Science of Science Maps, Tools, and Research

Katy Börner

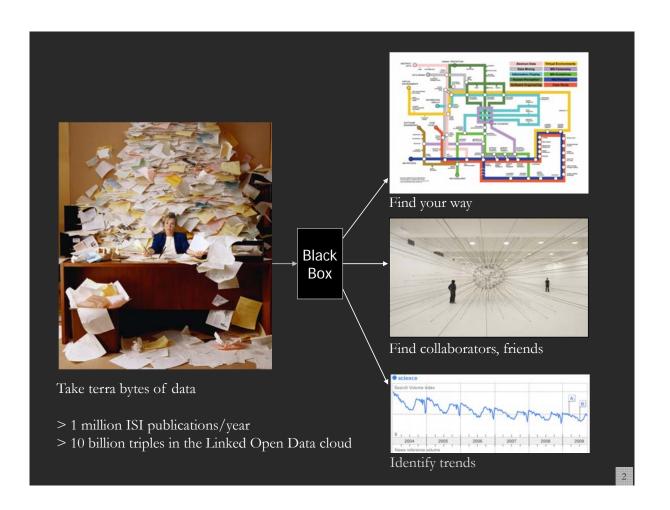
Cyberinfrastructure for Network Science Center, Director Information Visualization Laboratory, Director School of Library and Information Science Indiana University, Bloomington, IN katy@indiana.edu





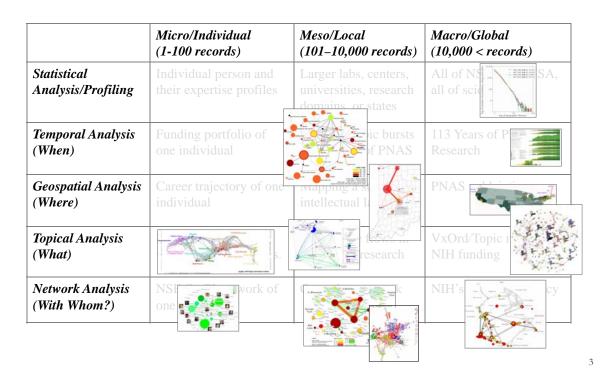
With special thanks to the members at the Cyberinfrastructure for Network Science Center, Kevin W. Boyack, the Mapping Science exhibit advisory board, and the VIVO Consortium

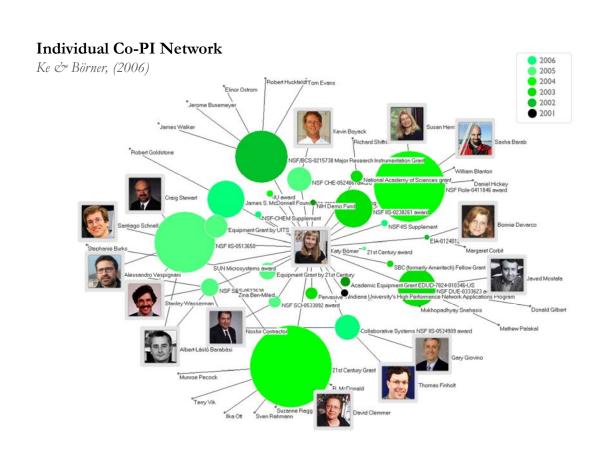
National Academy of Sciences, Washington, DC August 22, 2011

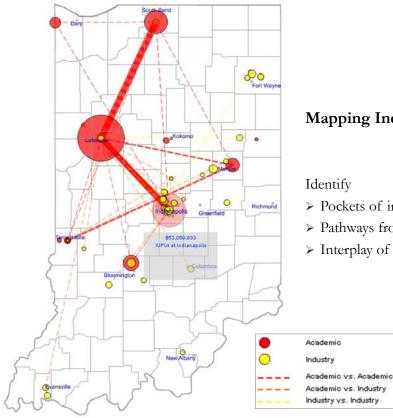




Type of Analysis vs. Level of Analysis







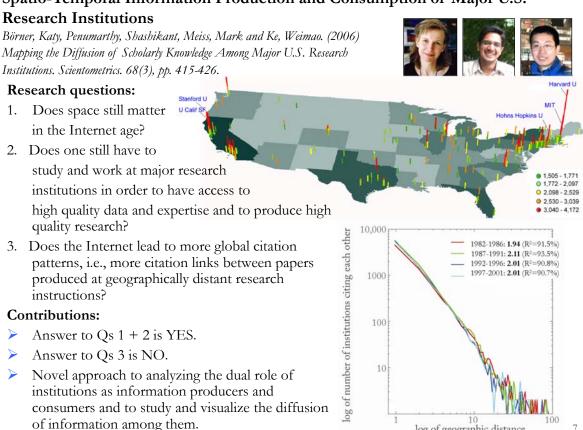
Mapping Indiana's Intellectual Space

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

5

Mapping Topic Bursts CLONING RNA ne expression regulation Co-word space of molecular weight the top 50 highly inhibition 96' frequent and bursty cultured polymerase chain reaction transfection words used in the activation top 10% most sequence homology EXPRESSION. models highly cited PNAS 92' & 94' comparative study induction nucleic acid hybridization publications in signal transduction CELL LINE 90' antibodies 1982-2001. GENES 2000' tumor cells structural inbred strains plasmids Mane & Börner. (2004) 99 MOLECULAR SEQUENCE DATA PNAS, 101(Suppl. 1): mutagenesis *5287-5290*. nucleic acid apoptosis growth-factor 95 Color Code base sequence 82 - 85 86 - 89 90 - 93 94 - 97 mino acid amino acid seguence circle size ~ burst weight n cicle color ~ burst onset ring color ~ year of max word count years of 2nd and 3rd burst are given in color sequence gene expression

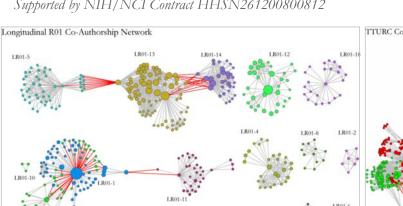
Spatio-Temporal Information Production and Consumption of Major U.S.

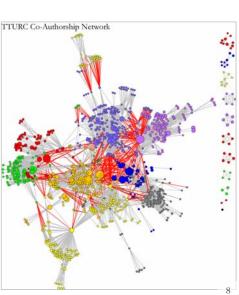


Mapping Transdisciplinary Tobacco Use Research **Centers Publications**

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

Zoss & Börner, forthcoming. Supported by NIH/NCI Contract HHSN261200800812





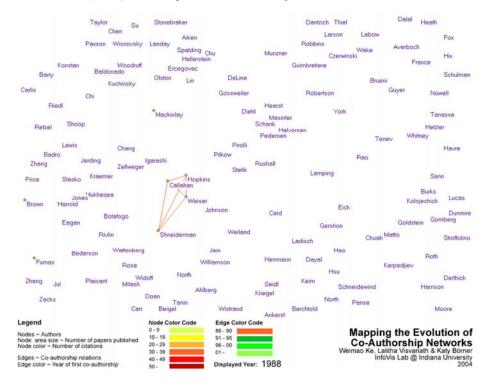
log of geographic distance

R01 & TTURC Project Information

120 ž 100 TTURC \$ in 100k

Mapping the Evolution of Co-Authorship Networks

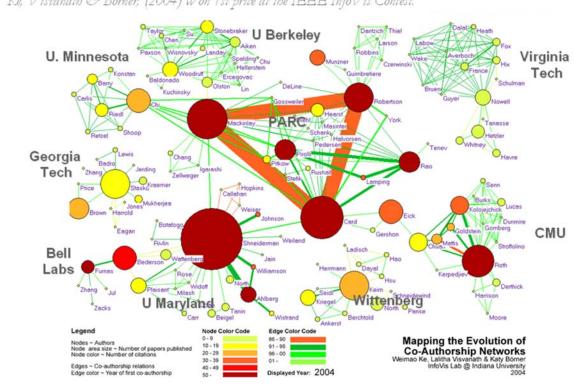
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



9

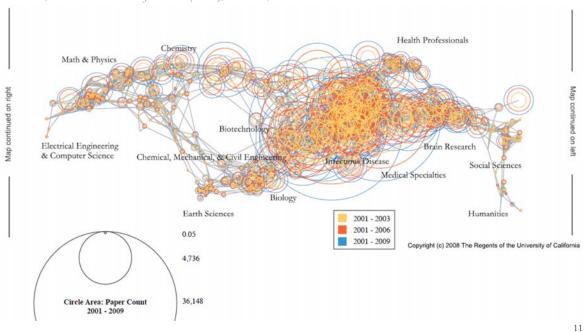
Mapping the Evolution of Co-Authorship Networks

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

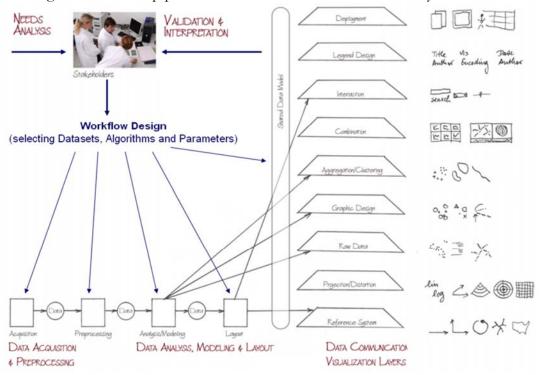


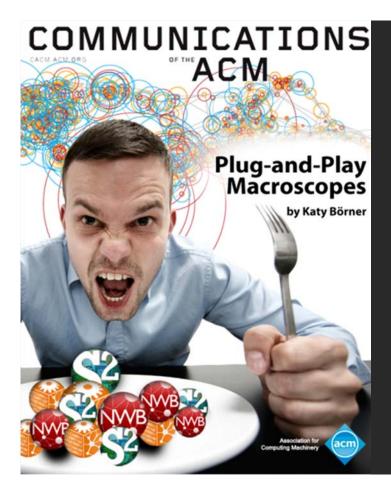
MEDLINE Publication Output by The National Institutes of Health (NIH) Using Nine Years of ExPORTER Data

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.



Needs-Driven Workflow Design using a modular data acquisition/analysis/ modeling/visualization pipeline as well as modular visualization layers.





Börner, Katy. (March 2011). Plug-and-Play Macroscopes. *Communications of the ACM*, 54(3), 60-69.

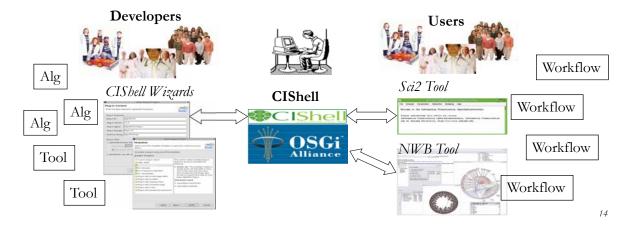
Video and paper are at http://www.scivee.tv/node/27704

13



OSGi & CIShell

- ➤ CIShell (http://cishell.org) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- ➤ It extends the Open Services Gateway Initiative (OSGi) (http://osgi.org), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- > Specifically, CIShell provides "sockets" into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.





CIShell Developer Guide

(http://cishell.wiki.cns.iu.edu)





@1 Added by Micah Linnemeier, last edited by Micah Linnemeier on Mar 16, 2011 (view change)

About the Cyberinfrastructure Shell

The Cyberinfrastructure Shell (CIShell) is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithm integration support is built in for Java and most other programming languages. Being Java based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License.

CIShell is the basis of Network Workbench, TexTrend, Scif and the upcoming EpiC tool.

CIShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

CIShell Features

A framework for easy integration of new and existing algorithms written in any programming language

Using CIShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and

Learn More...

- CIShell Papers
- CIShell Powered Tools
- Algorithms
- Plugins (coming soon)
- Misc. Tool Documentation
- · CIShell Web Services (coming soon)
- Screenshots

Getting Started...

- · Documentation & Developer Resources
- Download

Getting Involved...

· Contact Us

15



CIShell Portal (http://cishell.org)

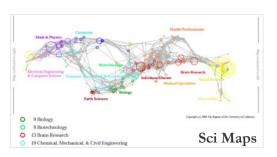


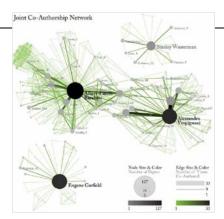


Sci² Tool - "Open Code for S&T Assessment"

http://sci2.cns.iu.edu

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

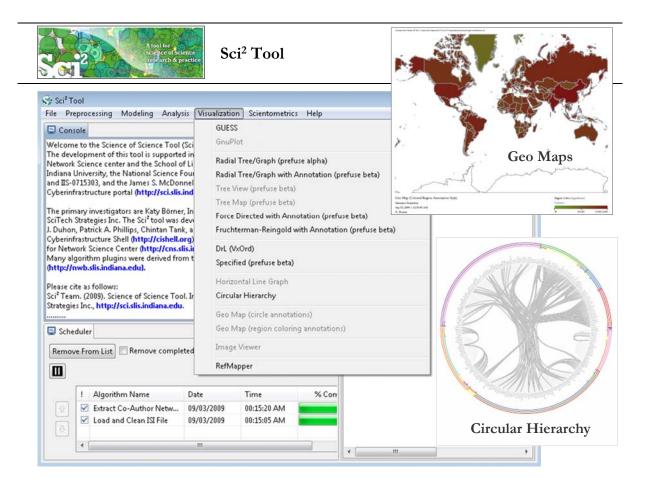


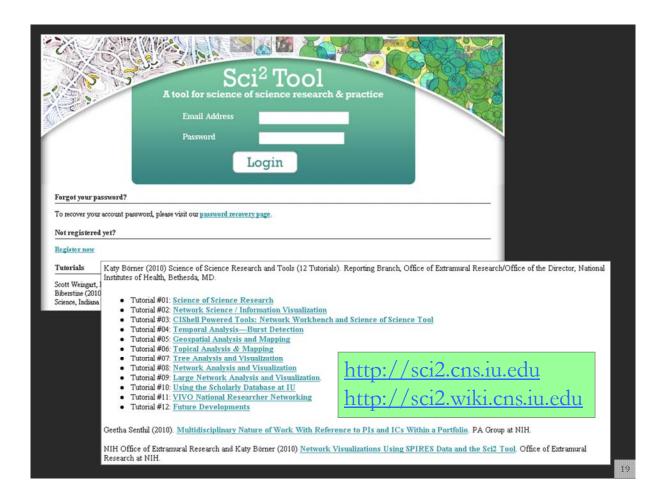


GUESS Network Vis



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: Algorithms

Modeling

Random Graph

HITS

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

Watts-Strogatz	Nearest Neighbor Degree
Small World	Strength vs Degree
Barabási-Albert Scale-Free	Degree & Strength
TARL	Average Weight vs End-point Degree
	Strength Distribution
Analysis	Weight Distribution
Network Analysis Toolkit (NAT)	Randomize Weights
Unweighted & Undirected	
Node Degree	Blondel Community Detection
Degree Distribution	
	HITS
K-Nearest Neighbor (Java)	Unweighted & Directed
Watts-Strogatz Clustering Coefficient	Node Indegree
Watts Strogatz Clustering Coefficient over K	Node Outdegree
	Indegree Distribution
Diameter	Outdegree Distribution
Average Shortest Path	
Shortest Path Distribution	K-Nearest Neighbor
Node Betweenness Centrality	Single Node in-Out Degree Correlation
Weak Component Clustering	Dyad Reciprocity
Global Connected Components	Arc Reciprocity
1	Adjacency Transitivity
Extract K-Core	
Annotate K-Coreness	Weak Component Clustering
	Strong Component Clustering

Weighted & Undirected

Clustering Coefficient



Sci² Tool: Algorithms cont.

Extract K-Core Annotate K-Coreness

HITS
PageRank
Weighted & Directed
HITS
Weighted PageRank

Textual

Burst Detection

Visualization

GnuPlot GUESS Image Viewer

DrL (VxOrd)

Specified (prefuse beta)

Horizontal Bar 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

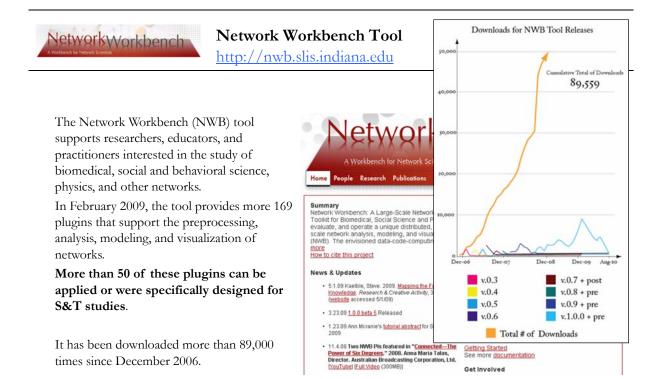
Extract Word Co-Occurrence Network
Extract Co-Author Network
Extract Reference Co-Occurrence
(Bibliographic Coupling) Network

Extract Document Co-Citation Network

Soon:

Database support for ISI and NSF data.

21



Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), Progress in Convergence - Technologies for Human Wellbeing (Vol. 1093, pp. 161-179), Annals of the New York Academy of Sciences, Boston, MA.

22

Computational Proteomics

What relationships exist between protein targets of all drugs and all disease-gene products in the human protein—protein interaction network?

Yildriim, Muhammed
A., Kwan-II Goh,
Michael E. Cusick,
Albert-László Barabási,
and Marc Vidal. (2007).
Drug-target Network.
Nature Biotechnology
25 no. 10: 1119-1126.

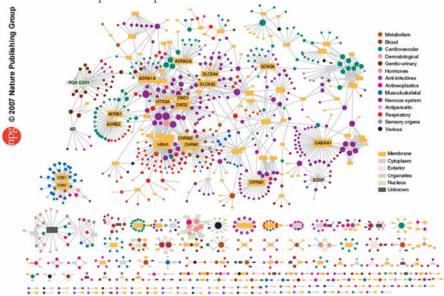




Figure 2 Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend, Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their Anatomical Therapeutic Chemical Classification, and the

23

Computational Economics Does the type of product that a country exports matter for subsequent economic performance? vegetables

C. A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann (2007) The Product Space Conditions the Development of Nations. Science 317, 482 (2007).



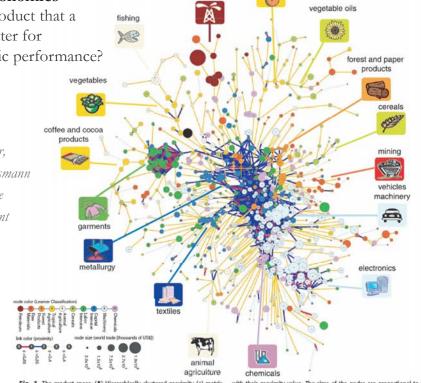


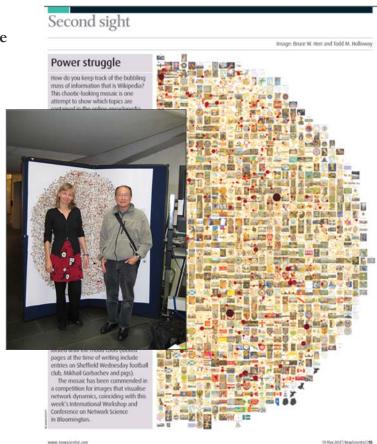
Fig. 1. The product space. (A) Hierarchically clustered proximity (φ) matrix representing the 775 SΠC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded

with their proximity value. The sizes of the nodes are proportional to world trade, and their colors are chosen according to the classification introduced by

Computational Social Science

Studying large scale social networks such as Wikipedia

Second Sight: An Emergent Mosaic of Wikipedian Activity, The New Scientist, May 19, 2007





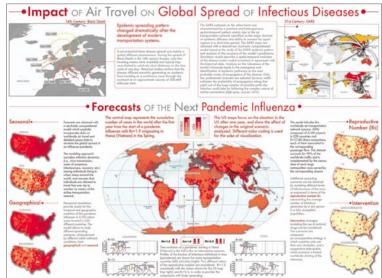
Computational Epidemics

Forecasting (and preventing the effects of) the next pandemic.

Epidemic Modeling in Complex realities, V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, Comptes Rendus Biologie, 330, 364-374 (2007).

Reaction-diffusion processes and metapopulation models in heterogeneous networks, V.Colizza, R. Pastor-Satorras, A.Vespignani, Nature Physics 3, 276-282 (2007).

Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions, V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, A. Vespignani, PloS-Medicine 4, e13, 95-110 (2007).





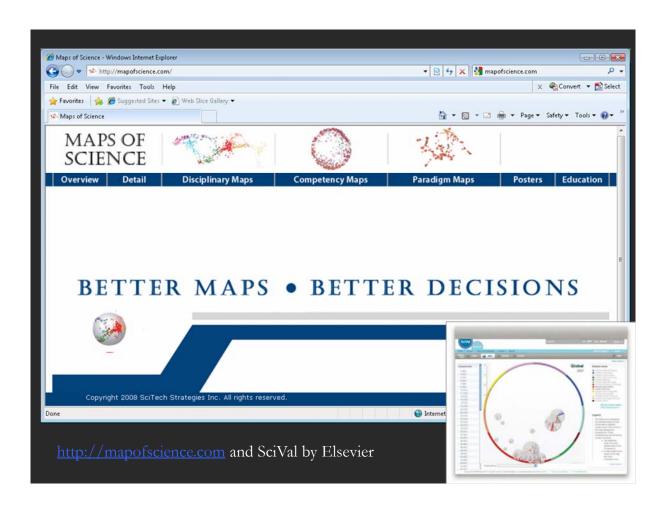


OSGi/CIShell Adoption

A number of other projects recently adopted OSGi and/or CIShell:

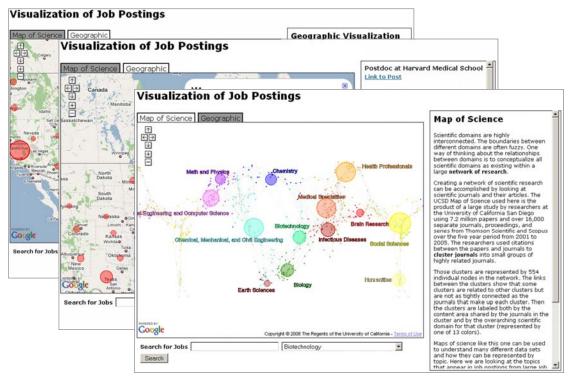
- Cytoscape (<u>http://cytoscape.org</u>) Led by Trey Ideker at UCSD is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).
- Taverna Workbench (http://taverna.org.uk) Developed by the myGrid team (http://mygrid.org.uk) led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.
- MAEviz (https://wiki.ncsa.uiuc.edu/display/MAE/Home) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.
- TEXTrend (http://textrend.org) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.
- DynaNets (<u>http://www.dynanets.org</u>) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
- > SISOB (<u>http://sisob.lcc.uma.es</u>) An Observatory for Science in Society Based in Social Models.

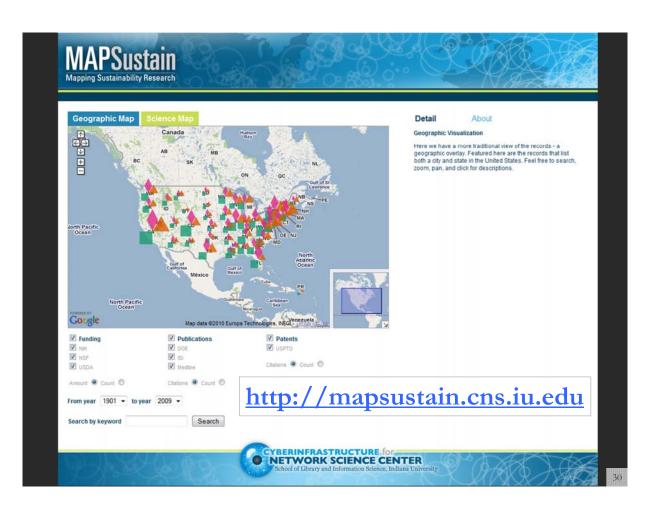
As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.



Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010)

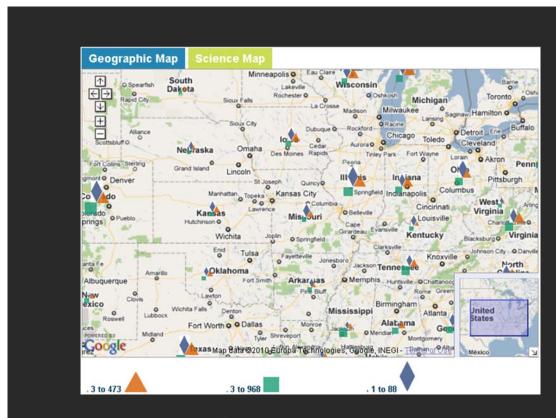




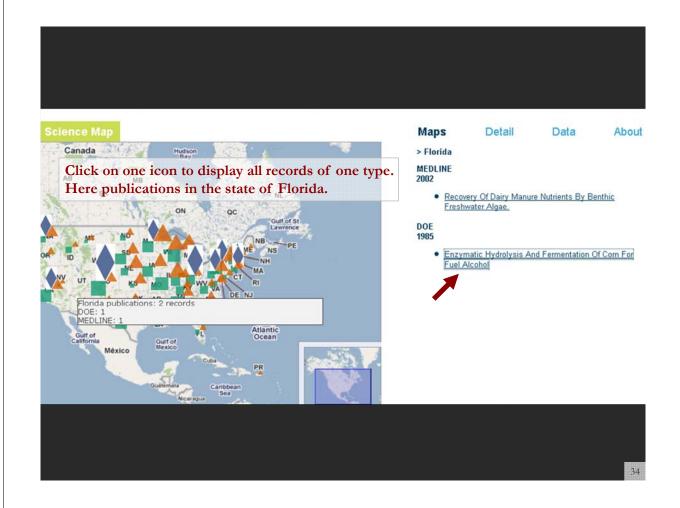


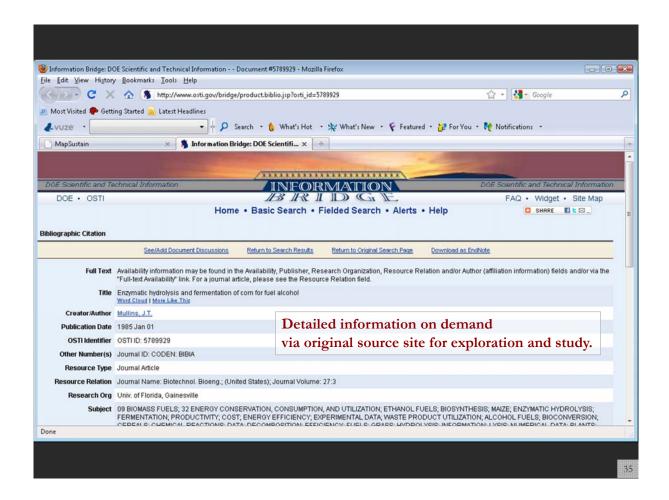
The geographic map at state level.

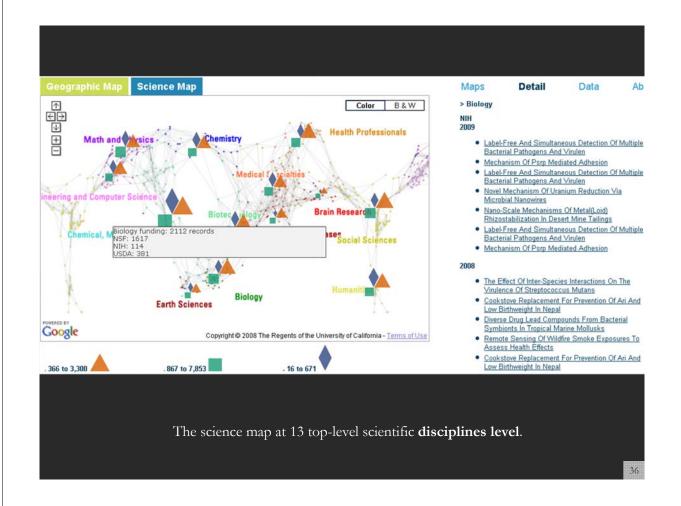


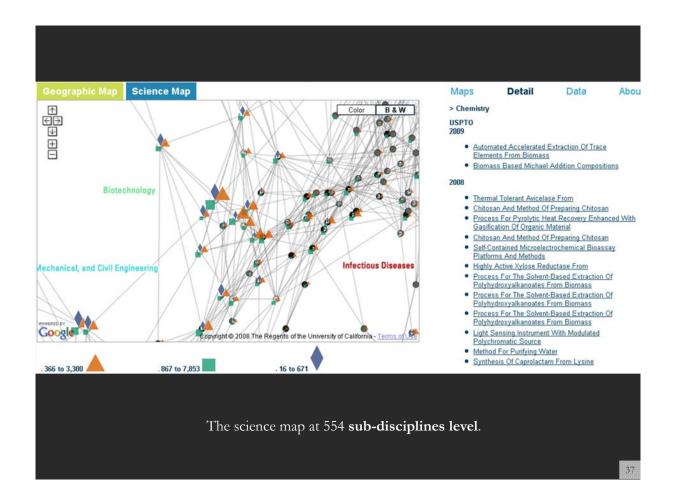


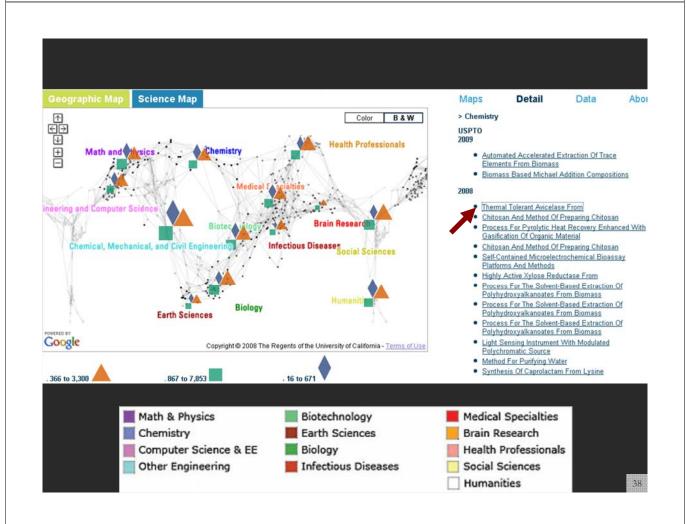


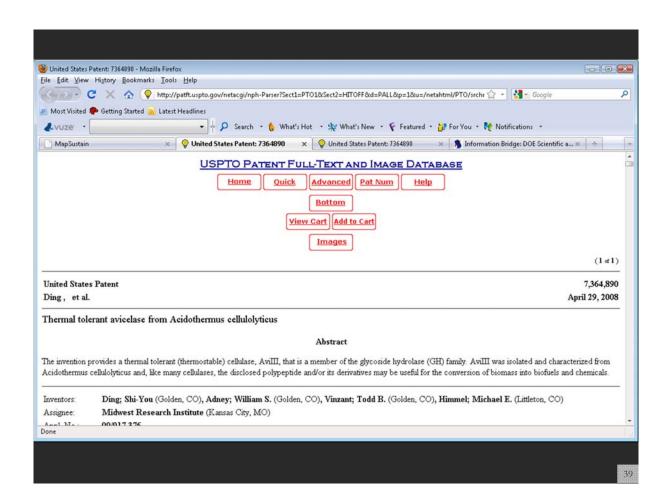


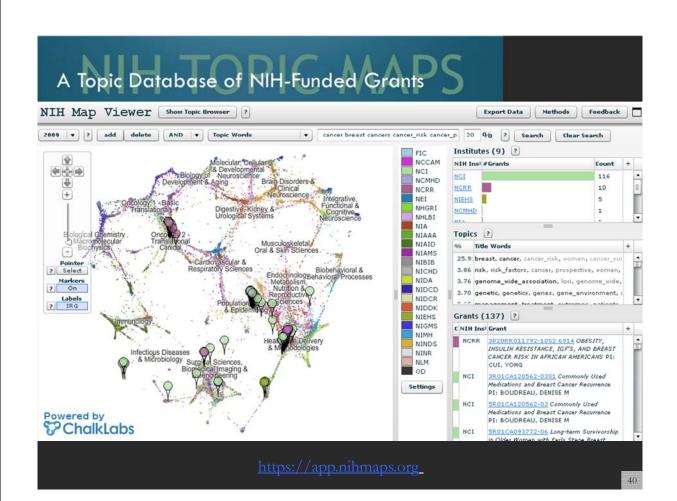










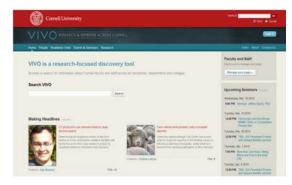


A Topic Database of NIH-Funded Grants NIH Topic Browser Show Map Viewer ? Export Data Methods Feedback Topics by NIH Institute Topics by Category 2009 | ▼ | ? | add | delete | AND | ▼ | Exact Text Search Clear Search 2009 Grants (137) Institutes (9) Col NIH Inst Project/Subproje Title Investigator(s) #1 Topic ▼ #1 Topic Worc + NIH Inst #Grants Count + 3P20RR011792- OBESITY, INSULIN RESISTANCE, IGF'S, AND BREAST NCRR CUI, YONG 686 cancer brea... NCI 116 1082 6914 CANCER RISK IN AFRICAN AMERICANS (50%) NCRR 10 3R01CA120562-Commonly Used Medications and Breast Cancer BOUDREAU, DENISE M NIEHS (42%) 0351 NCMHD Commonly Used Medications and Breast Cancer BOUDREAU, DENISE M 5R01CA120562-NIA 1 Recurrence (42%) cancer brea.. 5R01CA093772-Long-term Survivorship in Older Women with Early SILLIMAN, REBECCA A 686 NCCAM 1 (42%) NICHD 1 5R01CA064277- Shanghai Breast Cancer Study NINR 1 (41%) NHGRI Show Top 100 on Map Topics Similar Grants Topic Words Title Words Simil: C NIH Inst Grant 1R01CA129639-01A2 Genome-Wide Association Study of Radiation Exposure and Bilateral Breast Cancer PI: BERNSTEIN, JONINE LISA 25.91 6.51 NCI 686 437 risk risk factors cases cohort prospective high ris risk, risk factors, 3.86 6.46 NCI 1K07CA136758-01A1 Genetic variants in the PI3K pathway in 3.76 544 snps snp genome_wide_association cases genes genome_wide_ass nammographic density and breast cancer PI: THOMPSON, CHERYL L. genetic genes risk susceptibility polymorphisms (genetic, genetics, 3.70 6.31 NCI 5P50CA116199-05 UTMDACC SPORE in Breast Cancer PI: HORTOBAGYI, treatment patients management patient outcome management, trea conference meeting workshop symposium scienti th, conference, syr 1.64 235 6.02 NCI 2R01CA050385-21A1 Risk Factors for Breast Cancer in Younger Nurses PI: WILLETT, WALTER C. 1.63 351 community implementation community based he community, preve 5R01CA127617-02 Who Cares For Older Breast Cancer Surivors And How Does It Affect Quality? PI: MANDELBLATT, JEANNE 1.54 million disease treatment united_states public_he disease, treatmen https://app.nihmaps.org



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



Funded by \$12 million NIH award.

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.

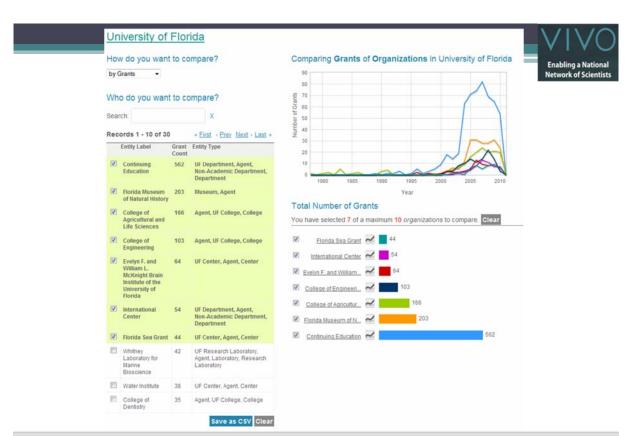




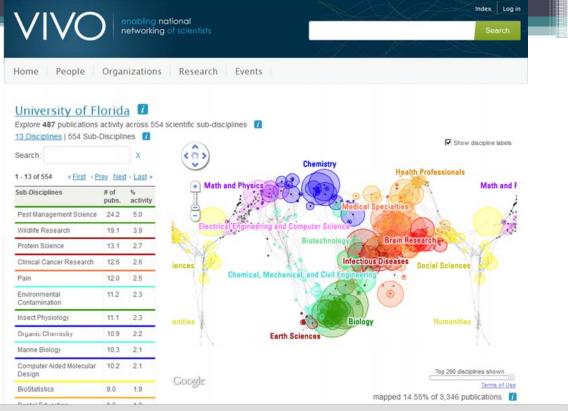




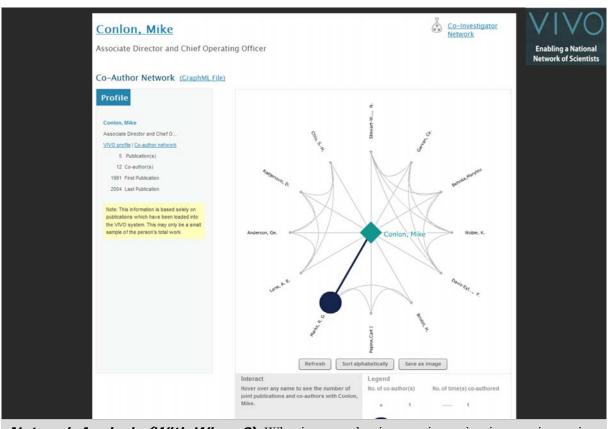




Temporal Analysis (When) Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level



Topical Analysis (What) Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)



Network Analysis (With Whom?) Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?





Second Annual VIVO Conference

August 24-26, 2011

Gaylord National, Washington D.C.

http://vivoweb.org/conference



Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org





Mapping Science Exhibit at MEDIA X, Wallenberg Hall, Stanford University http://scaleindependentthought.typepad.com/photos/scimaps

40

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)

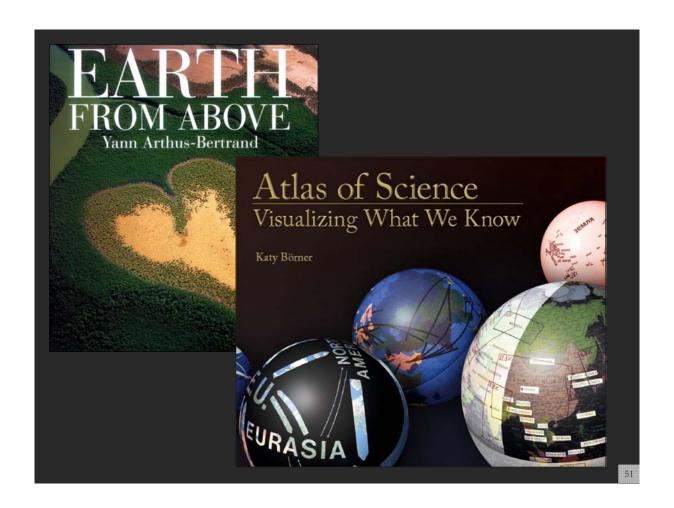
Towards Science Mapping Standards (2014)

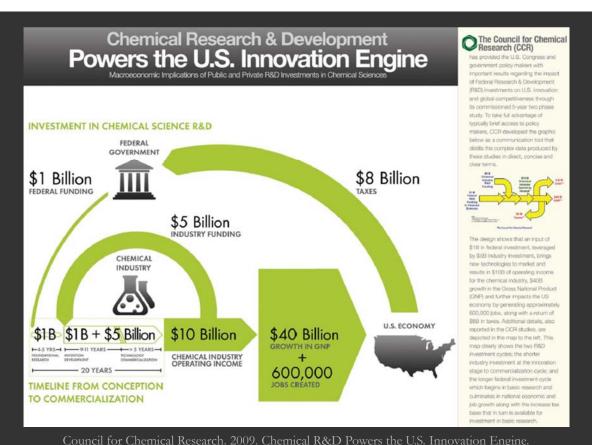
Exhibit has been shown in 72 venues on four continents. Currently at

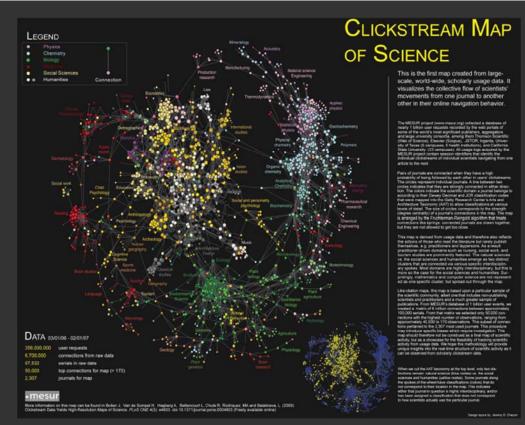
- NSF 10th Floor 4201 Wilson Boulevard, Arlington, VA
- Center of Advanced European Studies and Research, Bonn, Germany
- University of North Texas, Denton, Texas





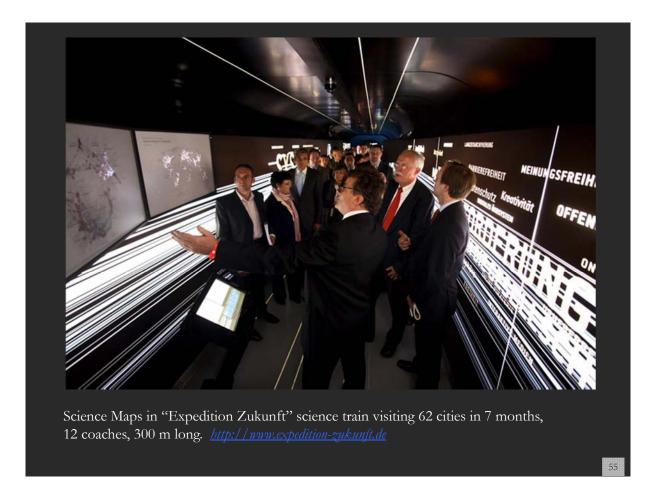






Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.







References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, 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., Volume 41, Chapter 12, pp. 537-607.

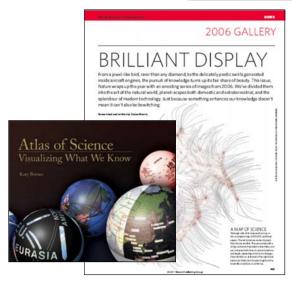
http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

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

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2011) **Models of Science Dynamics**. Springer Verlag.







57



All papers, maps, tools, talks, press are linked from http://cns.iu.edu

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience