Institution Profiling Systems at IU: VIVO et al.
Several slides are from a presentation to OVPR in 2010.

Chin Hua Kong – SLIS
Robert Light - SLIS
Katy Borner – SLIS

I would like to thank Ryan Cobine and David Cliff for their input.

Statewide IT Conference (SITC), IUB. Sept 25 at 11:30AM, Oak room, IMU.
Overview

• What is VIVO?
• How does it work?
• How have we implemented it at Indiana University?
• How is it used?
• Incentives & challenges
VIVO Collaboration:

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

What is VIVO?

An open-source semantic web application that enables the discovery of research and scholarship across disciplines in an institution.

Populated with detailed profiles of faculty and researchers; displaying items such as publications, teaching, service, and professional affiliations.

A powerful search functionality for locating people and information within or across institutions.
VIVO is a resource of Indiana University that provides information on:

- people
- departments
- facilities
- courses
- grants
- publications

vivo.iu.edu

VIVO harvests data from IU verified sources

Internal data sources:
- Faculty Systems (FAR > IUIE)
- HR System (HRMS > IUIE)
- Registrar System (SIS > IUIE)
- Research Data Systems (VPR>IUIE)
- Events and Seminars

Data stored as RDF triples using standard ontology

Faculty and unit administrators can then add additional information to their profile.

External data sources:
- Publication warehouses: e.g. PubMed, Web of Science
- Grant databases: e.g. NSF/NIH
- National Organizations: AAAS, AMA, etc.

VIVO data is available for reuse by web pages, applications, and other consumers both within and outside the institution.
**Linked Data Principles**

- Tim Berners-Lee:
  - Use URIs as names for things
  - Use HTTP URIs so that people can look up those names
  - When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
  - Include links to other URIs so that people can discover more things

- [http://www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)
- [http://linkeddata.org](http://linkeddata.org)
- [http://data.gov.uk/](http://data.gov.uk/)

---

**Storing Data in VIVO**

- Information is stored using the **Resource Description Framework (RDF)**.
- Data is structured in the form of “triples” as subject-predicate-object.
- Concepts and their relationships use a **shared ontology** to facilitate the harvesting of data from multiple sources.

![Diagram showing subject-predicate-object relationships]

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ying Ding</td>
<td>is member of</td>
<td>SLIS</td>
</tr>
<tr>
<td></td>
<td>has affiliations with</td>
<td>Cognitive Science</td>
</tr>
<tr>
<td></td>
<td>author of</td>
<td>VIVO Ontology Team</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Journal article</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Book</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Book chapter</td>
</tr>
</tbody>
</table>

---

**INDIANA UNIVERSITY**
VIVO Standard Ontology

- **Network Structure**: foaf:Person, foaf:Organization, vivo:InformationResources

- **Individual**
  - Teaching (vivo:TeacherRole, vivo:AdvisingRelationship)
  - Services (vivo:Service, vivo:CoreLaboratory, vivo:MemberRole)
  - Expertise (vivo:SubjectArea)

Linked Data: Local to National Scale

Local
- search
- browse
- visualize

National
- search
- browse
- share as RDF
- visualize

- Cornell University
- University of Florida
- Indiana University
- Ponce School of Medicine
- The Scripps Research Institute
- Washington University, St. Louis
- Weill Cornell Medical College

Text indexing

Filtered RDF
A VIVO profile will allow researchers to:

- **Map colleagues** by research area, authorship, and collaborations.
- **Showcase** credentials, expertise, skills, and professional achievements.
- **Connect** within research areas and geographic expertise.
- **Display** current research, and selected publications.
- **Publish** the URL or link the profile to other applications.

**Incentives**

- Federated Searching Across Domains
  - CTSA Federated Search
  - VIVO Federated Search
- NIH/NSF Biosketch Generation
- Mapped Data from IU Institutional Data Sources
  - 80/20
- Visualization and Scientometric Mapping Components
• Display visualizations of complex research networks and relationships.

Topical Analysis (What)  Science map overlays show where a person, department, or university publishes most in the world of science. (in work)
Topical Analysis (What) Science map overlays show where a person, department, or university publishes most in the world of science. (in work)

Current versions of VIVO do:

- Generate CVs and biosketches for faculty reporting or grant proposals - NIH/NSF.
- Incorporate external data sources for publications and affiliations.
- Link data to external applications and web pages.

Try it at http://vivo.ufl.edu/display/n25562
American Psychological Association VIVO
Bournemouth University Academic Pages
Brown University VIVO
Cornell University VIVO
Curtin University of Technology Metadata Hub
Duke University VIVO
Eindhoven University of Technology VIVO
Griffith University Research Hub (VIVO)
Hunter College VIVO
Indiana University VIVO
Johns Hopkins University VIVO
Massachusetts Institute of Technology VIVO
Medical College of Wisconsin VIVO
New York University Langone Medical Center VIVO
Northeastern University VIVO
Notre Dame University VIVO
Otago Research Data Registry
Penn State University VIVO
Queensland University of Technology VIVO
RPI Center for Nanoscale Innovation VIVO
SUNY Research in Academic Health VIVO
The Scripps Research Institute VIVO
UCLA VIVO

University of Arkansas for Medical Sciences VIVO
University of Cambridge VIVO
University of Colorado VIVO
University of Florida VIVO
University of Hawaii VIVO
University of Melbourne Find an Expert (VIVO)
University of Nebraska VIVO
University of New Mexico VIVO
University of North Texas VIVO
University of Pennsylvania VIVO
University of Virginia VIVO
University of Washington VIVO
University of Western Australia VIVO
USDA VIVO
Weill Cornell Medical College VIVO

INDIANA UNIVERSITY
Profiling Systems at IU

IU currently has four systems—more than any other university I know is able to afford.

Only VIVO and IndianaCTSI have exchanged data. SciVal and Pivot are commercial solutions that are not interoperable.

Other universities have conducted extensive market studies - see comparison of 45 systems at [http://en.wikipedia.org/wiki/Comparison_of_Research_Networking_Tools_and_Research_Profiling_Systems](http://en.wikipedia.org/wiki/Comparison_of_Research_Networking_Tools_and_Research_Profiling_Systems) - and decided to implement a combination of SciVal (to purchase publications) and VIVO (to be compliant with many other systems that expose their data using the VIVO ontology in support of national search and other services across profiling instances).

Many universities are using their Faculty Profiling System to compile large teams in response to funding solicitations. Purdue has connected their profiling system to their FAR like 'Digital Measures' system and will soon be able to analyze and visualize their impact in new ways.

I believe we all would benefit if IU commits to one Faculty Profiling System but it will take IU leadership to make this happen.

---

Data in VIVO at IU

VIVO development instance
- Bloomington faculty (source: IUIE data warehouse)
- IUPUI faculty (source: IUIE data warehouse)
- IU CTSI personnel (source: CTSI Portal, IUIE data warehouse)
- Inclusion of a faculty member entails basic HR data (appointments, rudimentary contact info), academic courses taught in last five years, and federal or federal pass-through grant awards for which they were PI or co-PI.
- IUCTSI personnel may additionally include a research overview and research area keywords.
- IUB and IUPUI organizational structure

VIVO production instance
- Bloomington faculty (source: IUIE data warehouse)
- IU CTSI personnel (source: CTSI Portal, IUIE data warehouse)
- Inclusion of a faculty member entails basic HR data (appointments, rudimentary contact info), academic courses taught in last five years, and federal or federal pass-through grant awards for which they were PI or co-PI.
- IUCTSI personnel may additionally include a research overview and research area keywords.
- IUB and IUPUI organizational structure

Provost Robel suggested to explore linking VIVO and FAR—this would considerably improve data quality and coverage.
We started to run analyses of teaching/funding/affiliation data from http://vivo.iu.edu to identify collaborations/connections/overlaps for the SLIS-SOIC merger.

The very same analyses might be valuable for other reorganization efforts.

There are about 10 science of science/scientometrics scholars at IUB and several of us would be interested to perform more detailed studies or to provide advise.
P30 Member Collaborations – Sponsored Project Co-Participation and Co-Authorship Network. Used in successful! P30 funding application. Shows the PI’s relationships with various P30 members, conveying that the PI was not only the formal center of the group but also the informal center and the person who exhibited the highest betweenness centrality. Contact: Jeffrey Horon, J.Horon@elsevier.com

Questions?

Chin Hua Kong
kongch@indiana.edu

Robert Light
lightr@indiana.edu

Katy Bomer
kty@indiana.edu

Thank you!
Senior Software Engineer/Research Analyst (3IT)  IU Job #6839
As Senior Software Engineer, you will perform research and programming for current and future externally funded research projects at the CNS Center. These projects include tools powered by the Cyberinfrastructure Shell (CIShell, http://cishell.org), an open-source software platform that supports the interchange of datasets and algorithms; MapIN, a map of Indiana's expertise and resources; and other online interactive maps and web sites. You will participate in the entire software development process, from the collection of user stories through planning, implementation, testing, deployment, and documentation. You will also be expected to participate in the training new developers, and the creation of educational material for workshops. As Senior Software Engineer, you will have a chance to help set the standards of our team in many areas, including code, teamwork, product direction, and process.

Software Developer (2IT)  IU Job #6862
As a Software Developer, you will work in a team of four to perform research and programming for current and future externally funded research projects at the CNS Center. The main focus will be on tools powered by the Cyberinfrastructure Shell (CIShell, http://cishell.org). CShell is an open-source software platform, built on Java and OSGi that allows developers and scientists to easily exchange datasets and algorithms, and bundle them into custom tools that serve the particular needs of research communities. You will participate in the entire software development process, from the collection of user stories through planning, implementation, testing, deployment, and documentation.