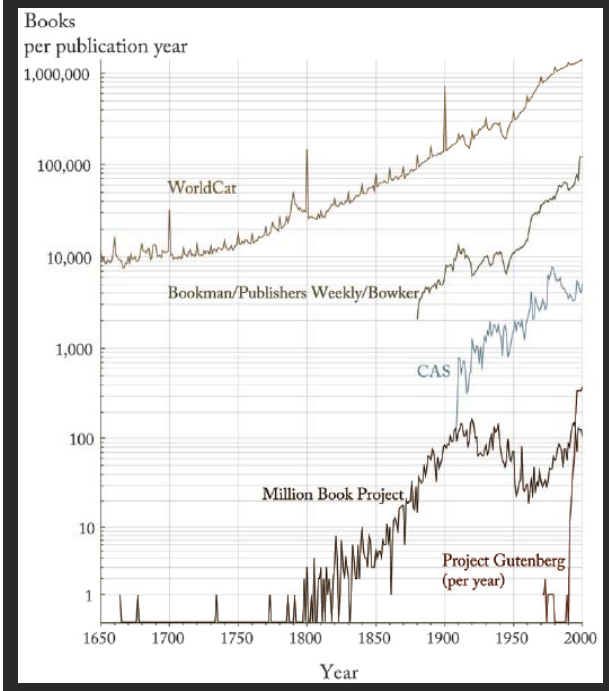
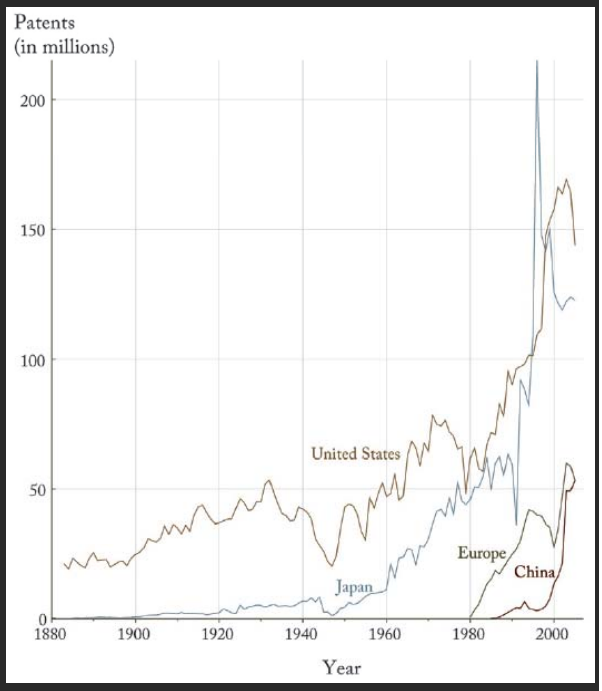
Envisioning (Biomedical) Science

Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

With special thanks to the members at the Cyberinfrastructure for Network Science Center, Mapping Science exhibit map makers and advisory board members, and the VIVO team.

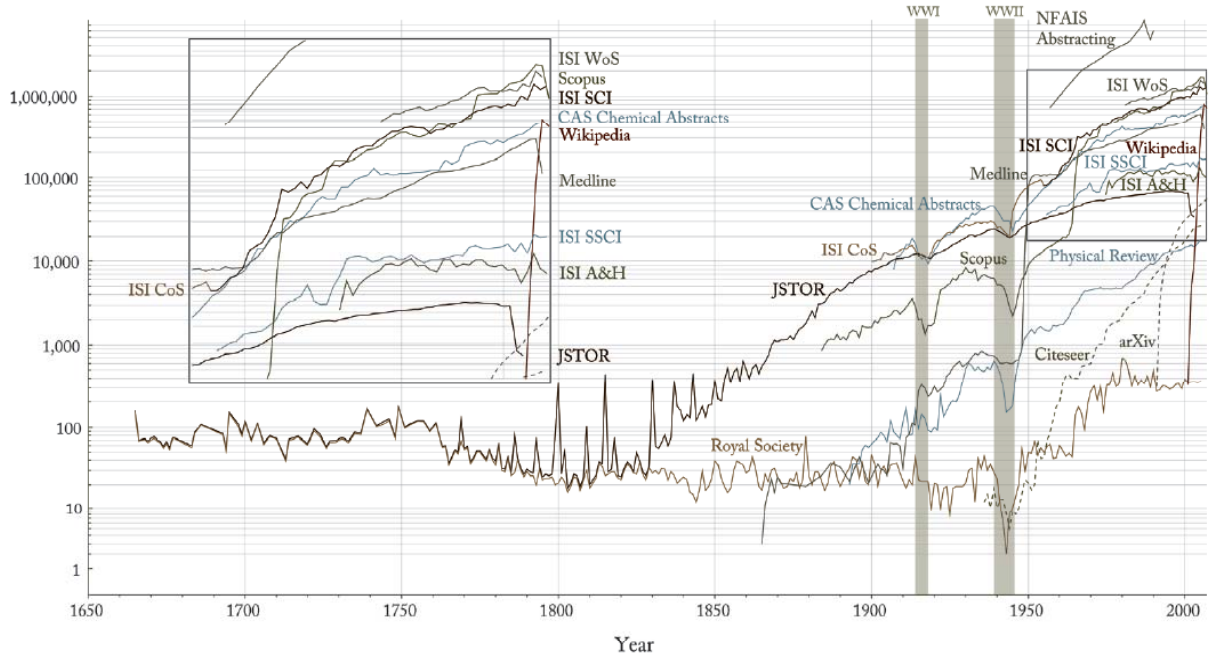
July 22, 2010
NIH Library, Bethesda, MD



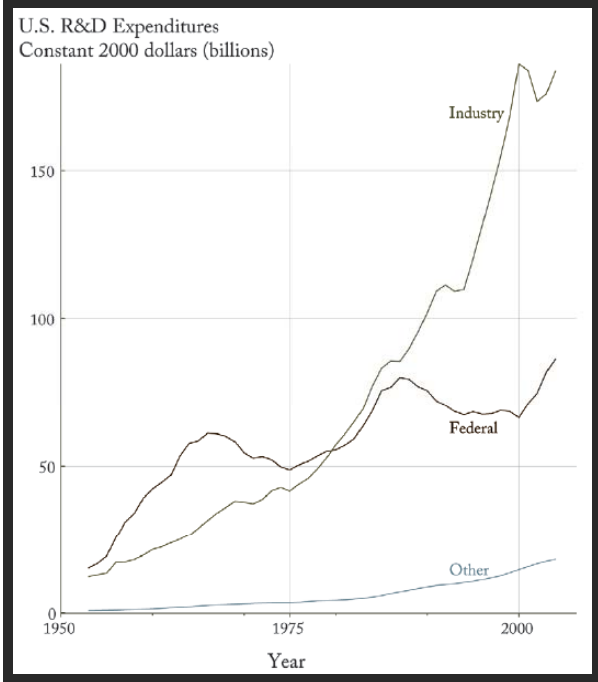


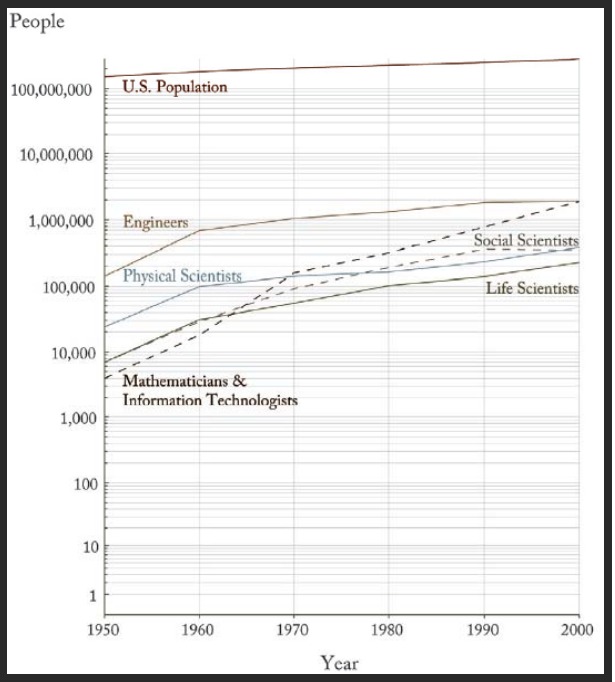
Börner, Katy (2010) Atlas of Science. MIT Press. http://scimaps.org/atlas

Papers & Wikipedia Entries



Börner, Katy (2010) Atlas of Science. MIT Press. http://scimaps.org/atlas





Börner, Katy (2010) Atlas of Science. MIT Press. http://scimaps.org/atlas

S&T Navigation, Management Tools that Different Stakeholders Want

Funding Agencies

- Need to monitor (long-term) money flow and research developments, identify areas for future development, stimulate new research areas, evaluate funding strategies for different programs, decide on project durations, funding patterns.

Scholars

- Want easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications **(research push).**

Industry

- Is interested in fast and easy access to major results, experts, etc. Influences the direction of research by entering information on needed technologies **(industry-pull)**.

Advantages for Publishers

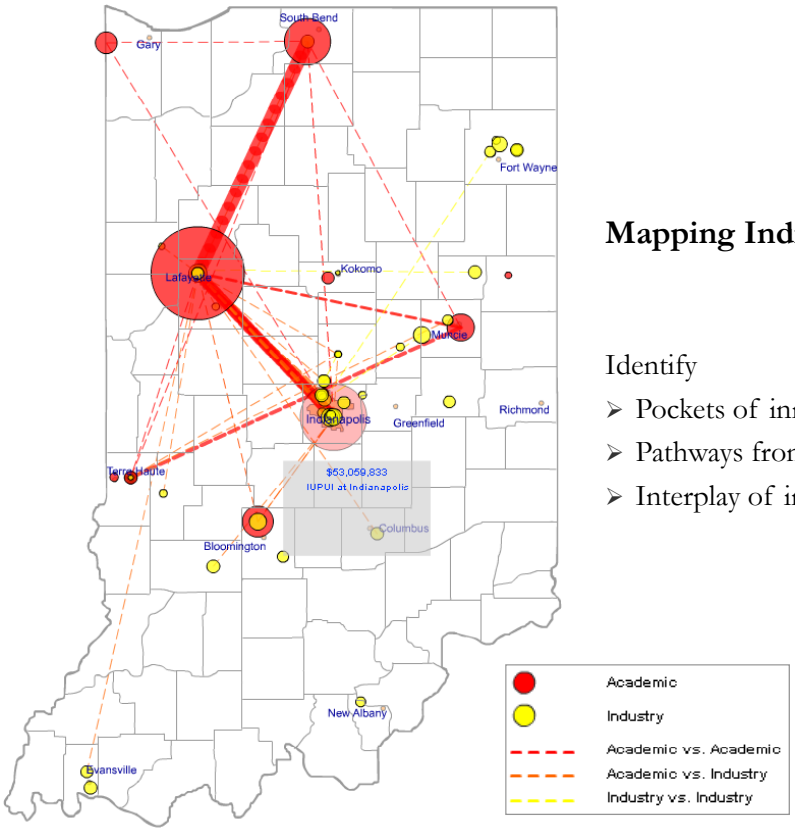
- Need easy to use interfaces to massive amounts of interlinked data. Need to communicate data provenance, quality, and context.

Society

- Needs easy access to scientific knowledge and expertise.

Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NS SA, all of sci
Temporal Analysis (When)	Funding portfolio of one individual	pic bursts of PNAS	113 Years of P Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a sta intellectual la	PNAS
Topical Analysis (What)	s.	e flows in research	VxOrd/Topic r NIH funding
Network Analysis (With Whom?)	NSF work of one	k	NIH's core competency



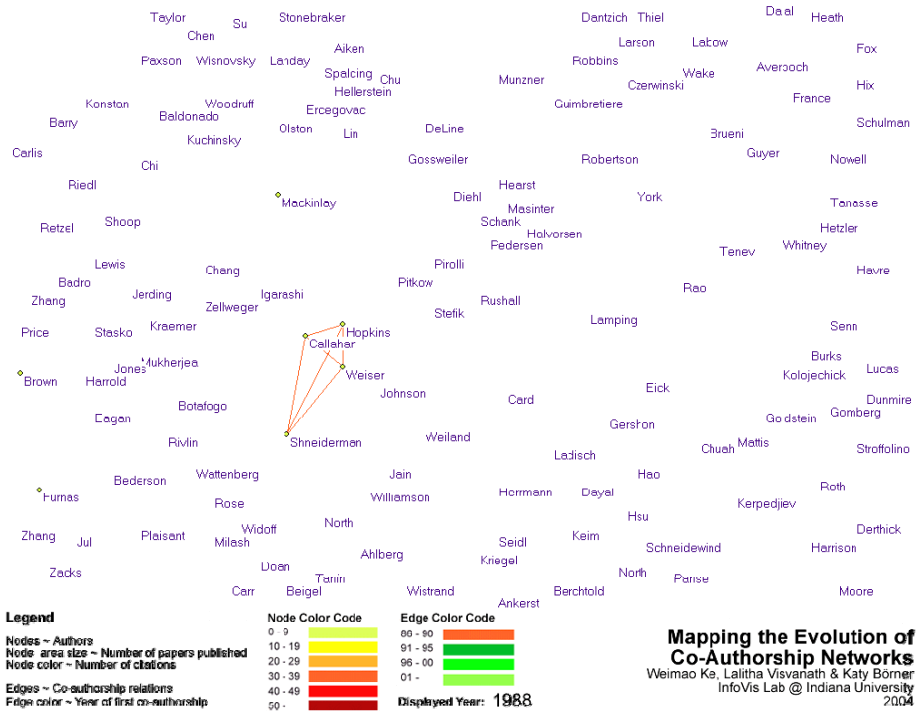
Mapping Indiana's Intellectual Space

Identify

- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

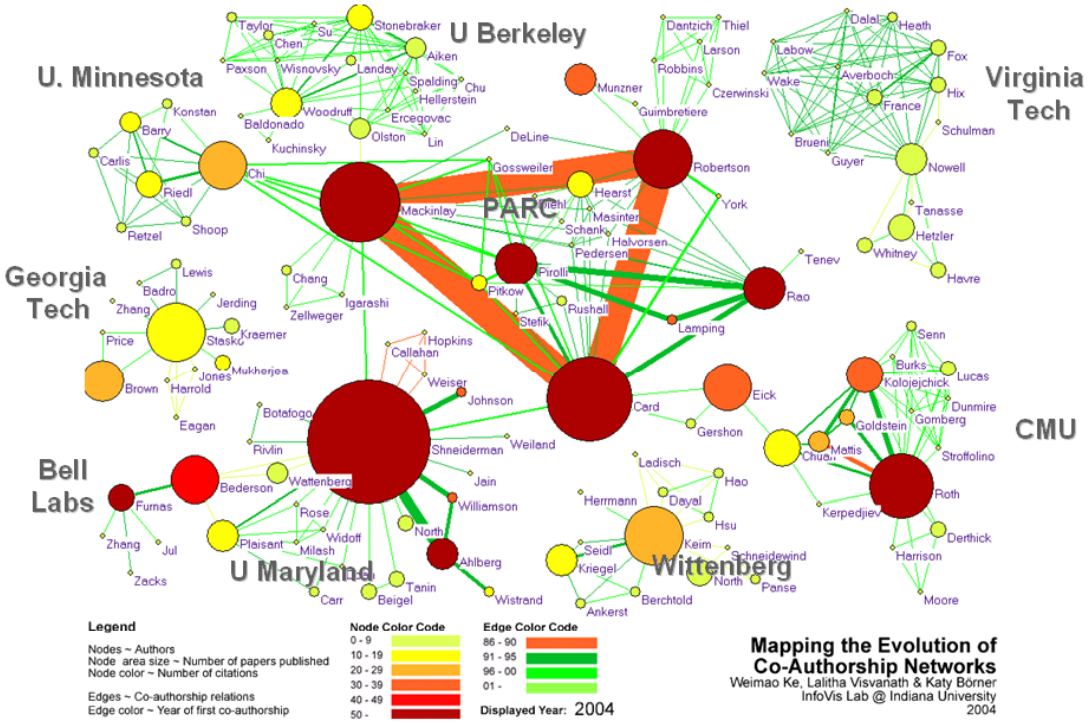
Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest



Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

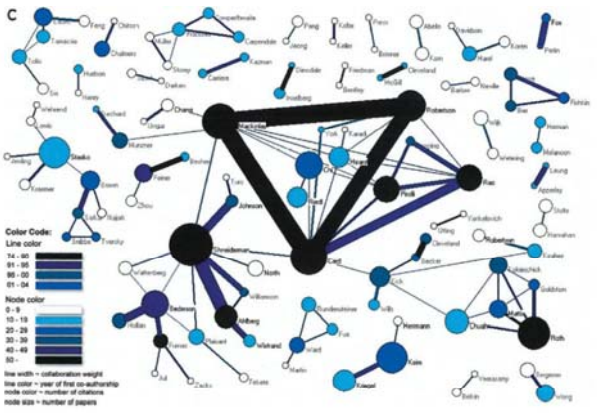
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

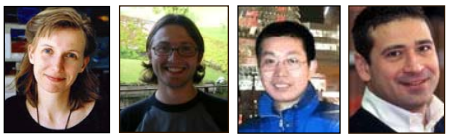
Research question:

- Is science driven by prolific single experts or by high-impact co-authorship teams?

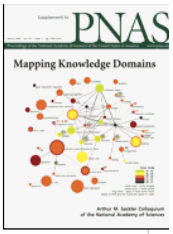
Contributions:

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.





Mapping Topic Bursts

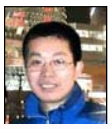


Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.

Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

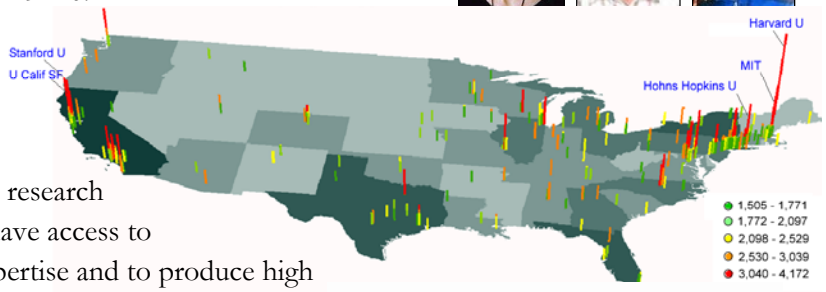


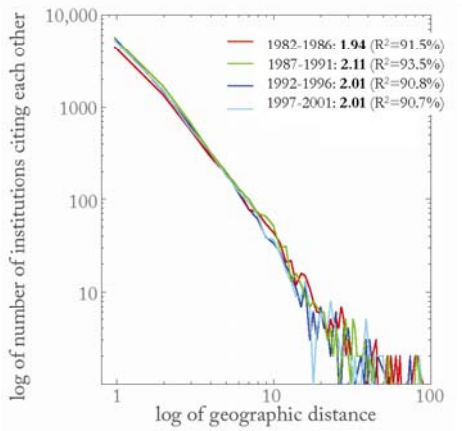


Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?





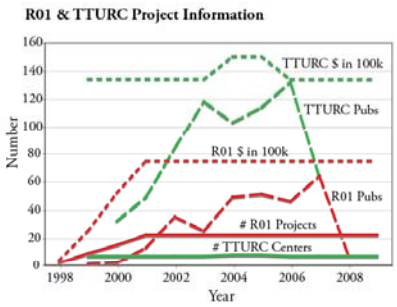
Contributions:

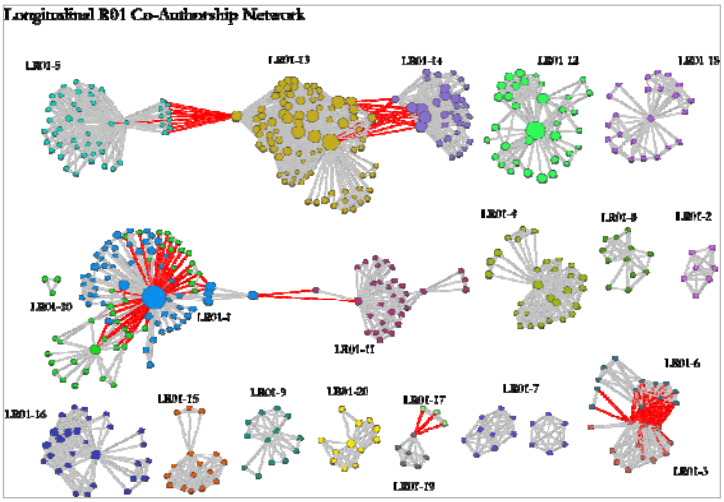
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

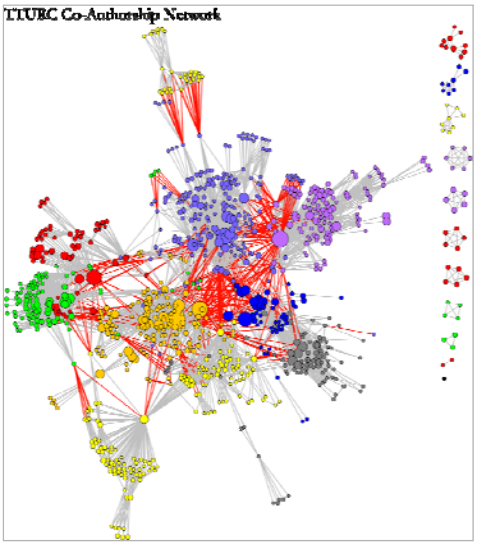
Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.



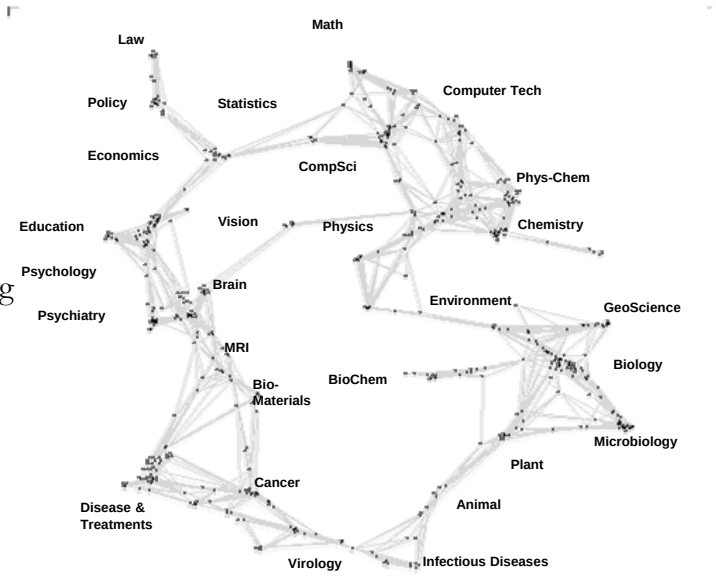




2002 Base Map of Science

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007). Mapping the Structure and Evolution of Chemistry Research. 11th International Conference on Scientometrics and Informetrics. pp. 112-123.

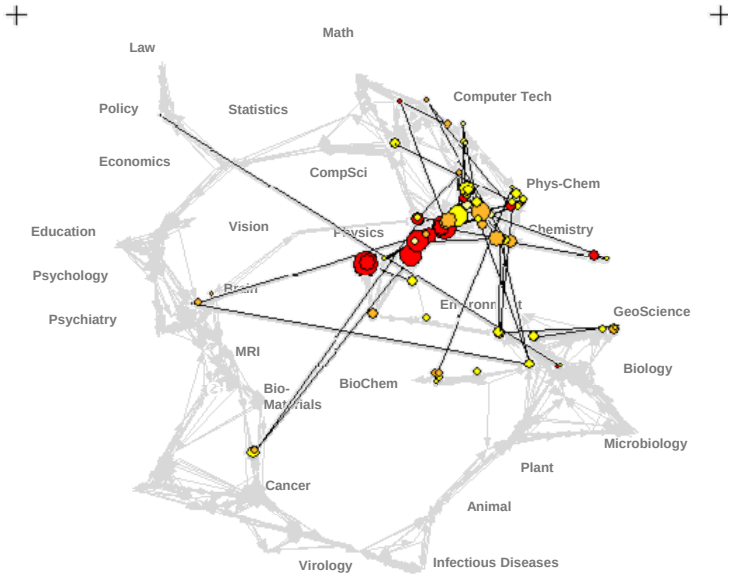
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y) positions for each journal cluster
  - by association, (x,y) positions for each journal



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

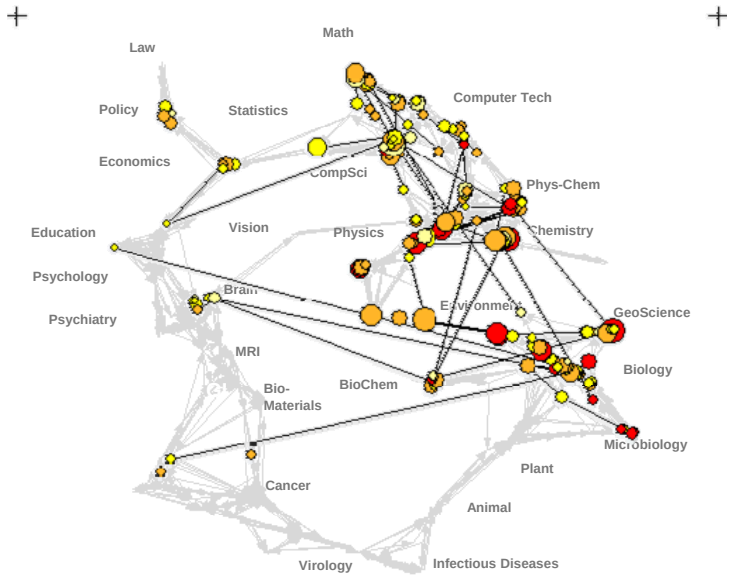
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

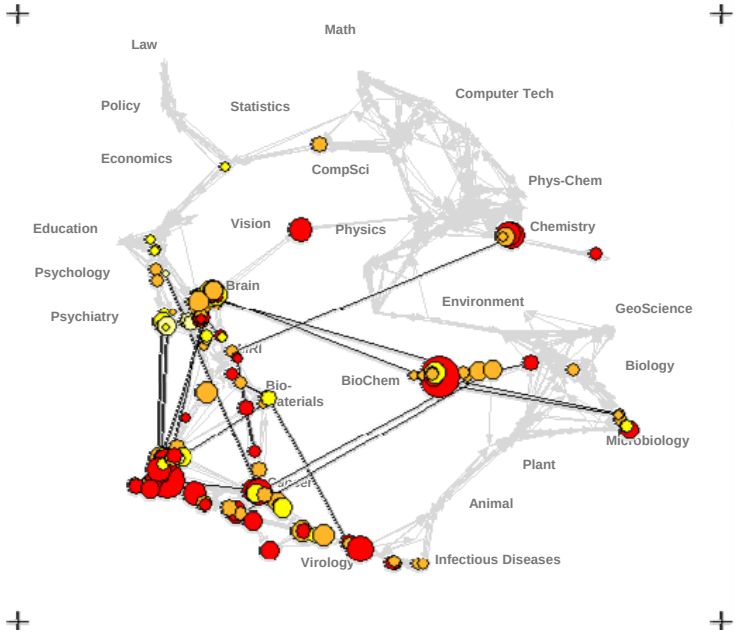
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

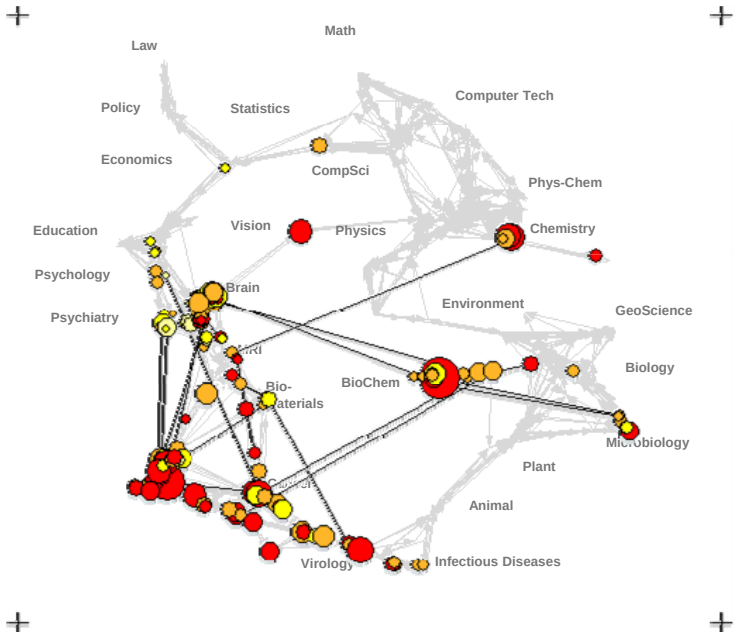
Funding Patterns of the National Institutes of Health (NIH)



Science map applications: Identifying core competency

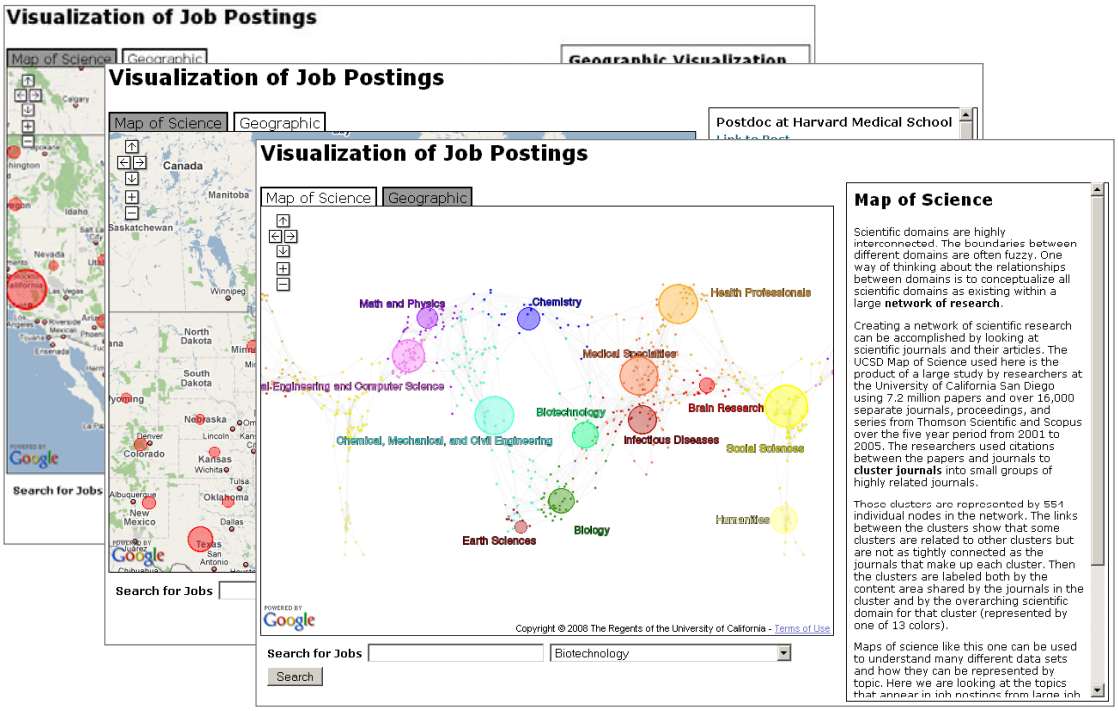
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)

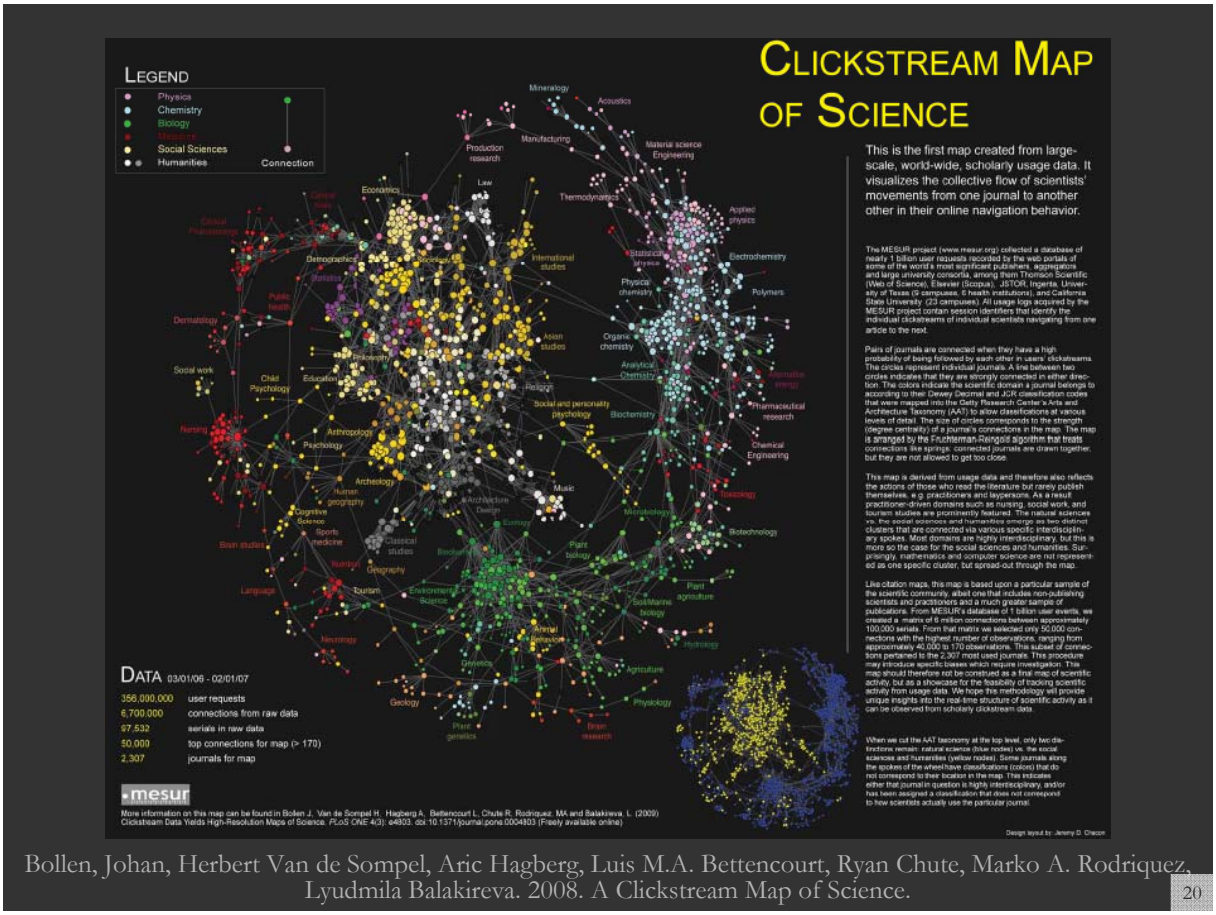


Where Are the Academic Jobs? Interactive Exploration of Job Advertisements in Geospatial and Topical Space

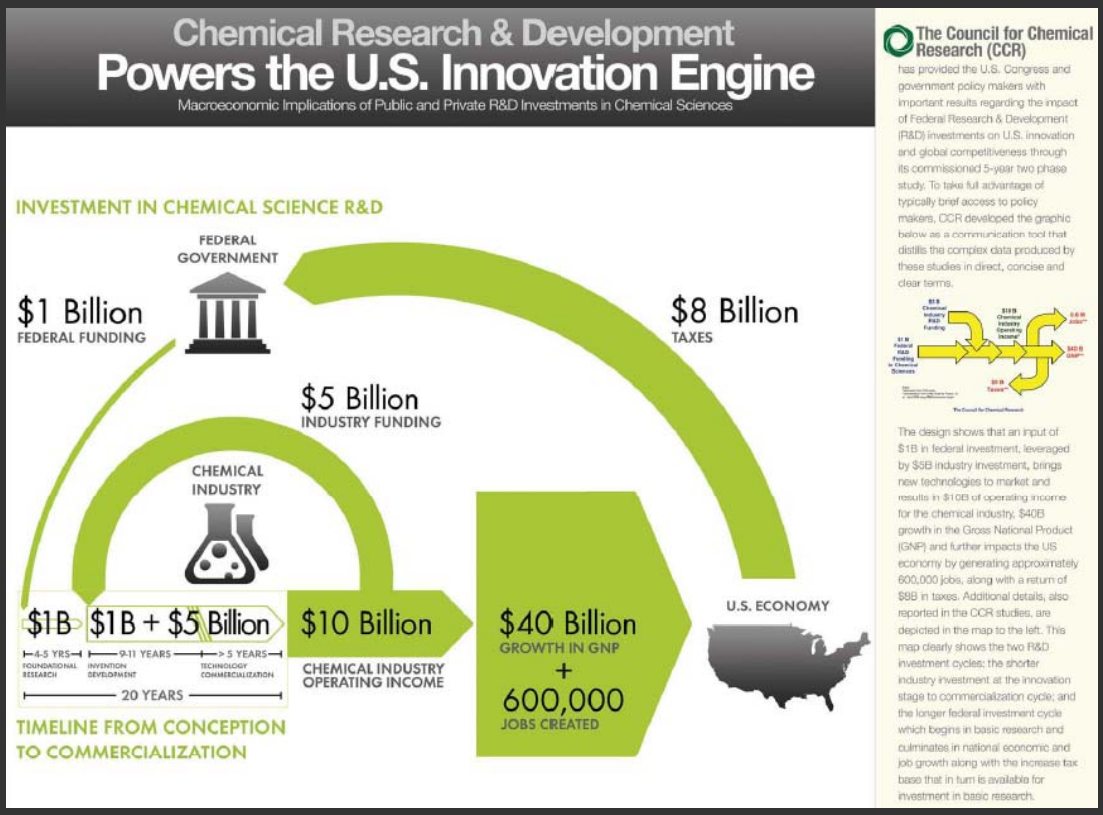
Angela Zoss, Michael Connover, Katy Börner (2010)



http://cns-nd3.slis.indiana.edu/mapjobs/geo



Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.



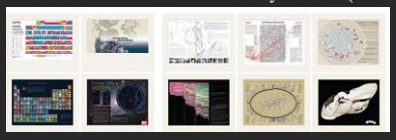
Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research.

Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org

The Power of Maps (2005)

The Power of Reference Systems (2006)



The Power of Forecasts (2007)

Science Maps for Economic Decision Makers (2008)

Science Maps for Science Policy Makers (2009)

Science Maps for Scholars (2010)
Science Maps as Visual Interfaces to Digital Libraries (2011)
Science Maps for Kids (2012)
Science Forecasts (2013)
How to Lie with Science Maps (2014)

Exhibit has been shown in 72 venues on four continents. Currently at
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA
- Marston Science Library, University of Florida, Gainesville, FL
- Center of Advanced European Studies and Research, Bonn, Germany
- Science Train, Germany.

Debut of 5th Iteration of Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, http://mediax.stanford.edu, http://scaleindependentthought.typepad.com/photos/scimaps

Illuminated Diagram Display

W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007) Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007.

Large-scale, high resolution prints illuminated via projector or screen.

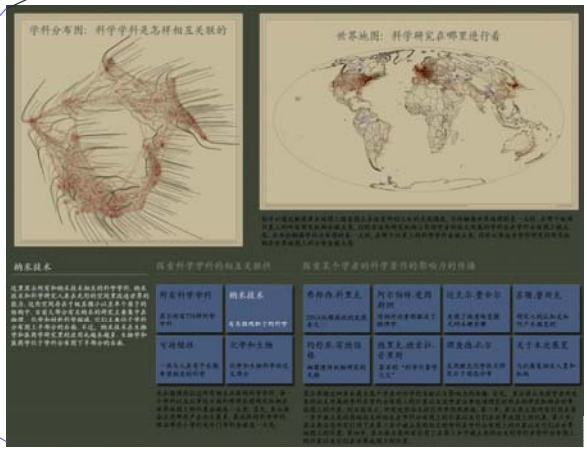
Questions:

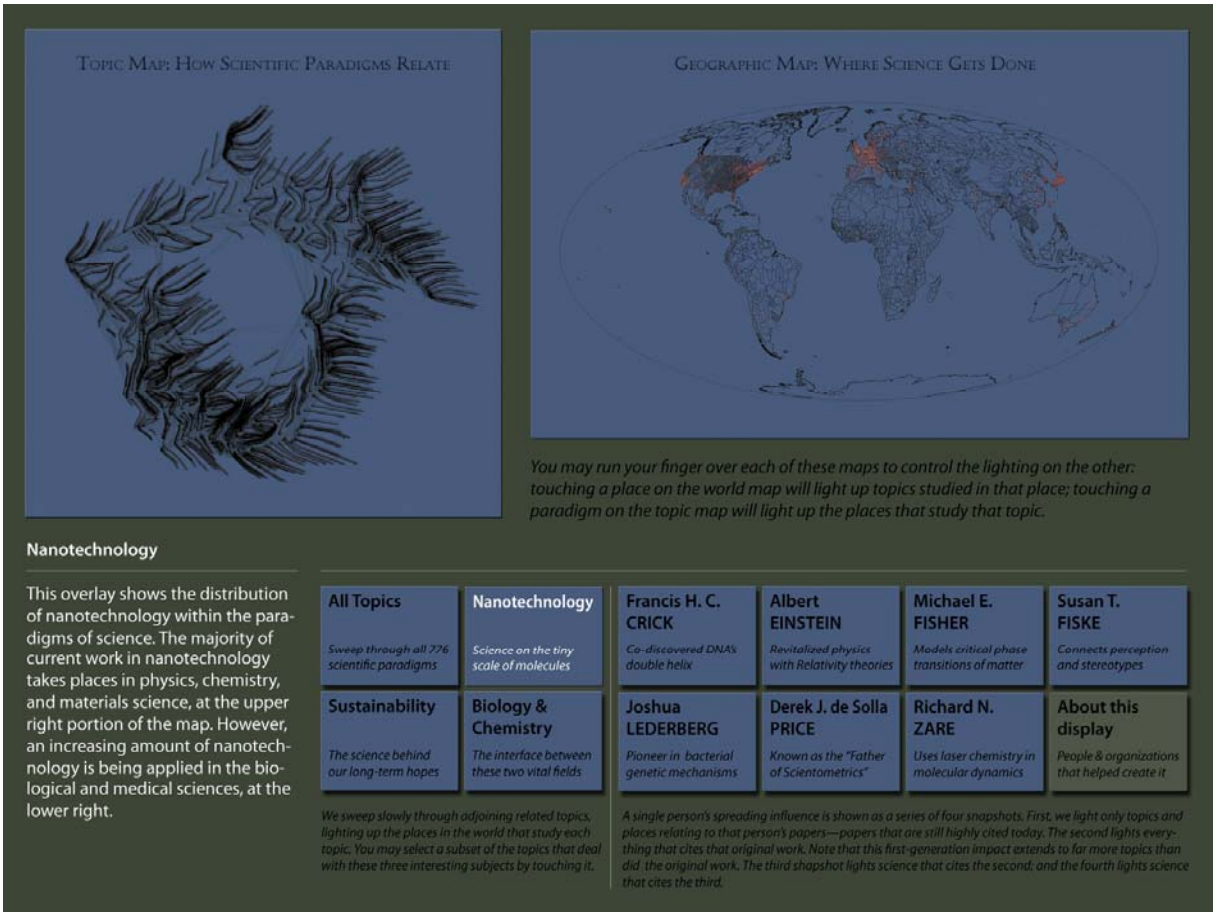
- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

Interactive touch panel.

Contributions:

- Interactive, high resolution interface to access and make sense of data about scholarly activity.





Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months
12 coaches, 300 m long
Opening was on April 23rd, 2009 by German Chancellor Merkel
http://www.expedition-zukunft.de

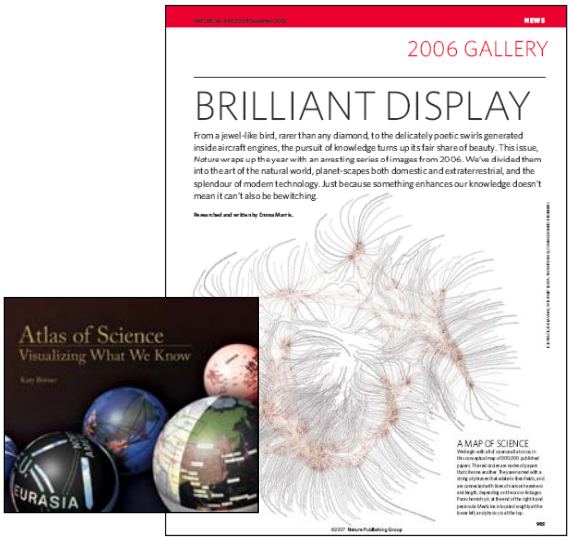
Computational Scientometrics References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255. http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

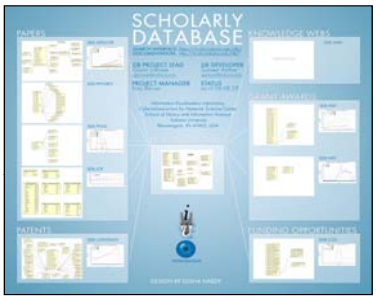
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

Börner, Katy (2010) Atlas of Science. MIT Press. http://scimaps.org/atlas



Computational Scientometrics Cyberinfrastructures



Scholarly Database: 23 million scholarly records
http://sdb.slis.indiana.edu

James S. McDonnell Foundation

VIVO Research Networking
http://vivoweb.org

Information Visualization Cyberinfrastructure
http://iv.slis.indiana.edu

Network Workbench Tool & Community Wiki
http://nwb.slis.indiana.edu

Science of Science (Sci²) Tool and CI Portal
http://sci.slis.indiana.edu

Epidemics Cyberinfrastructure
http://epic.slis.indiana.edu/

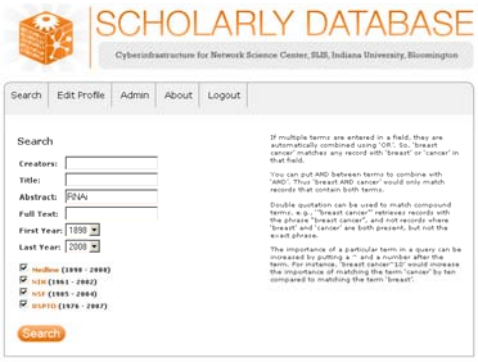
OSGi Alliance

CIShell cyberinfrastructure shell

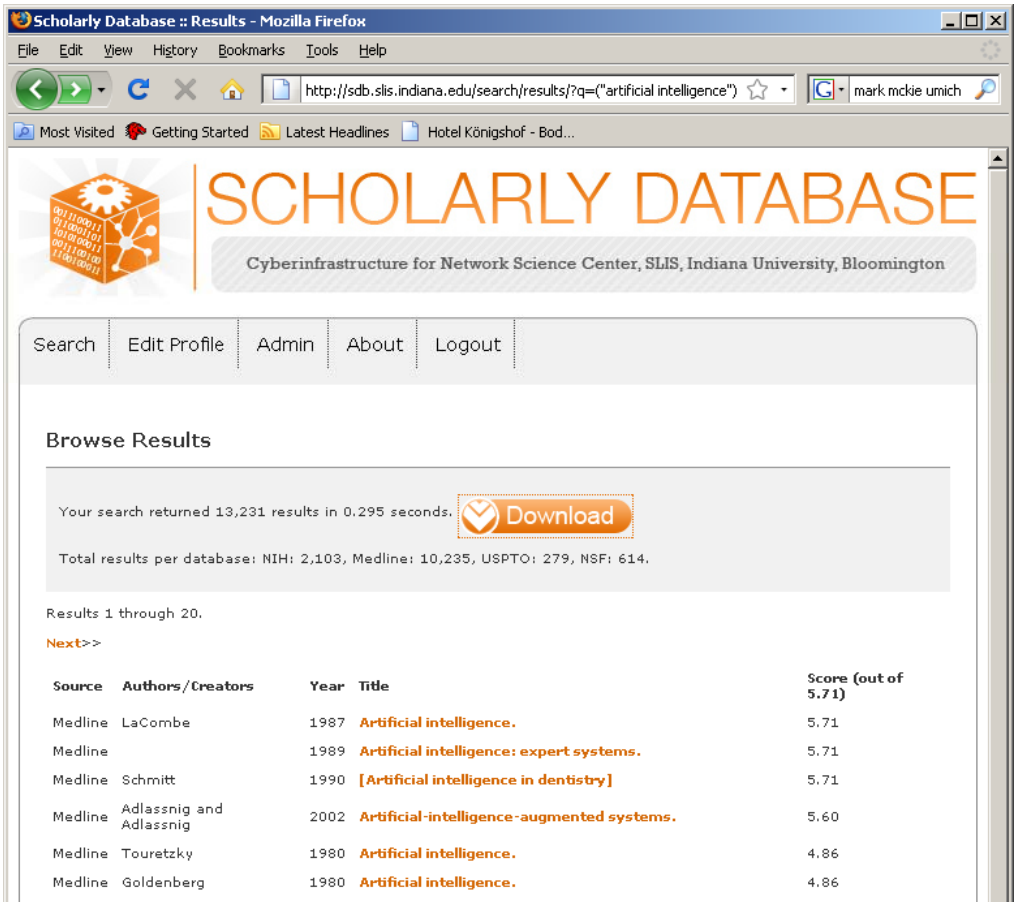
Scholarly Database: Web Interface

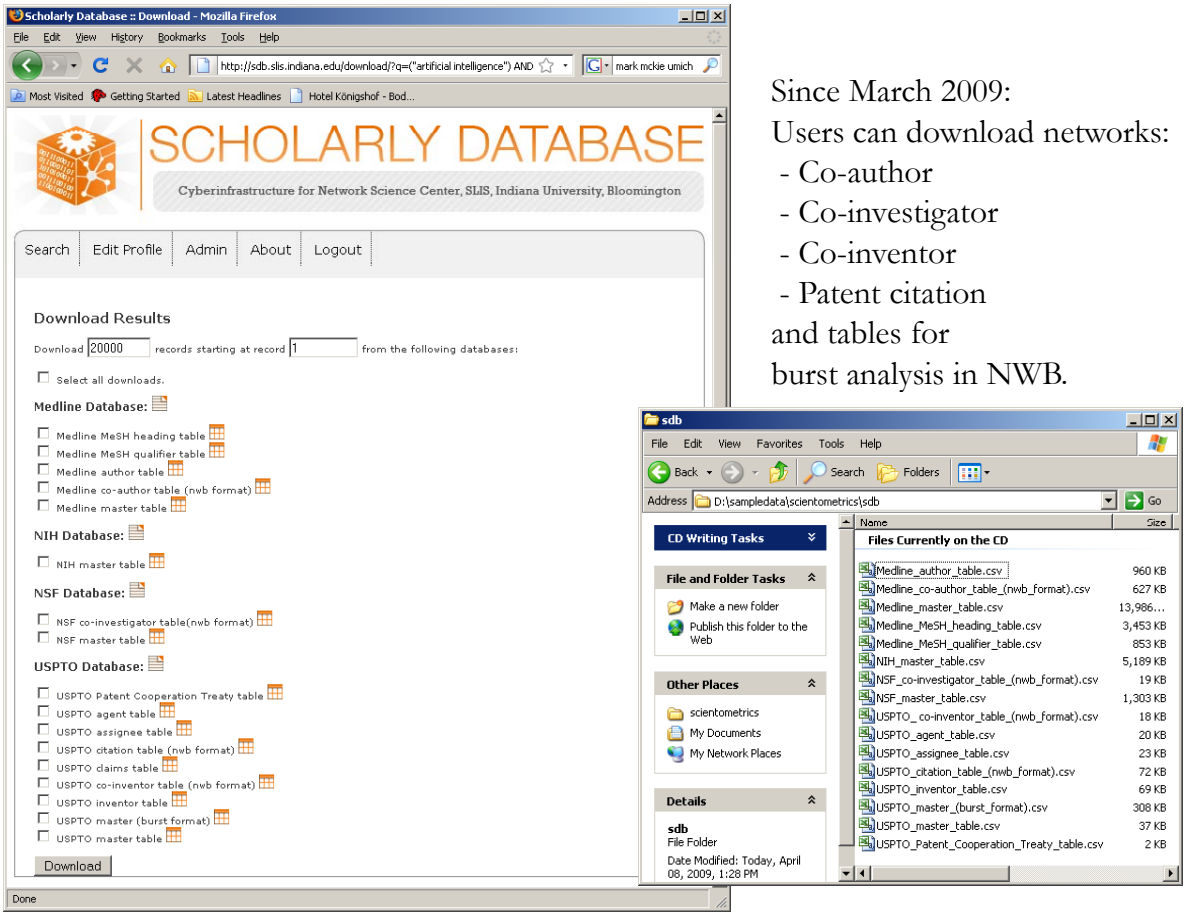
http://sdb.slis.indiana.edu

Supports federated search of 23 million publication, patent, grant records.
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.



Register for free access at http://sdb.slis.indiana.edu





Since March 2009:
Users can download networks:
- Co-author
- Co-investigator
- Co-inventor
- Patent citation

and tables for
burst analysis in NWB.

Scholarly Database: # Records, Years Covered

Datasets available via the Scholarly Database (* internally)

Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	17,764,826	1898-2008	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3, 875,694	1976-2008	Yes*	
NSF	174,835	1985-2004	Yes*	
NIH	1,043,804	1961-2002	Yes*	
Total	**23,167,642**	**1893-2006**	**4**	**3**

Aim for comprehensive time, geospatial, and topic coverage.

Mapping the Field of RNAi Research (SDB Data)

(section 5.2.7)

RNAi	
Time frame:	1865-2008
Region(s):	Miscellaneous
Topical Area(s):	RNAi
Analysis Type(s):	Co-Author Network, Patent-Citation Network, Burst Detection

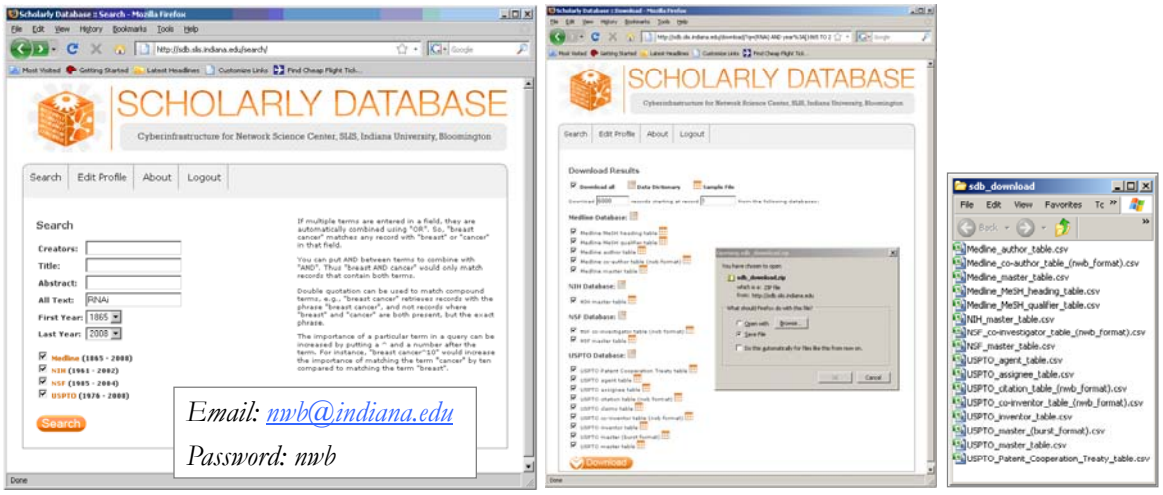
How many papers, patents, and funding awards exist on a specific topic?

Here we selected research on RNA interference (RNAi) is a system within living cells that helps to control which genes are active and how active they are.

The data for this analysis comes from a search of the Scholarly Database (SDB) (http://sdb.slis.indiana.edu/) for "RNAi" in "All Text" from MEDLINE, NSF, NIH and USPTO. A copy of this data is available in '**yoursci2directory*/sampledata/scientometrics/sdb/RNAi*'. The default export format is .csv, which can be loaded in the Sci2 Tool directly.

Mapping the Field of RNAi Research (SDB Data)

(section 5.2.7)



The **Scholarly Database** at Indiana University provides free access to 23,000,000 papers, patents, and grants. Since March 2009, users can also download networks, e .g., co-author, co-investigator, co-inventor, patent citation, and tables for burst analysis. For more information and to register, visit http://sdb.slis.indiana.edu.

Mapping the Field of RNAi Research (SDB Data)
(section 5.2.7)

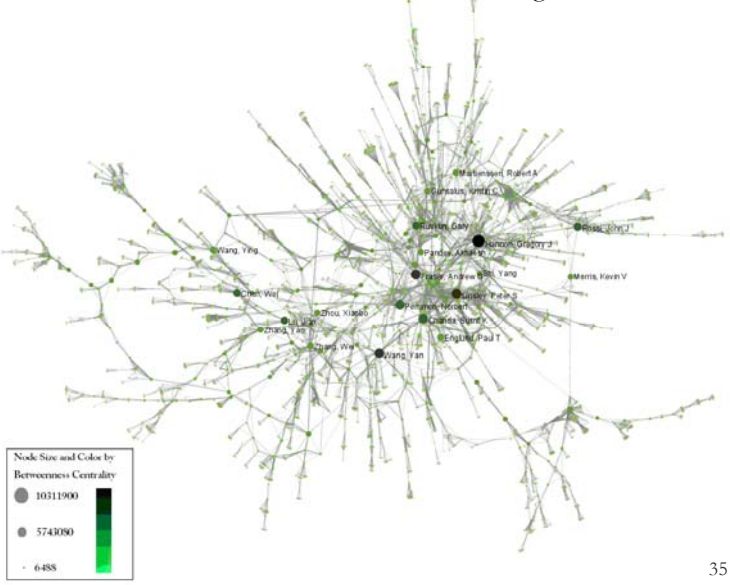
Co-Author Network

Load '**yoursci2directory*/sampledata/scientometrics/sdb/RNAi/Medline_co-author_table_(nwb_format).csv'* as a standard csv file. SDB tables are already pre-normalized, so now simply run '*Data Preparation > Text Files > Extract Co-Occurrence Network'* using the default parameters.

Network Analysis Toolkit (NAT):
21,578 nodes with 131 isolates,
77,739 edges.

Extract only the largest component by running '*Analysis > Networks > Unweighted and Undirected > Weak Component Clustering.'*

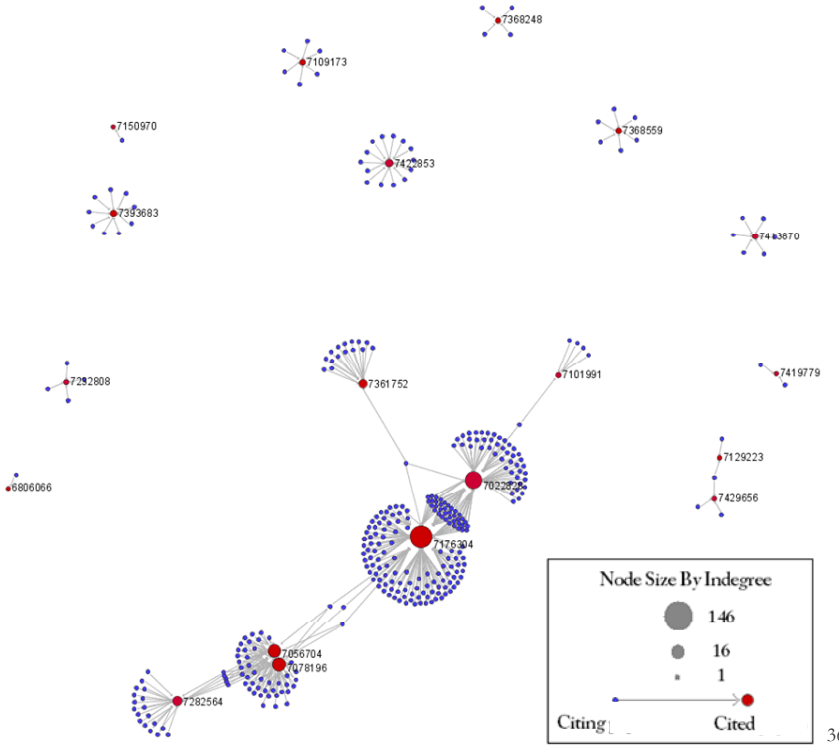
Visualize with *GUESS* using '*Layout > GEM'*.
Use a custom python script to color and size the network.



Mapping the Field of RNAi Research (SDB Data)
(section 5.2.7)

Patent Citation Network

To visualize the citation patterns of patents on RNAi, load '**yoursci2directory*/sampledata/scientometrics/sdb/RNAi/USPTO_citation_table_(nwb_format).csv'* as a standard csv file and follow the instructions in the tutorial.



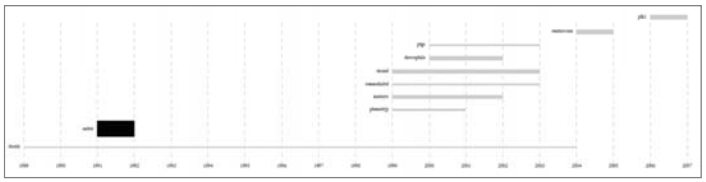
Mapping the Field of RNAi Research (SDB Data)
(section 5.2.7)

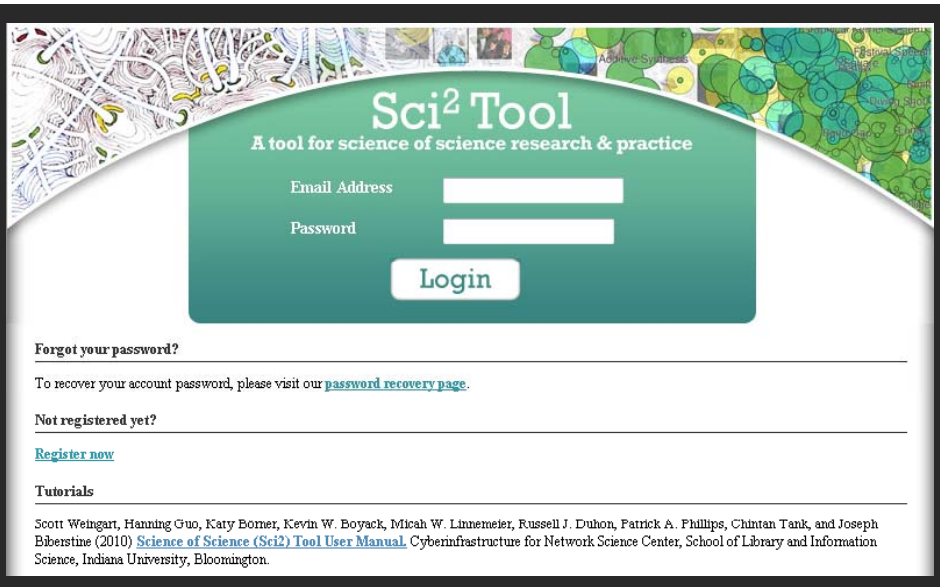
Topic Bursts

- Load '**yoursci2directory*/sampledat/scientometrics/sdb/RNAi/Medline_master_table.csv*'. This table includes full records of MEDLINE papers, and can be used to find bursting terms from MEDLINE abstracts dealing with RNAi.

Load the file as a standard csv and run '*Preprocessing* > *Topical* > *Normalize Text*' with the default separator and the "abstract" box checked. Run '*Analysis* > *Topical* > *Burst Detection*' with "date_cr_year" in the Date Column and "abstract" in the Text Column, leaving the rest of the values default.

Right click on "Burst detection analysis (date_cr_year, abstract): maximum burst level 1" in the Data Manager and view the file. There are more words than can easily be viewed with the horizontal bar graph, so sort the list by "Strength" and prune all but the strongest 10 words. Save the file as a new .csv and load it into the Sci2 Tool as a standard csv file.
Select the new table in the data manager and visualize it using '*Visualize* > *Temporal* > *Horizontal Bar Graph*.'





Katy Börner (2010) Science of Science Research and Tools (12 Tutorials). Reporting Branch, Office of Extramural Research/Office of the Director, National Institutes of Health, Bethesda, MD.

- Tutorial #01: Science of Science Research
- Tutorial #02: Network Science / Information Visualization
- Tutorial #03: CIShell Powered Tools: Network Workbench and Science of Science Tool
- Tutorial #04: Temporal Analysis—Burst Detection
- Tutorial #05: Geospatial Analysis and Mapping
- Tutorial #06: Topical Analysis & Mapping
- Tutorial #07: Tree Analysis and Visualization
- Tutorial #08: Network Analysis and Visualization
- Tutorial #09: Large Network Analysis and Visualization
- Tutorial #10: Using the Scholarly Database at IU
- Tutorial #11: VIVO National Researcher Networking
- Tutorial #12: Future Developments

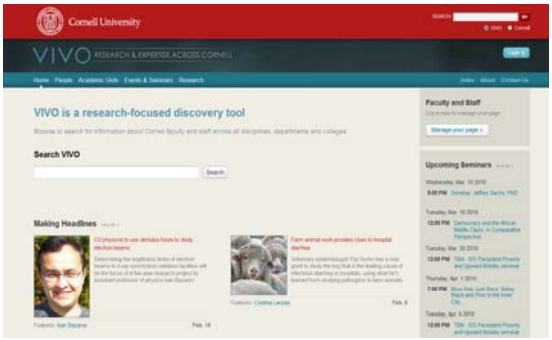
http://sci.slis.indiana.edu/sci2

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

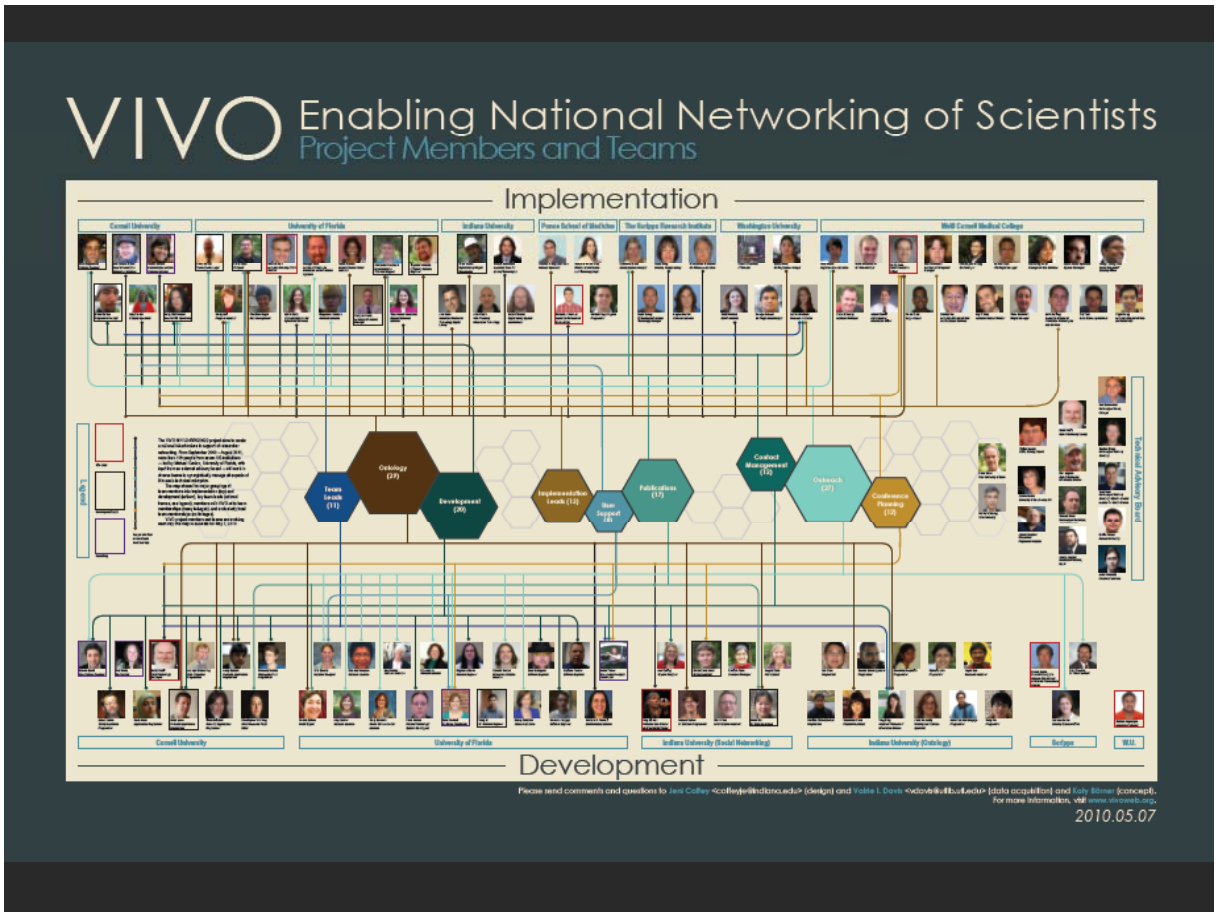
- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.

Soon:

- Simplify reporting tasks, e.g., generate biosketch, department report.



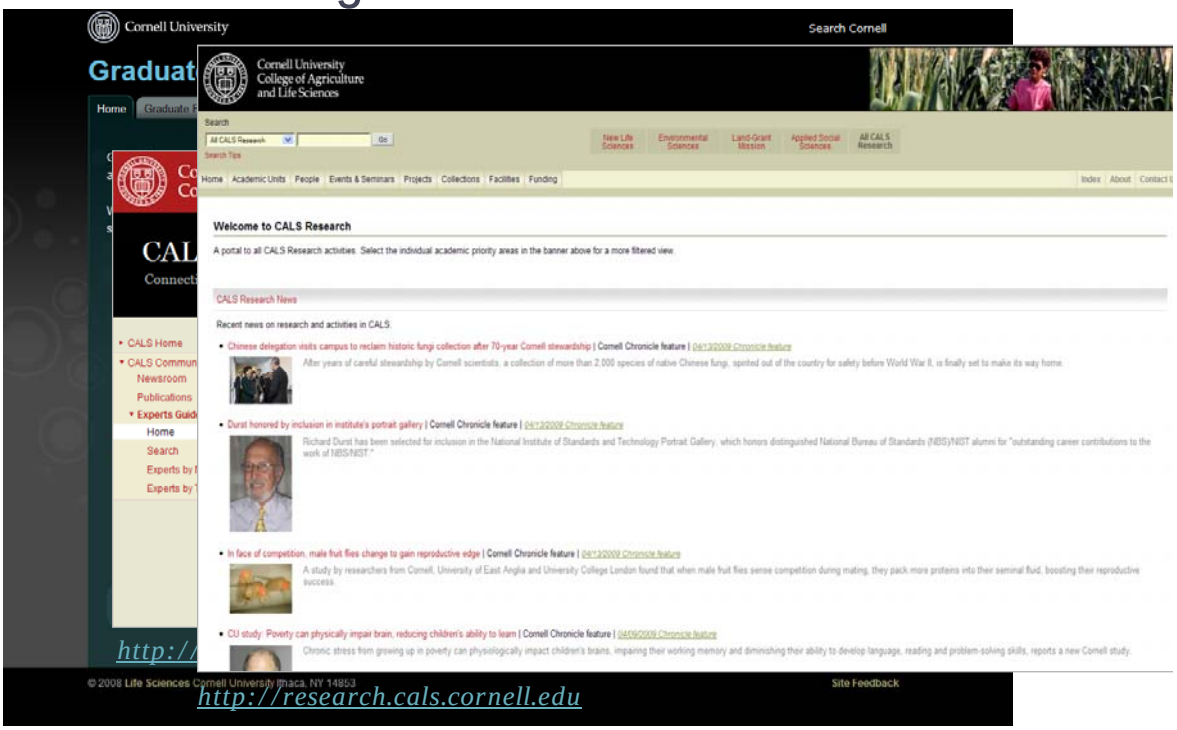
Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO 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 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://research.cals.cornell.edu

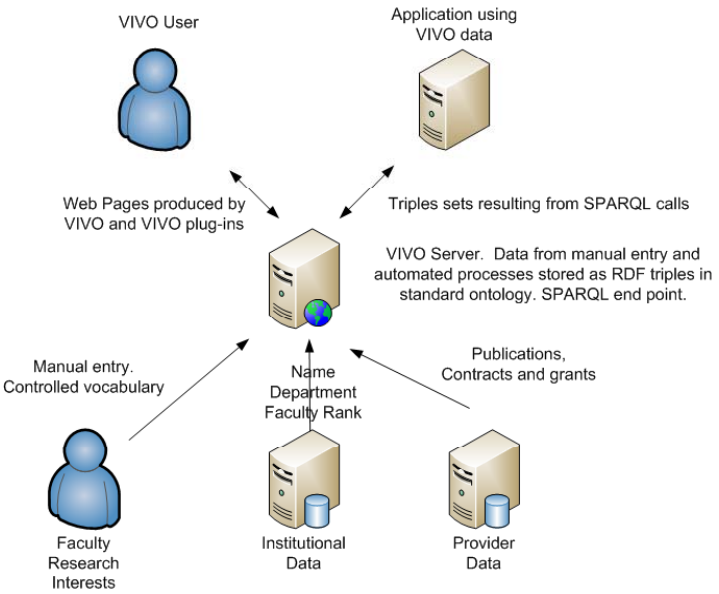
http://gradeducation.lifesciences.cornell.edu

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

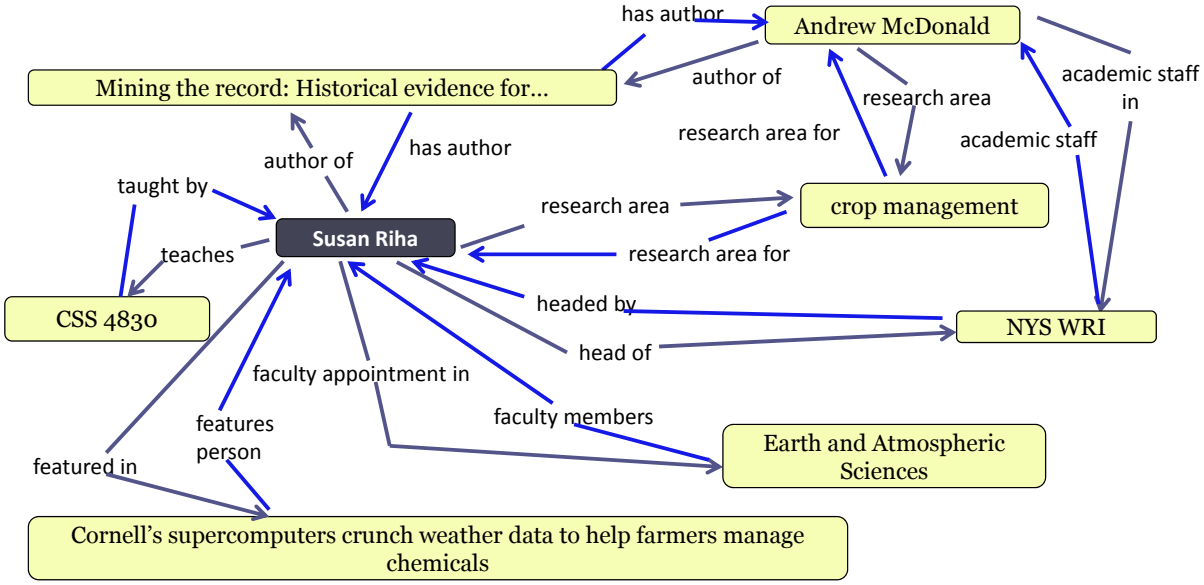
Institutional Architecture

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



Data Representation using RDF Triples

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



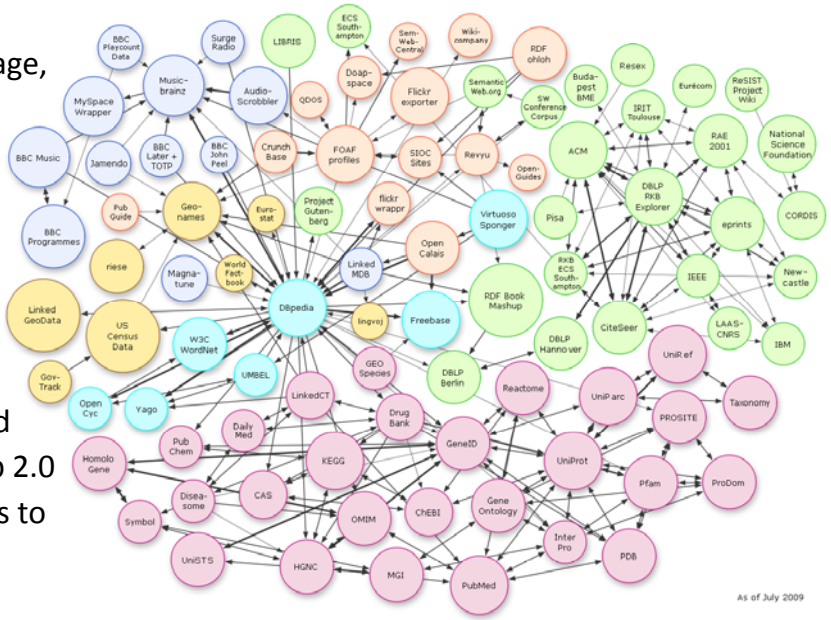
VIVO & Linked Open Data

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

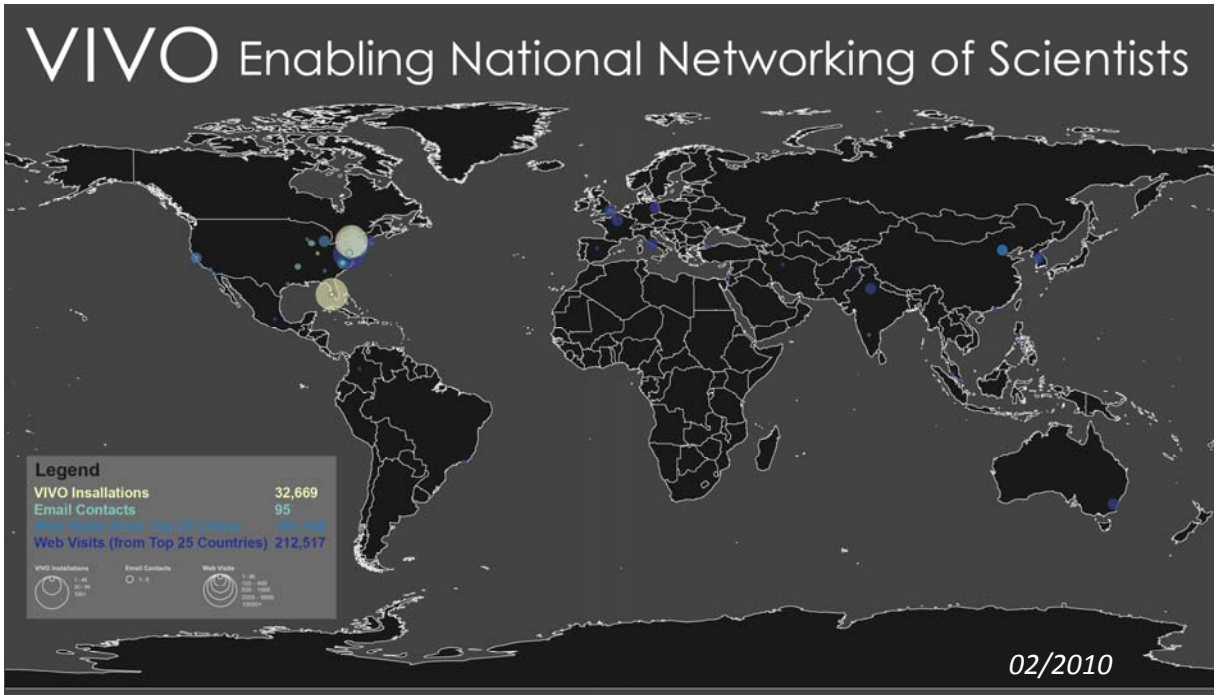
VIVO makes high coverage, high quality data from systems of record

- 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.



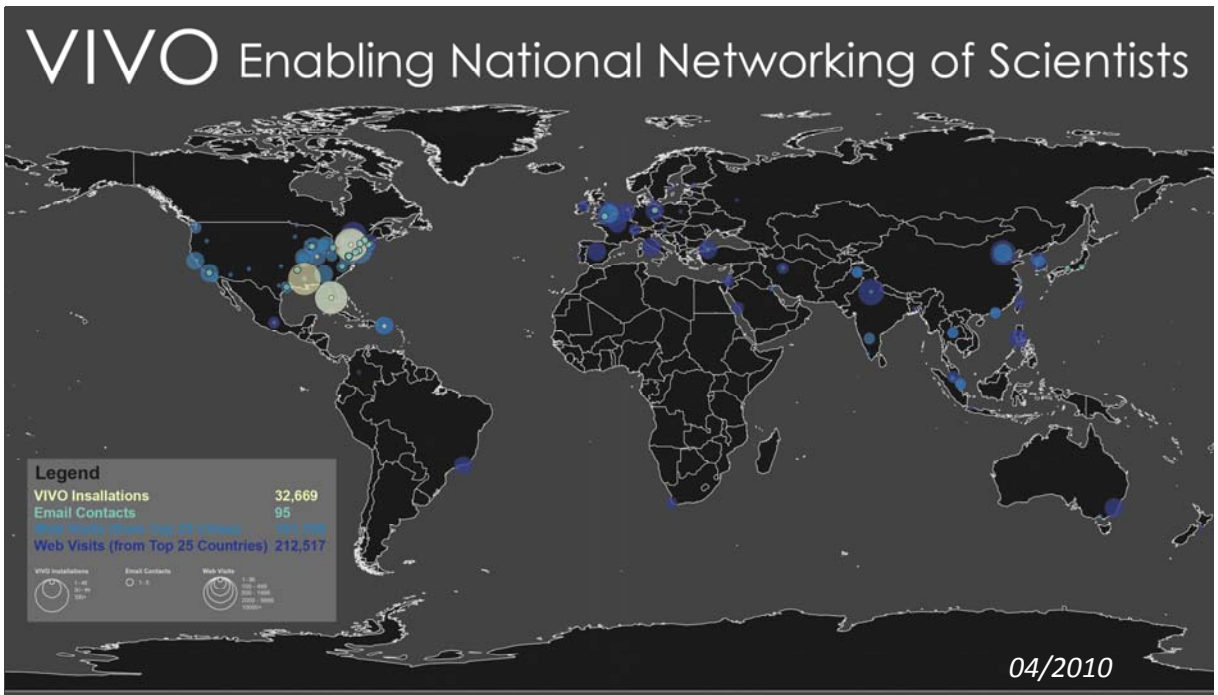
http://www4.wiwiss.fu-berlin.de/bizer/pub/lod-datasets_2009-07-14_colored.png



Visualization created by: Katy Börner (concept), Jeni Coffey (design), Kaveh Ekbia (ArcGIS) and Justin Peters (ArcGIS).

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

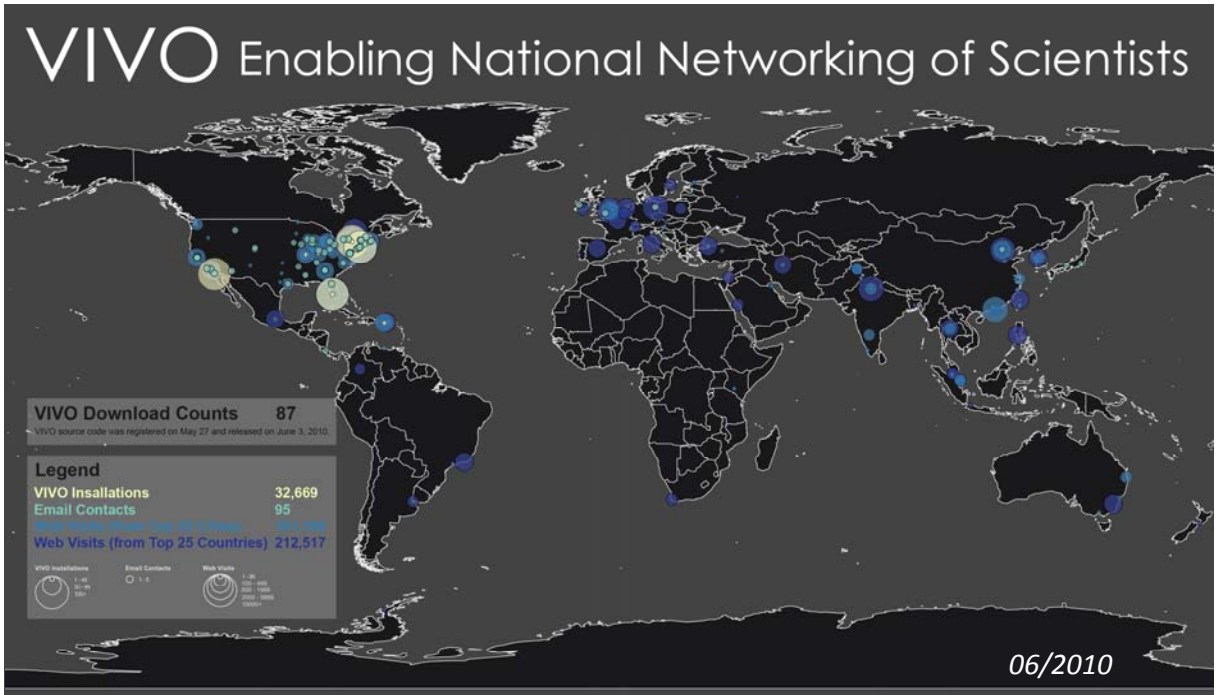
DRAFT



Visualization created by: Katy Börner (concept), Jeni Coffey (design), Kaveh Ekbia (ArcGIS) and Justin Peters (ArcGIS).

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

DRAFT



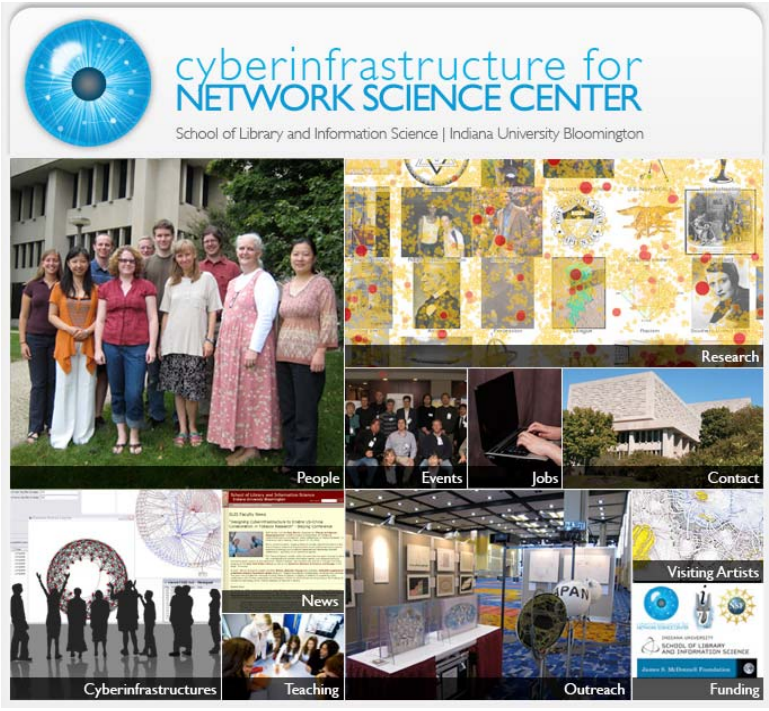
Visualization created by: Katy Börner (concept), Jeni Coffey (design), Kaveh Ekbia (ArcGIS) and Justin Peters (ArcGIS).

VIVO 1.0 source code was publicly released on April 14, 2010

DRAFT

87 Downloads by June 11, 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.



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