D. Krafft, V. Davis: Presenters

Co-Authors: J. Corson-Rikert, M.Conlon, M. Devare, B.Lowe, B.Caruso, K.Börner, Y.Ding, L.McIntosh; M. Conlon; and

VIVO Collaboration*
VIVO Collaboration:

**Cornell University:** Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly 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, Valrie 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 Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **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 Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.
Overview

• What is VIVO?
• How does it work?
• How do we implement it?
• What’s ahead?
Problem:

- Researchers often struggle to locate and communicate with collaborators across fields and outside rigidly defined organizational confines.

Solution:

- VIVO will help create the collaborations that are increasingly crucial in science by facilitating communication and collaboration across interdisciplinary and institutional boundaries NOT ONLY for scientists but also for administrators, students, faculty, donors, funding agencies, and the public.
What is VIVO?
VIVO is:

A **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.
A VIVO profile allows researchers to:

• Illustrate academic, social, and research networks.
• Showcase special skills or expertise.
• Establish connections and communities within research areas and geographic expertise.
• Summarize credentials and professional achievements.
• Publish the URL or link the profile to other applications.
Who can use VIVO?

- Looking for collaborators on large or multi-disciplinary grants
- Keeping abreast of new research

- Locating graduate programs, mentors, or advisors
- Looking for seminars or events

- Searching for specialized expertise
- Visualizing research activity within an institution

- Showcasing departmental activities
- Centralizing faculty information

...and many more!
VIVO as disseminator

Graduate Programs in the Life Sciences

Graduate programs at Cornell are organized by Fields. The first step in applying is identifying a Field that best matches your academic goals.

With over 30 Fields to choose from in the Life Sciences, let’s narrow it down by selecting a broad interest...

- applied biology
- biomedical sciences
- ecology, evolution, and integrative biology
- genetics, genomics, and bioinformatics
- molecular and cellular biology
- physical, chemical, and computational biology

...or see the full list of Fields instead

Already determined which Field is right for you?  Apply Now

http://gradeducation.lifesciences.cornell.edu/
How does it work?
VIVO harvests much of its data automatically from verified sources

- Reducing the need for manual input of data.
- Centralizing information and providing an integrated source of data at an institutional level.

Individuals may also edit and customize their profiles to suit their professional needs.
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.

Subject | Predicate | Object
--- | --- | ---
Jane Smith | is member of | Dept. of Genetics
Jane Smith | has affiliations with | College of Medicine
Jane Smith | author of | Genetics Institute
Jane Smith | author of | Journal article
Jane Smith | author of | Book
Jane Smith | author of | Book chapter

Jane Smith is member of the Dept. of Genetics and has affiliations with the College of Medicine, Genetics Institute, Journal article, and Book. She is the author of several works, including a Book, a Book chapter, and a Journal article.
Detailed relationships for a researcher

- Susan Riha
  - taught by
  - CSS 4830
  - faculty appointment in
    - Earth and Atmospheric Sciences
  - featured in
    - CSS 4830

- Andrew McDonald
  - author of
    - Mining the record: Historical evidence for...
  - research area for
    - crop management
  - academic staff in
    - NYS WRI
  - head of
    - Earth and Atmospheric Sciences

- Cornell’s supercomputers crunch weather data to help farmers manage chemicals
Query and explore

- By individual
  - Everything about an event, a grant, a person
- By type
  - Everything about a class of events, grants, or persons
- By relationship
  - Grants with PIs from different colleges or campuses
- By combinations and facets
  - Explore any publication, grant, or talk with a relationship to a concept or geographic location
  - Explore orthogonally (navigate a concept or geographic hierarchy)
By individual: a grant

<table>
<thead>
<tr>
<th>Research Grant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>has primary investigator</strong></td>
</tr>
<tr>
<td>Pell, Alice N.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>award administered by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell International Institute for Food, Agriculture, and Development (CIFAD)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>funded by</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Science Foundation (NSF)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>has co investigator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrett, Chris</td>
</tr>
<tr>
<td>Blume, Lawrence Edward</td>
</tr>
<tr>
<td>Lehmann, Christopher Johannes</td>
</tr>
<tr>
<td>Riha, Susan Jean</td>
</tr>
</tbody>
</table>

From Cornell OSP data warehouse (project id: 41944) on 2007-03-08T17:19:42.813-05:00
Browse by Type: seminar series

Seminar and Exhibit Series
Series sponsored by individual departments or academic units.

AAP Exhibition Series | exhibit series | Exhibition Calendar
Events and Exhibitions in the AAP Galleries, which include Hartell Gallery, Tjaden Gallery, and the Experimental Gallery.

Africana Colloquium Series | Seminar or Lecture Series | Africana Colloquium Series
Americanist Colloquium | seminar series | History department web page
Tuesdays, 4:30-6pm, McGraw Hall 215

Analysis Seminar | Seminar or Lecture Series | Analysis Seminar Series
The Cornell Analysis Seminar takes place Monday at 2:30-3:30 pm in 406 Malott Hall. All are welcome.

Andrew D. White Professor at Large Lecture Series | lecture series | A.D. White Professor at Large webpage

Animal Science Seminar Series | Seminar or Lecture Series | Department of Animal Science Seminar web page

Anthropology Colloquia | seminar series | Department of Anthropology Seminar Series
All Lectures are Free and Open, Fridays at 3:30 PM in 215 McGraw Hall. (Unless otherwise indicated).

Applied Economics and Management Seminars | seminar series | AEM seminars

Applied Microeconomics Seminars | seminar series | Applied Microeconomics seminars
All workshops are on Wednesdays at 4:00 p.m. in Uris Hall 498 unless otherwise noted.

Architecture Lectures | Seminar or Lecture Series | Architecture Lectures
Results by Topic: homeostasis
Advantages of an ontology approach

- Provides the key to meaning
  - Defines a set of *classes* and *properties* in a unique namespace
  - Embedded as RDF so data becomes self-describing
  - Definitions available via the namespace URI

- Helps align RDF from multiple sources
  - VIVO core ontology maps to common shared ontologies organized by domain
  - Local extensions roll up into VIVO core
Information flow with shared ontologies

Log boom on Williston Lake, British Columbia
Information flow without shared ontologies

Log jam, looking up the dalles, Taylor's Falls, St. Croix River, MN. Photo by F.E. Loomis
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)
VIVO enables authoritative data about researchers to join the Linked Data cloud
Challenges in our approach

- Granularity levels
- Terminologies
- Scalability
- Disambiguation
- Provenance
- Temporality

BUT - “I believe that a WEB of Semantic terminology (anchored in URIs qua RDF/OWL) will create a powerful network effect, where the wealth of knowledge (despite being open and uncurated) creates a wide range of new possibilities, without the need for expressive logics and powerful reasoners (that cannot scale to web sizes).” – Jim Hendler

Major VIVO Components

- A general-purpose, open source Web-based application leveraging semantic standards: ontology editor, data manager, display manager

- Customizations of this application for VIVO
  - Ontology
  - Display theming
  - Installation

- Additional new software to enable a distributed network of RDF by harvesting from VIVO instances or other sources
VIVO’s three functional layers

- End users
- Curators
- Ontology editing & data flow

- Search and browse interface
- Editing
  - Display, search and navigation setup
  - Curator editing
  - Ontology Editing
  - Data ingest
  - Data export
Local data flow

1. Local systems of record
2. National sources
3. Data ingest ontologies (RDF)
4. VIVO (RDF)
5. Shared as RDF

Interactive input:
- RDFa
- RDF harvest
- SPARQL endpoint

Peoplesoft
Grants DB
PubMed
Publishers

Researchers
Librarians
Administrative Staff
Self-Editors
From local to national

Local
- local sources
- nat’l sources

VIVO
- search
- browse
- visualize

Website

Data

Text indexing

National
- search
- browse
- share as RDF

Visualize

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

Linked Open Data
How do we implement it?
Start with the project participants

Institutions currently working to enable National Networking with VIVO

Work to enable National Networking is supported via stimulus funds from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH).
Enable a National Network

The National VIVO Network enables the discovery of research across institutions.

Data is imported, stored, and maintained by the individual institutions.

VIVO allows researchers to browse and ...enables networking on a national scale.
Participation: potential adopters and data providers

- Individual institutions – Columbia University, …
- Federal agencies – NIH (NIH RePORTER), NSF, …
- Consortia of schools – SURA, CTSA…
- eagle-i (“enabling resource discovery” U24 award)
- Publishers/vendors – PubMed, Elsevier, Collexis, ISI…
- Professional societies – AAAS, …
Participation: application developers and consumers

- Semantic Web community – DERI, others…
- Search Providers – Google, Bing, Yahoo, …
- Professional Societies – AAAS, …
- Producers, consumers of semantic web-compliant data
Facilitating adoption: Challenges

- Scientist vs. administrator
  - Individual vs. institutional

- Research vs. clinical community
  - Researchers: Early vs. established; enthusiastic, willing to participate
  - Clinicians: Less enthusiastic; “business” accumulated in other ways

- Innovators, early adopters
  - Excited; willing to experiment

- Early, late majority
  - Hesitant to adopt new technologies
  - VIVO takes work
Facilitating adoption & participation

- Support and dissemination through libraries
  - Neutral, trusted institutional entities
  - IT, information management and dissemination expertise
  - Subject experts; experience in translational/outreach roles
  - Knowledge of institutional, research, instruction landscape
  - Service ethic; tradition of academic support
  - Recognition of challenges posed by user communities…
  - Facilitate resolving data integration problems endemic to legacy systems
Librarians as facilitators: Our model

- Oversight of initial content development
  - Content types, ontology, interface refinement…
  - Negotiation with campus data stewards for publicly visible data

- Support and training: local and national level
  - Documentation, presentation/demo templates
  - Help-desk support, videos, tutorials, web site FAQs, etc.

- Communication/liaising
  - Maintain web site (http://www.vivoweb.org)
  - Engage with potential collaborators, participants
  - Create community of support via user forums
  - Usability: Feedback, new use cases from users to technical team

- Marketing
  - Demonstrations/exhibits, conferences, workshops, website
  - PR materials
What’s ahead?
Near-term developments

- Release 1 (as of Friday April 16th)
  - Deployment at 7 sites on production hardware
  - Mapping local data to R1 ontology
  - Batch data ingest (extract-transform-load)

- Upcoming Releases
  - Feedback on ontology; develop local ontologies
  - Improved usability
  - Open software (BSD license) and ontology for download

- Next 6-12 months
  - Expand data ingest framework
    - Publications | people | grants | courses
  - Visualization in-page and at site level
  - National network application design
  - Improvements to VIVO to support modularity & customizability
  - User support material development
  - Expand functionality to meet developing use cases
Driving future participation

- “User scenario”-based ontology and feature development
- Usability; robust user support
- Compelling proof of concept within consortium
  - Administrators, scientists, clinicians, students, staff…
- Addressing interest from other institutions, partners
  - [http://www.vivoweb.org](http://www.vivoweb.org)
  - Conferences and workshops: National VIVO Conference (8/12/10; NYC)
  - Individual demos; meetings
  - Exhibits, publicity
Network Analysis & Visualization

- At the individual investigator level
  - In-page, small graphs highlighting publication history, collaborations or co-authorship networks

- At the department or institutional level
  - Trends in research grants or publications
  - Collaboration networks across a larger group
  - Topical alignment with base maps of science

- At the network level
  - Patterns and clusters by geography, topic, funding agency, institution, discipline
Future versions of VIVO will:

- **Generate** CVs and biosketches for faculty reporting or grant proposals.
- **Incorporate** external data sources for publications and affiliations.
- **Display** visualizations of complex research networks and relationships.
- **Link** data to external applications and web pages.
- **Realize** the full potential of the semantic web.
Call to action

- Get involved as an:
  - adopter,
  - data provider, or
  - application developer

- Visit:
  - http://www.vivoweb.org

Thank you!  Questions?