalyst. In addition to information obtained from internal administrative systems, the primary data source for HC Profiles is PubMed. The software automatically adds publications to faculty profiles using an author name "disambiguation engine". Data mining algorithms extract MeSH terms and other information from the publications to generate interactive visualizations that illustrate a person's research interests and collaborators. HC Profiles is open to the public, and APIs enable anyone to repurpose the data for other websites.

Harvard Faculty Finder (HFF): The Harvard Provost Office requested their own university-wide version of Profiles RNS that would include all Harvard faculty. Because few faculty outside of biomedicine publish in PubMed, we purchased commercial publication data and added many new data sources (books, courses, patents, projects, etc.). In order to gain approval from each Harvard school for HFF, we had to remove most of the faculty "profiling" functionality from the website. The remaining tool is primarily used as a cross-school search engine, which guides users to local school or department websites that present additional information about the faculty.

FDA Profiles: The US Food & Drug Administration uses Profiles RNS internally to assemble committees to review medical devices. Custom data feeds import sensitive information about prior device reviews and committee members. As a result, FDA Profiles must be protected behind a local firewall and access is restricted to a small number of users.

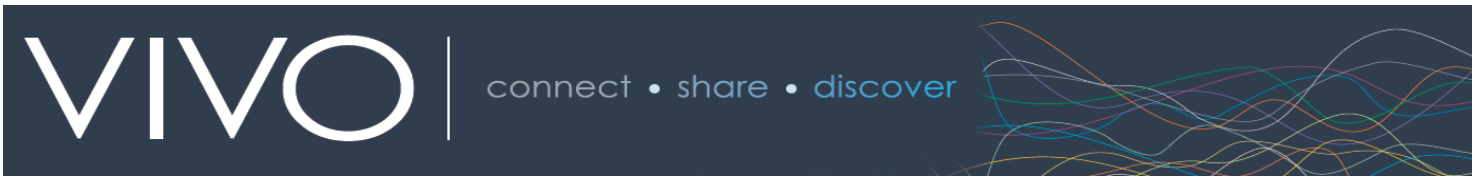
Undiagnosed Disease Network (UDN Profiles): The UDN is an NIH-funded collaboration between seven U.S. clinical sites to help patients with rare or hard-to-diagnose diseases. These patients often suffer for years without a diagnosis and are unable to find a physician who can treat them. The UDN performs a complete clinical evaluation of its patients, including genomic analyses, metabolic studies, and evaluation of environmental exposures. They then seek to match the patients with the best possible physicians within their network. UDN Profiles is a single website containing physicians from multiple UDN hospitals across the country, which enables users to search for individuals with expertise in rare or complex conditions.

ORCID iDs in VIVO: More than Just Another Piece of Flair!

Elizabeth Krznarich

Abstract: Since 2012, ORCID (orcid.org) has been pursuing its goal of addressing the name ambiguity problem in scholarly communications by maintaining a registry of unique and persistent researcher identifiers. ORCID also provides a transparent method of linking research activities and outputs to these identifiers by connecting to external systems and resources. Due to these linkages and the persistence of an individual's ORCID identifier throughout their career, ORCID possesses a unique capacity to serve as a hub connecting researchers to research activities and outputs.

With the recent release of version 1.7, VIVO is now among the growing set of applications and resources that support linkages to ORCID. By enabling the ORCID features within VIVO, local VIVO administrators can allow researchers to display a link to their ORCID identifier on their VIVO profile and (optionally) to display a link to their VIVO profile on their public ORCID record.



While new options and features to experiment with are often welcome additions to any software application, the real advantages of linking VIVO profiles to ORCID identifiers may not be immediately obvious. This poster will explores the potential benefits of linking VIVO profiles to ORCID identifiers, for both individuals and institutions, and provides relevant use cases.

Use cases to be addressed include:

Institutions: Tracking graduate outcomes for internal analysis and reporting purposes.

Institutions: Tracking early career progress for external reporting requirements.

Individuals: Collocating research activities and outputs from multiple institutions.

Applying System of Systems Engineering(SoSE) to VIVO?

Jing Wang

Abstract: System of systems (SoS) refers to a set of operational and managerial independent systems interacting with each other to provide capabilities which cannot be accomplished by any single system. SoS has its own characteristics and challenges, such as emergent behavior and evolutionary development. This presentation will:

Introduce the concepts and characteristics of system of systems (SoS) and system thinking practice. Discuss SoS challenges using SoS case studies such as Global Earth Observation SoS discuss how SoS Engineering process can be applied to networked research ecosystem in various areas such as community partnerships and SoS architecture.

VIVO at Brown University

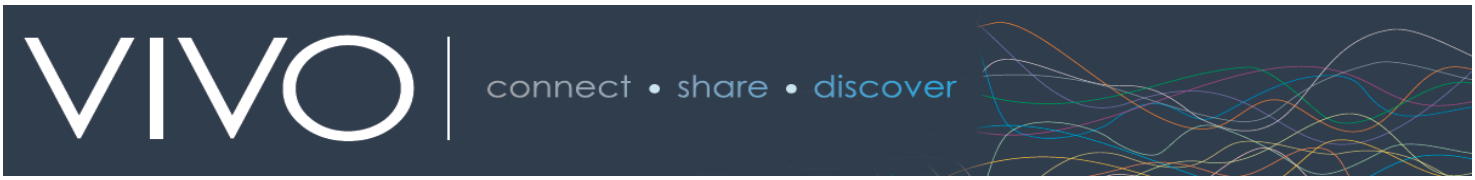
Steven McCauley, Jean Rainwater and Ted Lawless

Abstract: Brown University went live with its VIVO installation, “Researchers@Brown,” in May 2014. The platform serves as the main public portal for information about Brown faculty: their appointments, publications, areas of specialty, and more. In March 2015, one year later, Brown’s VIVO hosted nearly 60,000 unique users, making it one of Brown’s most popular websites.

Our poster will describe the current state of the VIVO project at Brown, and how we arrived here. We will cover staff support, surveying and working with faculty, and local modifications to the VIVO ontology. In addition, we have extended the base VIVO installation with a number of features. These include:

- A custom web-based editing interface, for faculty and their delegates to manage their profiles;
- Publication harvesting from CrossRef and PubMed;
- A JSON data feed pushing VIVO data to other Brown web services;
- Synchronizing local resources with external data sources, like Wikidata and OCLC’s FAST vocabulary service;
- A modified presentation layer.

We will provide an overview of these systems, our data sources and workflow, and near-term plans for future development. All together, this poster will provide a broad look at the particulars of the Brown VIVO installation, and the environment within which it operates.



Getting to Know You

Julia Trimmer and Carol Minton Morris

Abstract: Telling VIVO Stories introduces people, projects, ideas and innovation.

Using VIVO to Drive Article-Level Metrics Collection

Marianne Parkhill and Andrea Michalek

Abstract: The collection and aggregation of article-level metrics across dozens of sites for potentially millions of individual research artifacts is an extremely challenging technical problem. Plum Analytics makes it happen through the use of multiple unique identifiers that work together to not only collate multiple versions of the same artifact, but also to disambiguate authors and connect content with the right researchers in the PlumX metrics dashboards.

Research organizations work with multiple software applications to manage and keep track of grants, research output, researcher profiles and other data. Interoperability is key in order to reduce redundant metadata tracking and save staff time and effort.

Our poster will demonstrate how to use VIVO as a way of driving the collection, categorization and analysis of metrics.

We've found through more than three years of working with all types of third-party applications and research organization types that such interoperability, while challenging, can be a positive driver for adding value to work that's already been invested in such applications as institutional repositories.

A Cognitive Assistant for Cancer Team Science

Paul Thompson

Abstract: Douglas Englebart advocated the use of computers for "augmented cognition" in 1962. Even earlier artificial intelligence researchers had debated whether computers should be artificially intelligent in the same way as humans, or as engineered systems that would perform intelligently, but not necessarily in ways recognizably human. Development of a cognitive assistant is in keeping with Englebart's concept of augmented cognition. Recently IBM developed Watson, a question answering system that was able to outperform two human Jeopardy champions. IBM Watson is being used in several medical research settings for various tasks. For example, Sloan Kettering Cancer Institute developed a Watson cognitive assistant for diagnosing and treating cancer.

A cognitive assistant can provide recommendations. Two recommender system projects were presented at the 3rd National VIVO Conference in 2012, which could automatically assemble a team for team science based on meta-data available in the VIVO system. VIVO and its companion system, eagle-i, together represent researchers and their publications and resources in biomedical research laboratories. The concept of a reference interview has been discussed for decades in the field of librarianship. This concept has significant overlap with that of the role of a cognitive assistant. A patron consulting a reference librarian at a library often does not have a precise formulation of an information need. Through a discussion with the patron the reference librarian eventually determines what the patron's information need is. Several years ago Syracuse University had a virtual reference librarian project. Another library science concept worth noting in this connection is that of the Anomalous State of Knowledge, or ASK. In the 1970s Belkin and Oddy built a system based on this concept. A patron often does not know what he or she is seeking. He or she has a mental map, but there is a gap, or anomaly which needs to be filled by knowledge to be discovered in documents or other knowledge

resources. Often the patron cannot formulate such an inchoate need, but a mixed-initiative interaction system, much like the reference librarian in a human context, can help the patron clarify his, or her, need. Our project can also build on ideas from these and other biomedical Watson projects.

A Watson cognitive computing project relies on the data which is fed into Watson. Other biomedical research projects feed biomedical publications, medical records, and other biomedical knowledge into Watson. Our project will ingest similar data, but we will also take advantage of human curation applied to such data by biomedical librarians and other information specialists, e.g., references sources such as UptoDate. Instead of treating all journal articles equal as sources of biomedical knowledge, more weight will be given to sources such as meta reviews or systematic reviews. Furthermore, we will develop a collaborative user modeling frame work for information retrieval, or question answering, which will treat the queries of collaborating team scientists as additional data sources for Watson. Although our approach can support biomedical science teams more generally, we will consider the particular use case of a team of specialists working together to solve a particular difficult problem in cancer research.poster

Moving a Custom Application to use VIVO-Compliant Linked Open Data

Paul Friedman, Warren Kibbe and Violeta Ilik

Abstract: Your institution probably uses a Research Profiling System and/or Networking Tool[1]. It lets you know the structure of your institution and where a person is located within it, not matter how many times. It will tell you who wrote what, when, and with whom.

With your interest in VIVO, you understand the importance of Data Standards and the need to model your data according to an accepted ontology. However, the system you currently use is important to several people, organizations, and workflows that exist at your institution.

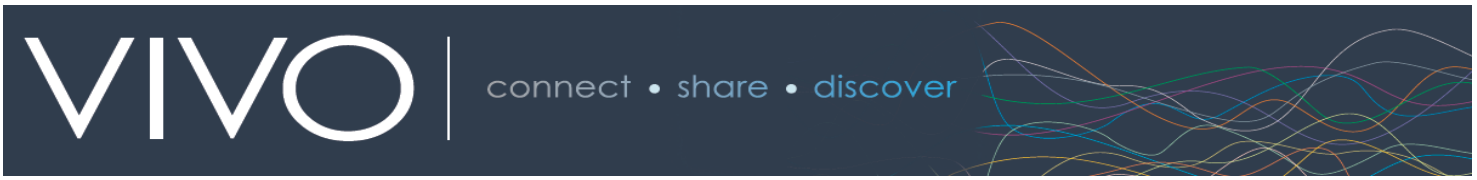
Here at Northwestern we have started transitioning from our proprietary system, LatticeGrid[2][3], to VIVO[4]. We also have a need to keep the functionality of the existing system. In moving from one application to another we have identified several pieces of functionality that do not necessarily rely on the proprietary data schema, nor do they even rely on the existing web framework. The first concrete example of this move from a proprietary data schema to the VIVO-ISF is the VIVO Visualization Library[5] project. In this project, we have uncoupled parts of the existing application from the proprietary data schema to use data from a repository using the VIVO-ISF ontology. The VIVO community can now use this previously internal functionality.

A Working Prototype for Vizualizing VIVO Linked Open Data - toward a New VIVO LOD Browser

Don Elsborg

Abstract: VIVO intially shipped with flash vizualizations which were limited to viewing particular subsets of local VIVO data. Other similar profiling applicaitons have similar visualizations which have a limited range of data which can be view.

This working prototype will actively demonstrate web based javascript browsing of VIVO data. All VIVO data. VIVO data which can cross multiple web sites, domains, and institutions. The visualization utilizes standard VIVO jsonld renderings of individuals and classgroups which VIVO can deliver "out of the box", hence additional services like SPARQL endpoints are not required.



The talk will explore the funcationality required to view data properties and classes, display VIVO images, utilize various graph displays such as sunburts and radial graphs. What VIVO data should be displayed to an end user. It's important to show real meaningful information in a simple manner that's easy to digest. How and where should this be done? What was required by the javascript developer to accomodate the VIVO model?

The prototype is a work in progress. The source code will be freely available and can be used to vizualize data in any VIVO system.

Reconceptualizing VIVO as an IR: Challenges and Opportunities in Thinking Outside of the 'Box'

Annie Gaines

Abstract: In 2012, the University of Idaho Library began implementing VIVO, an open-source Semantic Web application, both as a database to describe, visualize, and report university research activity as well as a discovery layer for its fledgling institutional repository.

The poster will detail some of the challenges librarians encountered developing this resource, while discussing the tools (such as OpenRefine, née Google Refine) and techniques they used for obtaining, editing, and uploading institutional data into the RDF-based VIVO system. Included will be hard-won tips on developing similar resources: such as "how NOT to anger entire academic departments in one email" and "Explaining RDF without sounding like a Martian."

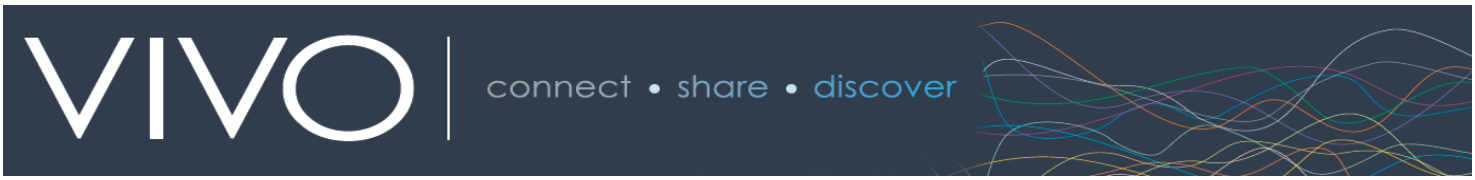
The focus on using VIVO as a front-end for an institutional repository created unique difficulties for the team, and this poster will describe the challenges and opportunities inherent in using VIVO in this manner.

A RoadMap Process for VIVO

Mike Conlon and Steering Group

Abstract: As an open source community, the VIVO Project relies on contributions of effort from its members. The VIVO strategic plan established the VIVO value proposition "VIVO provides an integrated view of the scholarly work of an organization" and two large scale goals supporting the value proposition: 1) Clarify and improve the value proposition; and 2) Promote a more open and networked research ecosystem. The roadmap process is designed to identify features and establish priorities that serve the value proposition and the strategic goals. The roadmap process follows basic principles of open source community development. First, it is open. All members of the community have an opportunity to participate in the roadmap process. Second, it serves the various constituents of the community. VIVO must serve end users, providing features that they can use in their daily work to advance scholarship. Data stewards and system administrators responsible for insuring local VIVO installations have quality data, and that the system accurately and effectively provides services to end-user must see improvement in VIVO relevant to their concerns. And finally, technologists, including developers, ontologists and committers must have environments and tools that support their productivity and creativity in enhancing VIVO.

In this poster we will present the VIVO roadmap process, describing its steps, its participants and the approach it takes to engaging the community and generating a prioritized set of features for future releases of VIVO.



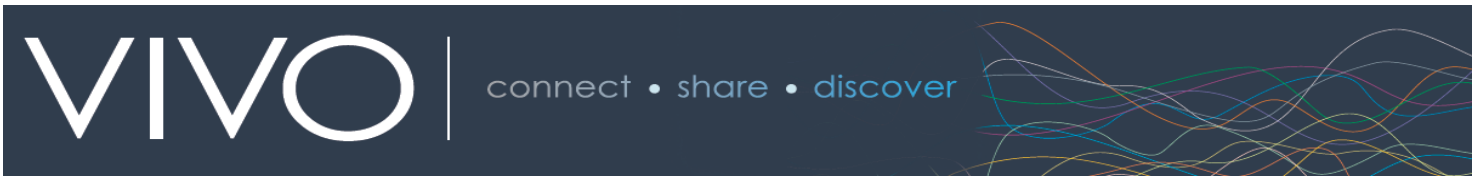
Framing and Implementing Researcher Services at the University of Pennsylvania

Manuel de La Cruz Gutierrez and Sarah Wipperman

Abstract: The University of Pennsylvania Libraries is building a comprehensive suite of researcher services. This poster will review the framework to be used and the implementation already under way. We will show how these services provide a synergy to be exploited for the benefit of the researchers and the university. At the same time, we argue this integration provides a more efficient way to deliver these specialized services in an institution like ours, and how the lessons learned could guide implementations at other institutions. Our framework for services entails three components: products, support, and benefits. Products are tools used in creating or managing research outputs and workflows throughout the research lifecycle. Examples of these include internally supported products, such as Symplectic Elements, VIVO, DMPTool, or BePress' Digital Commons; and externally used products, such as ResearchGate, Academia.edu, or MyNCBI. Support is defined as librarian activities aiding researchers directly - with or without a mediating product. Examples of these are copyright guidance, data management advice, grant support (e.g. compliance), and the dissemination of scholarly materials. Benefits are the value these products and services bring to the whole university's community and the individual researchers. For example, the university benefits from greater compliance by its researchers in terms of continuous funding. Finally, there are other clear benefits for both individual researchers and the university, such as the provision of metrics. These metrics provide a measure of the individual researcher's impact and, in an aggregate form, give university's administrators guidance for strategic initiatives. The Libraries' work on researcher services has reached a critical point of convergence in the last few years. While several products and research support have both been offered for many years now, their delivery has been piecemeal. The integration of tools, the increase in complexity and multidisciplinarity of the research enterprise, and the movement of research creation and dissemination toward mainly digital environments, have made holistic support for researchers by the Libraries a necessity. We will review what the process has been to implement these services in several different avenues. For example, VIVO was implemented two years ago to satisfy a requirement of a CTSA award to our institution. This requirement of making available a profile system for the researchers at the Perelman School of Medicine (PSOM) established a strong relationship between a cross-functional library team and PSOM. The school already had a homegrown database of faculty profiles invisible to outsiders, but the VIVO platform made their research outputs and expertise public and easily discoverable. Now we are in the process of implementing Symplectic Elements as the new internal profile system for PSOM. Its rollout will bring several benefits hitherto unavailable: data quality assurance through the use of canonical sources, updated bibliometrics and altmetrics for scholarly outputs, simplified workflow for depositing materials into our institutional repository, and the generation of reports of great value for institutional research and for compliance of large grant awards, among others. We envision great opportunities and challenges ahead in implementing these services. We have identified the need for cross-functional teams to be reformed into a more cohesive and independent library unit with more focused efforts by its librarians. We have also developed onboarding processes for rolling out new internal products to different schools within our distributed institution. Our ultimate goal is to provide researchers with a "one-stop shop" for all of their needs throughout the research lifecycle.

eagle-i and Profiles Integration: Leveraging the Integrated Semantic Framework to Connect Researchers and Resources

Tenille Johnson, Daniela Bourges, Sophia Cheng and Griffin Weber



Abstract: eagle-i (www.eagle-i.net) is a national network and open-source resource discovery tool funded by Harvard Catalyst. Its goal is to connect researchers with a variety of biomedical resources, such as animal models, cell lines, plasmids, software, instruments, and Core Facility services, while encouraging a culture of attribution for sharing. Two years ago, the eagle-i and VIVO ontologies were brought together and aligned under a common semantic framework under the VIVO-ISF to represent both people and the products of their research, including resources. Since the Profiles Research Networking Software also uses the VIVO ontology, it was a natural extension of these goals and efforts to incorporate information from eagle-i into Profiles RNS. eagle-i was designed with reusability in mind; its semantic architecture allowed us to present resource information in a way that was directly compatible with Profiles. The two aligned ontologies under the VIVO-ISF provided the backbone for data integration between eagle-i and Profiles. As a proof of concept, we began by integrating eagle-i content into Harvard Catalyst Profiles, the researcher networking tool for locating Harvard investigators. Using a call out to an eagle-i API, the Profiles database is refreshed nightly with information about any resources that a researcher has shared in eagle-i. HC Profiles then displays a short summary in that person's profile, including laboratory names, resource types, and number of resources. The latest version of the Profiles Research Networking Software now contains an optional eagle-i extension that will allow any institution running both eagle-i and Profiles RNS to connect researcher profiles to their resources in eagle-i.

Building a better VIVO

Kristi Holmes

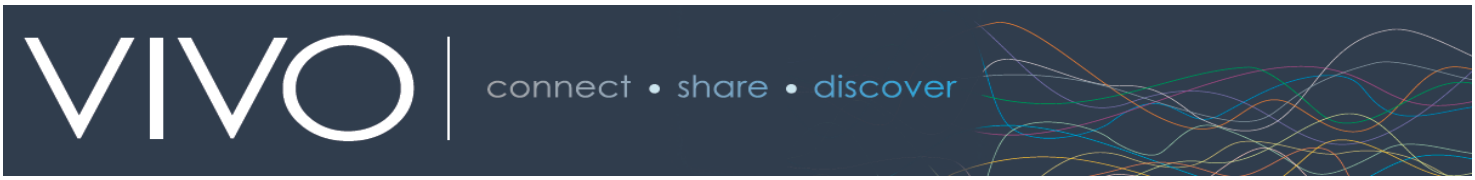
Abstract: This interactive poster will offer conference attendees an opportunity to think about the profile layout and features and brainstorm about how we can make it better. The suggestions will be delivered to the VIVO community as a set of feature requests.

Leveraging VIVO to Connect People, Instruments, and Data Semantically in the Geosciences: The UNAVCO Case Study

M. Benjamin Gross, Linda R. Rowan, Matthew Mayernik, Jon Corson-Rikert and Huda Khan

Abstract: UNAVCO is a non-profit university-governed consortium that facilitates geoscience research and education using geodesy. We plan to leverage and extend VIVO to enable the discovery of connections between people, data, and instrumentation. The UNAVCO case study is part of EarthCollab, a building block grant funded through the EarthCube Program at the National Science Foundation. The EarthCollab project includes a complementary case study at NCAR's Earth Observing Laboratory (EOL). VIVO provides an attractive alternative to developing a new application in-house. VIVO is open source, can be modified to suit our use case, and has an active implementation and development community.

VIVO has primarily been utilized as a semantic profiling application for people. As an extendable semantic framework, however, it can be used to profile datasets and instruments. Notably, the Deep Carbon Observatory and the Laboratory for Atmospheric and Space Physics (LASP) have used modified implementations of VIVO to manage and display datasets. UNAVCO utilizes a wide variety of instrumentation, such as high precision GPS receivers, terrestrial laser scanning instruments, borehole strainmeters and other geodetic tools. No single existing ontology can describe our instrument arsenal or the large amounts of real time and archived data they produce adequately. We plan to extend the VIVO ontology's limited support for datasets and instrumentation by incorporating parts of existing ontologies such as SWEET, DCAT, and GCIS, whenever possible.



UNAVCO is a geodetic facility that supports the research of external collaborators located at hundreds of institutions in the U.S. and abroad. We maintain over 2,000 instruments in the field. Additionally, we host and archive data, provide data services, have begun minting data DOIs, and offer instrument and data support to external researchers. We wish to use VIVO to display enhanced, up-to-date information about UNAVCO collaborators without the responsibility and resource commitment of being yet another canonical database store for their dynamic information, such as position titles and contact details. We will periodically ingest semantic data hosted at their home institution or at a third party (e.g. ORCID), which will be connected to our locally ingested information on datasets and instruments.

We have made significant progress ingesting datasets and publication data into VIVO. Next, we will finalize a set of ontological terms for instrument and dataset metadata. We will also develop a proof of concept implementation for sharing data between VIVO instances. Following the establishment of sustainable ingest methods, we will develop or adapt extensions to enhance discoverability of geoscience-specific information within VIVO.

Thomson Reuters and VIVO: Helping Institutions Successfully Populate and Utilize VIVO

Joey Figueroa and Julia Laurin

Abstract: Thomson Reuters, a corporate sponsor of the VIVO Durasapce project, helps institutions to successfully populate data into a VIVO instance and maximize its benefits as a research networking platform. Web of Science Core Collection offers accurate bibliographic metadata to populate VIVO publication lists. It covers over 12,600 peer-reviewed, scholarly journals from all disciplines (sciences, social science, arts & humanities), from around the world. It also covers over 160,000 conferences and 60,000 scholarly books. Over 7,000 institutions globally use Web of Science Core Collection as an essential research discovery and bibliometric analysis tool. It provides a valuable, validate collection of multidisciplinary, authoritative publication metadata for those implementing VIVO.

Additionally, Converis software helps institutions to easily create, validate, and maintain faculty profiles for use in administrative reporting workflows (e.g. promotion and tenure review, accreditation reporting, etc.), but also for public networks like VIVO. Converis enables institutions to easily combine their internal data (e.g. HR data, Institutional Repository data, legacy system data) with a variety of publication sources such as Web of Science Core Collection, PubMed, and ORCID. Converis outputs profiles into standard CVs formats such as NIH Biosketch, but also enables custom CV formats based on local needs. It minimizes the burden on faculty for maintaining profiles by automating reporting processes through customizable workflows. Converis data can be used for multiple purposes: ad hoc reporting on all faculty activities across campus; streamlining faculty review processes; and populating public web profiles, such as VIVO. Converis is highly configurable to an institution's specific needs.

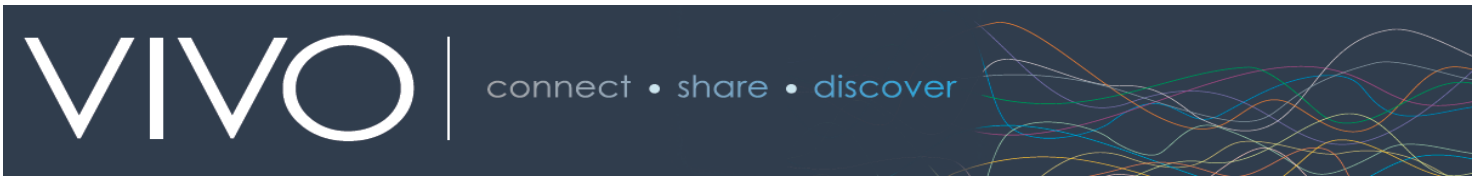
Microsoft Academic

Alex Wade

Abstract: Microsoft Academic is a new suite of experiences to facilitate discovery and awareness of academic research and the associated people, topics, journals, conferences, and universities. These experiences are fully integrated into and enhance Microsoft's existing discovery service, Bing, and our proactive digital assistant, Cortana. Come see more about how research content is made more discoverable by Bing and Cortana.

Accelerating Access – Making Open Access Policies Work from Day One

Jonathan Breeze and Graham Triggs



Abstract: Successful recruitment of published content into institutional repositories relies on three key components:

1) funder mandates or institutional policies requiring deposit of papers;

2) a means of ensuring that deposit occurs as soon as possible after a paper's acceptance or publication; and

3) an efficient, intuitive mechanism for helping faculty fulfill their deposit requirements with minimal effort.

Research management systems play an increasingly important role in the scholarly publishing ecosystem, helping collate information about scholars' publications and enabling institutions to effectively implement and monitor open access policies. This presentation examines how a research management system can be set up to provide faculty with the tools they need to easily comply with their institutional OA policy and to help repository managers track policy compliance rates across the institution.