VIVO

Enabling National Networking of Scientists

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 Enabling National Networking of Scientists

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.

VIVO Enabling National Networking of Scientists

Overview

- What is VIVO?
- How does it work?
- How do we implement it?
- What's ahead?

VIVO

Enabling National Networking of Scientists

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:

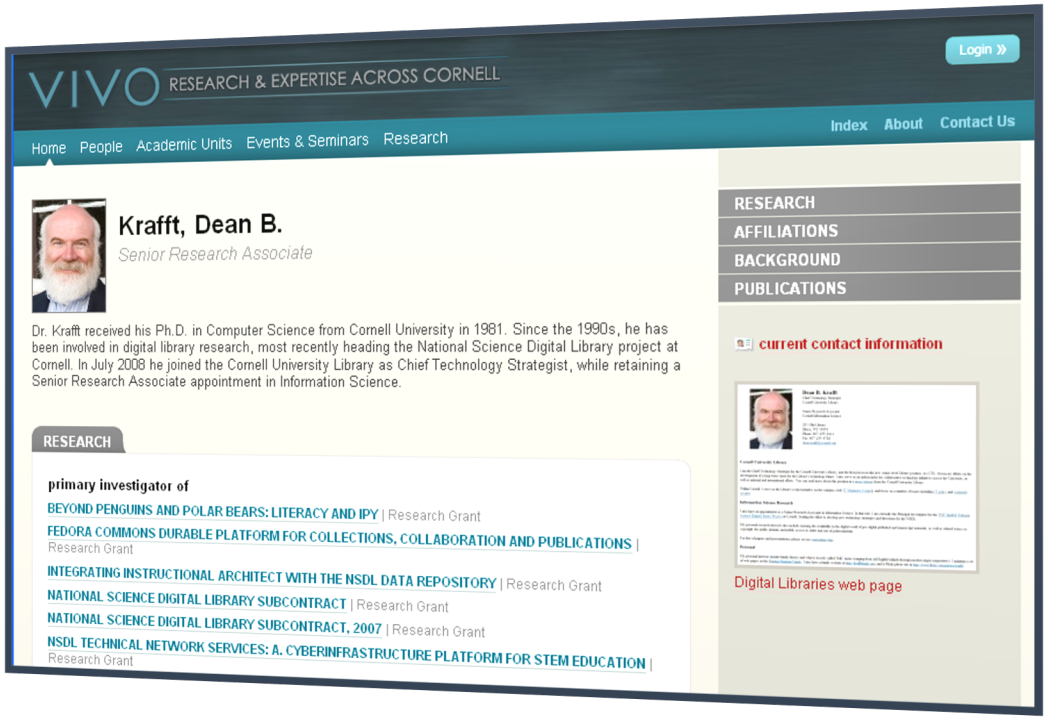
A **semantic web application** that enables the discovery of research and scholarship across disciplines in an institution.

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

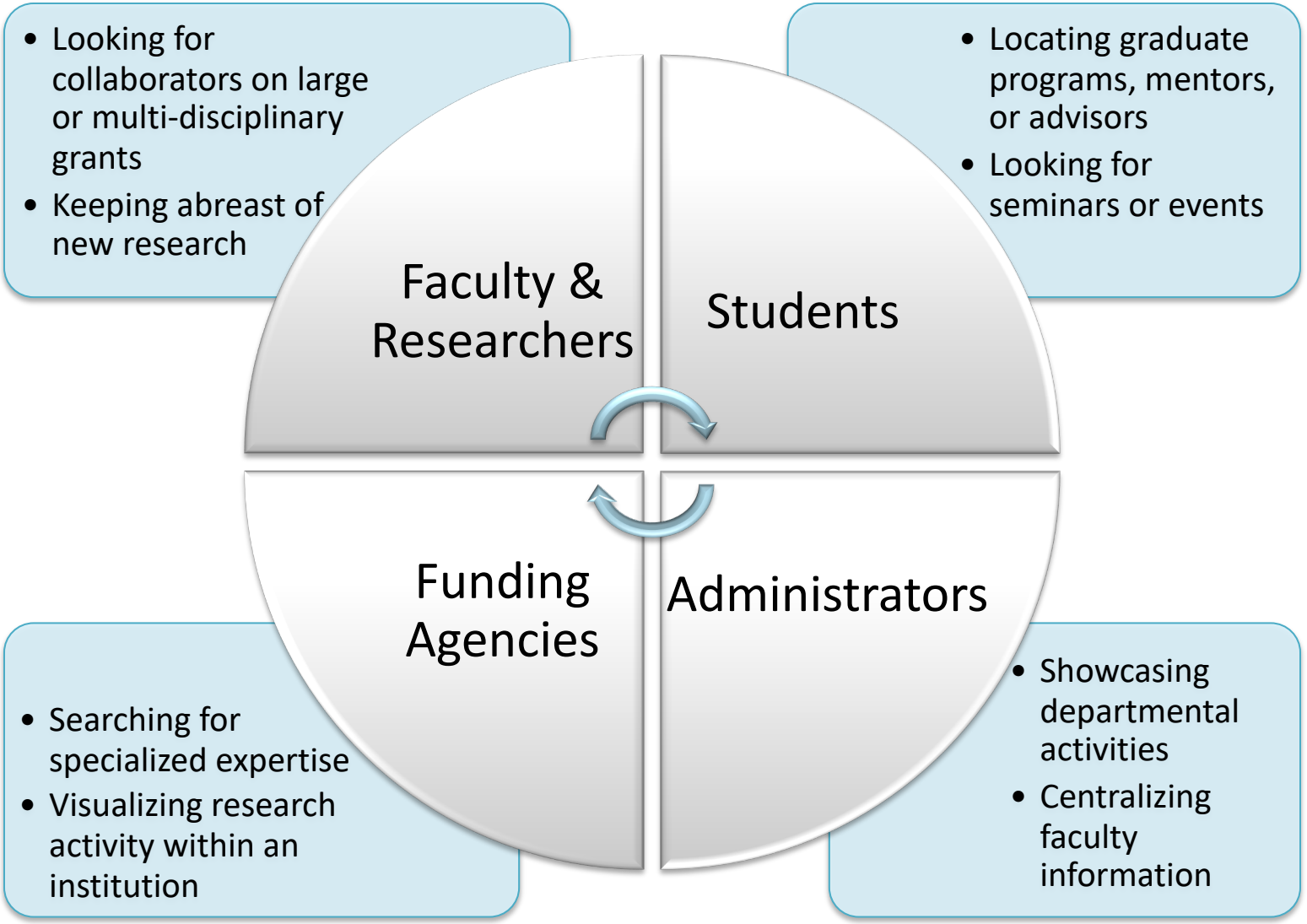
A **powerful search functionality** for locating people and information within or across institutions.

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.

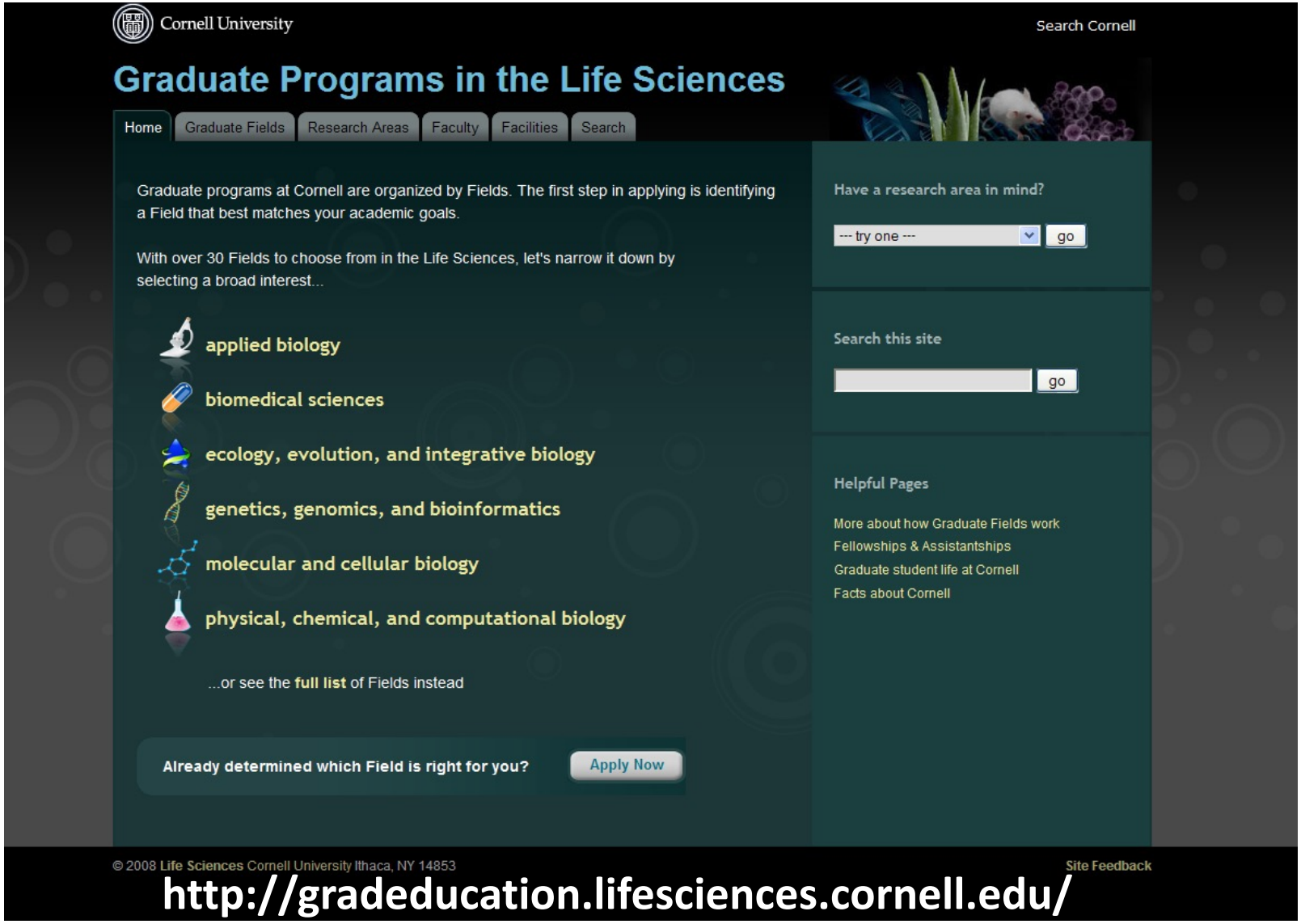


Who can use VIVO?



…and many more!

VIVO as disseminator



http://gradeducation.lifesciences.cornell.edu/

How does it work?

Data, Data, Data

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.

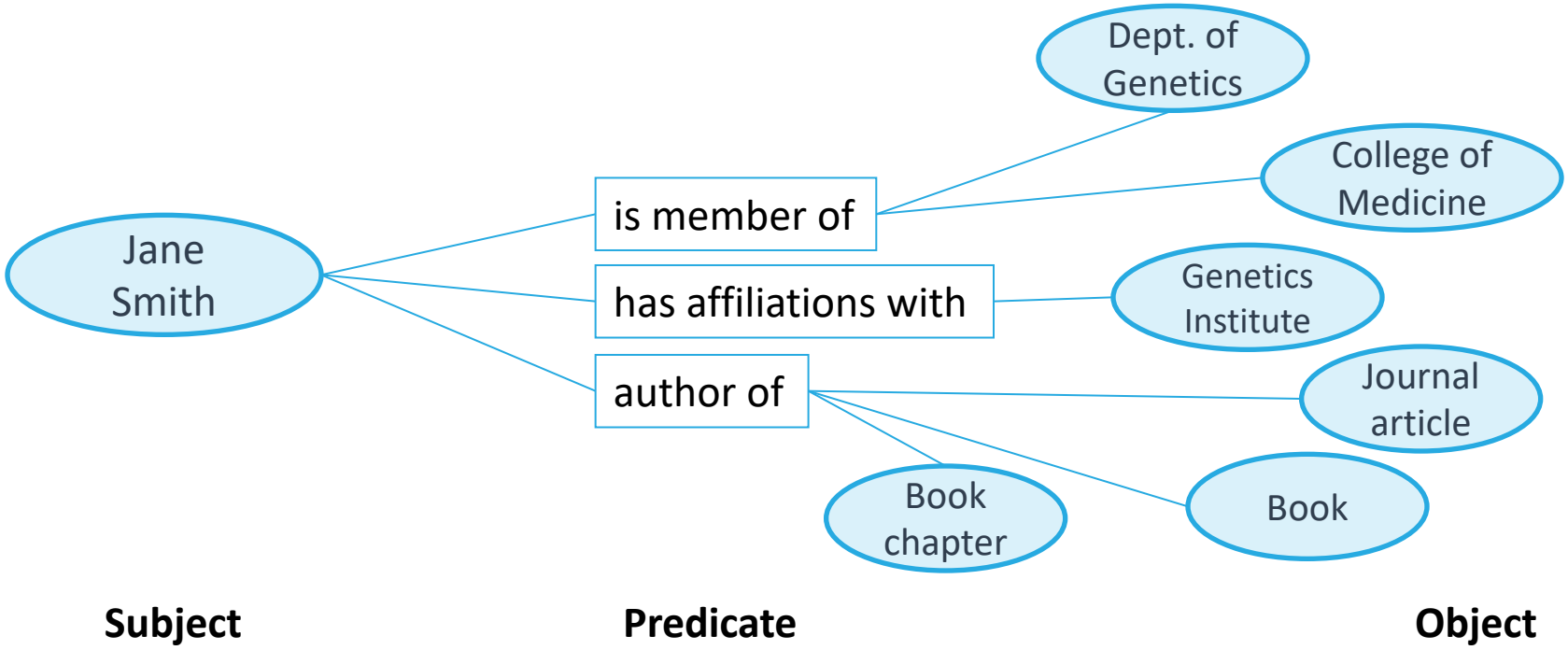
Internal data sources

External data sources

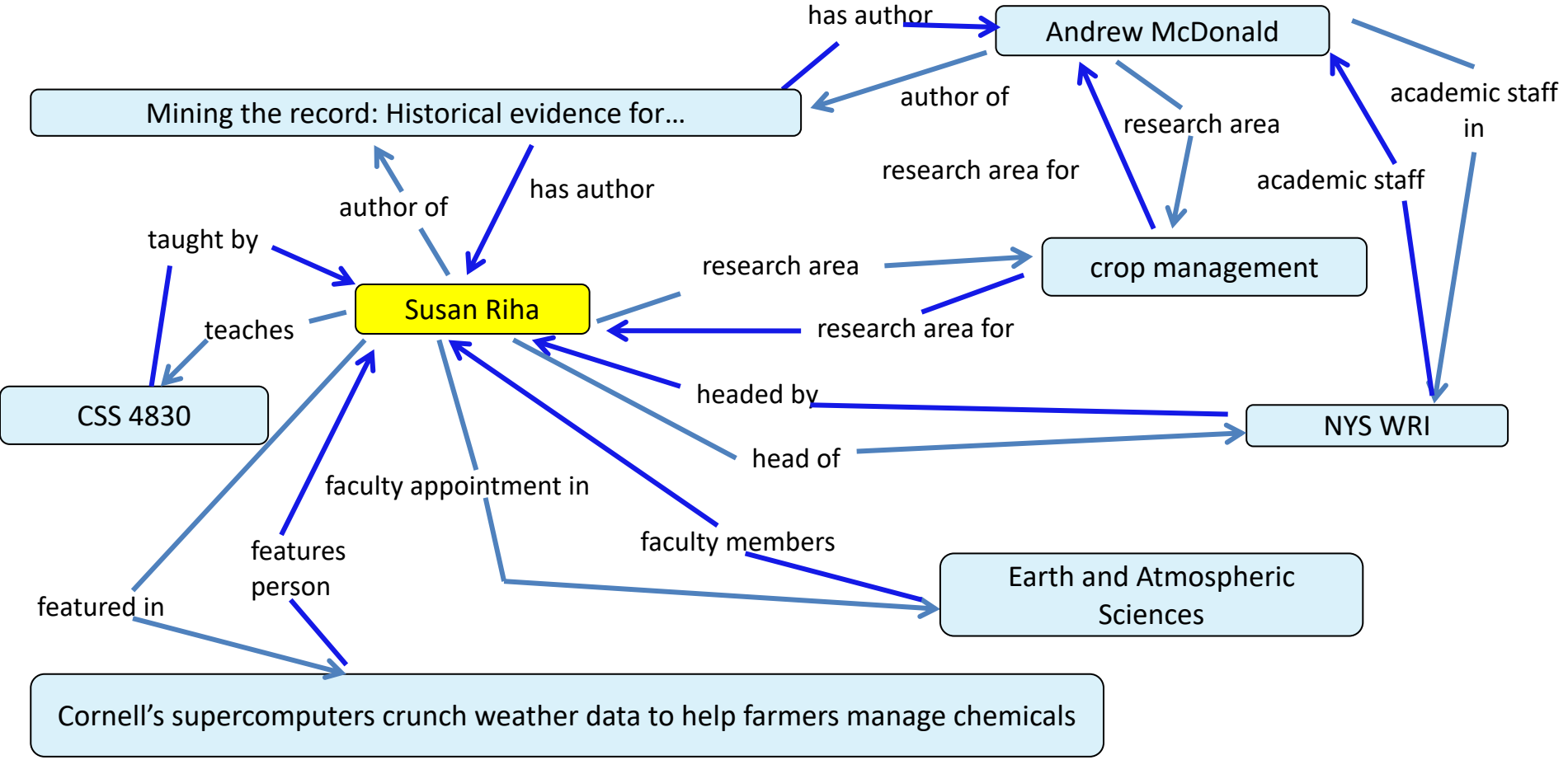
Individuals may also edit and customize their profiles to suit their professional needs.

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.



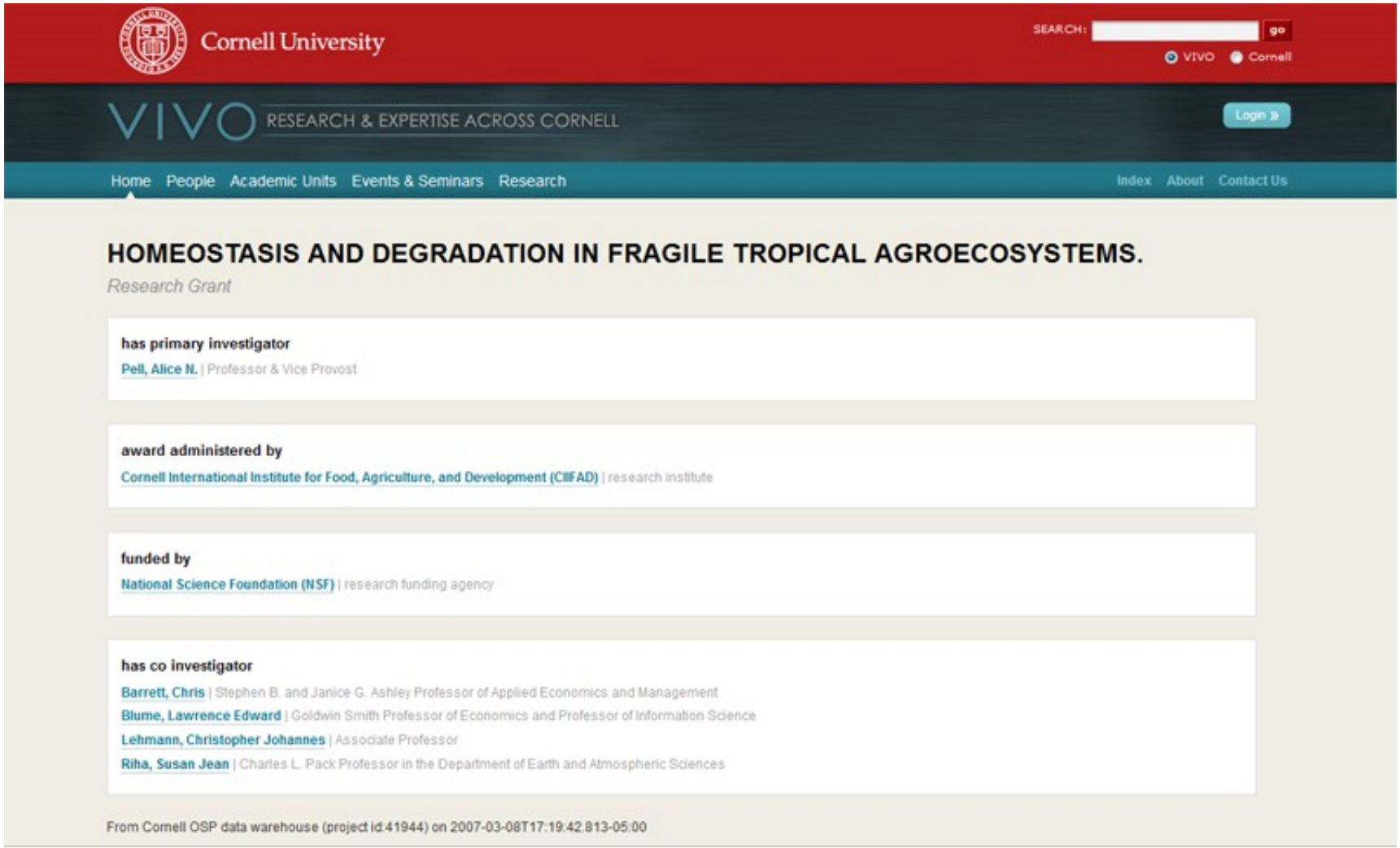
Detailed relationships for a researcher



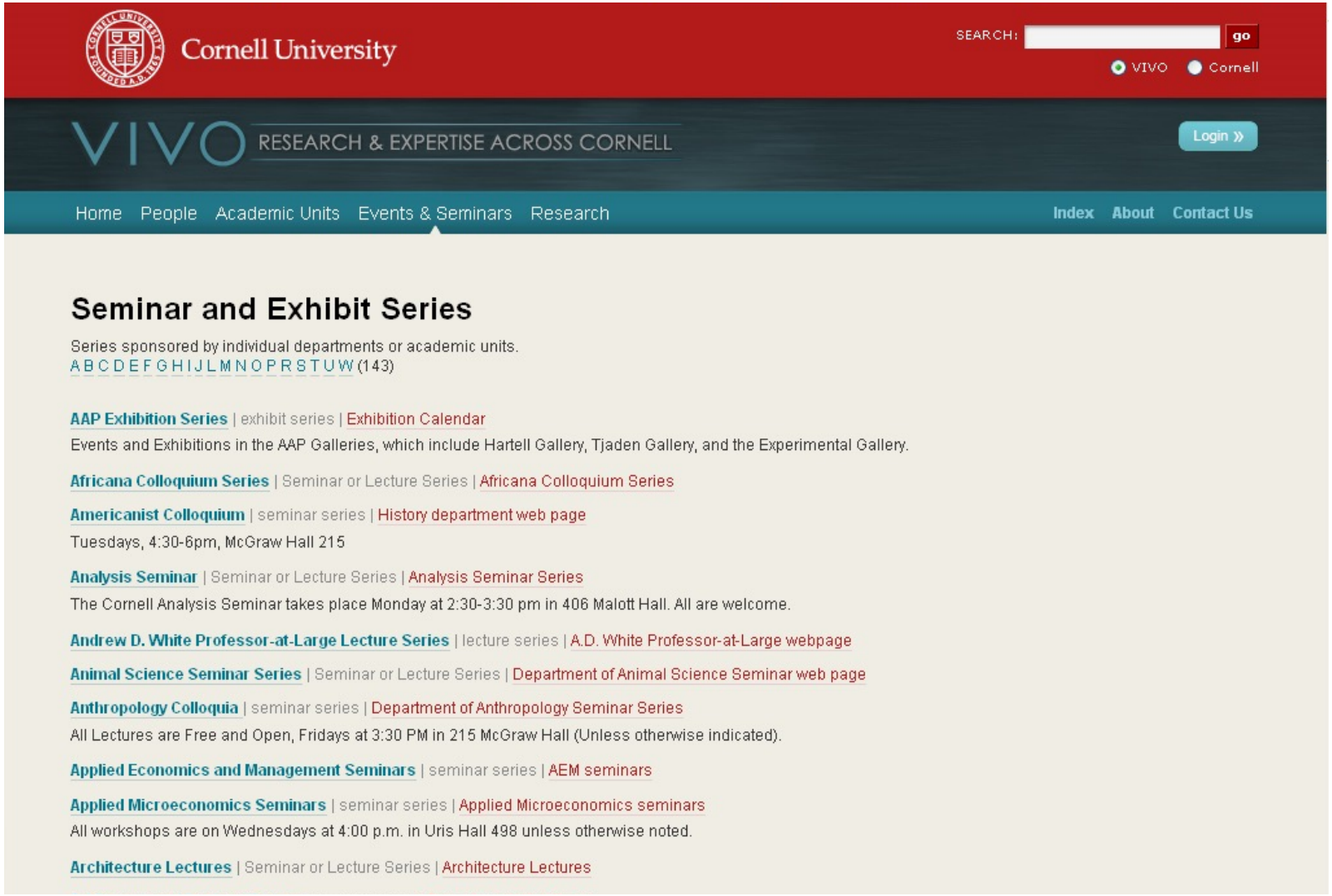
Query and explore

- 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)

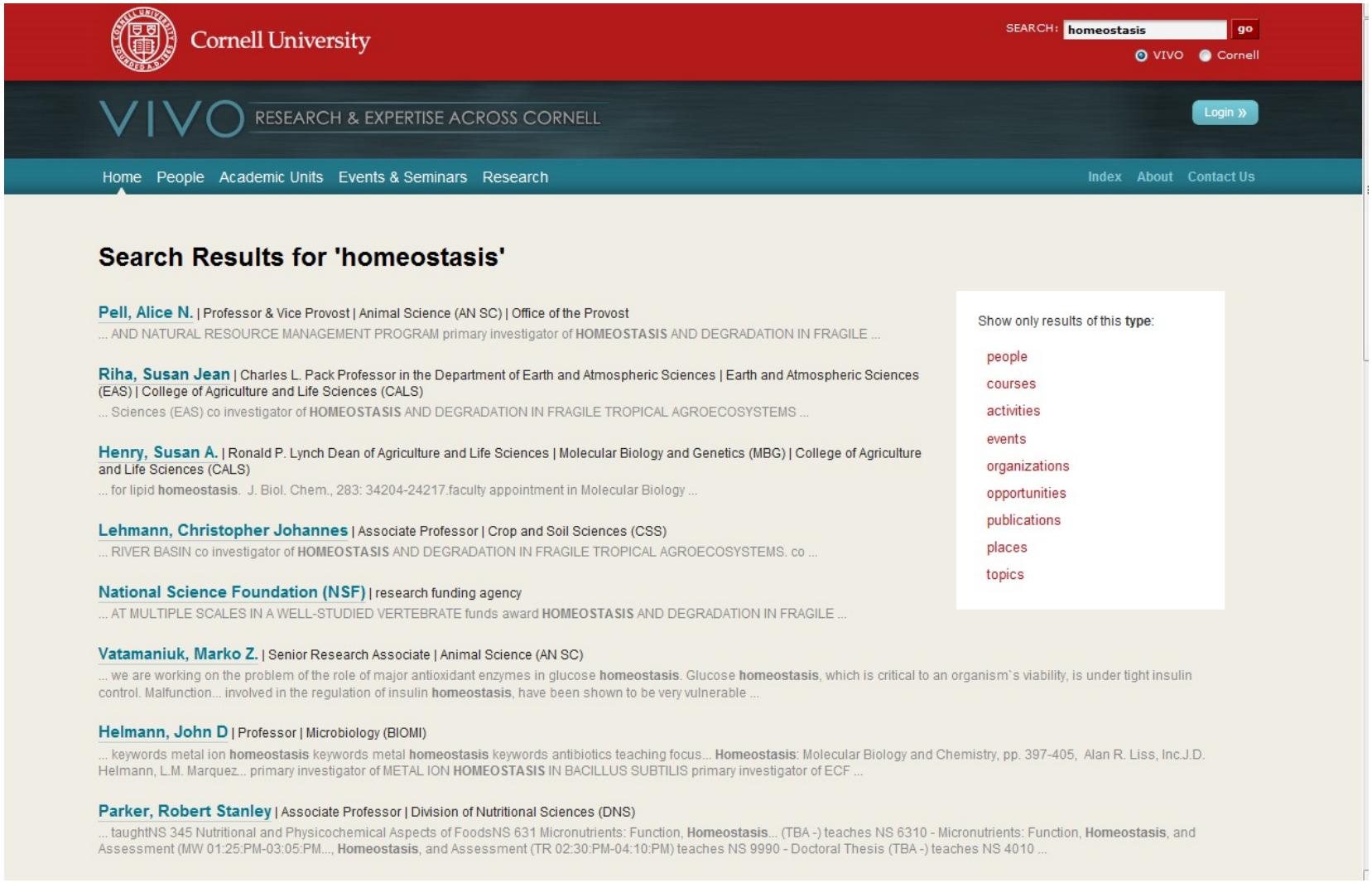
By individual: a grant



Browse by Type: seminar series



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

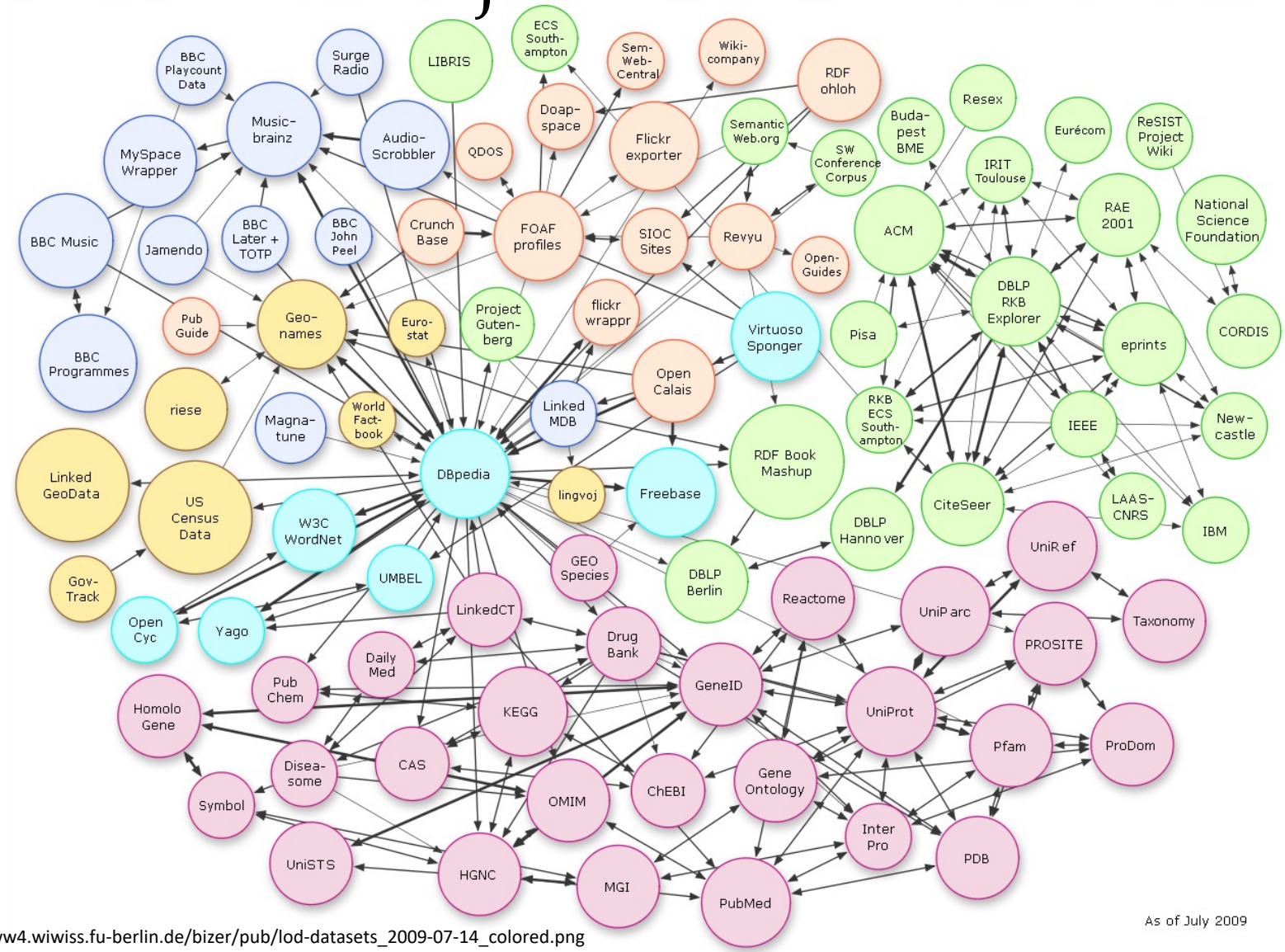
Linked Data principles

- 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://linkeddata.org

VIVO enables authoritative data about researchers to join the Linked Data cloud



As of July 2009

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

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

Keynote at 2003 Int’l Semantic Web Conference http://iswc2003.semanticweb.org/hendler_files/v3_document.htm

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

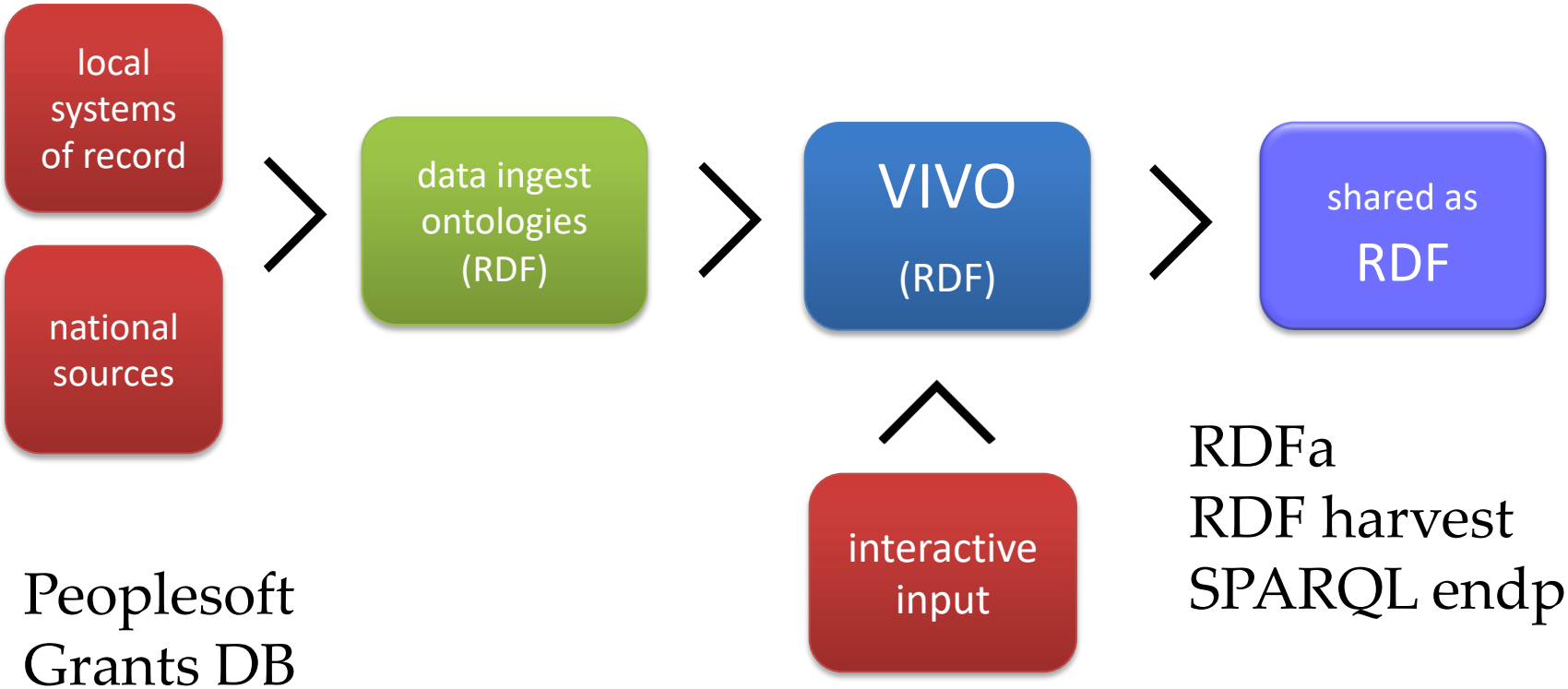
VIVO's three functional layers

end users

curators

ontology editing
& data flow

Local data flow

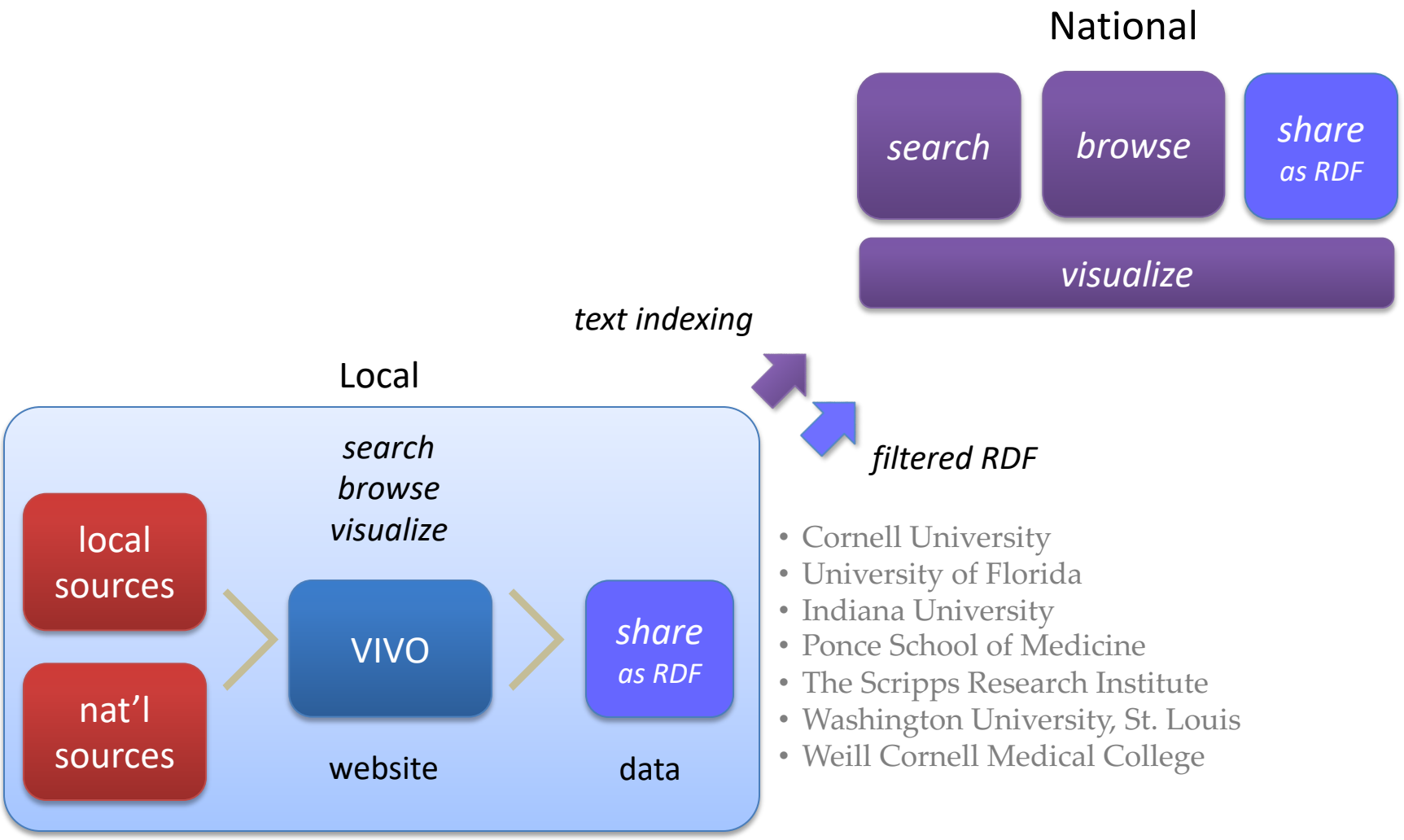


Peoplesoft
Grants DB
PubMed
Publishers

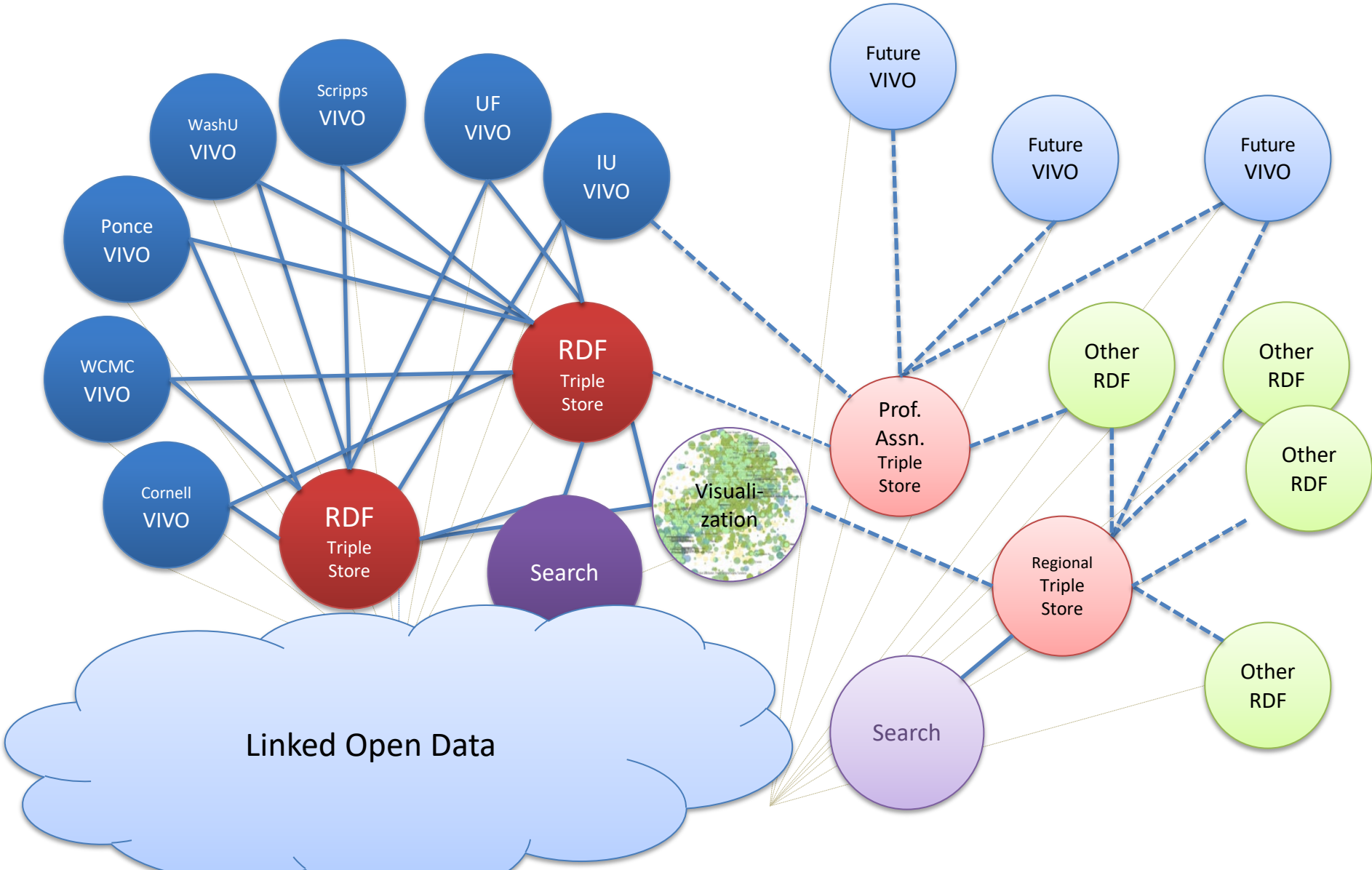
Researchers
Librarians
Administrative Staff
Self-Editors

RDFa
RDF harvest
SPARQL endpoint

From local to national



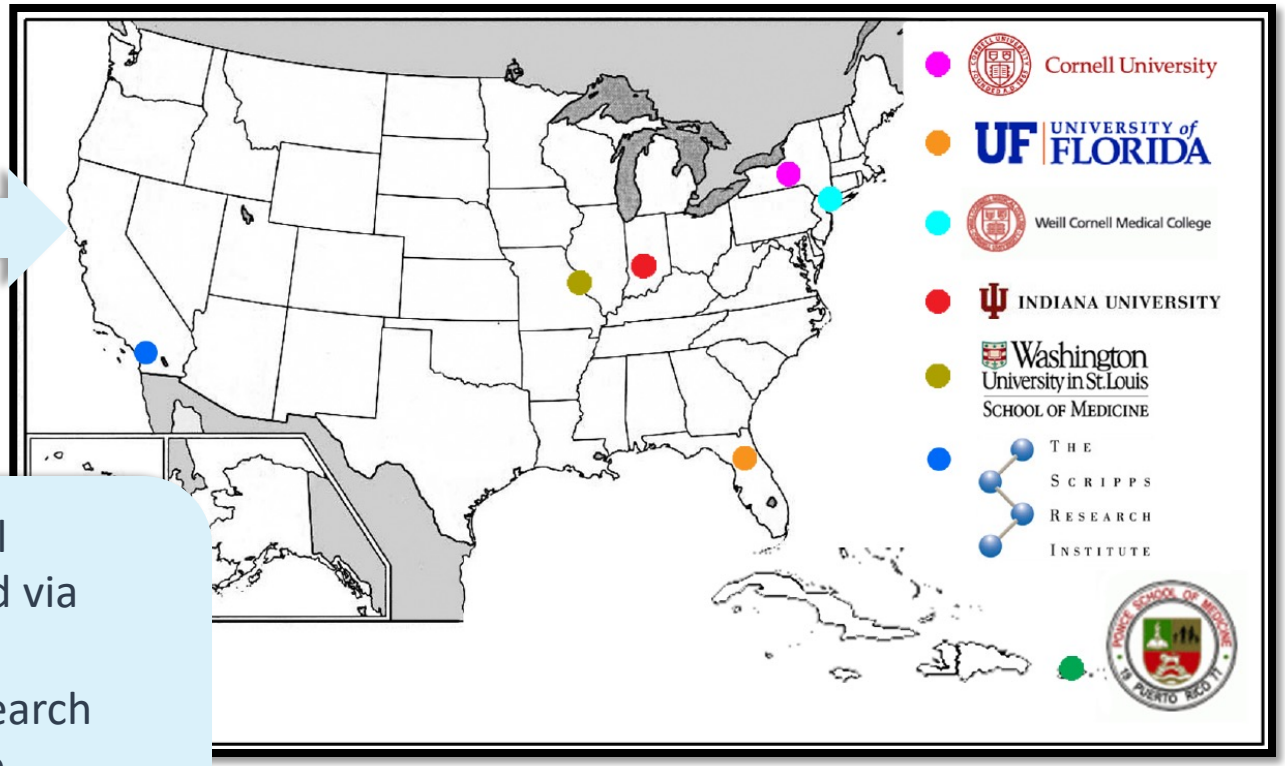
National networking



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

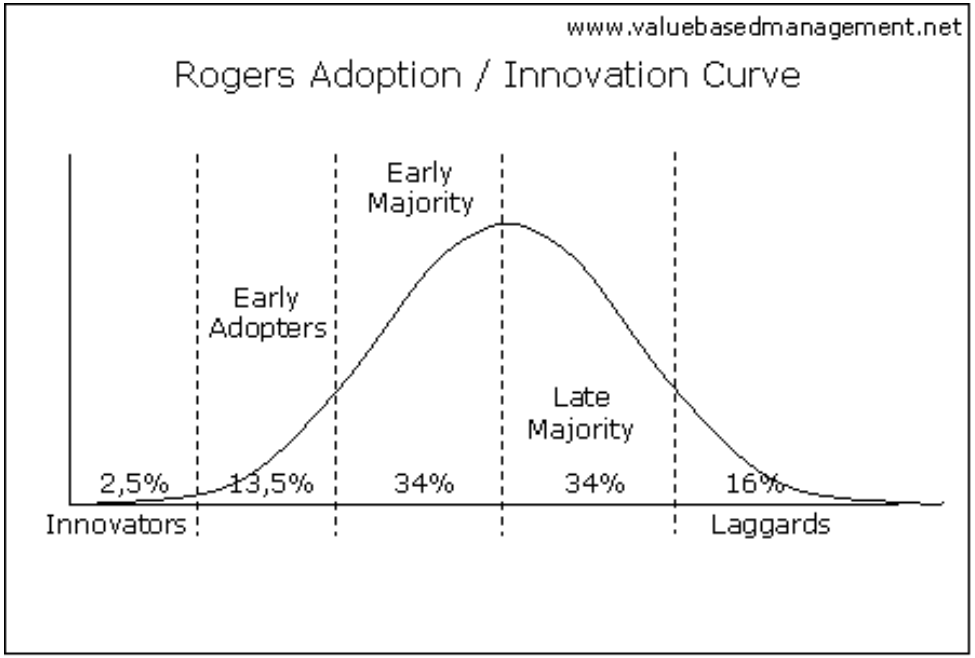
- Individual institutions – Columbia University, …
- Federal agencies – NIH (NIH RePORTER), NSF, …
- Consortia of schools – SURA, CTSA…
- eagle-i ("enabling resource discovery" U24 award)
- Publishers/vendors – PubMed, Elsevier, Collexis, ISI…
- Professional societies – AAAS, …

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

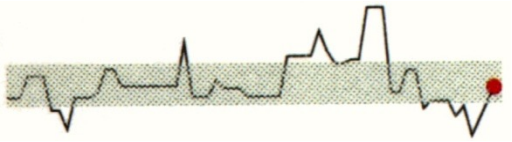
- Release 1 (as of Friday April 16th)
  - Deployment at 7 sites on production hardware
  - Mapping local data to R1 ontology
  - Batch data ingest (extract-transform-load)
- Upcoming Releases
  - Feedback on ontology; develop local ontologies
  - Improved usability
  - Open software (BSD license) and 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

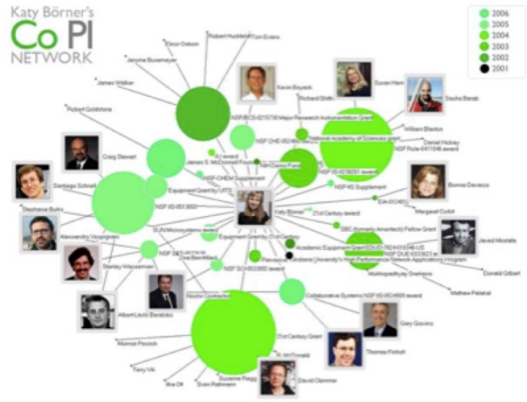
Driving future participation

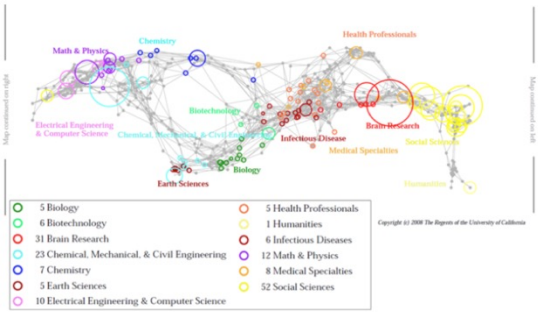
- "User scenario"-based ontology and feature development
- Usability; robust user support
- Compelling proof of concept within consortium
  - Administrators, scientists, clinicians, students, staff…
- Addressing interest from other institutions, partners
  - http://www.vivoweb.org
  - Conferences and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

Network Analysis & Visualization

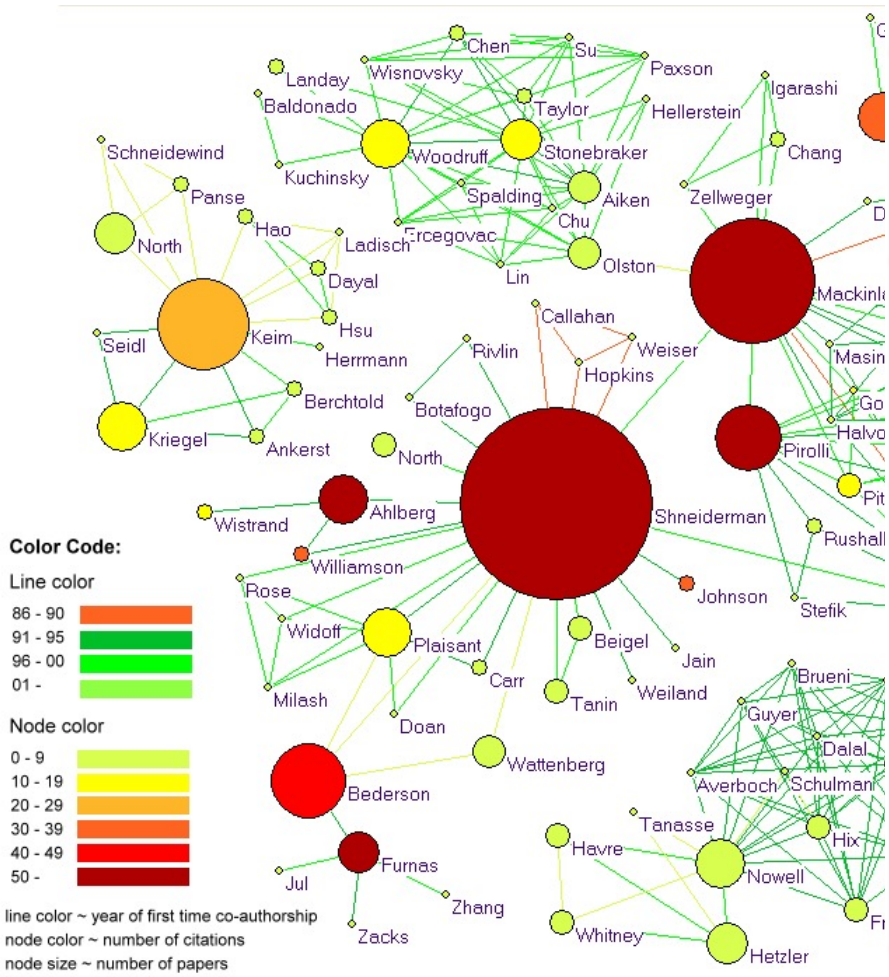
- At the individual investigator level
  - In-page, small graphs highlighting publication history, collaborations or co-authorship networks
- At the department or institutional level
  - Trends in research grants or publications
  - Collaboration networks across a larger group
  - Topical alignment with 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.

VIVO

Enabling National Networking of Scientists

Call to action

- Get involved as an:
  - adopter,
  - data provider, or
  - application developer
- Visit:
  - http://www.vivoweb.org

Thank you!

Questions?