

Plug-and-Play Microscopes That Empower Science

Katy Börner

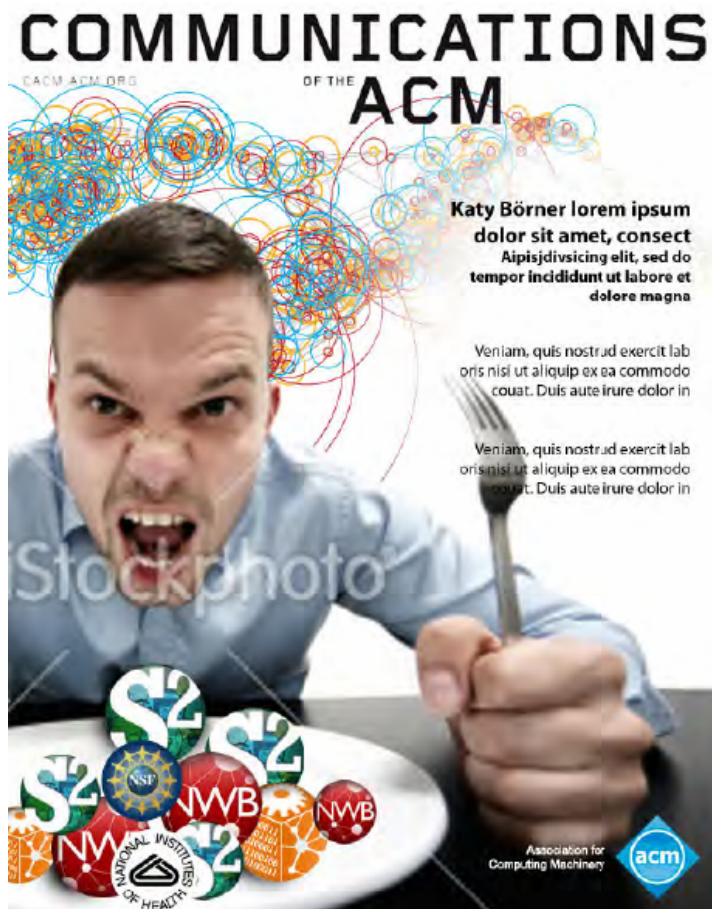
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu



With special thanks to the members at the Cyberinfrastructure for Network Science Center, the NWB team, the Sci2 team, the EpiC team, and all other teams that use OSGi/CIShell.

*Center for Bioinformatics and Computational Biology
National Institute of General Medical Sciences, National Institutes of Health
Natcher 45, second floor, Room 2As.10*

February 17, 2011



Börner, Katy. (March 2011).
Plug-and-Play Microscopes.
Communications of the ACM.



Designing “Dream Tools”

Many of the best micro-, tele-, and macrosopes are designed by **scientists keen to observe and comprehend what no one has seen or understood before.** Galileo Galilei (1564–1642) recognized the potential of a spyglass for the study of the heavens, ground and polished his own lenses, and used the improved optical instruments to make discoveries like the moons of Jupiter, providing quantitative evidence for the Copernican theory.

Today, scientists **repurpose, extend, and invent new hardware and software** to create **“macrosopes”** that may solve both local and global challenges.

The tools I will show you today **empower** me, my students, colleagues, and 100,000 others that downloaded them.

3



Macrosopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of data sets representing the structure and dynamics of complex systems. Analysis, navigation, and management of these continuously evolving data sets require a new kind of data-analysis and visualization tool we call a macroscope (from the Greek macros, or “great,” and skopein, or “to observe”) inspired by de Rosnay’s futurist science writings. Macrosopes provide a “vision of the whole,” helping us “synthesize” the related elements and enabling us to detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, **macrosopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend.**



Microscopes



Telescopes



Macrosopes

4



Goal of This Talk

Inspire computer scientists to implement software frameworks that **empower domain scientists** to assemble their own continuously evolving macrosopes, adding and upgrading existing (and removing obsolete) plug-ins to arrive at a set that is truly relevant for their work—with little or no help from computer scientists.

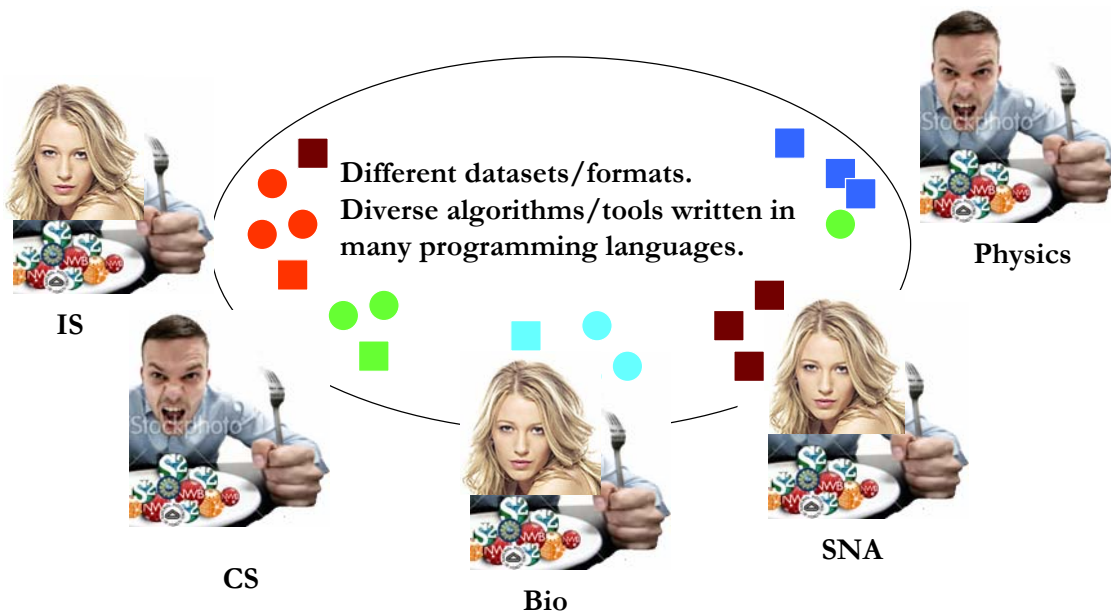
While microscopes and telescopes are physical instruments, **macrosopes resemble continuously changing bundles of software plug-ins.** Macrosopes make it easy to select and combine algorithm and tool plug-ins but also interface plug-ins, workflow support, logging, scheduling, and other plug-ins needed for scientifically rigorous yet effective work.

They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. **Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.**

5



Changing Scientific Landscape—Personal Observations



6



Changing Scientific Landscape—General Observations

Science becomes more data driven and computational but also collaborative and interdisciplinary. There is increased demand for tools that are easy to extend, share, and customize:

- *Star scientist* —> *Research teams*. Traditionally, science was driven by key scientists. Today, science is driven by collaborating co-author teams, often comprising experts from multiple disciplines and geospatial locations.
- *Users* —> *Contributors*. Web 2.0 technologies empower users to contribute to Wikipedia and exchange images, videos, and code via Flickr, YouTube, and SourceForge.net.
- *Disciplinary* —> *Cross-disciplinary*. The best tools frequently borrow and synergistically combine methods and techniques from different disciplines of science, empowering interdisciplinary and/or international teams to collectively fine-tune and interpret results;
- *Single specimen* —> *Data streams*. Microscopes and telescopes were originally used to study a single specimen at a time. Today, many researchers must make sense of massive data streams comprising multiple data types and formats from different origins; and
- *Static instrument* —> *Evolving cyberinfrastructure*. The importance of hardware instruments that are static and expensive tends to decrease relative to software tools and services that are highly flexible and evolving to meet the needs of different sciences. Some of the most successful tools and services are decentralized, increasing scalability and fault tolerance.

7



Related Work

Google Code and SourceForge.net provide special means for developing and distributing software

- In August 2009, SourceForge.net hosted more than 230,000 software projects by two million registered users (285,957 in January 2011);
- In August 2009 ProgrammableWeb.com hosted 1,366 application programming interfaces (APIs) and 4,092 mashups (2,699 APIs and 5,493 mashups in January 2011)

Cyberinfrastructures serving large biomedical communities

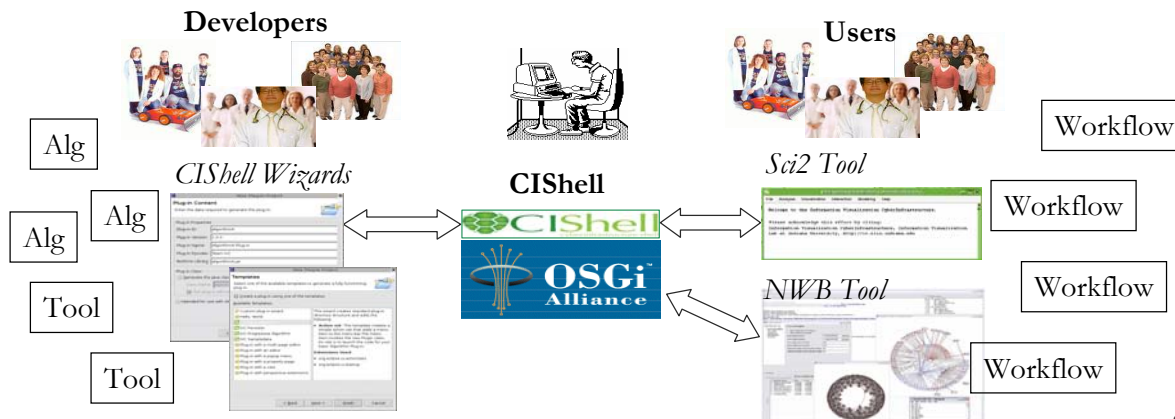
- Cancer Biomedical Informatics Grid (caBIG) (<http://cabig.nci.nih.gov>)
- Biomedical Informatics Research Network (BIRN) (<http://nbirn.net>)
- Informatics for Integrating Biology and the Bedside (i2b2) (<https://www.i2b2.org>)
- HUBzero (<http://hubzero.org>) platform for scientific collaboration uses
- myExperiment (<http://myexperiment.org>) supports the sharing of scientific workflows and other research objects.

Missing so far is a **common standard** for

- the design of **modular, compatible algorithm and tool plug-ins** (also called “modules” or “components”)
- that can be **easily combined into scientific workflows** (“pipeline” or “composition”),
- and packaged as **custom tools**.

8

- 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.



9

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 integrate it into cishell with no additional coding.

A well-defined pool of algorithms and datasets CIShell clearly defines how algorithms and datasets are integrated into the system to create a pool of algorithms and data. An application may then query for algorithms in this pool and execute them. Many applications/tools can be built and customized for different user groups by utilizing the same pool of algorithms.

Leveraging open standards CIShell avoids re-inventing wheels by building on other standards for its specification and reference implementations. It benefits most from the [Eclipse](#) family of projects (in particular, the [Rich Client Platform](#) and [Equinox](#)) and the Open Services Gateway Initiative ([OSGi](#)). All CIShell algorithms are integrated as OSGi services and can be used by any OSGi compliant system (including any Eclipse 3.0 or newer based products).

Choose the way you work CIShell offers reference applications that build on the pool of algorithms defined by CIShell. Scripting and a Graphical User Interface (GUI) are offered initially with a remoting (peer-to-peer and client-server) architecture, a web front-end, and other interfaces planned. We invite other toolkit developers to build their own applications on top of CIShell's algorithm pool.

Open source, community-driven project

CIShell is released under the [Apache 2.0 License](#). Community input is welcome to create a piece of software that advances science and education.

Algorithm Developer's Guide

Overview

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](#).

This guide attempts to aid algorithm developers in creating algorithms for CIShell (and applications built on CIShell).

This guide tries to contain all the information a new developer needs, but where necessary, it may cite the [CIShell 1.0 Specification \(API\)](#) or the [OSGi Service Platform Specification, Release 4 \(API\)](#). While the guide tries to make beginning algorithm development easier, the CIShell Specification has the last word on how the CIShell Platform works.

Table of Contents

1. [CIShell Basics](#)
2. Getting Started
 1. [Tutorial 0: Setting Up the Development Environment](#)
 2. [Tutorial 1: Creating a Hello World Java Algorithm](#)
 3. [Tutorial 2: Practical Java Algorithm Development](#)
 4. [Tutorial 3: Integrating a Non-Java Program As An Algorithm](#)
 5. [Mini-Tutorial: Integrating 3rd-party libraries](#)
 6. [Where to Learn More](#)
3. Reference
 1. [How Algorithms Work: A guide to algorithm plugins in CIShell](#)
 2. [Accessing the OSGi Console in CIShell tools](#)

11



The screenshot shows the CIShell Powered Tools Portal. At the top, there is a green header with a power button icon and the text "CIShell Powered Tools Portal". Below this, there is a section for "Cyberinfrastructure Shell (CIShell)" with a description and a "Visit the CIShell wiki" button. A "Learn more about existing CIShell-powered tools below." section follows, featuring two tool cards: "Network Workbench Tool (NWB)" and "Science of Science Tool (Sci²)".

Cyberinfrastructure Shell (CIShell)
CIShell supports the plug-and-play of datasets and algorithms and their bundling into custom tools that serve the specific needs of a user group or research community. It has been applied to develop diverse custom tools, see below. Feel free to take plugins from any of these tools to design your personal dream tool.

Visit the **CIShell wiki** to learn more about using CIShell as a platform for your tool!

Provided by the [Cyberinfrastructure for Network Science Center](#) at Indiana University.

Learn more about existing CIShell-powered tools below.

Network Workbench Tool (NWB)
The NWB Tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks. It comes with a 77-page [user manual](#).

Gallery


Science of Science Tool (Sci²)
The Sci² Tool was specifically developed for science policy makers and researchers that study science by scientific means. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels. There exists a [112-page user manual](#) and 24 hours of [NIH tutorials](#) in this tool.

12

Science of Science Cyberinfrastructures



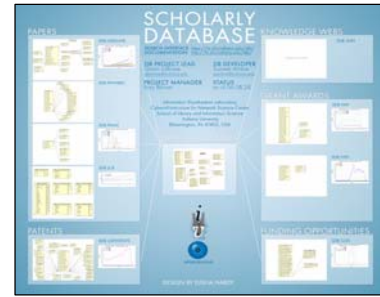
Scholarly Database: 25 million scholarly records

<http://sdb.slis.indiana.edu>



VIVO Research Networking

<http://vivoweb.org>



Information Visualization Cyberinfrastructure

<http://iv.slis.indiana.edu>



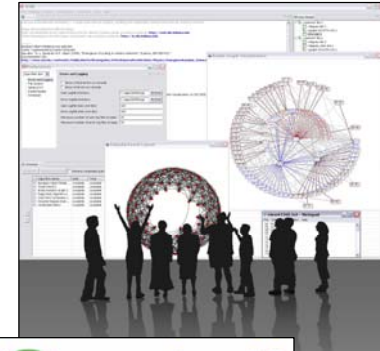
Network Workbench Tool & Community Wiki

<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/>



13



Network Workbench Tool

<http://nwb.slis.indiana.edu>

The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

In February 2009, the tool provides more than 169 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

More than 50 of these plugins can be applied or were specifically designed for S&T studies.

It has been downloaded more than 65,000 times since December 2006.

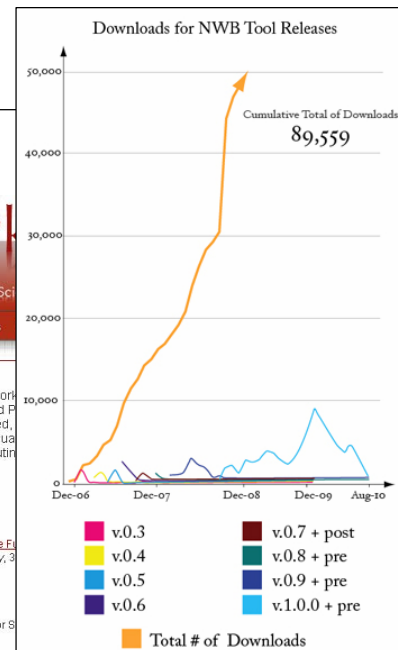
Summary
Network Workbench: A Large-Scale Network Toolkit for Biomedical, Social Science and F evaluate, and operate a unique distributed, scale network analysis, modeling, and visua (NWB). The envisioned data-code-computin [more](#)
[How to cite this project](#)

News & Updates

- 5.1.09 Kaelble, Steve. 2009. [Mapping the F Knowledge. Research & Creative Activity](#), 3 (website accessed 5/1/09)
- 3.23.09 [1.0.0 beta 5](#) Released
- 1.23.09 Ann Mccranie's [tutorial abstract](#) for S 2009
- 11.4.08 [Two NWB PIs featured in "Connected—The Power of Six Degrees."](#) 2008. Anna Maria Talas, Director. Australian Broadcasting Corporation, Ltd. [YouTube](#) [Full Video](#) (300MB)

[Getting Started](#)
See more [documentation](#)

[Get Involved](#)



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, Mibail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179), *Annals of the New York Academy of Sciences*, Boston, MA.

14

Computational Proteomics

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

Yildirim, 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.



© 2007 Nature Publishing Group

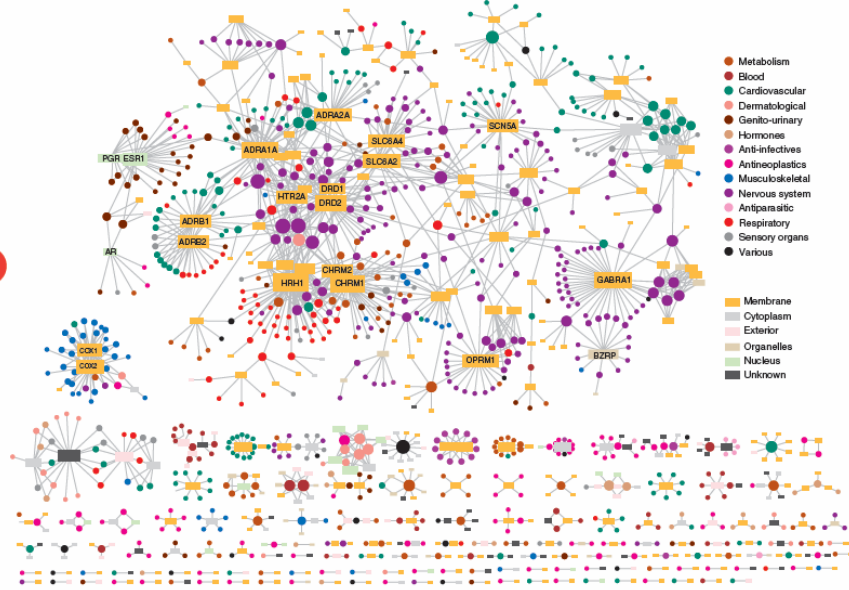


Figure 2 Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug (the protein) has. 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 cellular component obtained from the Gene Ontology database.

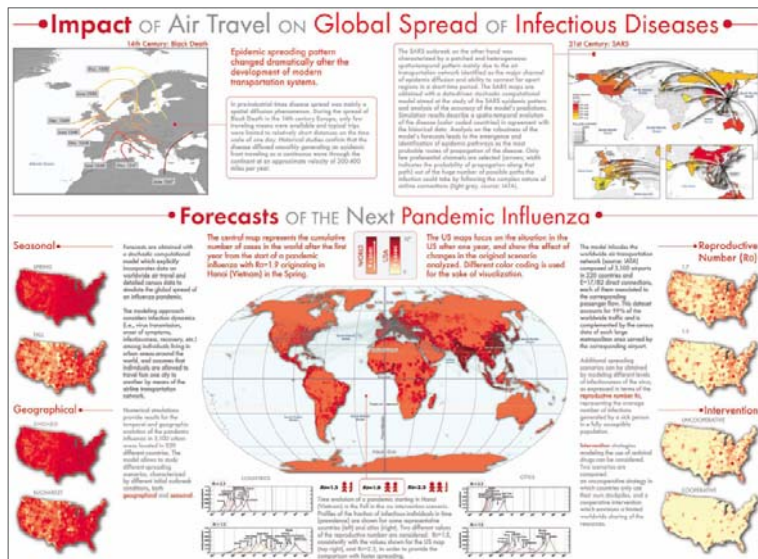
Computational Epidemics

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

Epidemic Modeling in Complex realities, V. Colizza, A. Barrat, M. Barthélemy, 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. Barthélemy, A.-J. Valleron, A. Vespignani, PloS-Medicine 4, e13, 95-110 (2007).





Forgot your password?

To recover your account password, please visit our [password recovery page](#).

Not registered yet?

[Register now](#)

Tutorials

Scott Weingart, I
Biberstine (2010)
Science, Indiana

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://sci2.cns.iu.edu>
<http://sci2.wiki.cns.iu.edu>

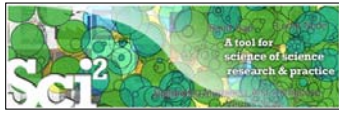
Geetha Senthil (2010). [Multidisciplinary Nature of Work With Reference to PIs and ICs Within a Portfolio](#). PA Group at NIH.

NIH Office of Extramural Research and Katy Börner (2010) [Network Visualizations Using SPIRES Data and the Sci2 Tool](#). Office of Extramural Research at NIH.



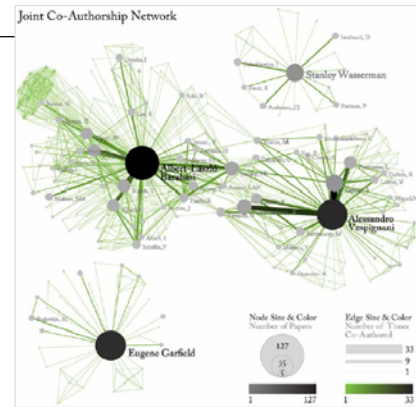
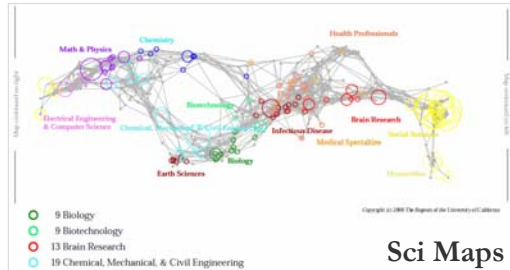
Type of Analysis vs. Level of Analysis

	<i>Micro/Individual (1-100 records)</i>	<i>Meso/Local (101–10,000 records)</i>	<i>Macro/Global (10,000 < records)</i>
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications
Topical Analysis (What)	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	VxOrd/Topic maps of NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-author network	NSF's core competency

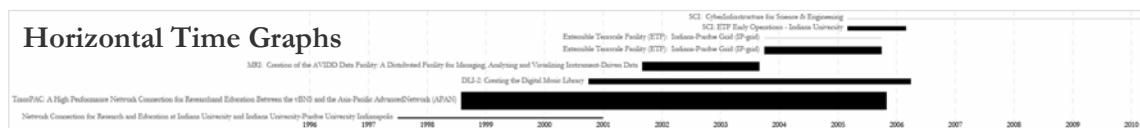


Sci² Tool – “Open Code for S&T Assessment”

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



GUESS Network Vis



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



Sci² Tool

Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

Welcome to the Science of Science Tool (Sci²). The development of this tool is supported in Network Science center and the School of Li Indiana University, the National Science Foundation (NSF-0715303), and the James S. McDonnell Cyberinfrastructure portal (<http://sci.slis.indiana.edu>).

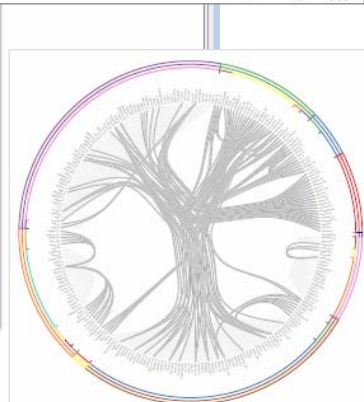
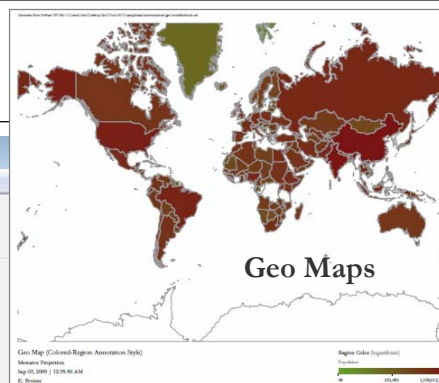
The primary investigators are Katy Börner, In SciTech Strategies Inc. The Sci² tool was developed by J. Duhon, Patrick A. Phillips, Chintan Tank, a Cyberinfrastructure Shell (<http://cishell.org>) for Network Science Center (<http://cns.slis.indiana.edu>). Many algorithm plugins were derived from the Network Science Center (<http://nwb.slis.indiana.edu>).

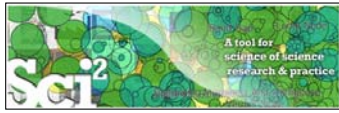
Please cite as follows:
Sci² Team. (2009). Science of Science Tool. In SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

Remove From List Remove completed

!	Algorithm Name	Date	Time	% Complete
<input checked="" type="checkbox"/>	Extract Co-Author Network	09/03/2009	00:15:20 AM	100%
<input checked="" type="checkbox"/>	Load and Clean ISI File	09/03/2009	00:15:05 AM	100%





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

21



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 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:

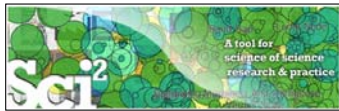
Database support for ISI and NSF data.

22

Contents

Contents.....	35
1 Introduction	35
2 Getting Started	
2.1 Download, Install, Uninstall.....	
2.2 User Interface.....	
2.2.1 Menus.....	
2.2.2 Console.....	
2.2.3 Data Manager.....	
2.2.4 Scheduler.....	
2.3 Data Formats.....	
2.4 Saving Visualizations for Publication.....	
2.5 Sample Datasets.....	
3 Algorithm and Tool Plugins.....	
3.1 Sci ² Tool Plugins.....	
3.2 Load, View, and Save Data.....	
3.3 Memory Allocation.....	
3.3.1 Windows and Linux.....	
3.3.2 Mac.....	
3.4 Memory Limits.....	
4 Workflow Design	
4.1 Overview.....	
4.2 Data Acquisition and Preparation.....	
4.2.1 Datasets: Publications.....	
4.2.2 Datasets: Funding.....	
4.2.3 Datasets: Scholarly Database.....	
4.3 Database Loading and Manipulation.....	
4.4 Summaries and Table Extractions.....	
4.5 Statistical Analysis/Profiling.....	
4.6 Temporal Analysis (When).....	
4.6.1 Burst Detection.....	
4.6.2 Slice Table by Time.....	
4.7 Geospatial Analysis (Where).....	
4.8 Topical Analysis (What).....	
4.8.1 Word Co-Occurrence Network.....	
4.9 Network Analysis (With Whom?).....	35
4.9.1 Network Extraction.....	35
4.9.2 Compute Basic Network Characteristics.....	
4.9.3 Network Analysis.....	
4.9.4 Network Visualization.....	
4.10 Modeling (Why?).....	
4.10.1 Random Graph Model.....	
4.10.2 Watts-Strogatz Small World.....	
4.10.3 Barabási-Albert Scale Free Model.....	
5 Sample Workflows	
5.1 Individual Level Studies - Micro.....	
5.1.1 Mapping Collaboration, Publication and Funding Profiles of One Rese.....	
5.1.2 Time Slicing of Co-Authorship Networks (ISI Data).....	
5.1.3 Funding Profiles of Three Researchers at Indiana University (NSF Data).....	
5.1.4 Studying Four Major NetSci Researchers (ISI Data).....	
5.1.5 Studying Four Major NetSci Researchers (ISI Data) using Database.....	
5.2 Institution Level Studies - Meso.....	
5.2.1 Funding Profiles of Three Universities (NSF Data).....	
5.2.2 Funding Profiles of Three Universities (NSF Data) Using Database.....	
5.2.3 Mapping CTSAs Centers (NIH RePORTER Data).....	
5.2.4 Biomedical Funding Profile of NSF (NSF Data).....	
5.2.5 Mapping Scientometrics (ISI Data).....	
5.2.6 Burst Detection in Scientometrics (ISI Data).....	
5.2.7 Mapping the Field of RNAi Research (SDB Data).....	
5.3 Global Level Studies - Macro.....	
5.3.1 Geo USPTO (SDB Data).....	
6 Sample Science Studies & Online Services.....	
6.1 Science Dynamics.....	
6.1.1 Mapping Topics and Topic Bursts in PNAS (2004).....	
6.2 Local Impact-Output / ROI Studies.....	
6.2.1 Indicator-Assisted Evaluation and Funding of Research: Visualizing the Influence of Grants on the Number and Citation Counts of Research Papers (2003).....	87
6.2.2 Mapping Transdisciplinary Tobacco Use Research Centers Publications (forthcoming).....	88
6.3 Local and Global Science Studies.....	89
6.3.1 Mapping the Evolution of Co-Authorship Networks (2004).....	89
6.3.2 Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams (2005).....	90
6.3.4 Mapping the Diffusion of Information Among Major U.S. Research Institutions (2006).....	92
6.3.5 Research Collaborations by the Chinese Academy of Sciences (2009).....	93
6.3.6 Mapping the Structure and Evolution of Chemistry Research (2009).....	94
6.3.7 Science Map Applications: Identifying Core Competency (2007).....	95
6.4 Modeling Science.....	96
6.4.1 113 Years of Physical Review: Using Flow Maps to Show Temporal and Topical Citation (2008).....	96
6.4.2 The Simultaneous Evolution of Author and Paper Networks (2004).....	97
6.5 Accuracy Studies.....	98
6.5.1 Mapping the Backbone of Science (2005).....	98
6.5.2 Toward a Consensus Map of Science (2009).....	99
6.6 Databases and Tools.....	100
6.6.1 The Scholarly Database and Its Utility for Scientometrics Research (2009).....	100
6.6.2 Reference Mapper.....	101
6.6.3 Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool (2009).....	102
6.7 Interactive Online Services.....	103
6.7.1 The NIH Visual Browser: An Interactive Visualization of Biomedical Research (2009).....	103
6.7.2 Interactive World and Science Map of S&T Jobs (2010).....	104
7 Extending the Sci² Tool	105
7.1 CShell Basics.....	105
7.2 Read New Data.....	105
7.3 Creating and Sharing New Algorithm Plugins.....	105
7.4 Tools That Use OSGI and/or CShell.....	106
8 Relevant Datasets and Tools.....	107
8.1 Datasets.....	107
8.2 Network Analysis Tools.....	108
9 References	111

See Sci² Tool Wiki
<http://sci2.wiki.cns.in.edu>

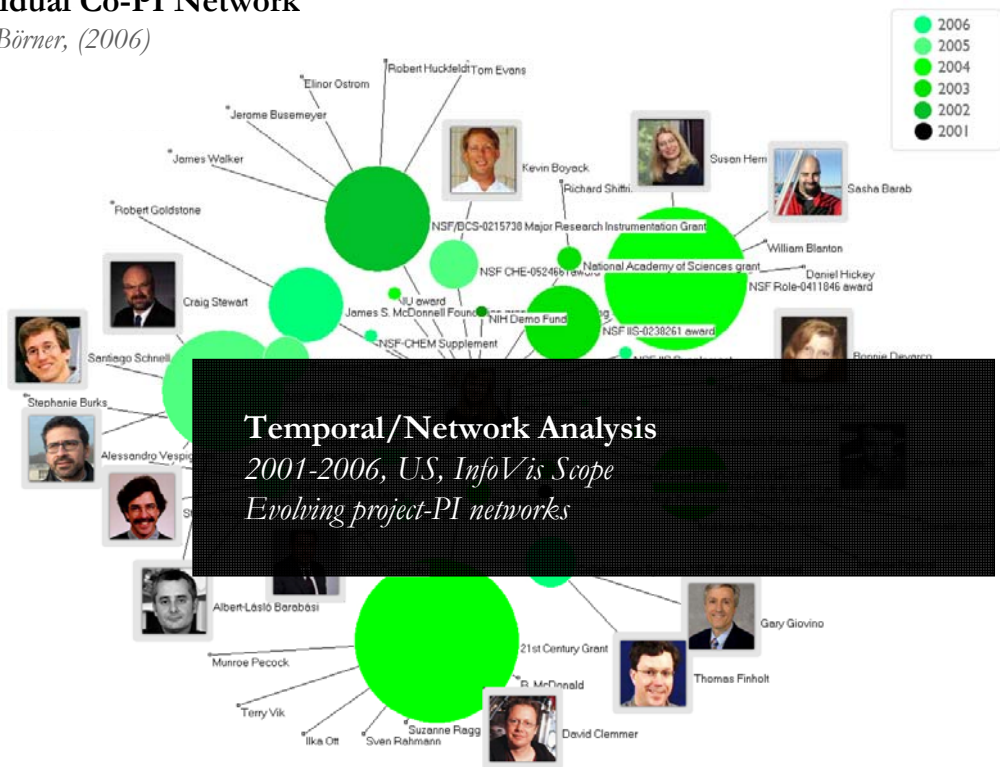


Type of Analysis vs. Level of Analysis

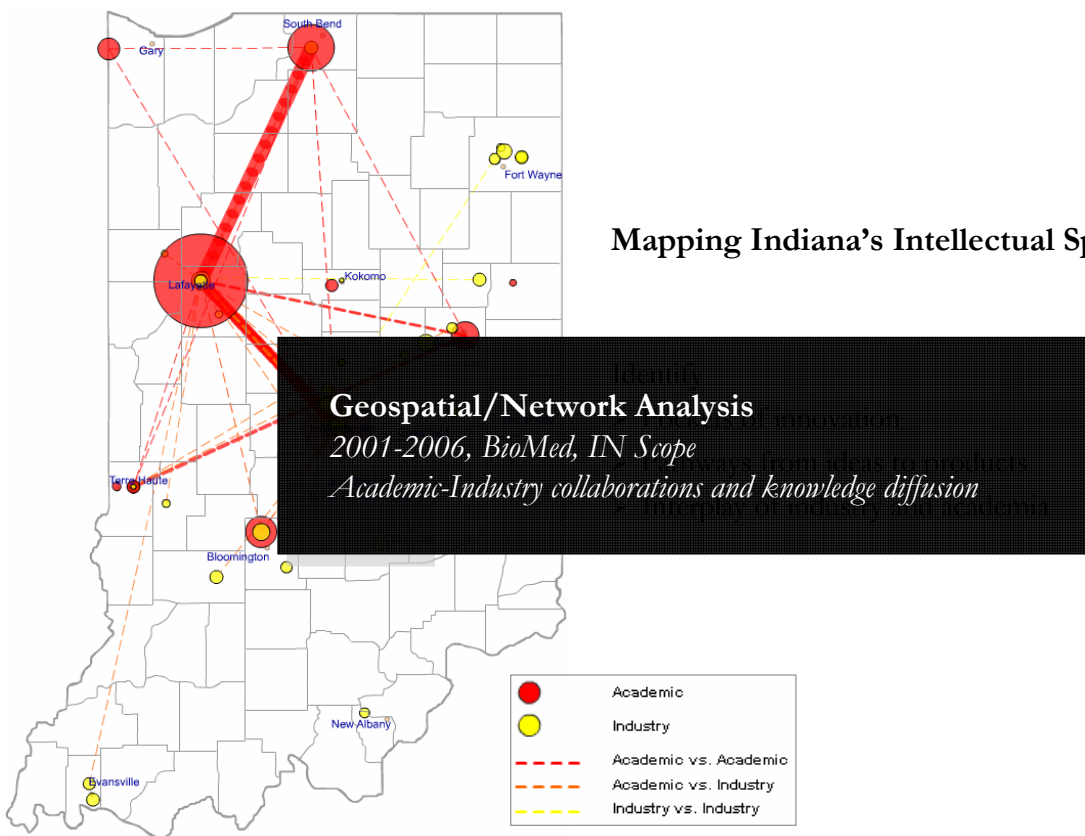
	Micro/Individual (1-100 records)	Meso/Local (101-10,000 records)	Macro/Global (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSI, all of sci SA,
Temporal Analysis (When)	Funding portfolio of one individual	Topic bursts of PNAS	113 Years of Physical Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a scientist's intellectual l	PNAS
Topical Analysis (What)			VxOrd/Topic in NIH funding
Network Analysis (With Whom?)	NSI's network of	NIH's network of	NIH's network of

Individual Co-PI Network

Ke & Börner, (2006)



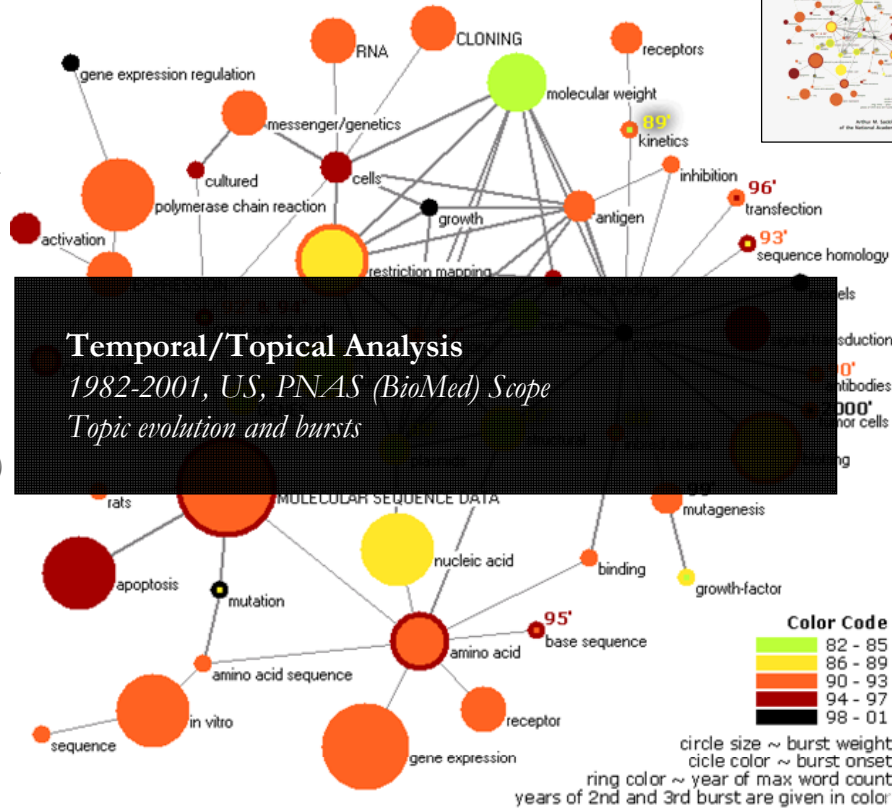
Mapping Indiana's Intellectual Space



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.



27

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



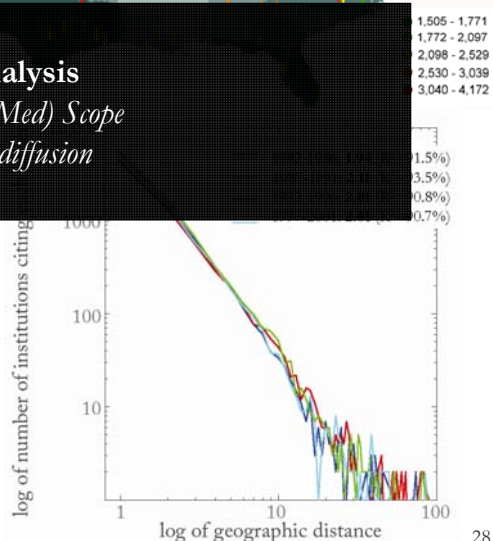
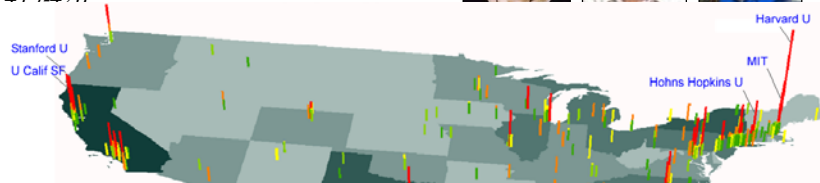
Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at research institutions in order to produce high quality data and quality research?
3. Does the Internet change citation patterns, i.e., more research produced at geographically distant research institutions?

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.

Temporal/Geospatial Analysis
1982-2001, US, PNAS (BioMed) Scope
Citation impact and knowledge diffusion

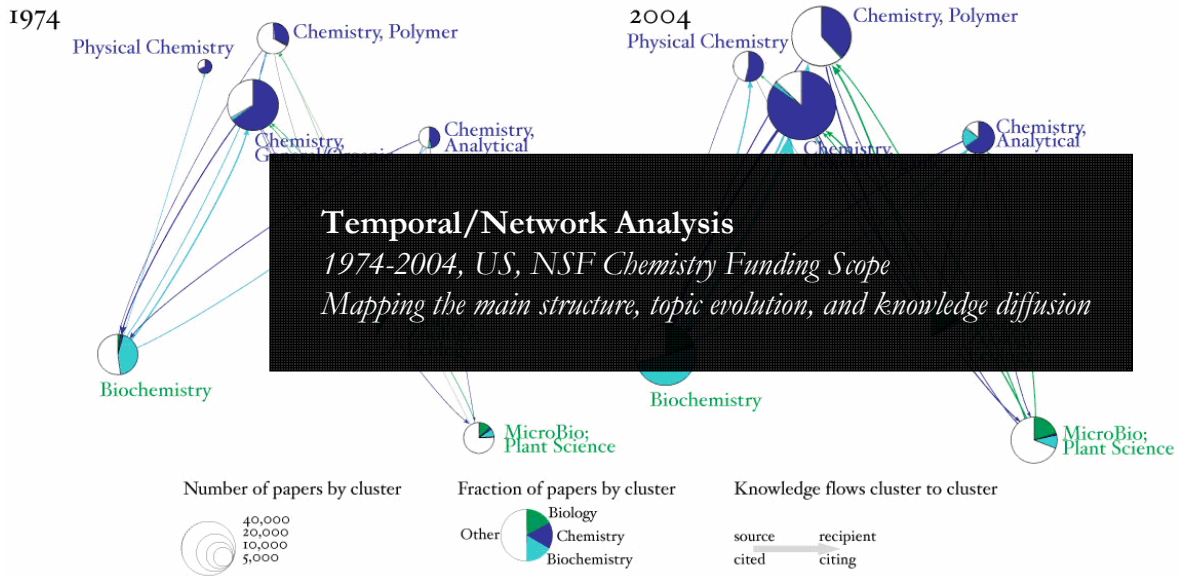


28

Topical Composition and Knowledge Flow Patterns in Chemistry Research for 1974 and 2004

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

Chemistry - Biology Interface

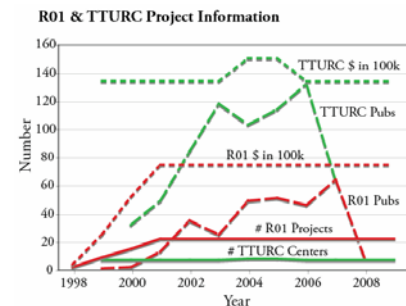


29

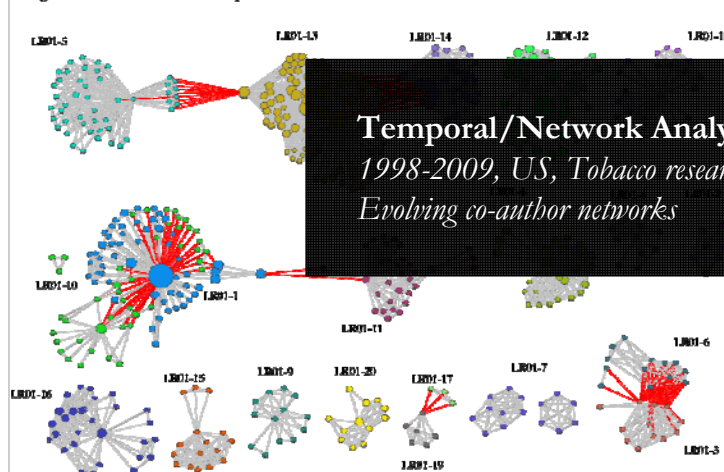
Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

Zoss & Börner, forthcoming.



Longitudinal R01 Co-Authorship Network



TTURC Co-Authorship Network

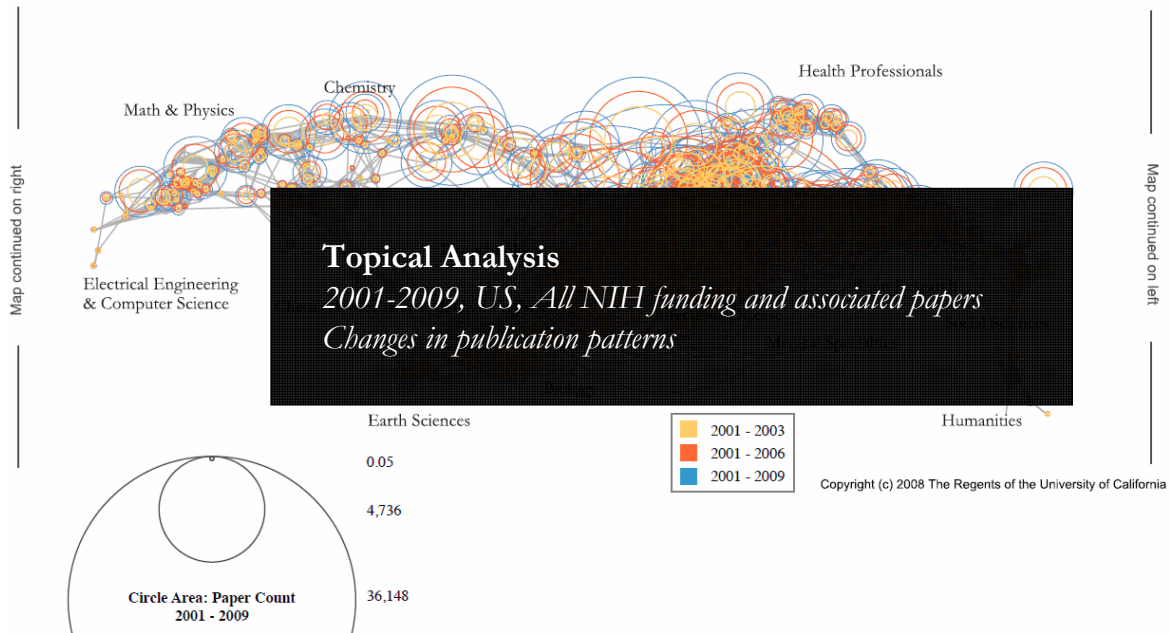


Temporal/Network Analysis
 1998-2009, US, Tobacco research scope
 Evolving co-author networks

30

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

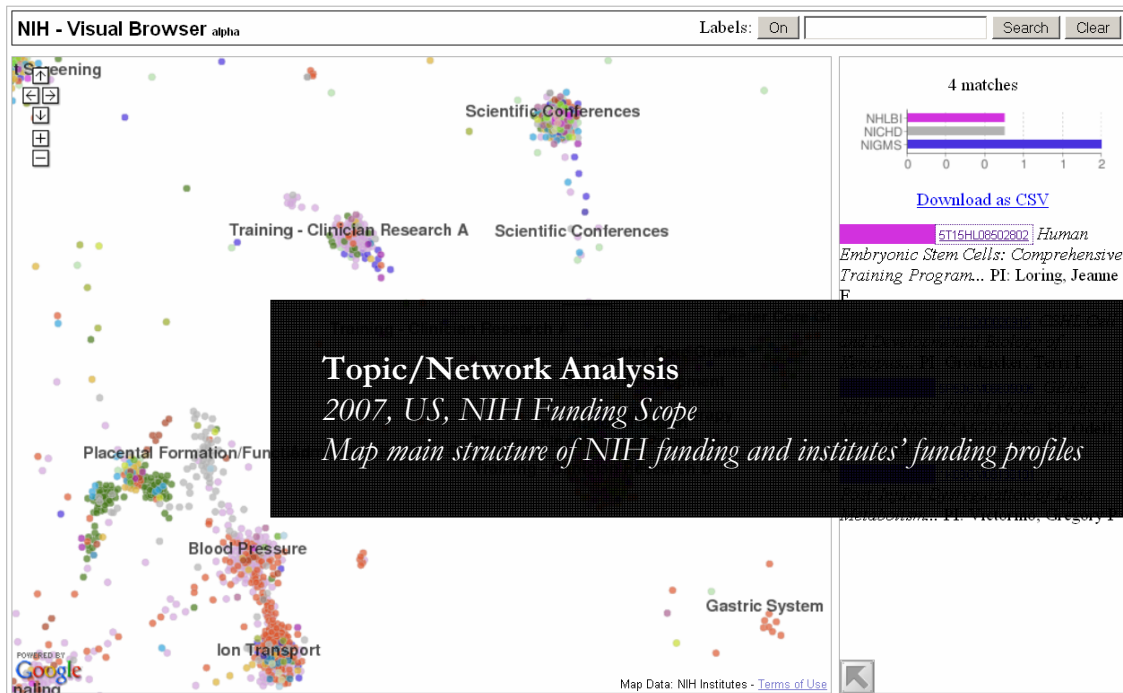
Katy Börner, Nianli Ma, Joseph R. Biberstine, Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Robin M. Wagner, Rediet Berhane, Hong Jiang, Susan E. Ivey, Katrina Pearson and Carl McCabe, Reporting Branch, Division of Information Services, Office of Research Information Systems,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



31

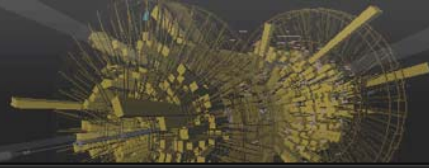
Interactive Science Map of NIH Funding

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



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

32



Grants Awarded

Discover Grants by clicking on map or using the options below.

United States

NIH/NSF

Year(s): 2000-2009

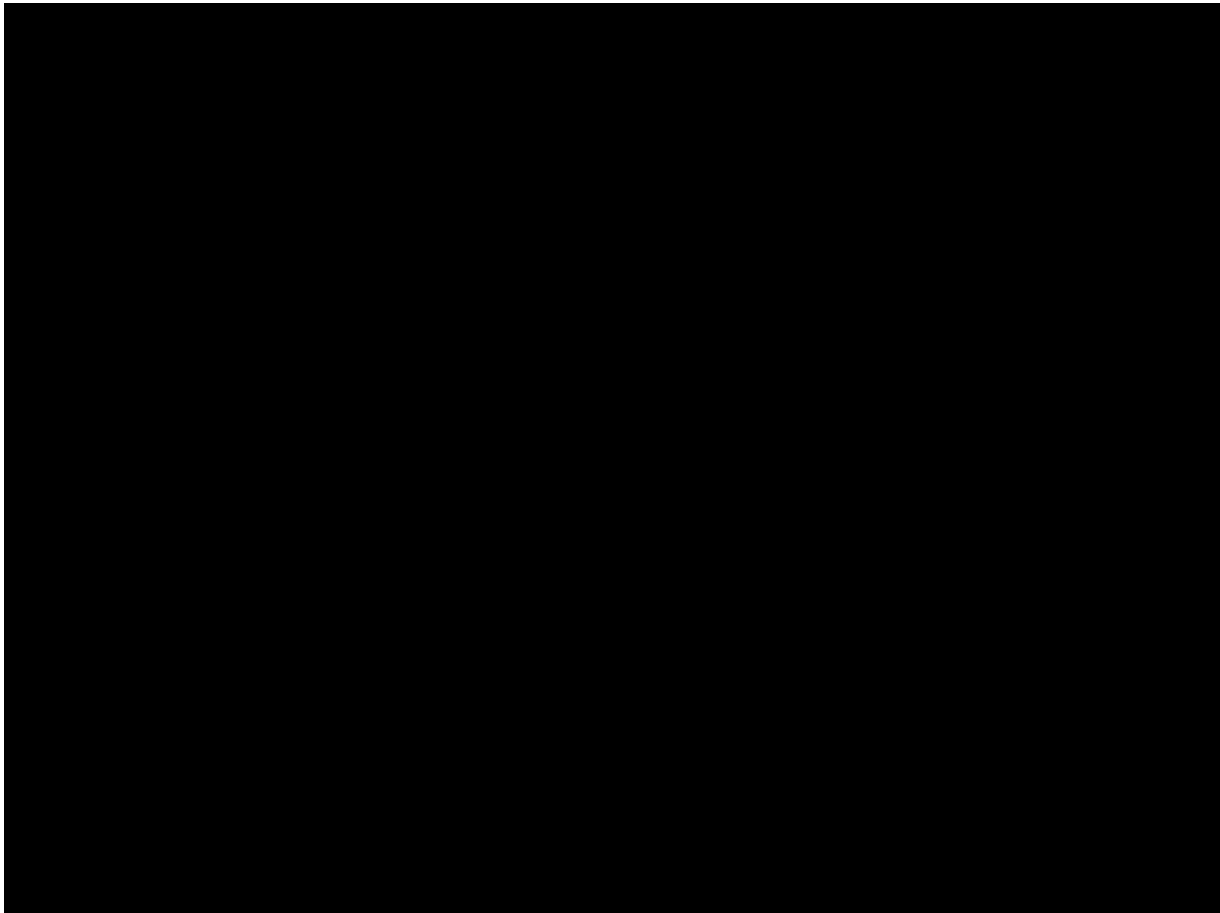
Amount: 0.00-130.00m

Refine results by selecting institutions or topics:

- Top Research Institutions

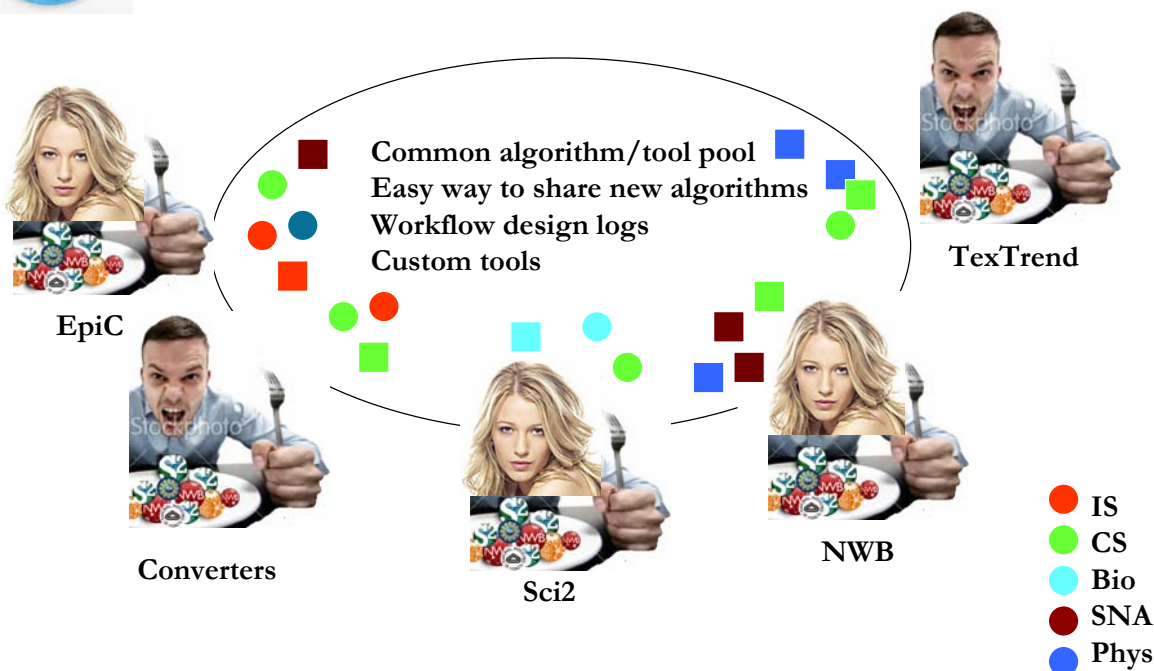


<http://rd-dashboard.nitrd.gov>





Changing Scientific Landscape—Personal Observations Cont.



35



OSGi/CIShell Adoption

CIShell/OSGi is at the core of different CIs and a total of 169 unique plugins are used in the

- **Information Visualization** (<http://iv.slis.indiana.edu>),
- **Network Science (NWB Tool)** (<http://nwb.slis.indiana.edu>),
- **Scientometrics and Science Policy (Sci² Tool)** (<http://sci.slis.indiana.edu>), and
- **Epidemics** (<http://epic.slis.indiana.edu>) research communities.

Most interestingly, a number of other projects recently adopted OSGi and one adopted CIShell:

Cytoscape (<http://www.cytoscape.org>) lead by Trey Ideker, 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.sourceforge.net>) lead by Carol Goble, University of Manchester, UK 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 Shawn Hampton, NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

TEXTrend (<http://www.textrend.org>) lead by George Kampis, Eötvös University, Hungary develops a framework for the easy and flexible integration, configuration, and extension of plugin-based components in support of natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corporuses with an inherently temporal component.

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.

Acknowledgements

- Micah Linnemeier and Russell J. Duhon Bruce W. Herr II, George Kampis, Gregory J. E. Rawlins, Geoffrey Fox, Shawn Hampton, Carol Goble, Mike Smoot, Yanbo Han for stimulating discussions and comments.
- The Cyberinfrastructure for Network Science Center (<http://cns.iu.edu>), the Network Workbench team (<http://nwb.cns.iu.edu>), and Science of Science project team (<http://sci2.cns.iu.edu>) for their contributions toward the work presented here.
- Software development benefits greatly from the open-source community. Full software credits are distributed with the source, but I would especially like to acknowledge Jython, JUNG, Prefuse, GUESS, GnuPlot, and OSGi, as well as Apache Derby, used in the Sci2 tool.

This research and development is based on work supported by National Science Foundation grants SBE-0738111, IIS-0513650, IIS-0534909 and National Institutes of Health grants R21DA024259 and 5R01MH079068.

37

References

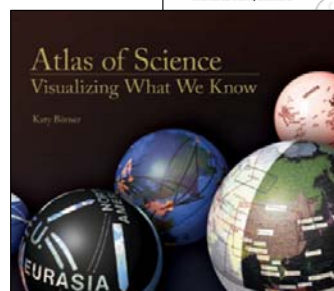
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255.
<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

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>

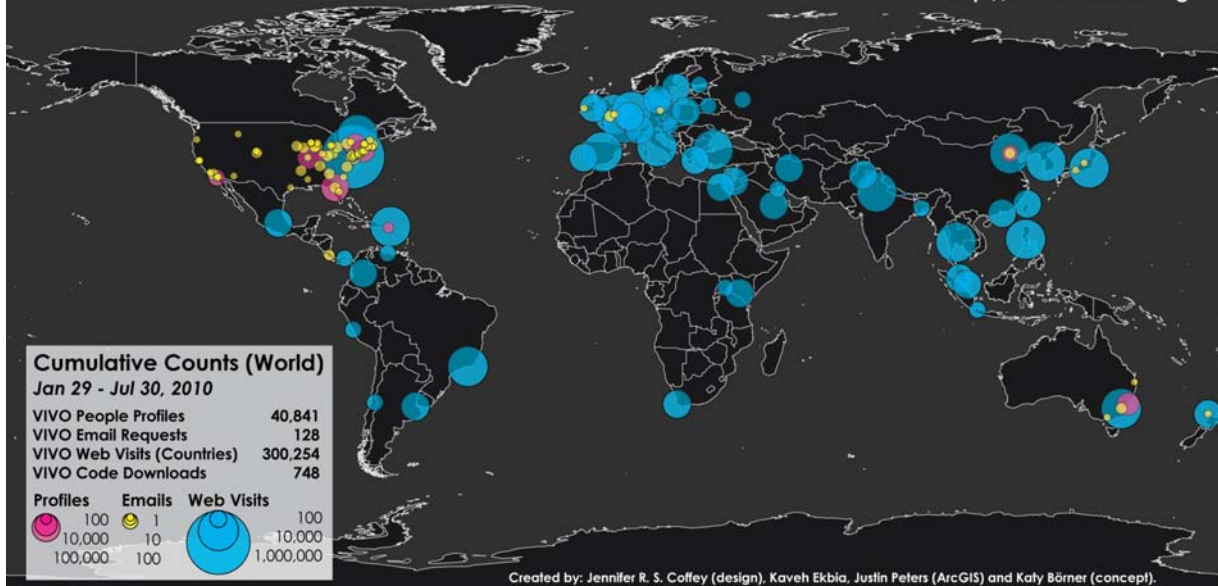
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2011) *Models of Science Dynamics*. Springer Verlag.



38

VIVO Enabling National Networking of Scientists

<http://www.vivoweb.org>



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

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.

39

cyberinfrastructure for NETWORK SCIENCE CENTER
School of Library and Information Science | Indiana University Bloomington

People, Events, Jobs, Contact, Research, News, Teaching, Outreach, Funding, Visiting Artists, Cyberinfrastructures

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

40