Development Overview

August 13, 2010
New York Hall of Science

VIVO: Enabling National Networking of Scientists is supported by NIH Award U24 RR029822.
VIVO Collaboration:

**Cornell University**: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, Elizabeth Hines, Huda Khan, Brian Lowe, Joseph McEnerney, Holly Mistlebauer, Stella Mitchell, Anup Sawant, Christopher Westling, Tim Worrall, Rebecca Younes.


**Indiana University**: Katy Borner (IU PI), Kavitha Chandrasekar, Bin Chen, Shanshan Chen, Jeni Coffey, Suresh Deivasigamani, Ying Ding, Russell Duhon, Jon Dunn, Poornima Gopinath, Julie Hardesty, Brian Keese, Namrata Lele, Micah Linnemeier, Nianli Ma, Robert H. McDonald, Asik Pradhan Gongaju, Mark Price, Yuyin Sun, Chintan Tank, Alan Walsh, Brian Wheeler, Feng Wu, Angela Zoss.

**Ponce School of Medicine**: Richard J. Noel, Jr. (Ponce PI), Ricardo Espada Colon, Damaris Torres Cruz, Michael Vega Negrón.

**The Scripps Research Institute**: Gerald Joyce (Scripps PI), Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaeleen Trimarchi.

**Washington University School of Medicine in St. Louis**: Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Caerie Houchins, George Joseph, Sunita B. Koul, Leslie D. McIntosh.

**Weill Cornell Medical College**: Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Mark Bronnimann, Adam Cheriff, Oscar Cruz, Dan Dickinson, Richard Hu, Chris Huang, Itay Klaz, Kenneth Lee, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Vinay Varughese, Virgil Wong.
VIVO development overview

I. VIVO development to date and goals for the coming year

II. VIVO virtual appliances and the Harvester

III. Collaborative development opportunities
Presenters

- Intro - Jon Corson-Rikert
- VIVO application improvements - Brian Caruso
- Semantic infrastructure & directions – Brian Lowe
- User interface goals & vision - Nick Cappadona
- Visualization – Micah Linnemeier
- Packaging & virtualization – Narayan Raum
- Harvester – Stephen Williams
- Collaboration opportunities – Chris Barnes
VIVO conceptual architecture

- Local systems of record
- National sources

Harvester

VIVO (RDF)

- Interactive input
- Verified Information Sources

Linked Data
SPARQL

Shared as RDF

HR/Peoplesoft
Grants DB
Courses
PubMed
Publishers

Researchers
Librarians
Administrative Staff
Self-Editors
From local to national

Local

- local sources
- nat’l sources

VIVO

- search
- browse
- visualize

Local to national

- search
- browse
- visualize
- share as RDF

Exemplar

- 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
Developments to date by the application team

- Linked data
- Editing support via custom forms for the 1.0 and 1.1 ontology
- Integration of Freemarker templating engine
- Addition of support for ajax features in custom forms
- Rewrite of image upload
- Rewrite of login
- Improvements to full text indexing
- Improvements to our development and release process
- Automated testing using Selenium
- Continuous integration
- Build script refactoring for maintainability
Future developments by the application team

- Self-editing
- Authentication
- Authorization (frequently called "roles")
- Improvements to navigation configurability
- Adding features to support the customization of editing and display
- Improvements to website search
- Improvements to support for external taxonomies and controlled vocabularies
National network search

- Linked data is in place
- Aggregator service
- National network search site that uses aggregator services
Semantic infrastructure and directions

- VIVO Core ontology and local extensions
- Ontology evolution
- Graph management
- Triple store scalability
- Reasoning and inter-ontology connections
VIVO Core ontology

- VIVO core ontology should be as stable as possible
- But it will never be perfect
- Needs to evolve in response to available data
- Migrate VIVO data and update local extension as core ontology evolves
Local extensions to VIVO Core

- Accommodate additional data for institutional needs
- Map up to common ontology for inter-institutional exchange
- Make it easy to create or import local extensions of the VIVO local ontology
Graph management

- Public sees a big network of linked data

- Behind the scenes, multiple graphs
  - Any statement can reference a subject, predicate, or object stored in a different graph

- Manage different data sources and ontologies separately

- Improve ability to update and edit different types of data
Triple store scalability

- Semantic Web apps love lots of memory
- But we want to integrate lots of data
- Optimize VIVO to keep ontologies in memory but most data in secondary storage
  - Multiple graphs!
Reasoning: additional triples

• Semantic Web technologies enable reasoning

• Assert as little as possible; let the rest be inferred
  ▫ Important - removed when no longer true

• Support as many popular ontologies as possible through mapping rules

• “Materialize” only the most important inferences directly in the VIVO application
Challenges in the semantic approach

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

VIVO approach

- Make it easy to enter structured data
- Address trust via authoritative sources
- Address privacy via focus on public data

# User interface – Goal & definitions

- Make the user’s experience as simple, productive, efficient and enjoyable as possible

<table>
<thead>
<tr>
<th>User Interface (UI)</th>
<th>system by which a person (user) interacts with a machine (human-computer interaction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Interface Design</td>
<td>design of such systems with focus on the user’s experience and interaction</td>
</tr>
<tr>
<td>Usability</td>
<td>the ease with which a user can employ a particular tool in order to achieve a particular goal</td>
</tr>
<tr>
<td>User Experience (UX)</td>
<td>a person’s perception in response to use or anticipated use of a system or product</td>
</tr>
</tbody>
</table>
UI design components

- Functionality Requirements
- User Analysis
- Information Architecture
- Visual Design
- Prototyping
- Usability Testing
Our process

1. Identify our users
2. Understand our users
3. Identify key tasks
4. Analyze interface elements in existing systems
5. Design and prototype interfaces
6. Apply visual design to prototypes
7. Test our designs on our users
8. Listen, analyze and learn (Identify Patterns)
9. Apply what we've learned

wash, rinse and repeat
User Interface
Deliverables
Visualization

- Introduction
- Completed Work
- In-Development
- Future Work
Visualization

- Introduction
- Completed Work
  - Co-Author visualization
  - Sparklines
  - VIVO world activity map
- In-Development
- Future Work
Borner, Katy

Person

This information is based solely on publications which have been loaded into the VIVO system. The data is not comprehensive.

General Statistics

36 publication(s) from 2001 to 2010 (.CSV File)

80 co-author(s) from 2001 to 2010 (.CSV File)
Co-Author Network (GraphML File)

15 co-author(s)
35 co-author link(s)

Russell Gonzalez, Sara A
Physical Sciences Librarian
VIVO profile | Co-author network

9 Publication(s)
15 Co-author(s)
1998 First Publication
2010 Last Publication
VIVO 1.0 source code was publicly released on April 14, 2010.  87 downloads by June 11, 2010.  917 downloads on July 16, 2010.

The more institutions adopt VIVO, the more high quality data will be available to understand, navigate, manage, utilize, and communicate progress in science and technology.

Cumulative Counts (World)
Jan 29 - Jul 30, 2010

- VIVO People Profiles: 40,841
- VIVO Email Requests: 128
- VIVO Web Visits (Countries): 300,254
- VIVO Code Downloads: 748

Profiles: 100, 1, 10, 100, 100,000
Emails: 1, 10, 10,000
Web Visits: 100, 10,000, 1,000,000

Created by: Jennifer R. S. Coffey (design), Kaveh Ekbia, Justin Peters (ArcGIS) and Katy Börner (concept).
Visualization

- Introduction
- Completed Work
- In-Development
  - SciMaps for VIVO
  - Comparsion vis
- Future Work
School Comparison Visualization

How do you want to compare schools?

Compare by:
- Grants
- Publications
- Item 3
- Item 4

Select schools to compare:
- School of Law
- School of Computer Science
- School of Art
- School of Music
- School of Folklore
- School of Informatics
- School of Education
- School of Journalism
- School of Libraries
- School of Information Science

Total Number of Publications
You have selected 7 of a maximum 10 schools to compare.

<table>
<thead>
<tr>
<th>School</th>
<th>Year 2010</th>
<th>Year 2011</th>
<th>Year 2012</th>
<th>Year 2013</th>
<th>Year 2014</th>
<th>Year 2015</th>
<th>Year 2016</th>
<th>Year 2017</th>
<th>Year 2018</th>
<th>Year 2019</th>
<th>Year 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>School of Art</td>
<td>812</td>
<td>823</td>
<td>834</td>
<td>845</td>
<td>856</td>
<td>867</td>
<td>878</td>
<td>889</td>
<td>890</td>
<td>891</td>
<td>892</td>
</tr>
<tr>
<td>School of Music</td>
<td>877</td>
<td>888</td>
<td>899</td>
<td>900</td>
<td>911</td>
<td>922</td>
<td>933</td>
<td>944</td>
<td>955</td>
<td>966</td>
<td>977</td>
</tr>
<tr>
<td>School of Informatics</td>
<td>2070</td>
<td>2081</td>
<td>2092</td>
<td>2103</td>
<td>2114</td>
<td>2125</td>
<td>2136</td>
<td>2147</td>
<td>2158</td>
<td>2169</td>
<td>2180</td>
</tr>
<tr>
<td>School of Education</td>
<td>2421</td>
<td>2432</td>
<td>2443</td>
<td>2454</td>
<td>2465</td>
<td>2476</td>
<td>2487</td>
<td>2498</td>
<td>2509</td>
<td>2520</td>
<td>2531</td>
</tr>
<tr>
<td>School of Journalism</td>
<td>1836</td>
<td>1847</td>
<td>1858</td>
<td>1869</td>
<td>1880</td>
<td>1891</td>
<td>1902</td>
<td>1913</td>
<td>1924</td>
<td>1935</td>
<td>1946</td>
</tr>
</tbody>
</table>
Visualization

- Introduction
- Completed Work
- In-Development
- Future Work
  - Institution level visualization/reports
  - Cross-site analysis (national level)
  - Topic analysis
VIVO virtual appliance

- Purpose
  - Don’t have a server?
  - Don’t want to install Java, Tomcat, Apache or VIVO source right away?
- Evaluate VIVO with sample data
- Proof of concept for potential adopters
- Training and workshops
- Tool development testing
- Upgrade testing
- Production at smaller schools
Different flavors

- **Standard version**
  - Pre-loaded tools
- **Marketing version**
  - Pre-loaded institutional data
  - Harvester available for test ingest of PubMed data
- **Upgrades**
  - Will upgrade versions with VIVO source
  - Tools upgraded as well
Installing the appliance

• ~2 GB file download
• ~10 minutes from download to running instance
• Oracle VirtualBox
  ▫ Mac, Windows, Linux compatible
  ▫ Snapshots for testing/rollback VIVO changes
• VMWare ESX/Desktop
• http://vivo.sourceforge.net
Using the appliance
VIVO harvests much of its data automatically from verified sources

• Reduces the need for manual input of data
• Provides an integrated and flexible source of publicly visible data at an institutional level

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

- Collection of tools for VIVO administrators to manipulate rdf models and import xml and csv data into their vivo installation
Data harvesting

- VIVO Harvester vision and implementation
  - Software design philosophies
    - Unix small tools
    - Daemons
  - Data source target selection
    - National importance
    - Methods applicable to other data sources
    - Ease of integration into VIVO
    - Vision of VIVO
  - Development timeline
    - Constant customer feedback
    - Agile development cycle
Harvester design

- External Data Store
- Fetch
- Translate
- Transfer
- Utilities
- RDF
  - Score
  - Qualify
- Raw Data
- VIVO
Targets for harvesting data

- Human Resources
  - Grants Management
- Faculty Reporting
  - Project Management
- Local Data Sources
- National Data Aggregators
  - CiteSeerX beta
  - Grantsfire
- National Data Repositories
  - PubMed.gov
  - GRANTS.GOV™
- National Organizations
  - NSF
  - National Institutes of Health
  - AAAS

VIVO | enabling national networking of scientists
Challenges

- Entity identification
  - Type of identification
    - Author disambiguation
    - Paper De-duping
  - Algorithms
    - Machine learning
    - Data classification
    - Natural language processing
  - User feedback - accept/reject
  - Unique identifiers
Get involved with VIVO

- As an:
  - adopter,
  - data provider, or
  - application developer

- Open source code (BSD) and ontology available at

http://vivoweb.org
Get involved with VIVO

Goals:
- Open development community
- Full transparency
- Design documents
- Goals / future direction
- Documentation
- Bugs
- Source code access
- Ontology access
- Utilities and community
- developed tools for VIVO

http://sourceforge.net/projects/vivo/
Get involved with VIVO

- irc.freenode.net #vivo
- Ask questions
- Get answers
- Submit feedback
- Collaborate with other VIVO community developers

http://webchat.freenode.net/?channels=VIVO
Get involved with VIVO

- Collaborate
- Use our Linked Data
- Consume VIVO data for your own tools

VIVO data in DRUPAL portal

From SPARQL EndPoint
Thanks!

Questions?