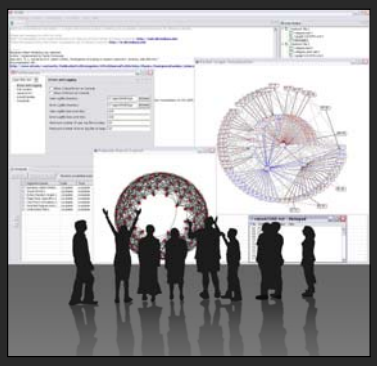
Towards a Macroscope for Science Policy Decision Making

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

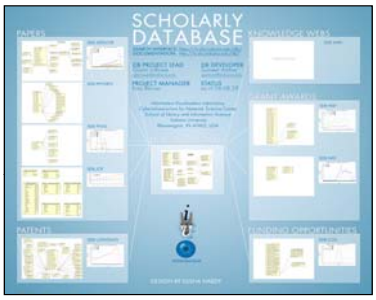
Kevin Boyack
SciTech Strategies Inc.
kboyack@mapofscience.com



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.

October 19, 2010
AAAS

Computational Scientometrics Cyberinfrastructures



Scholarly Database: 25 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

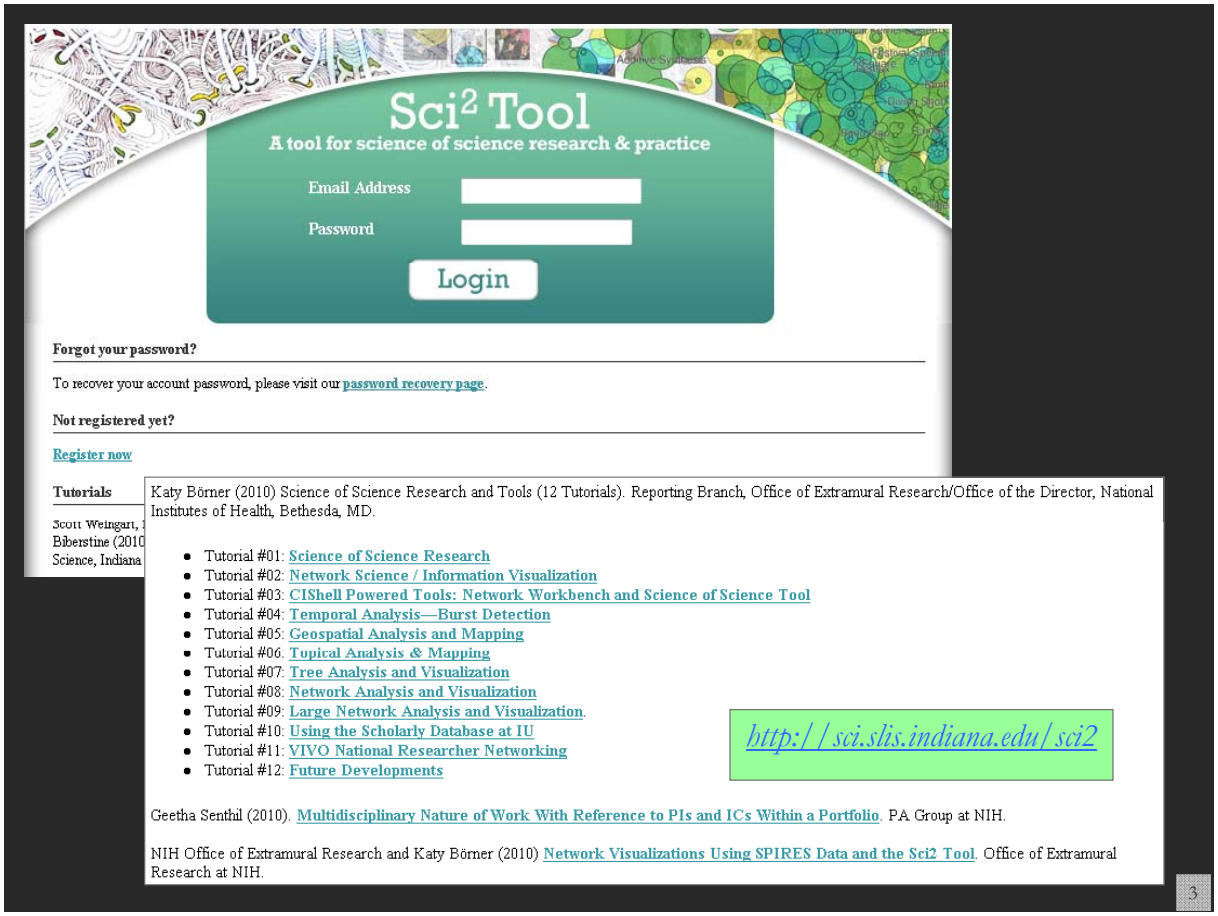
Network Workbench Tool & Community Wiki
http://nwb.slis.indiana.edu

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

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

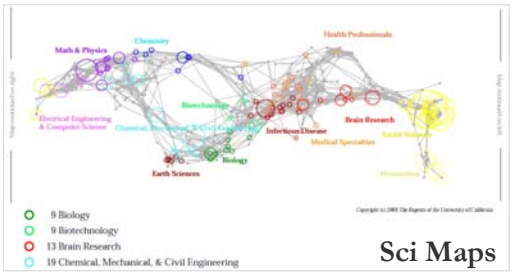
OSGi Alliance

CIShell cyberinfrastructure shell

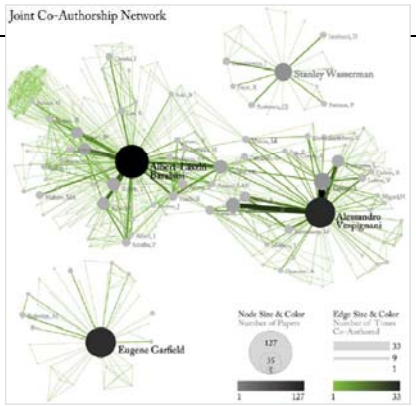


Sci² Tool – "Open Code for S&T Assessment"

OSGi/CIShell powered tool with NWB plugins and many new scientometrics and visualizations plugins.



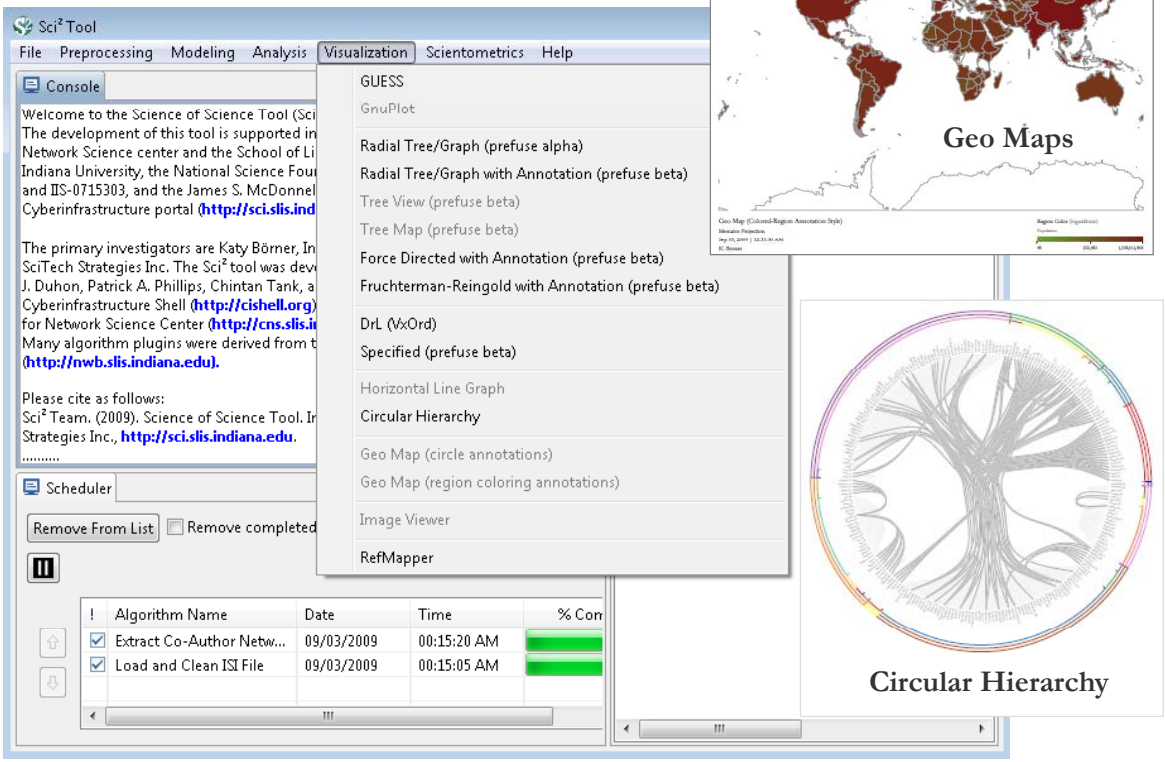
Sci Maps



GUESS Network Vis

Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Duhon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of ISSI 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17 . Vol. 2, pp. 619-630.

Sci² Tool



Sci² Tool: Algorithms

See https://nwb.slis.indiana.edu/community

Preprocessing
Extract Top N% Records
Extract Top N Records
Normalize Text
Slice Table by Line

Extract Top Nodes
Extract Nodes Above or Below Value
Delete Isolates

Extract top Edges
Extract Edges Above or Below Value
Remove Self Loops
Trim by Degree
MST-Pathfinder Network Scaling
Fast Pathfinder Network Scaling

Snowball Sampling (in nodes)
Node Sampling
Edge Sampling

Symmetrize
Dichotomize
Multipartite Joining

Geocoder

Extract ZIP Code

Modeling
Random Graph
Watts-Strogatz
Small World
Barabási-Albert Scale-Free
TARL

Analysis
Network Analysis Toolkit (NAT)
Unweighted & Undirected
- Node Degree
- Degree Distribution
- ---
- K-Nearest Neighbor (Java)
- Watts-Strogatz Clustering Coefficient
- Watts Strogatz Clustering Coefficient over K
- ---
- Diameter
- Average Shortest Path
- Shortest Path Distribution
- Node Betweenness Centrality
- ---
- Weak Component Clustering
- Global Connected Components
- ---
- Extract K-Core
- Annotate K-Coreness
- ---
- HITS

Weighted & Undirected
- Clustering Coefficient
- Nearest Neighbor Degree
- Strength vs Degree
- Degree & Strength
- Average Weight vs End-point Degree
- Strength Distribution
- Weight Distribution
- Randomize Weights
- ---
- Blondel Community Detection
- ---
- HITS

Unweighted & Directed
- Node Indegree
- Node Outdegree
- Indegree Distribution
- Outdegree Distribution
- ---
- K-Nearest Neighbor
- Single Node in-Out Degree Correlations
- ---
- Dyad Reciprocity
- Arc Reciprocity
- Adjacency Transitivity
- ---
- Weak Component Clustering
- Strong Component Clustering
- ---

Sci² Tool: Algorithms cont.

See https://nwb.slis.indiana.edu/community

Extract K-Core
Annotate K-Coreness

HITS
PageRank

Weighted & Directed
HITS
Weighted PageRank

Textual
Burst Detection

Visualization
GnuPlot
GUESS
Image Viewer

Radial Tree/Graph (prefuse alpha)
Radial Tree/Graph with Annotation (prefuse beta)
Tree View (prefuse beta)
Tree Map (prefuse beta)
Force Directed with Annotation (prefuse beta)
Fruchterman-Reingold with Annotation (prefuse beta)

DrL (VxOrd)
Specified (prefuse beta)

Horizontal Line Graph
Circular Hierarchy
Geo Map (Circle Annotation Style)
Geo Map (Colored-Region Annotation Style)
Science Map (Circle Annotation)

Scientometrics
Remove ISI Duplicate Records
Remove Rows with Multitudinous Fields
Detect Duplicate Nodes
Update Network by Merging Nodes

Extract Directed Network
Extract Paper Citation Network
Extract Author Paper Network

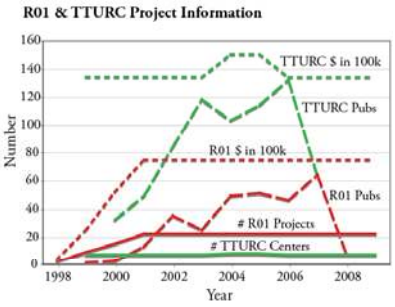
Extract Co-Occurrence Network
Extract Word Co-Occurrence Network
Extract Co-Author Network
Extract Reference Co-Occurrence (Bibliographic Coupling) Network
---
Extract Document Co-Citation Network

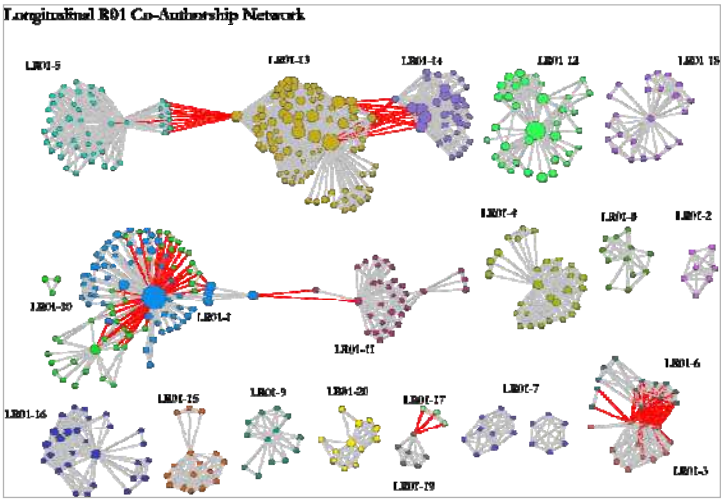
NEW:
Database support for ISI and NSF data.

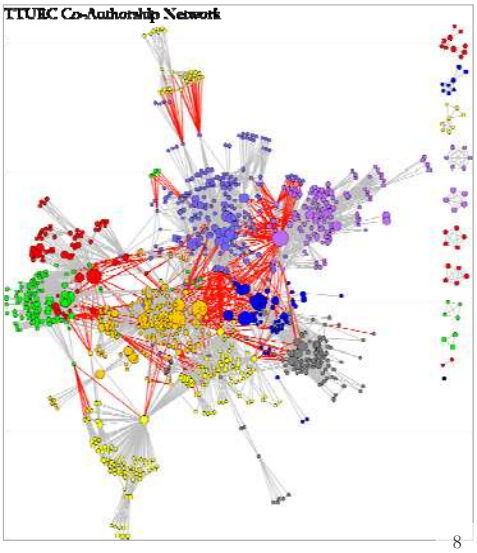
Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

Zoss & Börner, forthcoming.







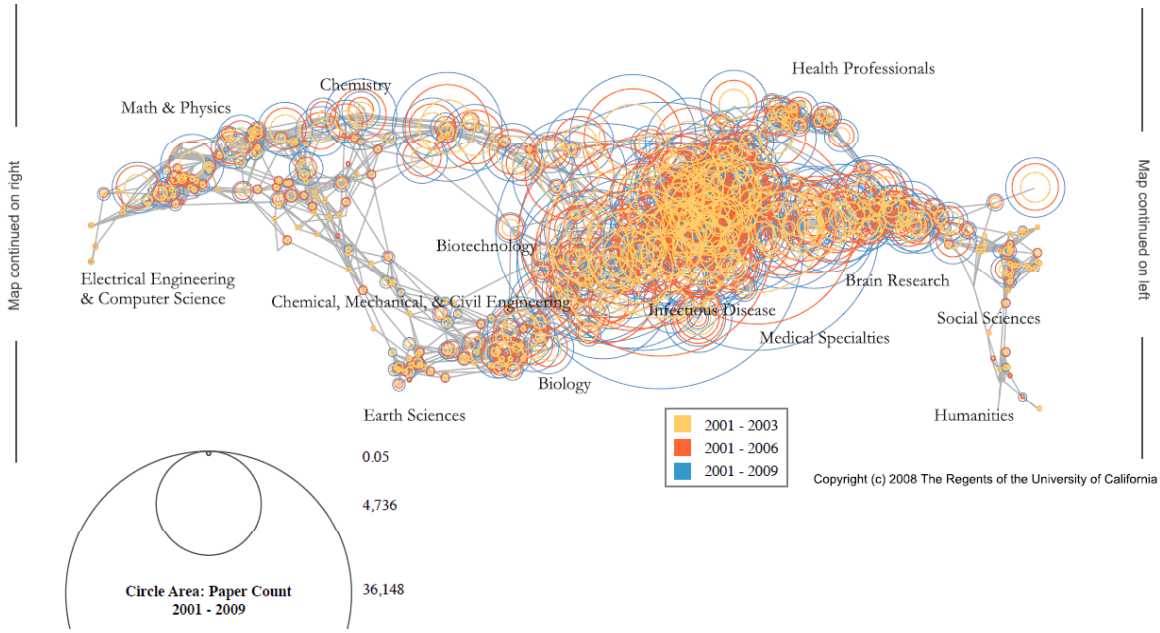
Interactive Science Map of NIH Funding

Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).

http://scimaps.org/maps/nih/2007

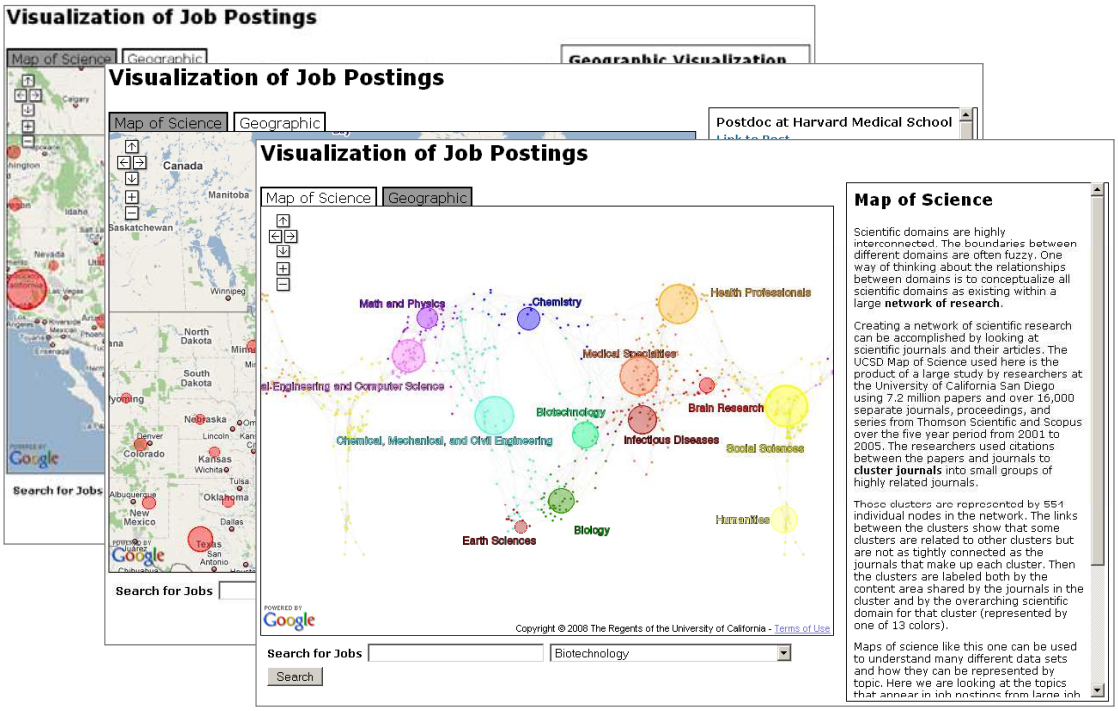
MEDLINE Publication Output by The National Institutes of Health (NIH) Using Nine Years of ExPORTER Data

Katy Börner, Nianli Ma, Joseph R. Biberstine, Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Robin M. Wagner, Rediet Berhane, Hong Jiang, Susan E. Ivey, Katrina Pearson and Carl McCabe, Reporting Branch, Division of Information Services, Office of Research Information Systems, Office of Extramural Research, Office of the Director, National Institutes of Health (NIH), Bethesda, MD.



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

Angela Zoss, Michael Connover, Katy Börner (2010)

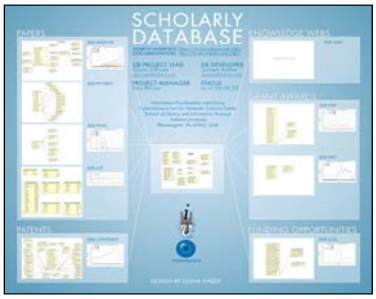


http://cns-nd3.slis.indiana.edu/mapjobs/geo

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.	research	VxOrd/Topic r NIH funding
Network Analysis (With Whom?)	NSF work of one	k	NIH's agora compatibility

Computational Scientometrics Cyberinfrastructures



Scholarly Database: 25 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

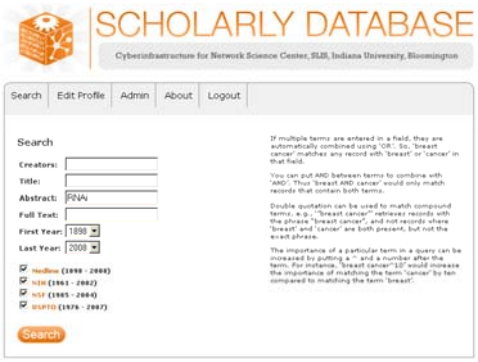
Epidemics Cyberinfrastructure
http://epic.slis.indiana.edu/

OSGi Alliance CIShell cyberinfrastructure shell

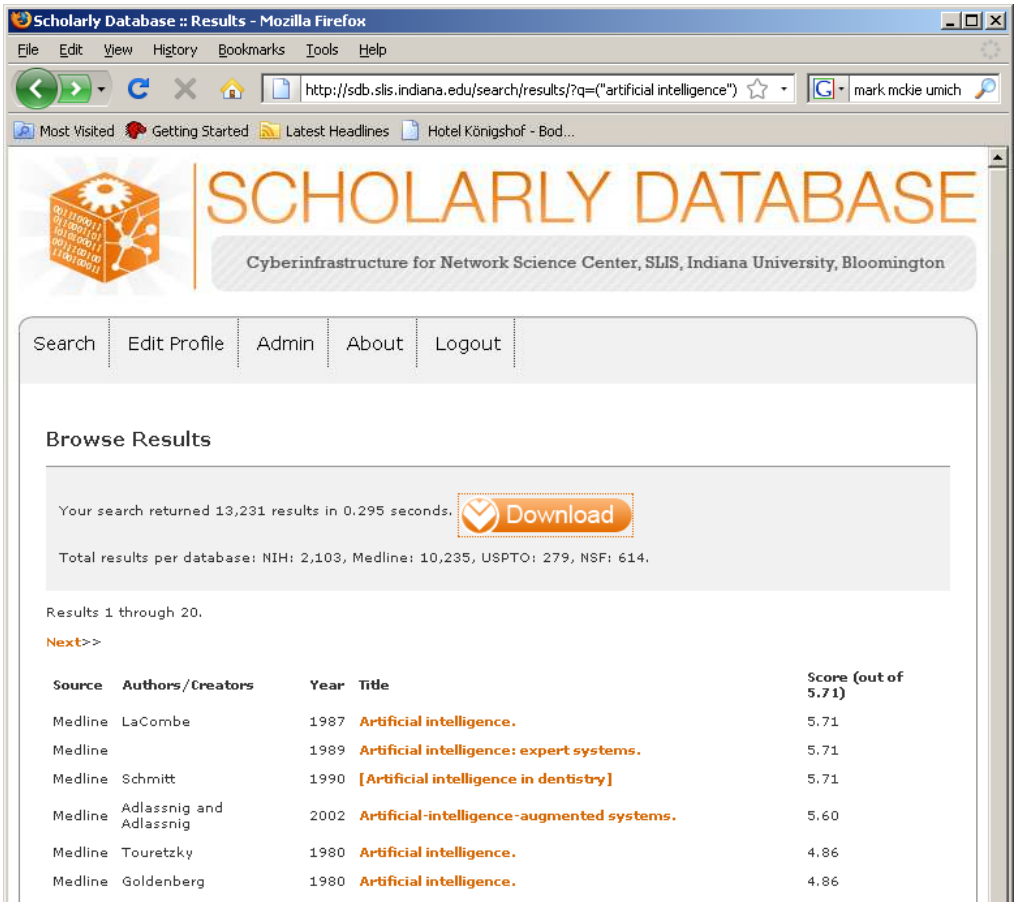
Scholarly Database: Web Interface

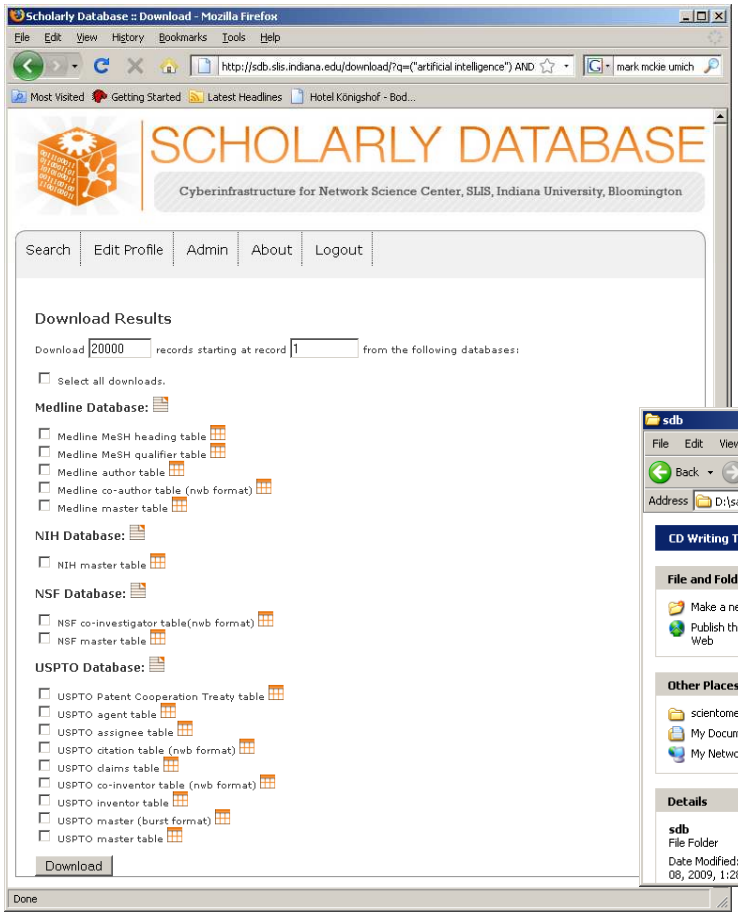
http://sdb.slis.indiana.edu

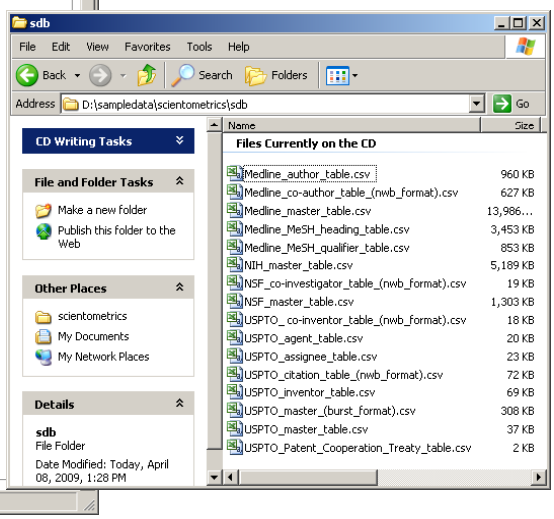
Supports federated search of 25 million publication, patent, grant records.
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.



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

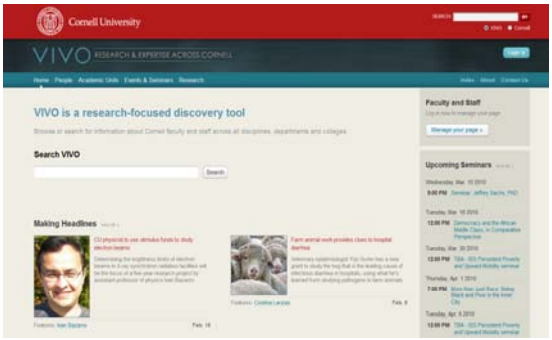
and tables for
burst analysis in NWB.

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

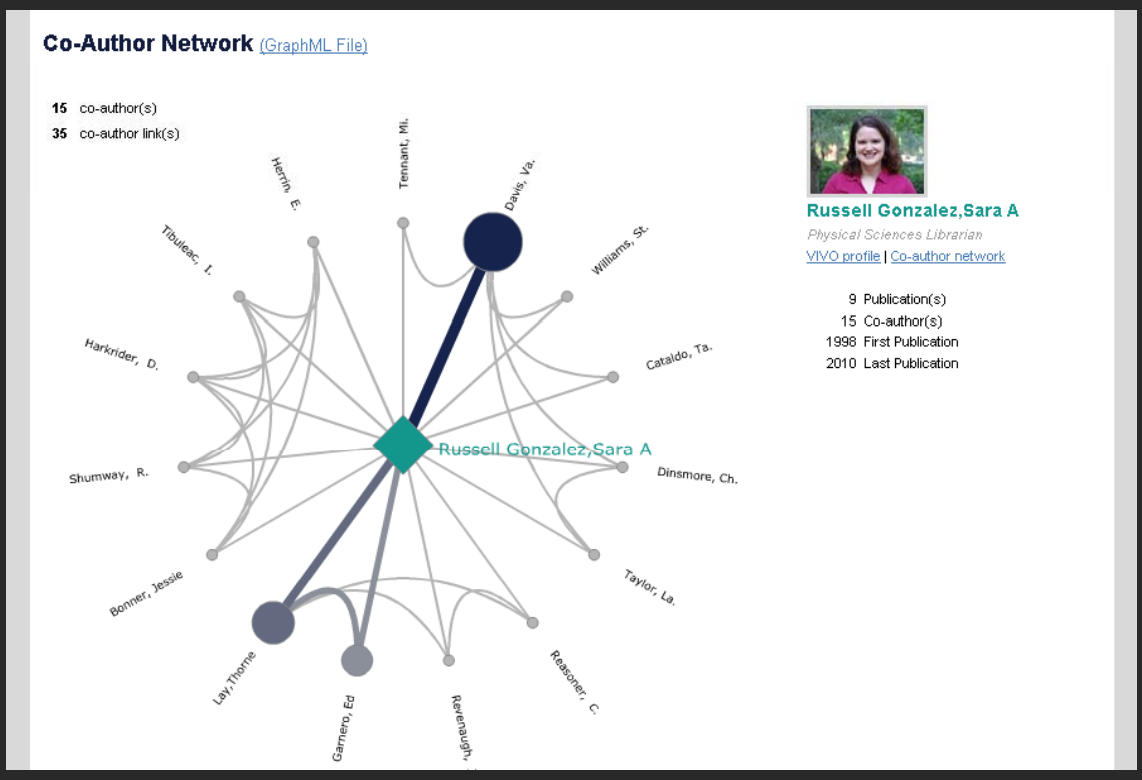
- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.

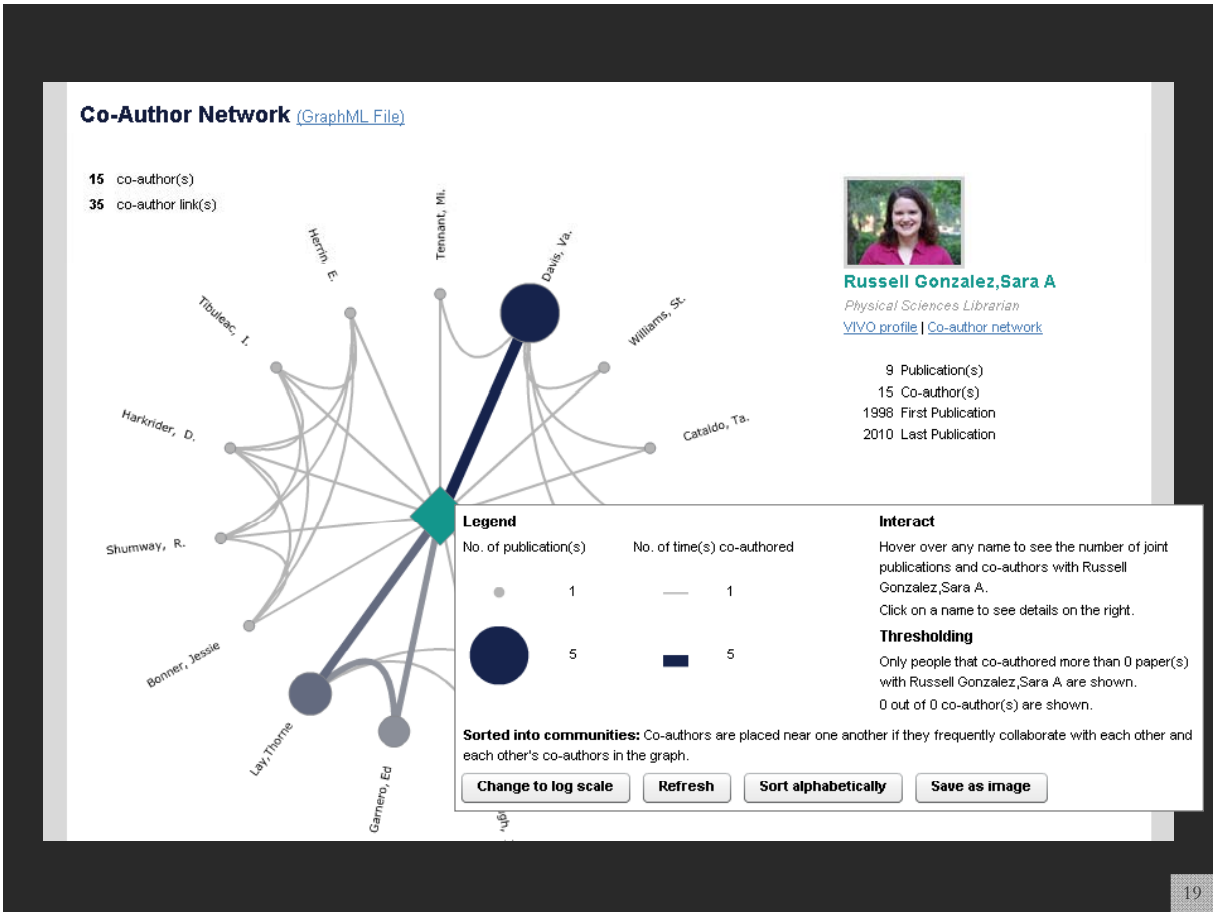
Soon:

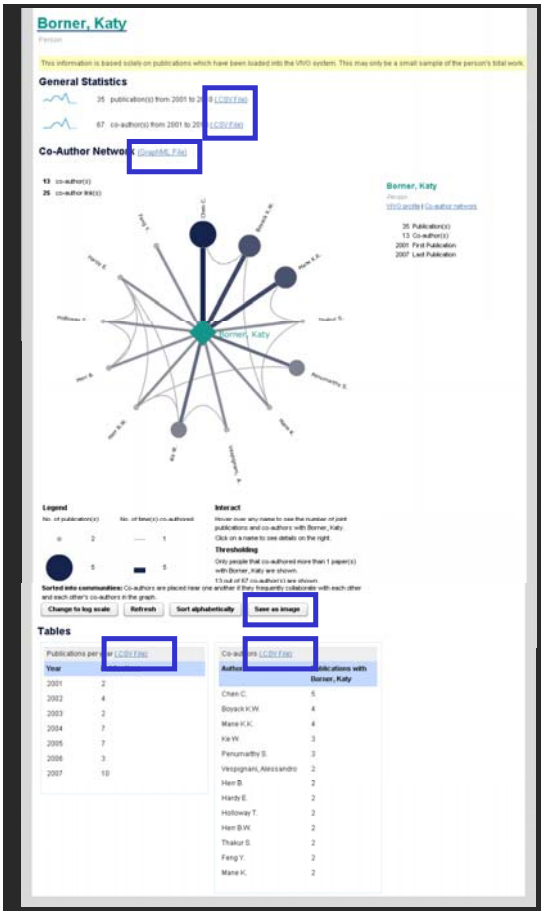
- Simplify reporting tasks, e.g., generate biosketch, department report.



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.







Download Data

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)

Save as Image (.PNG file)

Tables

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

http://vivo-vis.slis.indiana.edu/vivo1/visualization?uri=http%3A%2F%2Fvivoweb.org%2Fontology%2Fcore%2FPerson72&vis=person_level&render_mode=standalone

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

Year	Publications
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10
2010	1

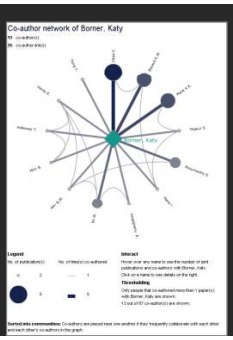
80 co-author(s) from 2001 to 2010 (.CSV File)

Year	Count	Co-Author(s)
2001	1	Chen C.
2002	3	Chen C.; McMahon T.; Feng Y.
2003	2	Chen C.; Boyack K.W.
2004	17	Sengupta A.; Penumarthy S.; Thakur S.; Sooriamurthi R.; Maru J.T.; Shiffrin R.M.; Mane K.; Moor K.A.;

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

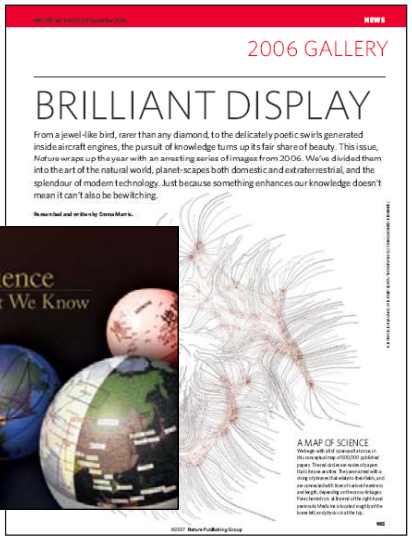
Computational Scientometrics References

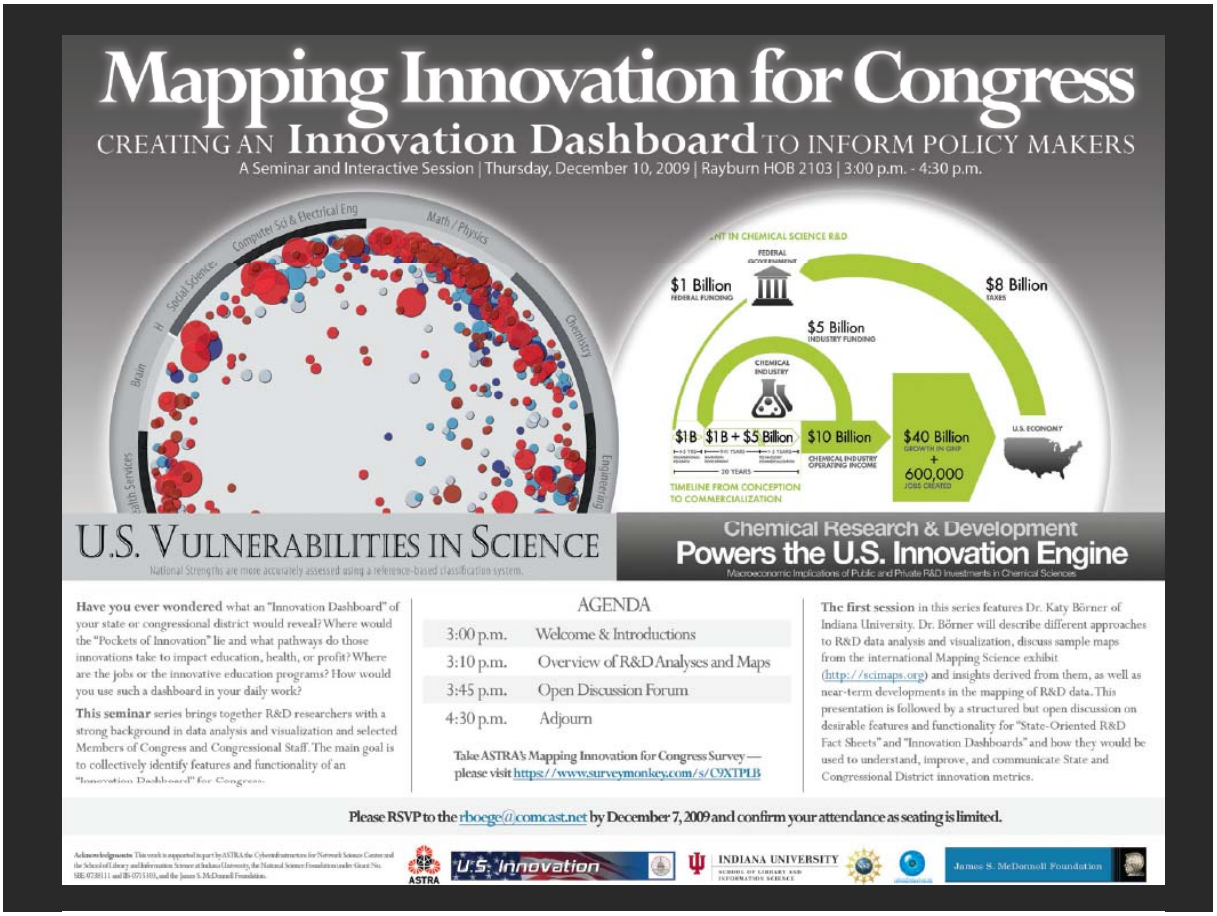
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255. http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

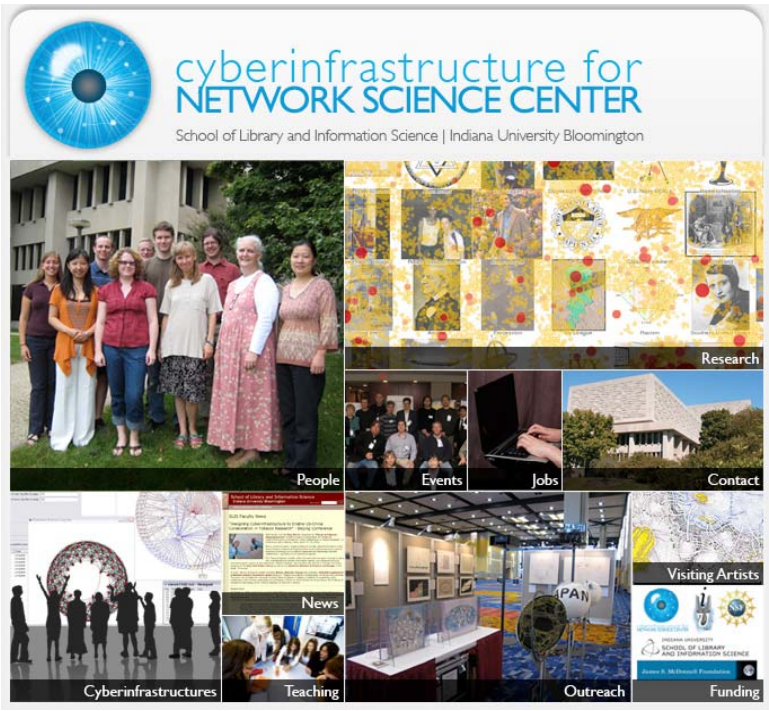
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

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







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