Science of Science Research and Tools
Tutorial #11 of 12

Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
http://info.slis.indiana.edu/~katy

With special thanks to Kevin W. Boyack, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank Nianli Ma, Hanning Guo, Mark A. Price, Angela M. Zoss, and Scott Weingart

Invited by Robin M. Wagner, Ph.D., M.S.
Chief Reporting Branch, Division of Information Services
Office of Research Information Systems, Office of Extramural Research
Office of the Director, National Institutes of Health

Suite 4090, 6705 Rockledge Drive, Bethesda, MD 20892
9:30a-11:30a, July 28, 2010

12 Tutorials in 12 Days at NIH—Overview

1. Science of Science Research — **1st Week**
2. Information Visualization
3. CIShell Powered Tools: Network Workbench and Science of Science Tool

4. Temporal Analysis—Burst Detection — **2nd Week**
5. Geospatial Analysis and Mapping
6. Topical Analysis & Mapping

7. Tree Analysis and Visualization — **3rd Week**
8. Network Analysis
9. Large Network Analysis

10. Using the Scholarly Database at IU — **4th Week**
11. **VIVO National Researcher Networking**
12. **Future Developments**

12 Tutorials in 12 Days at NIH—Overview

[#11] VIVO National Researcher Networking

- Motivation
- Users, Their Needs, and Usage Scenarios
- Development
- Implementation
- Usage
- Outlook
- Exercise: Identify Promising VIVO Collaborations

Recommended Reading

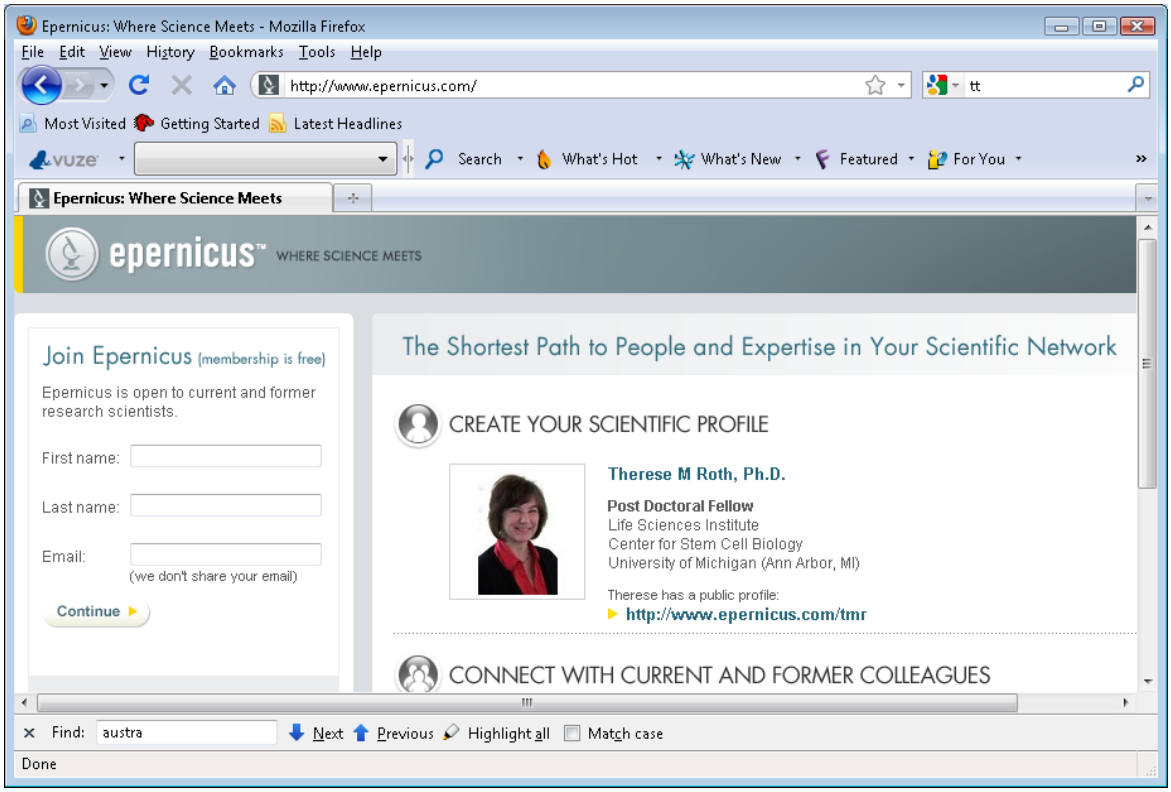
VIVO home page, http://vivoweb.org

VIVO Conference in NYC in August 2010, http://conferences.dce.ufl.edu/vivo

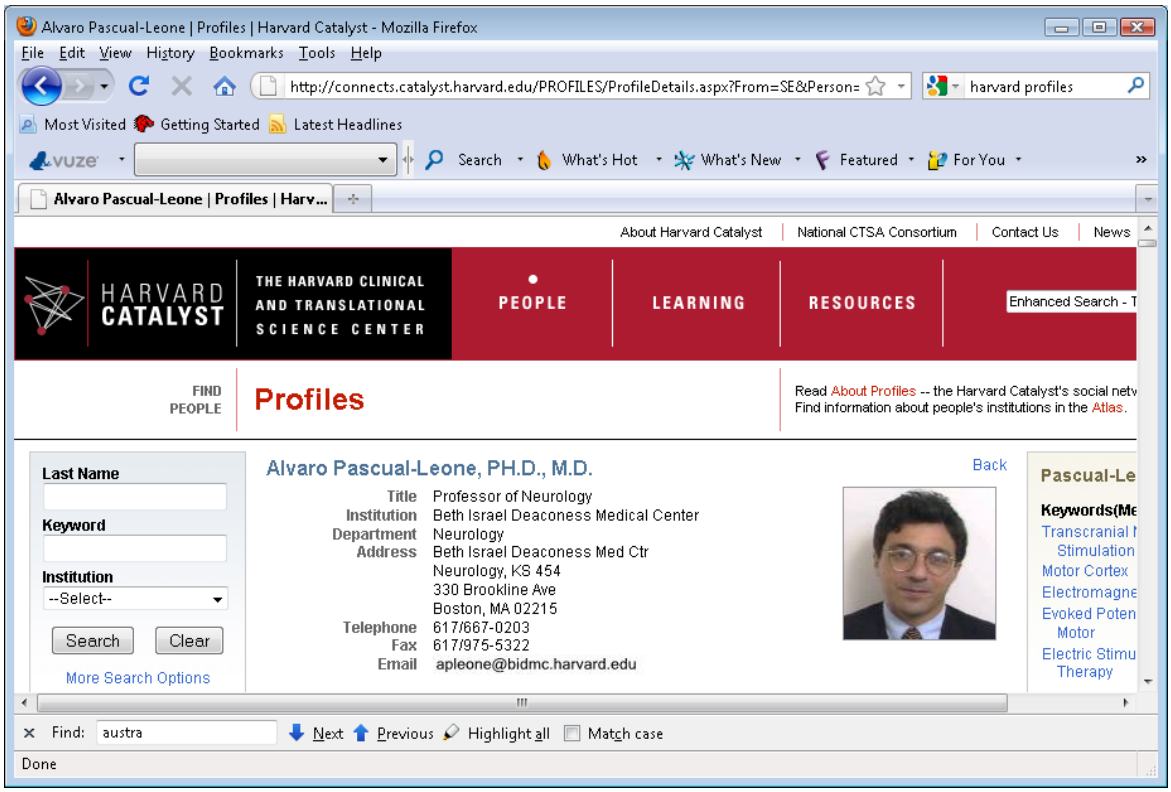
[#11] VIVO National Researcher Networking

- **Motivation**
- **Users, Their Needs, and Usage Scenarios**
- **Development**
- **Implementation**
- **Usage**
- **Outlook**
- **Exercise: Identify Promising VIVO Collaborations**

How many of you use

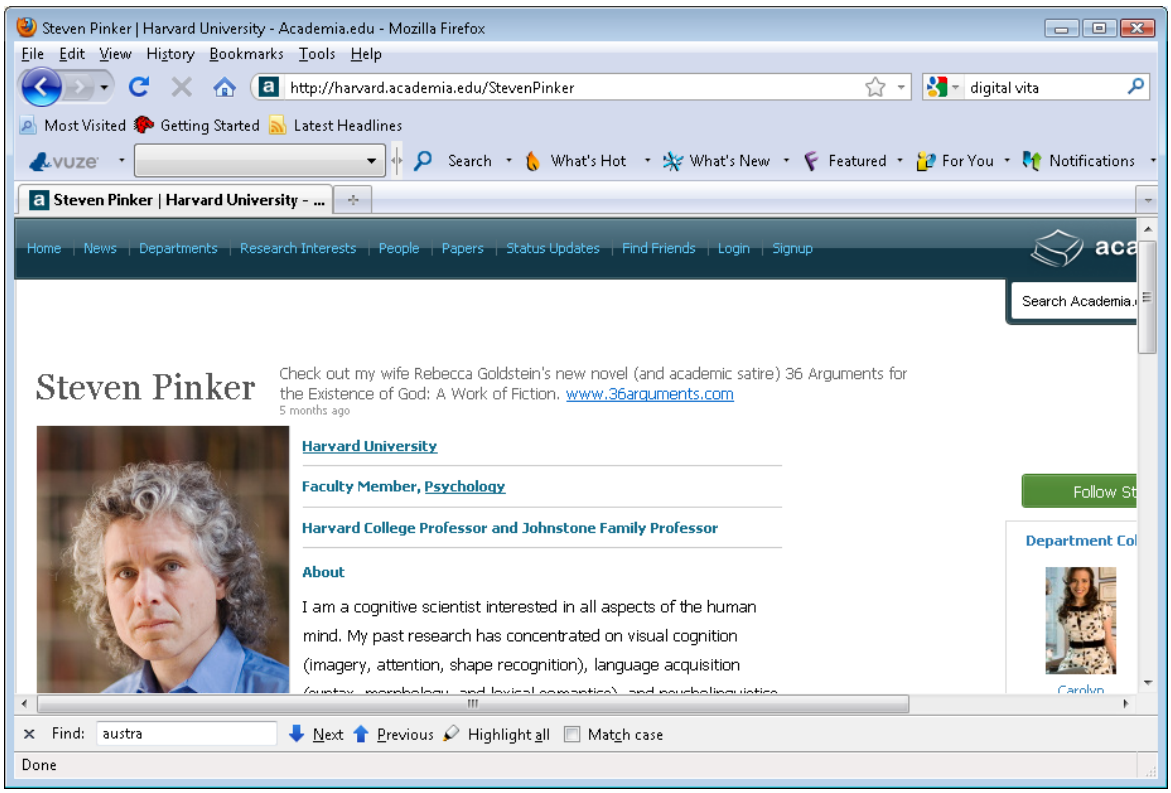


How many of you use

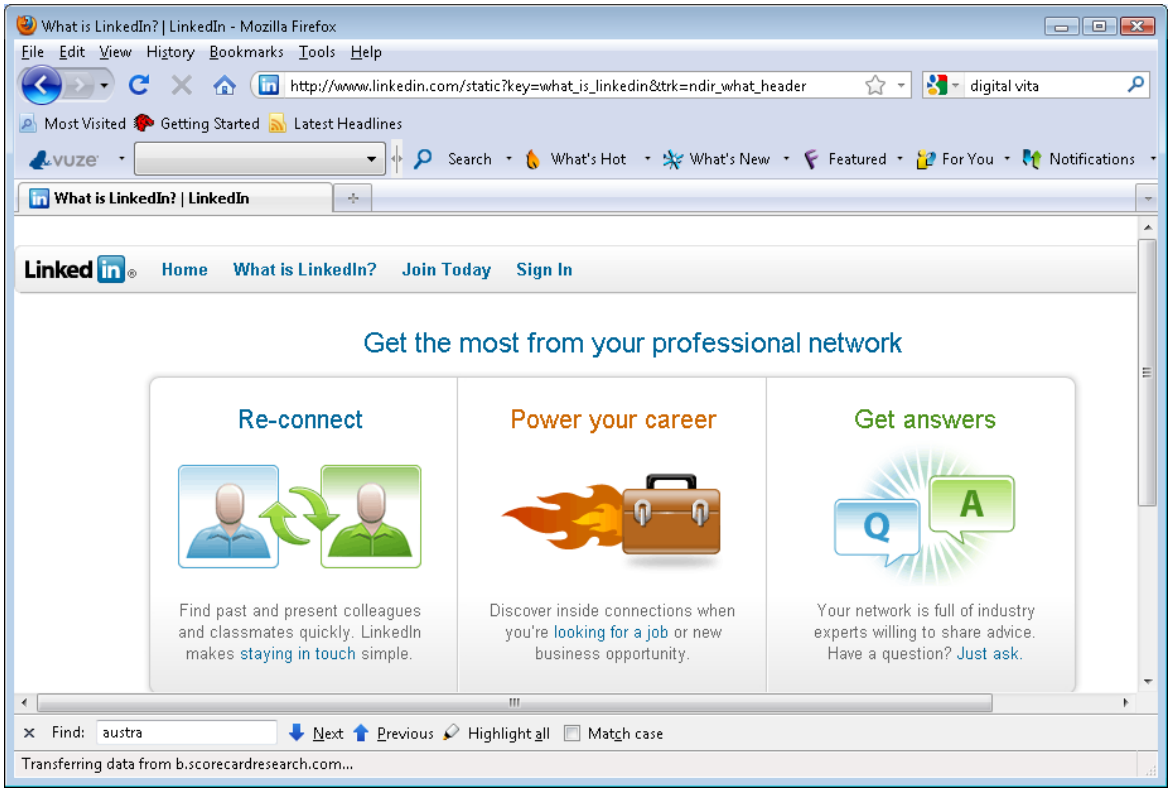


How many of you use

How many of you use



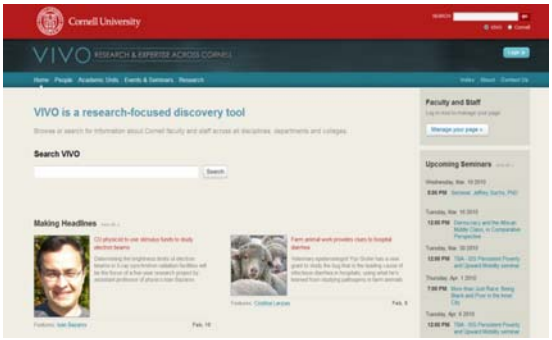
How many of you use



How many of you use FaceBook? What other social networking sites do you use?

VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Open data & code, accessible by anybody to continually improve and upgrade its quality and utility for the whole scientific community.



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.

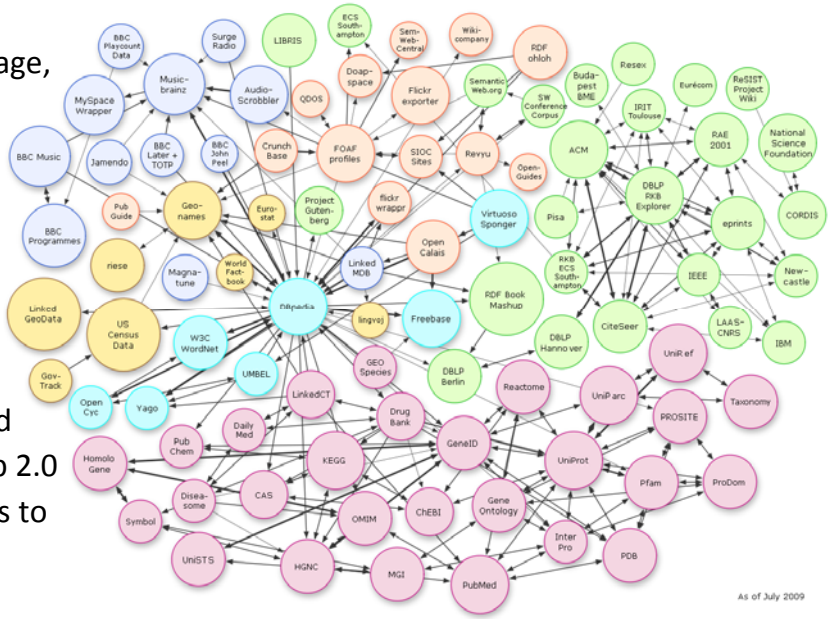
VIVO & Linked Open Data

2010 National VIVO Conference August 12&13, NYC
http://conferences.dce.ufl.edu/vivo

VIVO makes high coverage, high quality data from systems of record

- available online
- for free, and
- in machine readable format.

VIVO ontology is aligned with many existing Web 2.0 and scholarly ontologies to ease interoperability.



http://www4.wiwiss.fu-berlin.de/bizer/pub/lod-datasets_2009-07-14_colored.png

[#11] VIVO National Researcher Networking

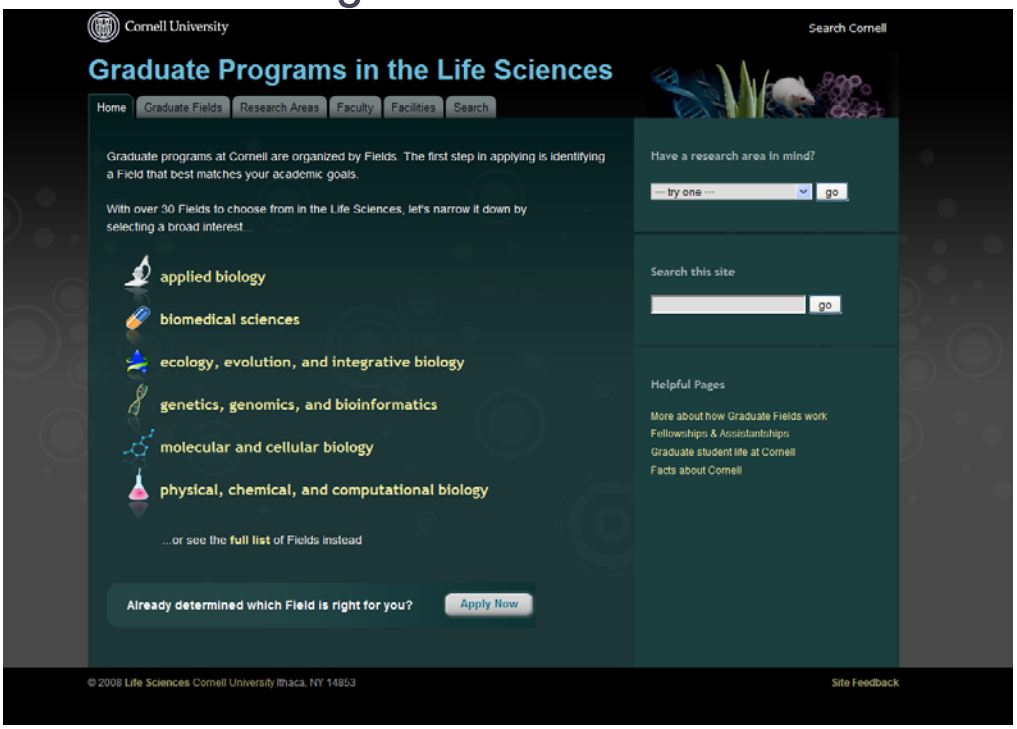
- Motivation
- Users, Their Needs, and Usage Scenarios
- Development
- Implementation
- Usage
- Outlook
- Exercise: Identify Promising VIVO Collaborations

VIVO
Enabling a National Network of Scientists

VIVO Users and Needs

- Faculty/Researchers
  - Customize profile created via feeds; find potential collaborators, "people like me"; discovery via high search rankings; info on activity of colleagues…
- Students
  - Create profiles; easily find mentors + collaborators; locate facilities, events, funding opportunities…
- Administrators
  - Quickly find cross-disciplinary expertise (research area; geography); centralize public data from diverse sources; easily repurpose information for consumers; improve faculty collaboration within or across departments and institutions…
- Funding, donor, legislative agencies
  - Discover projects, grants, expertise (e.g. for review panels; targets for funding)…
- General public
  - Find expertise, learn about research in a region/institution…

VIVO Web Pages

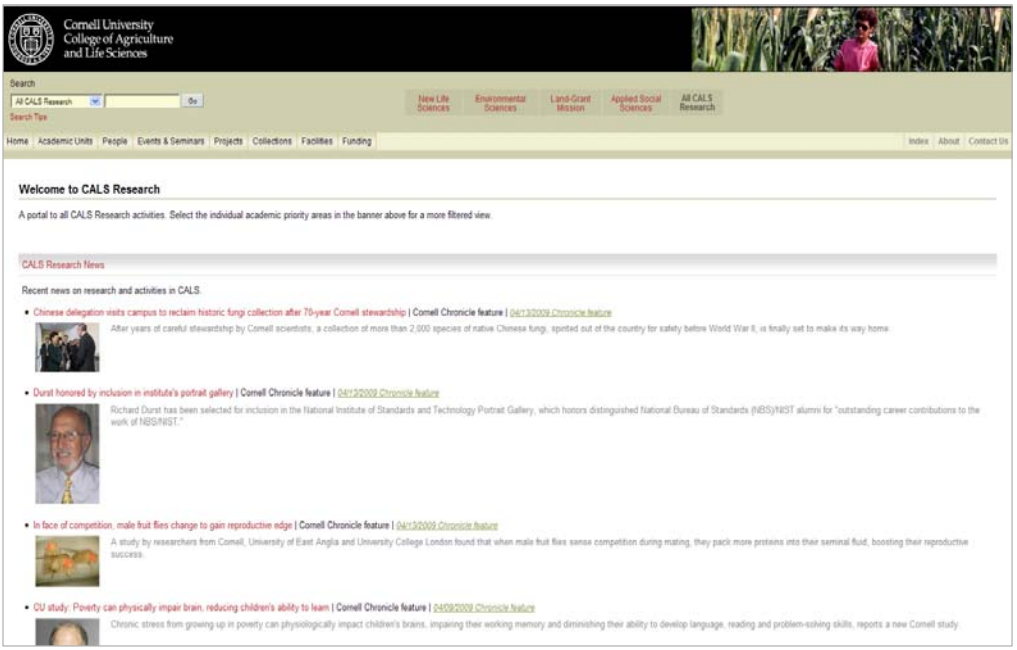


http://gradeducation.lifesciences.cornell.edu

VIVO Web Pages

http://cals-experts.mannlib.cornell.edu/

VIVO Web Pages



http://research.cals.cornell.edu

[#11] VIVO National Researcher Networking

- Motivation
- Users, Their Needs, and Usage Scenarios
- **Development**
- Implementation
- Usage
- Outlook
- Exercise: Identify Promising VIVO Collaborations

VIVO's Three Functional Layers

end users

curators

ontology editing
& data flow

Local Data Flow

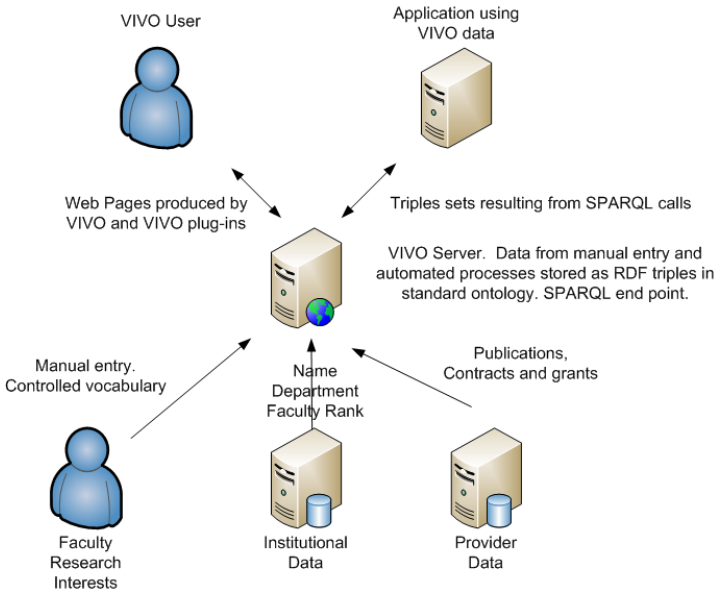
Peoplesoft
Grants DB
PubMed
Publishers

RDFa
RDF harvest
SPARQL endpoint

Researchers
Librarians
Administrative Staff
Self-Editors

Institutional Architecture

- Three sources of VIVO information
  - User data
  - Institutional data
  - Provider data
- Two formats for output
  - Web Pages for users
  - Resource Description Framework for applications

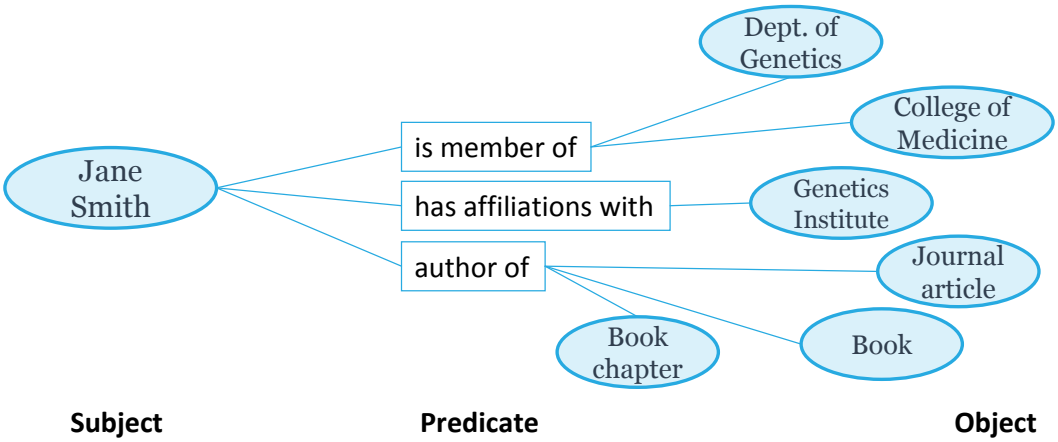


VIVO Data Providers & Users

- Eagle-i ("enabling resource discovery" U24 award)
- Federal agencies – NIH (NIH RePORTER), NSF, USDA, …
- Search Providers – Google, Bing, Yahoo, …
- Professional Societies – AAAS, …
- Publishers/vendors – PubMed, Elsevier, Collexis, ISI…
- Semantic Web community – DERI, …
- Consortia of schools – SURA, CTSA…
- Producers, consumers of semantic web-compliant data

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.



Advantages of an Ontology Approach

Provides the key to meaning

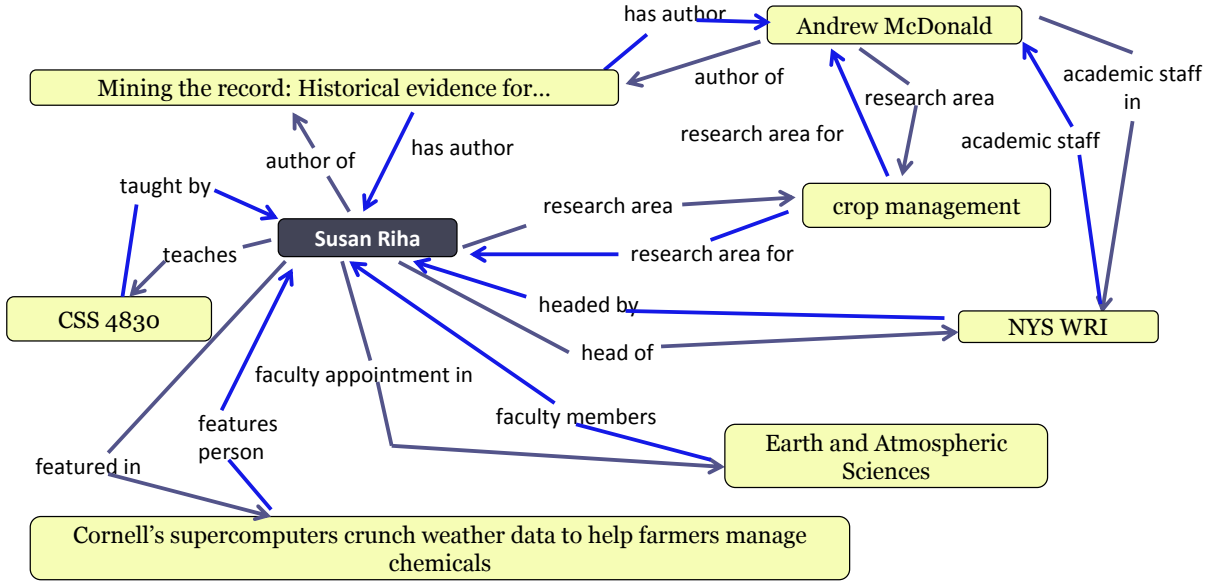
- Defines a set of *classes* and *properties* in a unique namespace
- Embedded as RDF so data becomes self-describing
- Definitions available via the namespace URI

Helps align RDF from multiple sources

- VIVO core ontology maps to common shared ontologies organized by domain
- Local extensions roll up into VIVO core

Data Representation Using RDF Triples

Detailed relationships for a researcher at Cornell U.
Open source code (BSD) and ontology available at http://vivoweb.org.

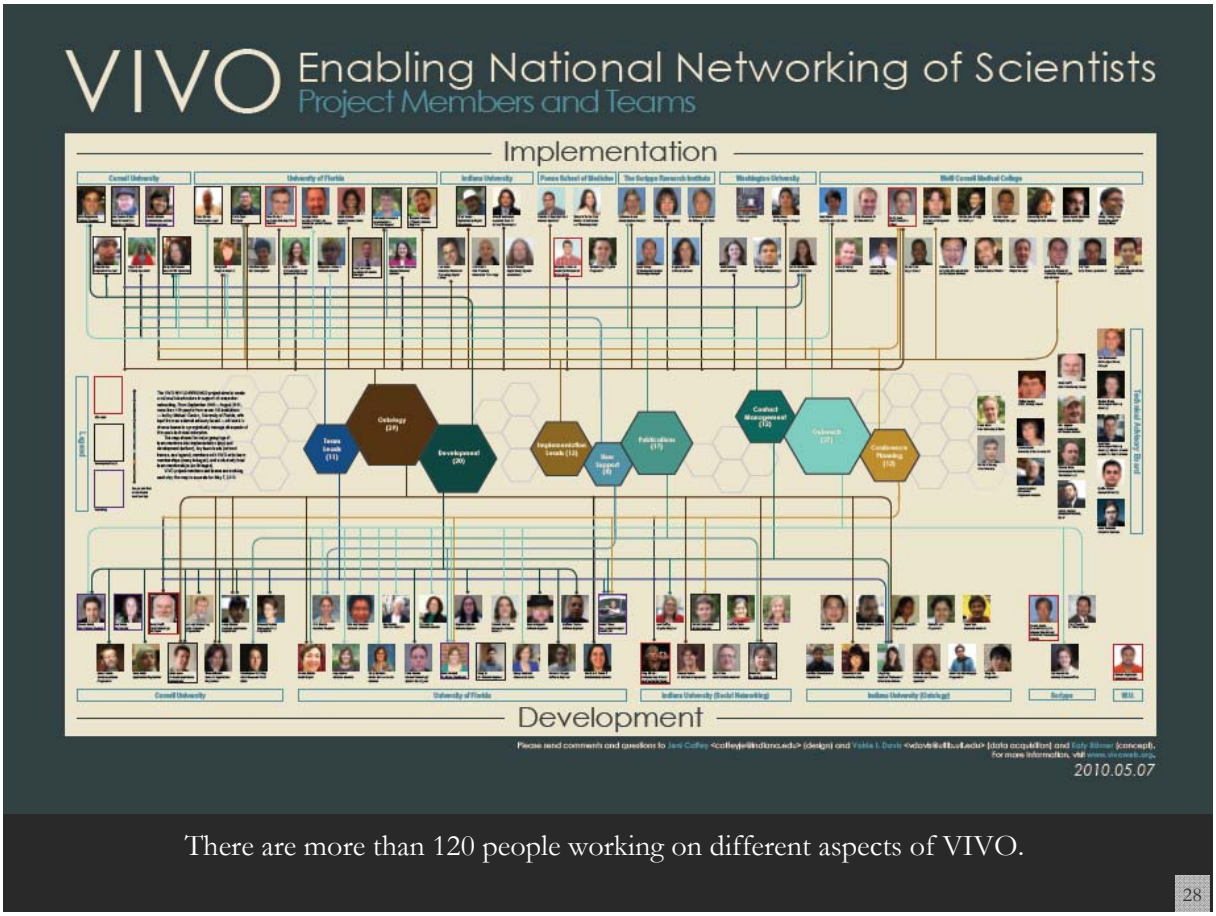


Query and explore

- 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)

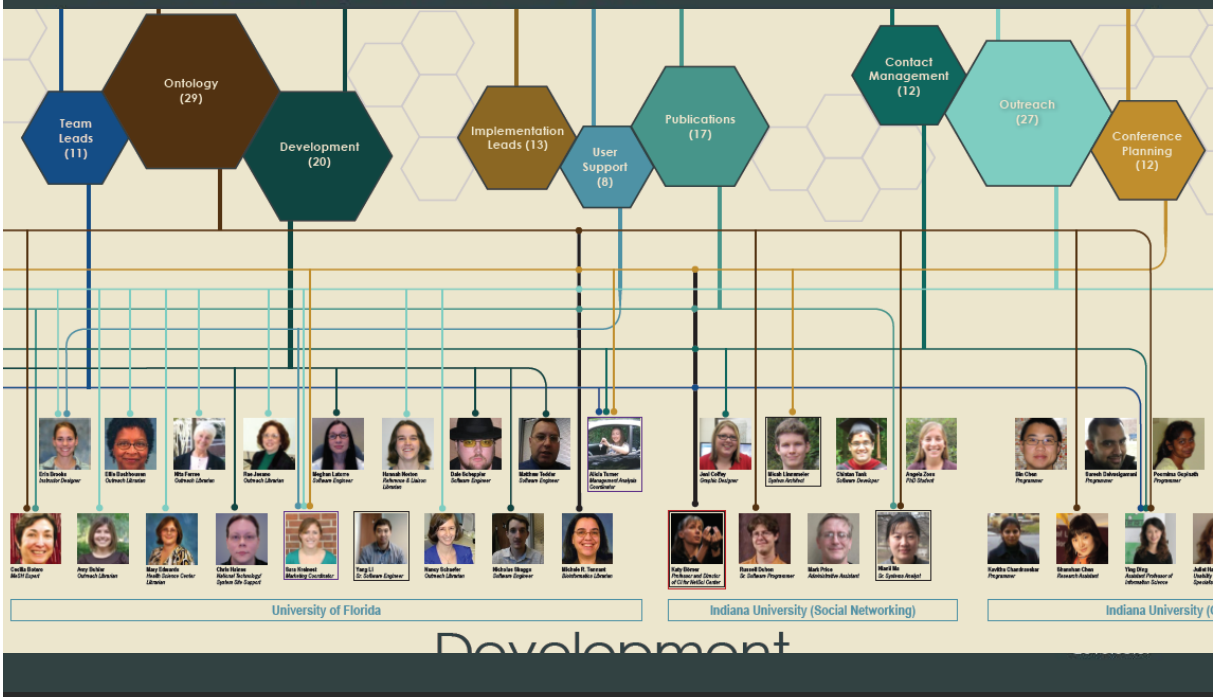
[#11] VIVO National Researcher Networking

- Motivation
- Users, Their Needs, and Usage Scenarios
- Development
- **Implementation**
- Usage
- Outlook
- Exercise: Identify Promising VIVO Collaborations



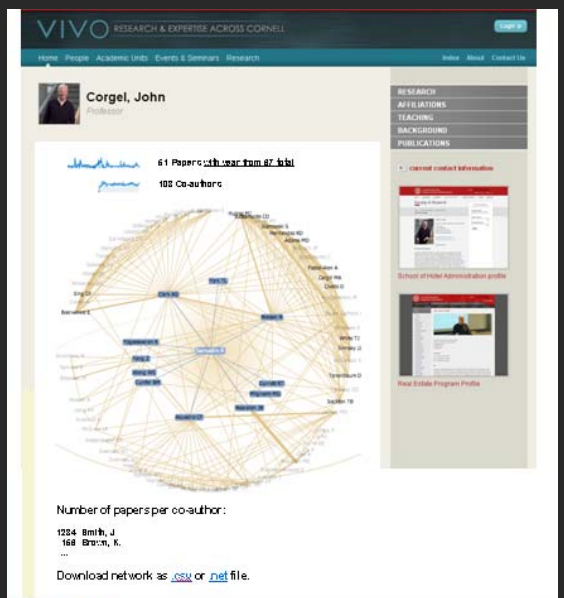
There are more than 120 people working on different aspects of VIVO.

VIVO Enabling National Networking of Scientists
Project Members and Teams

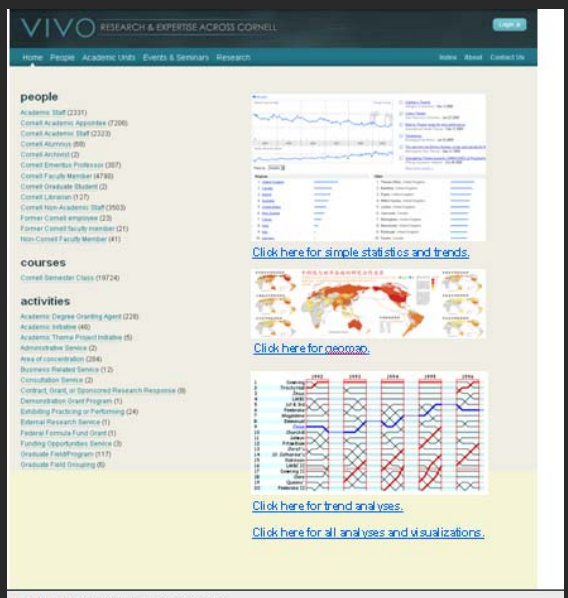


The subsequent slides focus on the Social Networking subproject.

VIVO Initial Drafts of Individual Level and Institution Level Visualizations

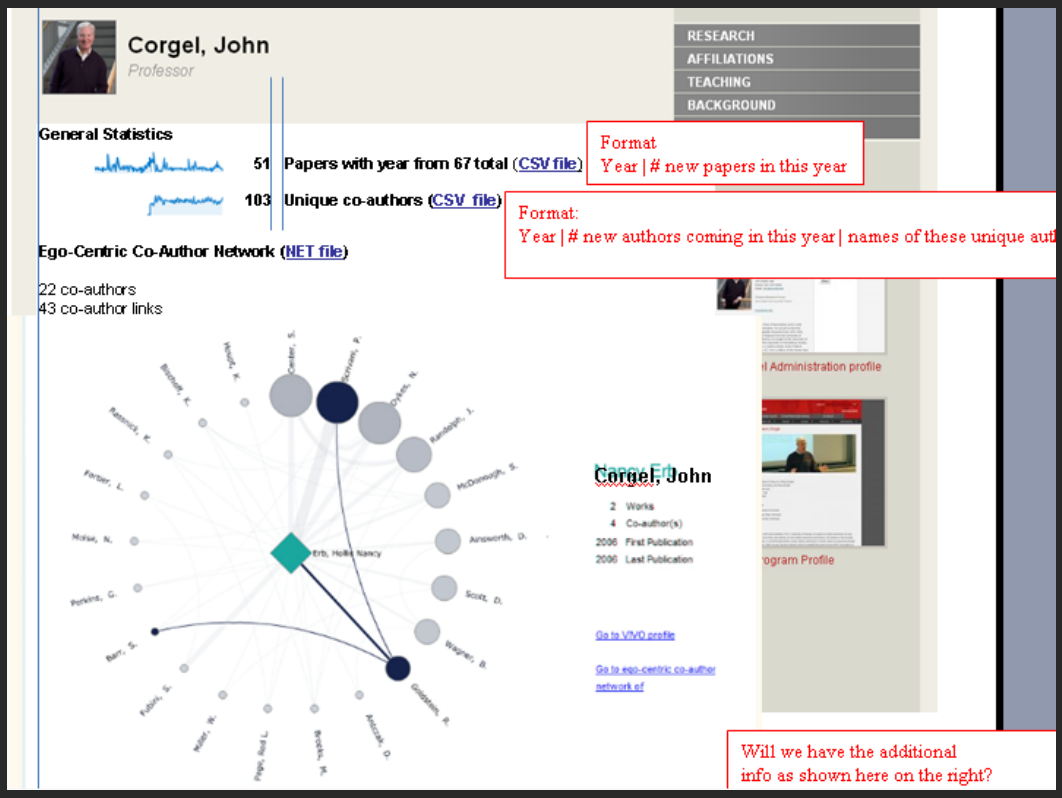


Individual Level VIVO Visualization – Co-Author Network
Simple co-author network for this person using code similar to http://www.chrisharrison.net/projects/clusterhull/index.html, tabular data, link to csv or network file for further processing in NWB Tool.
Sparklines covers all years in which publications exist but at least 10 years. Both are aligned right (today). Same max number of papers/co-authors.
Only local data, here Cornell, is considered.



Institution Level VIVO Visualization
Available from the VIVO Index page are statistics such as publications/funding/courses, # of linkages, e.g., co-author, paper-citation, paper-author, etc., # downloads over time are plotted, geo and science map overlays are generated, network layouts with well defined basemaps, e.g., bump charts or two lists of nodes in a bimodal network are written into a .ps file, converted into a .pdf file and are available for print. Temporal animation of growth corresponds to multiple pages (one per year) with identical reference system.
Only local data, here Cornell, is considered. We would like to know what vis users request.

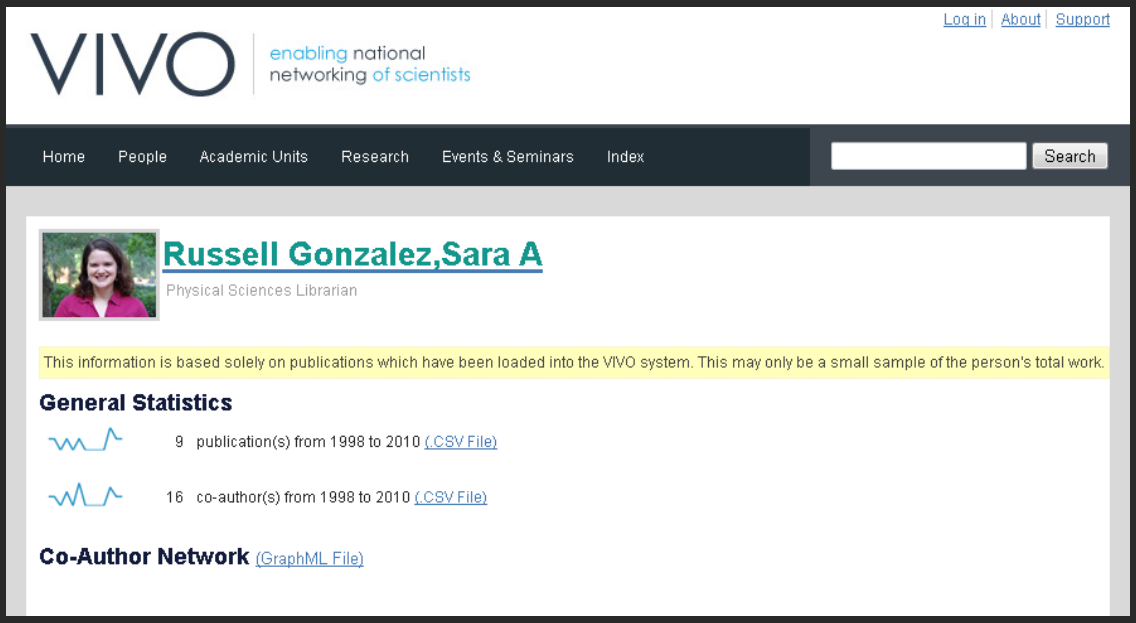
VIVO Later Draft of Individual Level Co-Author Visualization



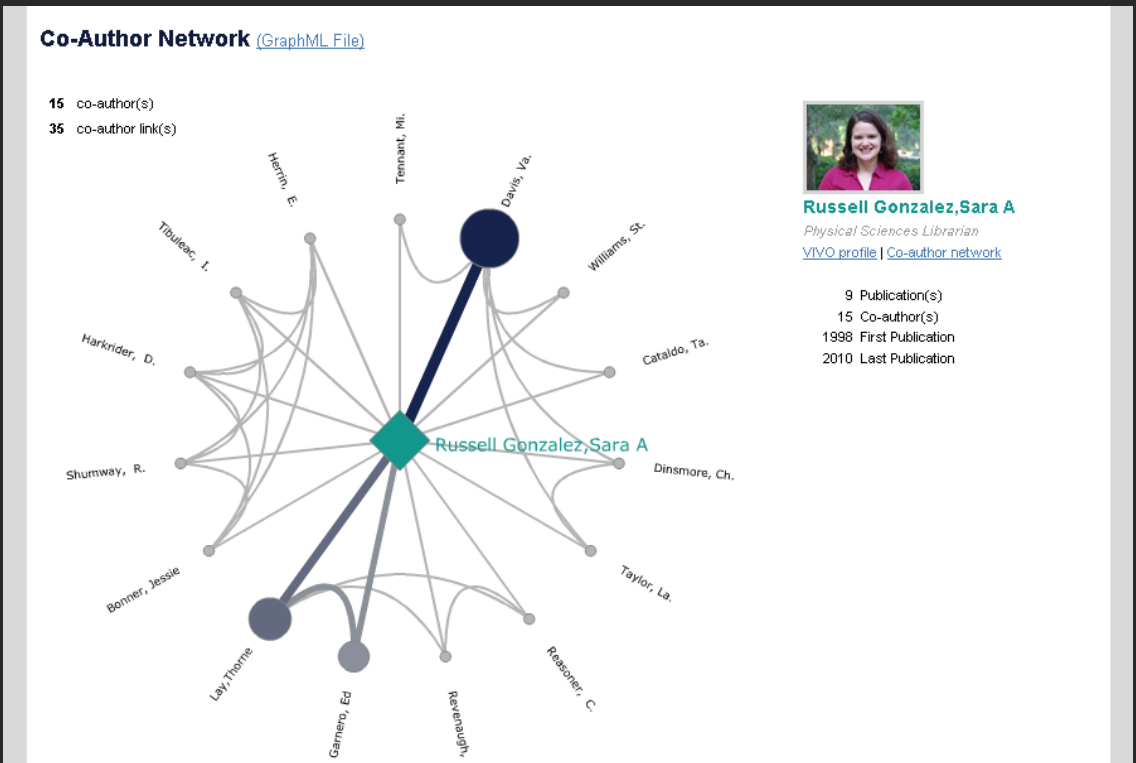
[#11] VIVO National Researcher Networking

- Motivation
- Users, Their Needs, and Usage Scenarios
- Development
- Implementation
- Usage
- Outlook
- Exercise: Identify Promising VIVO Collaborations

VIVO Release 1 v. 1.1: Individual Level Co-Author Visualization

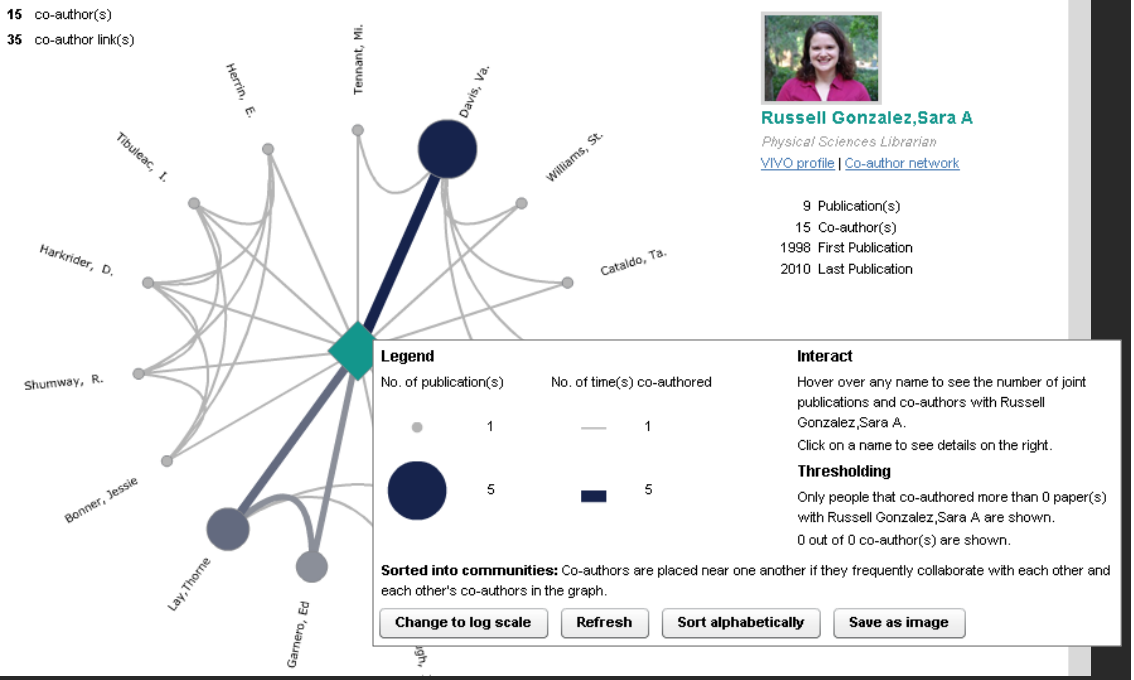


http://qa.vivo.ufl.edu/admin/visQuery?uri=http%3A%2F%2Fvivo.ufl.edu%2Findividual%2Fn22286&vis=person_level&render_mode=standalone



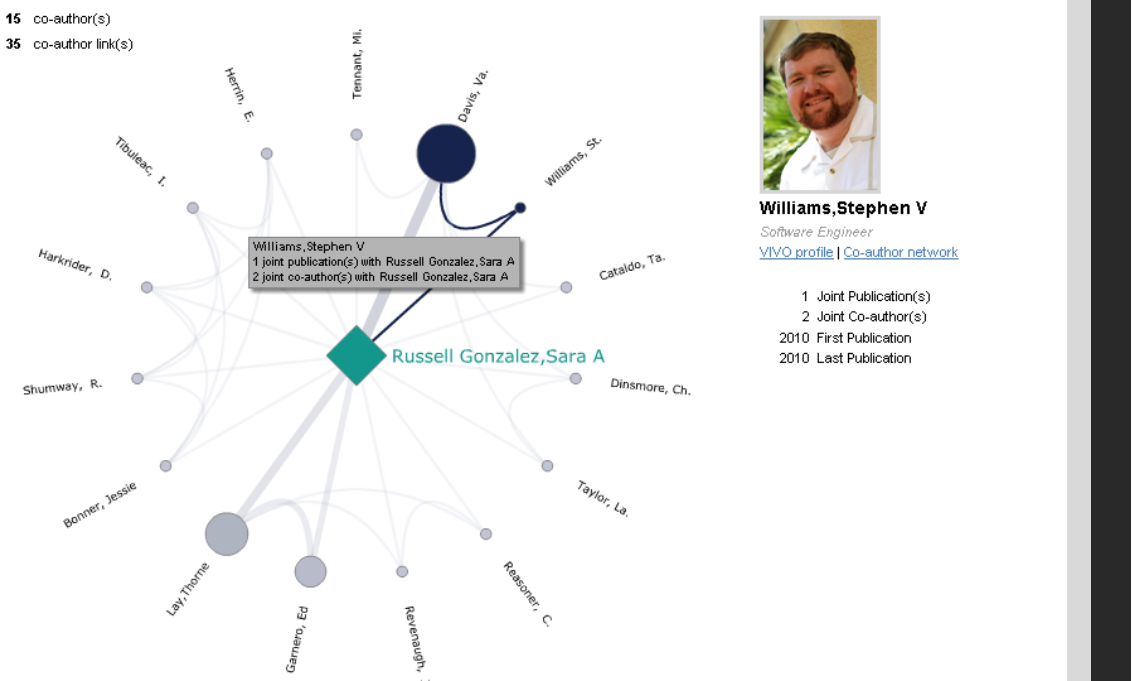
http://qa.vivo.ufl.edu/admin/visQuery?uri=http%3A%2F%2Fvivo.ufl.edu%2Findividual%2Fn22286&vis=person_level&render_mode=standalone

Co-Author Network (GraphML File)

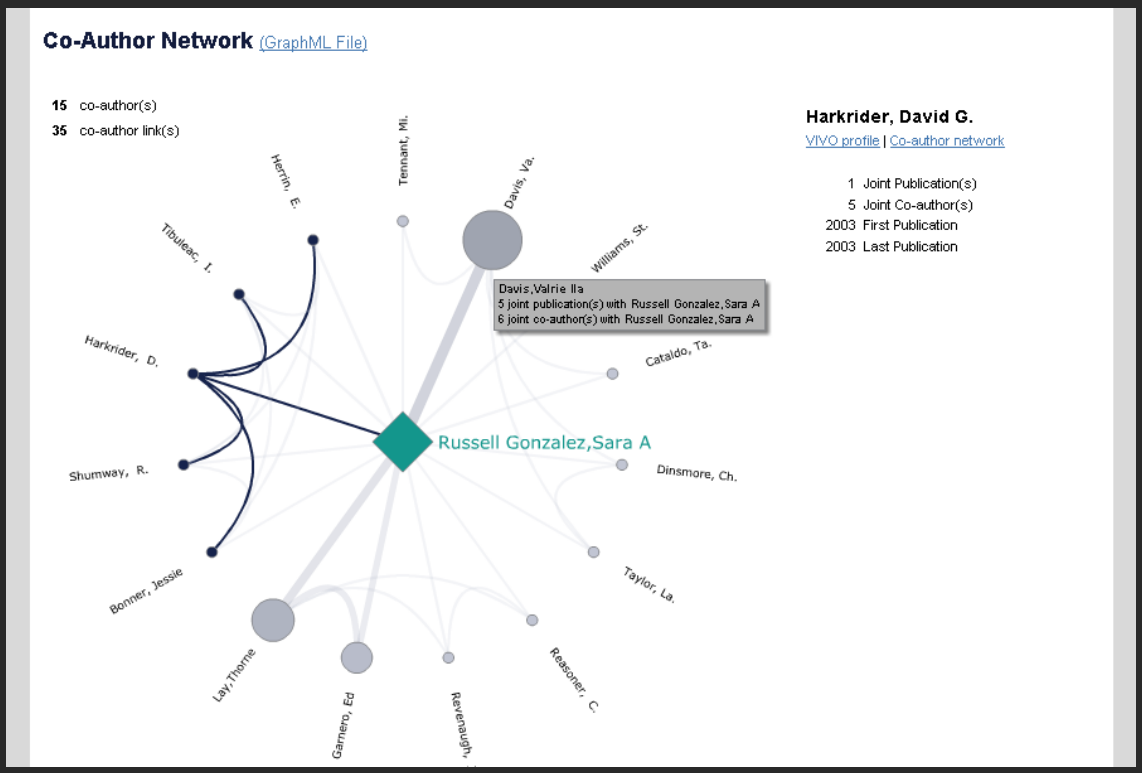


http://qa.vivo.ufl.edu/admin/visQuery?uri=http%3A%2F%2Fvivo.ufl.edu%2Findividual%2Fn22286&vis=person_level&render_mode=standalone

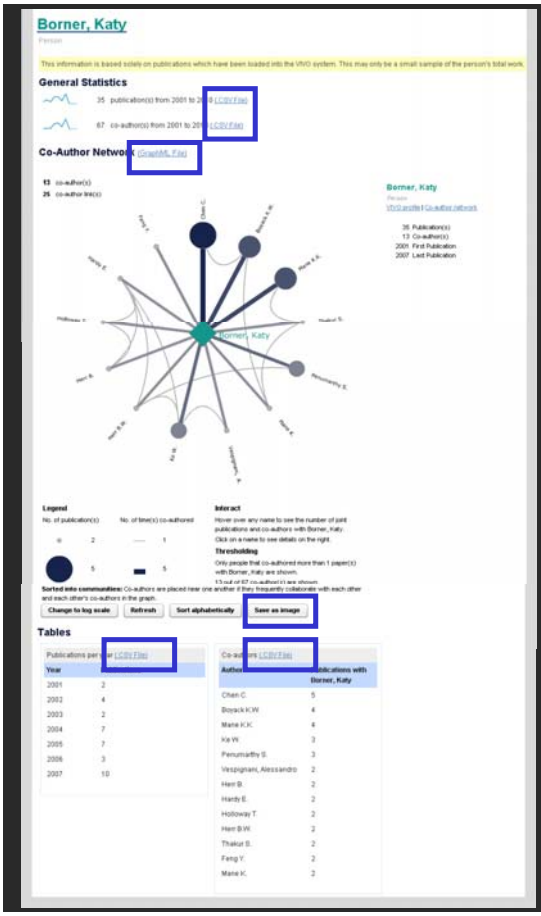
Co-Author Network (GraphML File)



http://qa.vivo.ufl.edu/admin/visQuery?uri=http%3A%2F%2Fvivo.ufl.edu%2Findividual%2Fn22286&vis=person_level&render_mode=standalone



http://qa.vivo.ufl.edu/admin/visQuery?uri=http%3A%2F%2Fvivo.ufl.edu%2Findividual%2Fn22286&vis=person_level&render_mode=standalone



Download Data

General Statistics

- 35 publication(s) from 2001 to 2010 (.CSV File)
- 67 co-author(s) from 2001 to 2010 (.CSV File)

Co-Author Network

(GraphML File)

Save as Image (.PNG file)

Tables

- Publications per year (.CSV File)
- Co-authors (.CSV File)

http://vivo-vis.slis.indiana.edu/vivo1/admin/visQuery?uri=http%3A%2F%2Fvivoweb.org%2Fontology%2Fcore%2FPerson72c&vis=person_level&render_mode=standalone

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

Year	Publications
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10

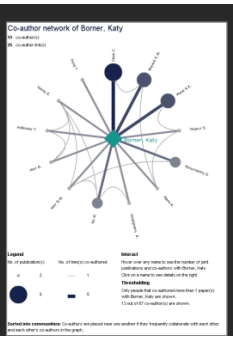
67 co-author(s) from 2001 to 2010 (.CSV File)

Year	"Number	"Co-Author(s)"
2001	1	Chen C.
2002	3	Feng Y.; McMahon T.; Chen C.
2003	2	Chen C.; Boyack K.W.
2004	17	Lee G.J.; Jones R.J.; Martins, Emilia; Moor K.A.; Menzel S.; Sooriamurthi R.; Ord T.; Boyack K.W.; Stamm S.;

Co-author network (GraphML File)

```
<?xml version="1.0" encoding="UTF-8"?>
    <graphml xmlns="http://graphml.graphdrawing.org/xmlns"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
  http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
<key id="label" for="node" attr.name="label" attr.type="string" />
<key id="number_of_authored_works" for="node" attr.name="number_of_authored_works" attr.type="int" />
<key id="num_unknown_publication" for="node" attr.name="num_unknown_publication" attr.type="int" />
<key id="num_latest_publication" for="node" attr.name="num_latest_publication" attr.type="int" />
<key id="latest_publication" for="node" attr.name="latest_publication" attr.type="int" />
<key id="profile_url" for="node" attr.name="profile_url" attr.type="string" />
```

Save as Image (.PNG file)

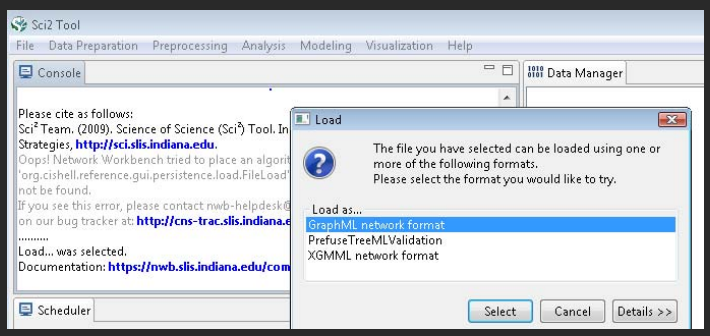


Publications per year (.CSV File), see top file.

Co-authors (.CSV File)

Co-Author	"Count"
Andrienko G.	1
Andrienko N.	1
Ben-Miled Z.	1
Blackwell A.	1
Boyack K.W.	4
Bozicevic M.	1
Brodbeck D.	1
Burkhard R.A.	1
Chen C.	5

Co-Author Network (GraphML File)

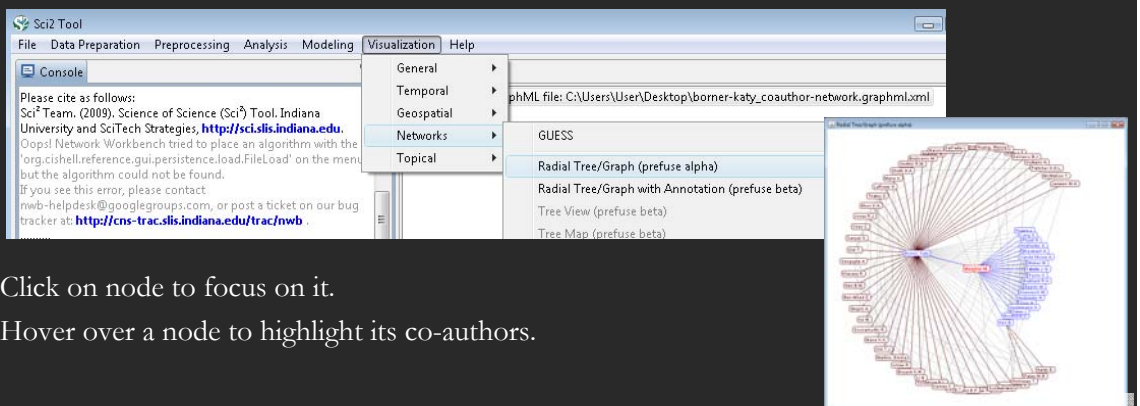


Network Analysis Toolkit
Nodes: 68
Edges: 299

Average degree: 8.7941
Density (disregarding weights): 0.1313

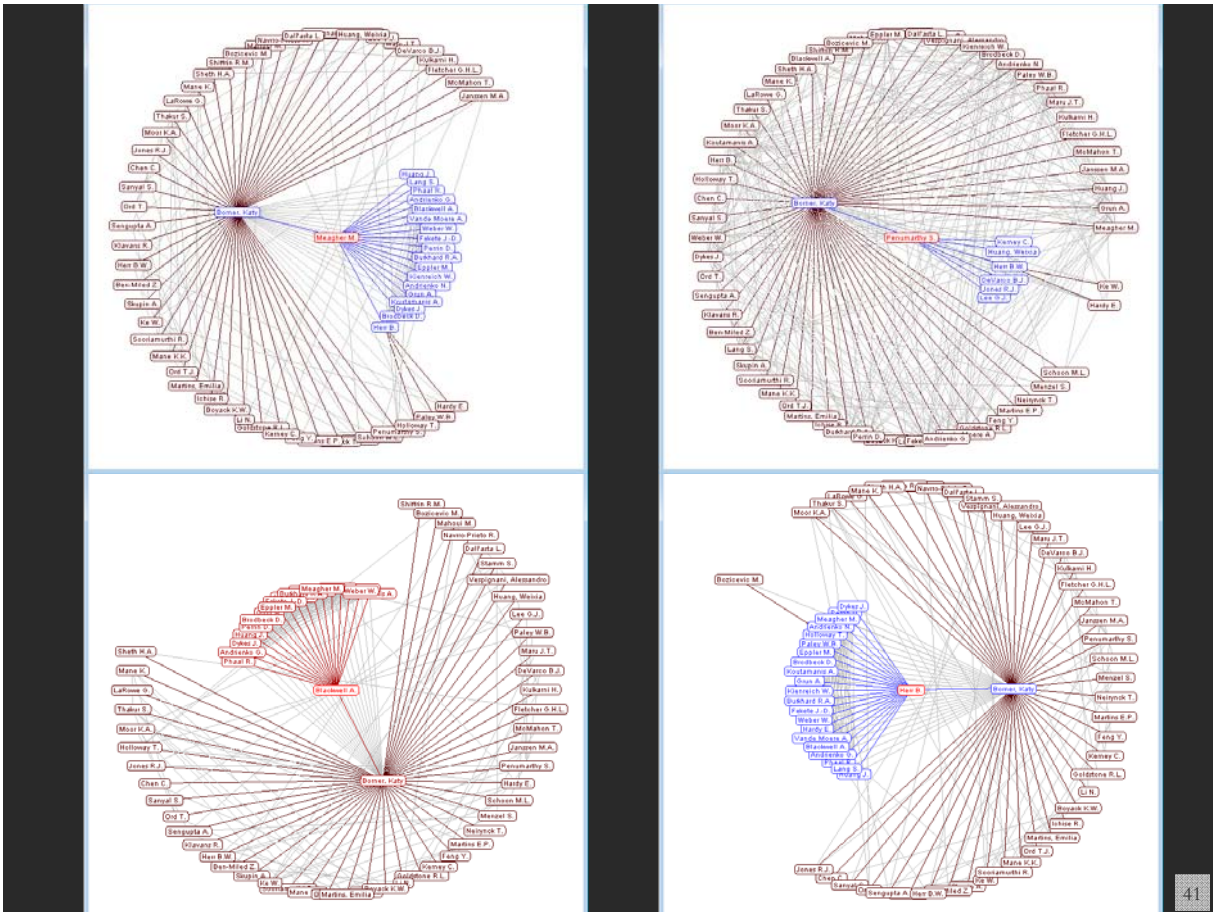
Additional Densities by Numeric Attribute
densities (weighted against standard max)
num_earliest_collaboration: 0.1343
num_latest_collaboration: 0.1335
number_of_coauthored_works: 0.1409
num_unknown_collaboration: -0.1313
earliest_collaboration: 263.3332
latest_collaboration: 263.3398

Visualize the file using Radial Graph layout.



Click on node to focus on it.

Hover over a node to highlight its co-authors.



[#11] VIVO National Researcher Networking

- Motivation
- Users, Their Needs, and Usage Scenarios
- Development
- Implementation
- Usage
- **Outlook**
- Exercise: Identify Promising VIVO Collaborations

VIVO Institution Level Visualizations

Institution Level VIVO Visualization
Available from the VIVO Index page are statistics such as publications/funding/courses, # of linkages, e.g., co-author, paper-citation, paper-author, etc., # downloads over time are plotted, geo and science map overlays are generated, network layouts with well defined basemaps, e.g., bump charts or two lists of nodes in a bimodal network are written into a .ps file, converted into a .pdf file and are available for print. Temporal animation of growth corresponds to multiple pages (one per year) with identical reference system.
Only local data, here Cornell, is considered. We would like to know what vis users request.

Institution Level analyses and visualization will be available from the VIVO Index page and comprise statistics such as

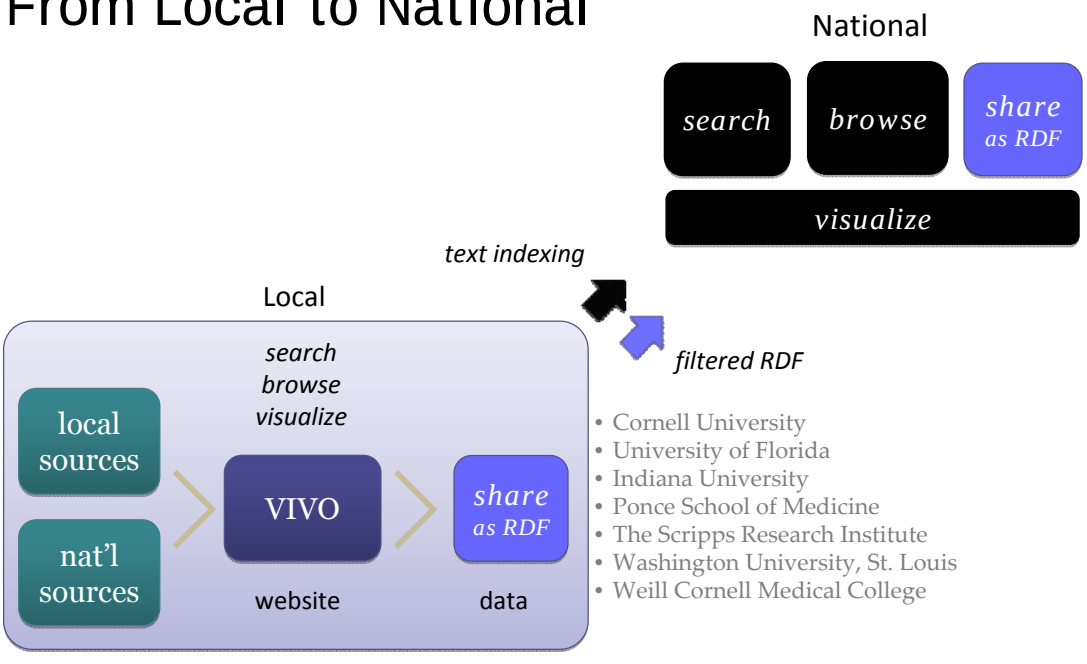
- publications/funding/courses,
- # of linkages, e.g., co-author,
- paper-citation, paper-author, etc.,
- # downloads over time are plotted.

Geospatial and science map overlays as well as network layouts with well defined base maps, e.g., two lists of nodes in a bimodal network will be written into a PDF file for viewing and printing.

Temporal animation of growth corresponds to multiple pages (one per year) with identical reference system.

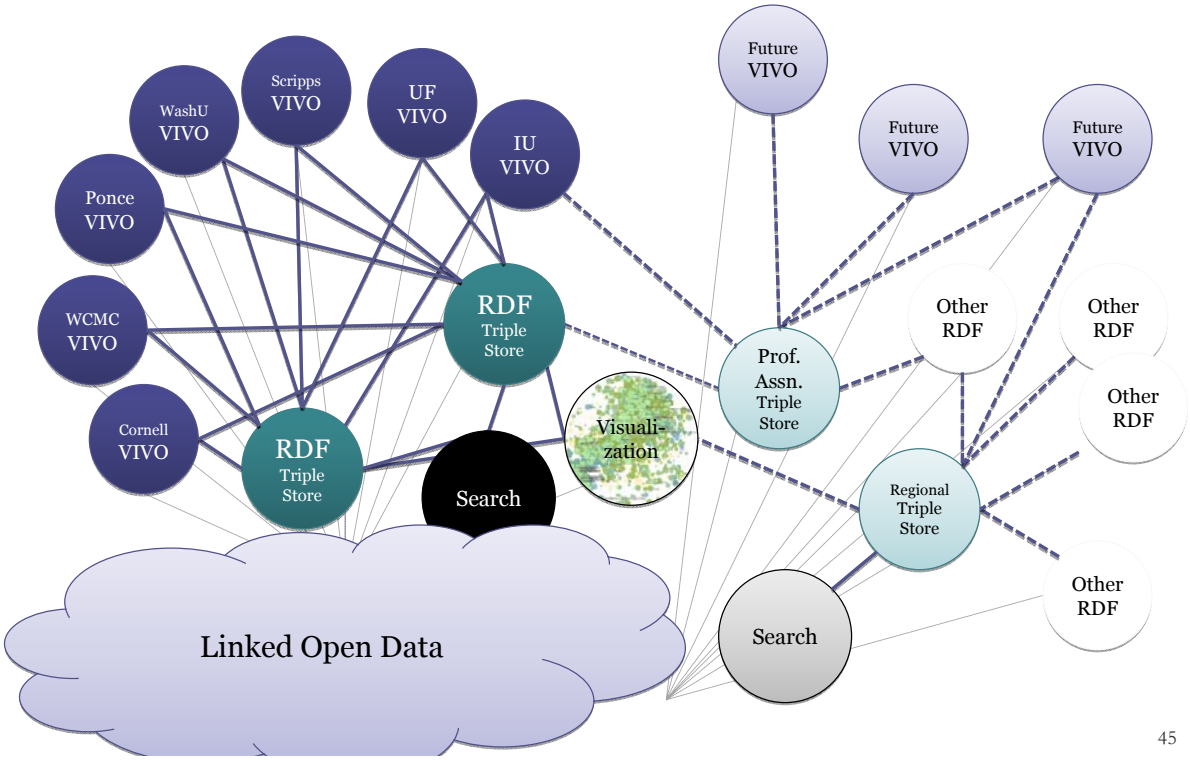
From Local to National

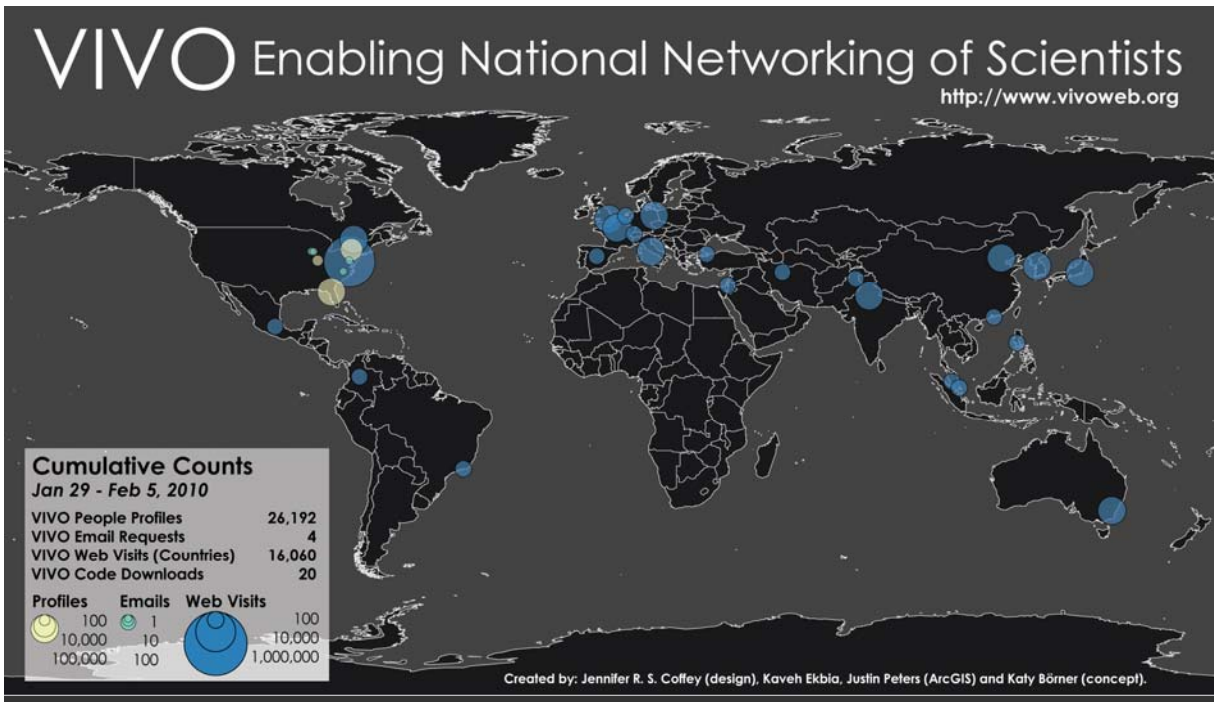
VIVO
Enabling a National Network of Scientists



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

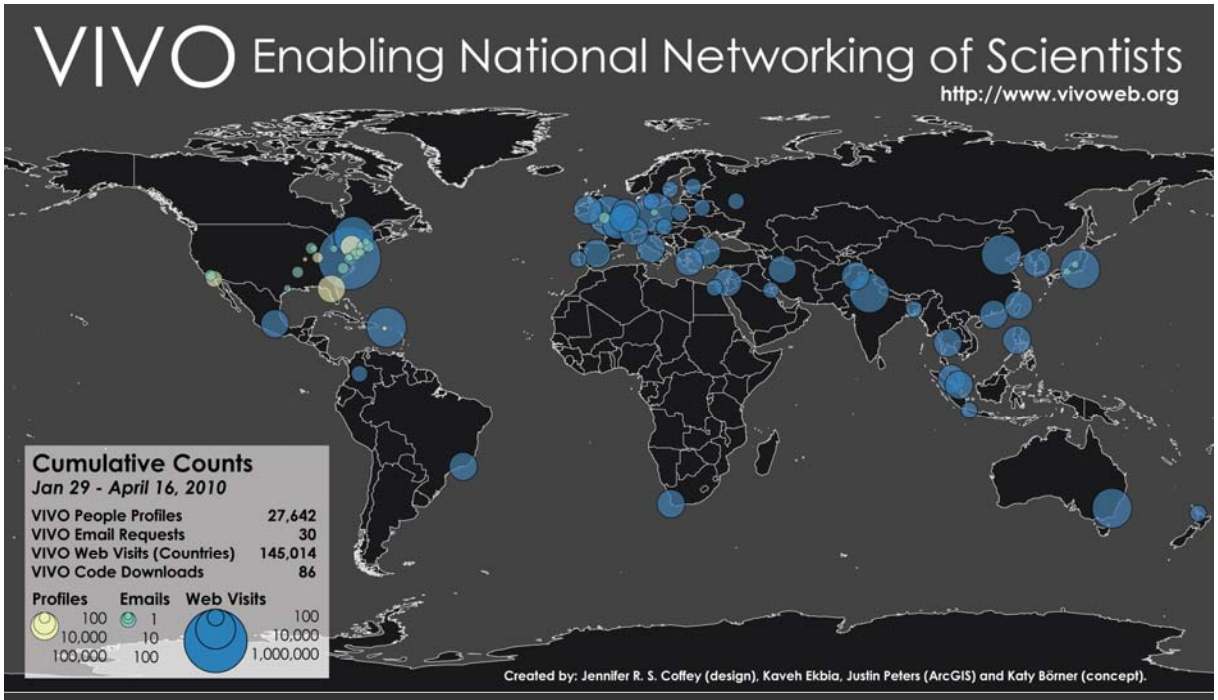
National Networking





The National Research Network: VIVO: Enabling National Networking of Scientists NIH U24RR029822
Start: Sept 2009
PI: Michael Conlon, University of Florida
Award amount: $12,300,000

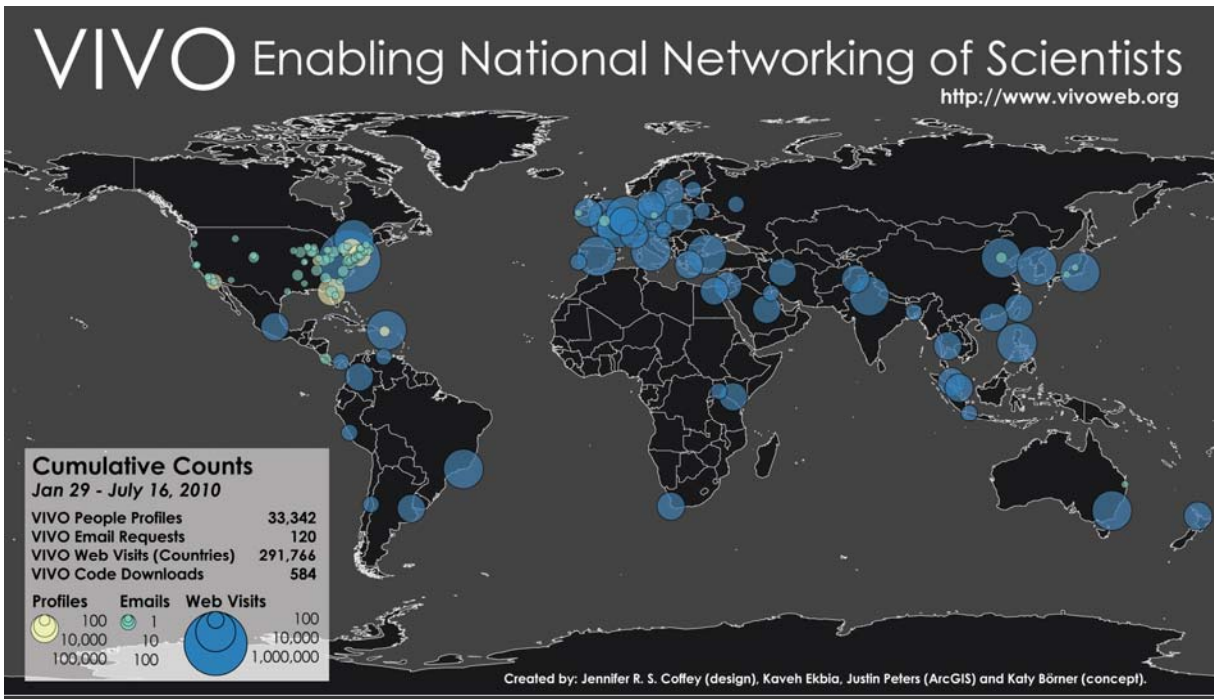
DRAFT



Shown are the
- Number of people profiles in the 7 different installation sites.
- Email contacts by data and service providers as well as institutions interested to adopt VIVO.
- The number of visitors on http://vivoweb.org

Circles are area size coded using a logarithmic scale.

DRAFT



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.

DRAFT

Email this

Register Online

CONTENTS

- Welcome
- Preliminary Agenda
- Program Committee
- Keynote Speakers
- Conference Speakers
- Sponsorship
- Call for Submissions
- Workshops and Tutorials
- Hotel & Transportation Information
- New York Hall of Science
- Map of Conference Venue
- Conference Save the Date

REGISTRATION & LOGISTICS:
Andy Campbell
tel:
(352) 294-0847
fax: (352) 392-5437

VIVO | Enabling National Networking of Scientists

Welcome

Workshops are filling up fast – reserve your seat now!

The first annual VIVO National Conference is on August 12 & 13, 2010 at the New York Hall of Science.

The VIVO National Conference, Enabling National Networking of Scientists, will bring together scientists, developers, publishers, funding agencies, research officers, students and those supporting the development of team science.

This two-day conference will begin with workshops and tutorials for those new to VIVO, those implementing VIVO at their institutions, and those wishing to develop applications using VIVO. Keynote speakers (to be announced) will present regarding the Semantic Web, Linked Open Data and the role of VIVO in support of team science. Panelists will discuss adoption and implementation findings. Feedback sessions will engage participants in requirements gathering and brainstorming regarding future network services. Presenters will discuss mapping, social networking, crowd sourcing, support for societies and other national network applications.

For more information about the VIVO project and national networking, please visit http://vivoweb.org

For content information about the VIVO conference, please click the following link: VIVO

http://conferences.dce.ufl.edu/vivo

Email this

Register Online

CONTENTS

- Welcome
- Preliminary Agenda
- Program Committee
- Keynote Speakers
- Conference Speakers
- Sponsorship
- Call for Submissions
- Workshops and Tutorials
- Hotel & Transportation Information
- New York Hall of Science
- Map of Conference Venue
- Conference Save the Date

REGISTRATION & LOGISTICS:

VIVO | Enabling National Networking of Scientists

Workshops and Tutorials

VIVO Data Analysis and Visualization Services: How to Program, Extend, and Utilize

Instructors: Micah Linnemeier, Chintan Tank, Nianli Ma, and Katy Börner
Cyberinfrastructure for Network Science Center, Indiana University

Through the VIVO project, high quality academic data from systems of record becomes available in a common format through Semantic Web technologies. Data that was previously difficult to access and combine becomes available to anyone, creating a unique opportunity for academic and industry stakeholders to utilize this data in conjunction with their own areas of expertise. The most compelling uses of VIVO data might come from 3rd-party developers creating analyses, applications, and services that meet their specific needs.

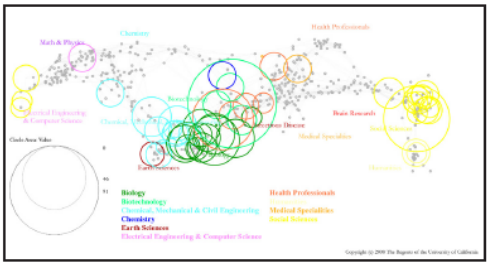
This hands-on workshop aims to empower participants to understand, access, and utilize VIVO data for administrative, commercial, or research purposes. It starts with a brief overview of techniques and workflows used to analyze and visualize temporal, geospatial, topical, and network datasets at a micro, meso, and macro level. Emphasis is on the design of insightful visualizations. Next, we will present the general VIVO architecture and explain and demonstrate different options to access and work with VIVO data and to use or extend VIVO code drawing on Indiana University's experience with VIVO service development. All services and applications are documented at a level of detail that makes it easy for others to replicate and extend them.

Main goal of the Social Networking team is to empower others to write useful VIVO applications and services.

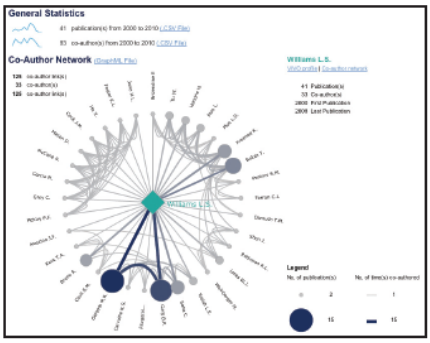
http://conferences.dce.ufl.edu/vivo

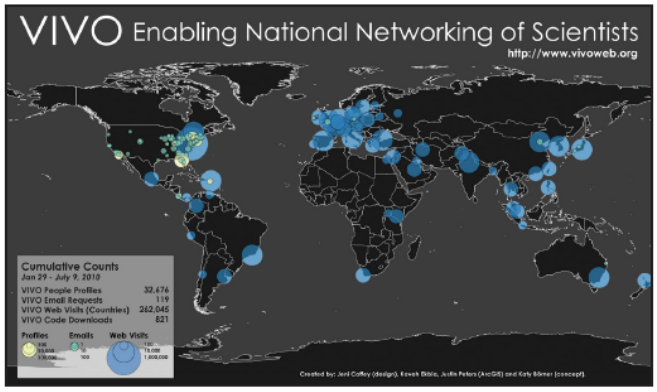
Last but not least, we will showcase different data analyses and visualizations of VIVO data at the individual, institution, and national level such as:

- **Individual level.** Statistics and ego-centric scholarly networks on VIVO Profile pages.
- **Institutional level.** Analyzes and visualizations of funding intake and publication output for departments and centers accessible via the VIVO Index page. Download of relevant data in tabular and network formats for further analysis using the Network Workbench tool.
- **National level.** Visualization of VIVO installations and their profile holdings together with web page access and general VIVO information requests. Plus, services that use VIVO URIs to access data across different VIVO instances.



The workshop concludes with a general question and answer session.





http://conferences.dce.ufl.edu/vivo

[#11] VIVO National Researcher Networking

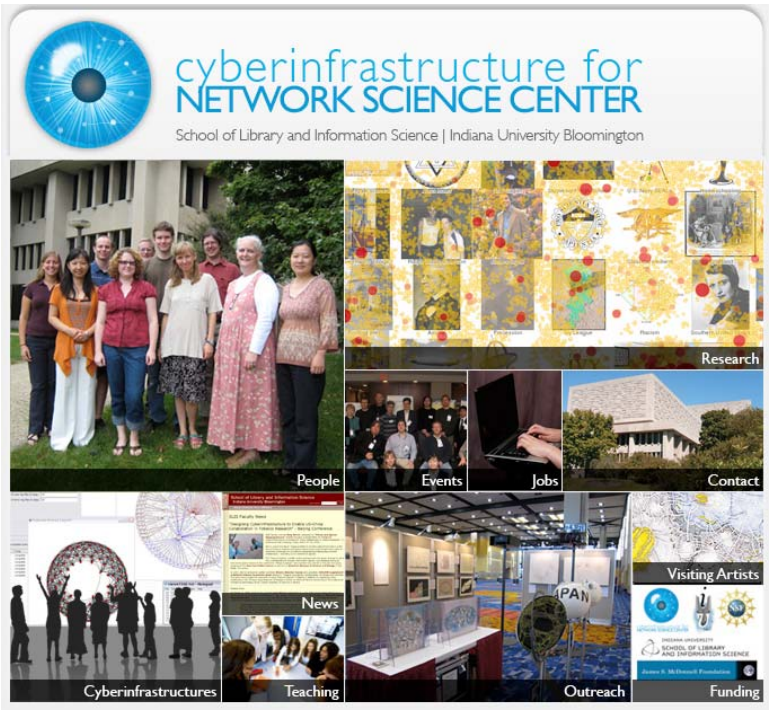
- Motivation
- Users, Their Needs, and Usage Scenarios
- Development
- Implementation
- Usage
- Outlook
- Exercise: Identify Promising VIVO Collaborations

Exercise

Please identify promising VIVO usages and/or collaborations.

Document it by listing

- Project title
- User, i.e., who would be most interested in the result?
- Insight need addressed, i.e., what would you/user like to understand?
- Data used, be as specific as possible.
- Analysis algorithms used.
- Visualization generated. Please make a sketch with legend.



All papers, maps, cyberinfrastructures, talks, press are linked from http://cns.slis.indiana.edu