

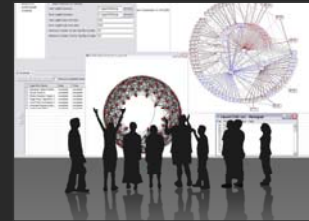
## Open Data and Open Code for S&T Assessment

**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](mailto:katy@indiana.edu)

With special thanks to Kevin W. Boyack, Micah Linnemeier,  
Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank  
Nianli Ma, Angela M. Zoss, Hanning Guo, Mark A. Price,  
Scott Weingart

*Northwestern Institute on Complex Systems (NICO) Annual Conference  
Northwestern University, IL  
September 3, 2009*



### Overview

#### Science of Science Studies

**Science of Science Cyberinfrastructure** (<http://sci.slis.indiana.edu>):

- **Scholarly Database (SDB)** (<http://sdb.slis.indiana.edu>) that provides free access to 23 million scholarly records
- **Sci<sup>2</sup> Tool** which reads SDB data and supports the identification of activity bursts, the extraction and display of co-author/inventor/investigator networks, and topical analysis, among others.

#### Mapping Science Exhibit



## Overview

### Science of Science Studies

Science of Science Cyberinfrastructure (<http://sci.slis.indiana.edu>):

- **Scholarly Database (SDB)** (<http://sdb.slis.indiana.edu>) that provides free access to 23 million scholarly records
- **Sci<sup>2</sup> Tool** which reads SDB data and supports the identification of activity bursts, the extraction and display of co-author/inventor/investigator networks, and topical analysis, among others.

### Mapping Science Exhibit

## Computational Scientometrics: Studying Science by Scientific Means

- Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology*, 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/](http://www.pnas.org/content/vol101/suppl_1/)
- Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology*, 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, Ma, Nianli, Dubon, Russell Jackson & Zoss, Angela. (2009). **Science & Technology Assessment Using Open Data and Open Code**. *IEEE Intelligent Systems*. Vol. 24(4), 78-81, *IEEE Computer Systems*..
- **Places & Spaces: Mapping Science** exhibit, see also <http://scimaps.org>.

## Computational Scientometrics Opportunities

### Advantages for Funding Agencies

- Supports monitoring of (long-term) money flow and research developments, evaluation of funding strategies for different programs, decisions on project durations, funding patterns.
- Staff resources can be used for scientific program development, to identify areas for future development, and the stimulation of new research areas.

### Advantages for Researchers

- Easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications (**research push**).
- More time for research and teaching.

### Advantages for Industry

- Fast and easy access to major results, experts, etc.
- Can influence the direction of research by entering information on needed technologies (**industry-pull**).

### Advantages for Publishers

- Unique interface to their data.
- Publicly funded development of databases and their interlinkage.

### For Society

- Dramatically improved access to scientific knowledge and expertise.

## Process of Computational Scientometrics

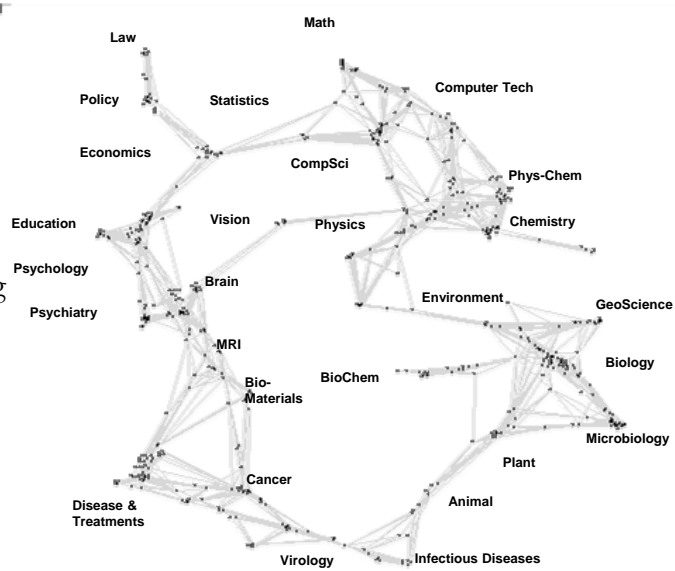
DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	
SEARCHES ISI INSPEC Eng Index Medline ResearchIndex Patents etc.	COMMON CHOICES Journal Document Author Term	COUNTS/FREQUENCIES Attributes (e.g. terms) Author citations Co-citations By year  THRESHOLDS By counts	SCALAR (unit by unit matrix) Direct citation Co-citation Combined linkage Co-word / co-term Co-classification  VECTOR (unit by attribute matrix) Vector space model (words/terms) Latent Semantic Analysis (words/terms) incl. Singular Value Decomposition (SVD)  CORRELATION (if desired) Pearson's R on any of above	DIMENSIONALITY REDUCTION Eigenvector/ Eigenvalue solutions Factor Analysis (FA) and Principal Components Analysis (PCA) Multi-dimensional scaling (MDS) LSA, <b>Topics</b> Pathfinder networks (PFNet) Self-organizing maps (SOM) includes SOM, ET-maps, etc.	INTERACTION Browse Pan Zoom Filter Query Detail on demand  ANALYSIS
BROADENING By citation By terms				CLUSTER ANALYSIS  SCALAR Triangulation Force-directed placement (FDP)	

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003) *Visualizing Knowledge Domains*. In Blaise Cronin (Ed.), [Annual Review of Information Science & Technology, Volume 37](#), Medford, NJ: Information Today, Inc./ American Society for Information Science and Technology, chapter 5, pp. 179-255.

## Latest '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.

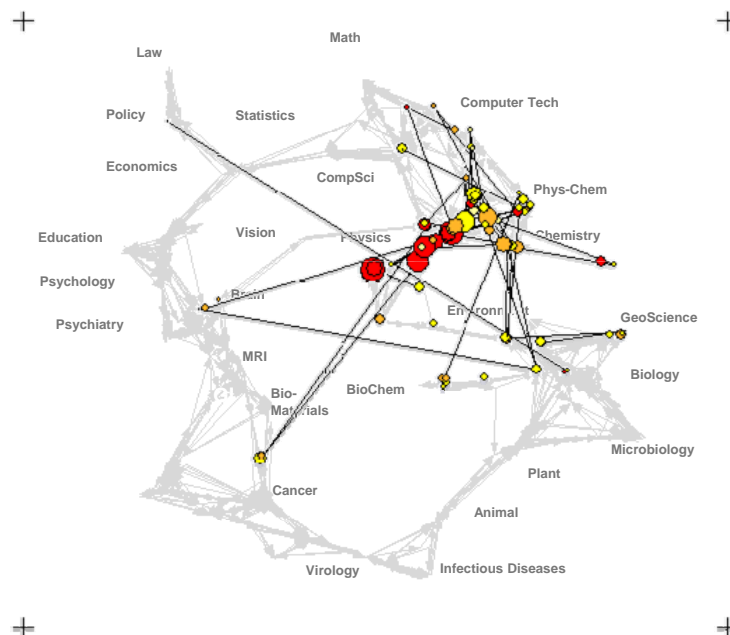
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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

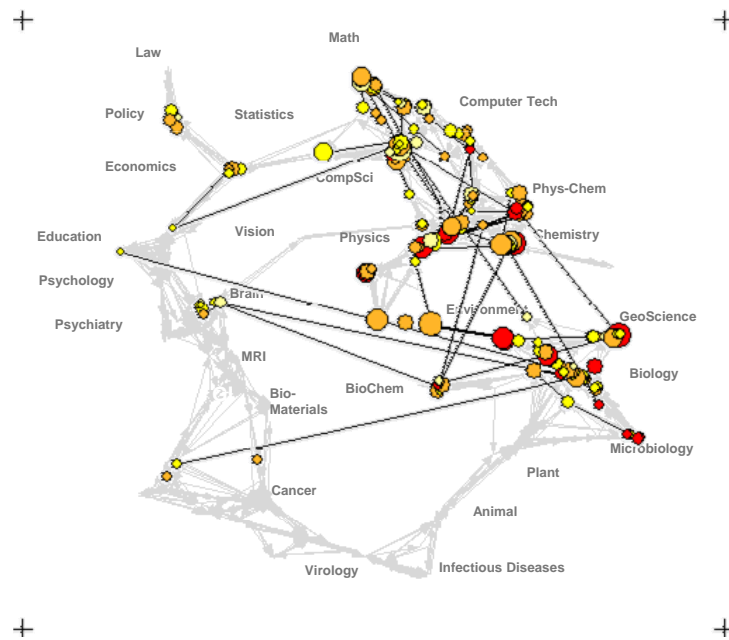
Funding patterns of the US Department of Energy (DOE)



## Science map applications: Identifying core competency

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

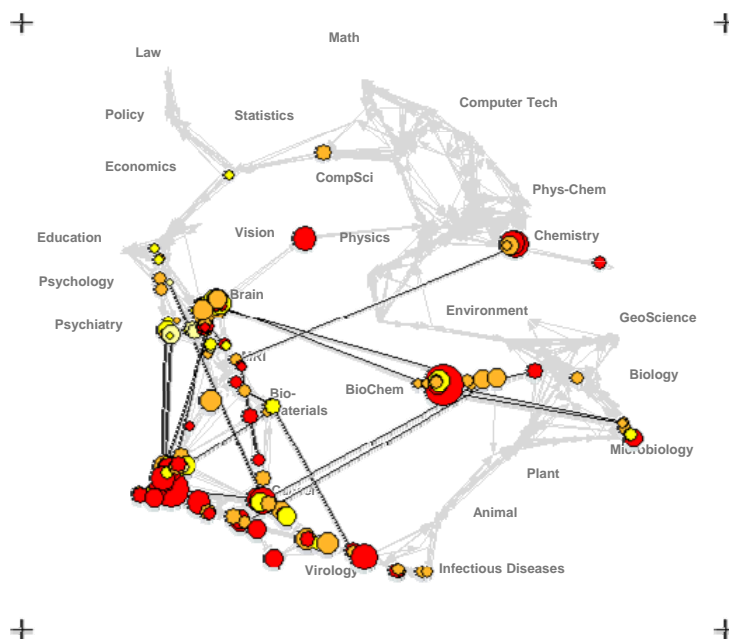
### Funding Patterns of the National Science Foundation (NSF)



## Science map applications: Identifying core competency

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

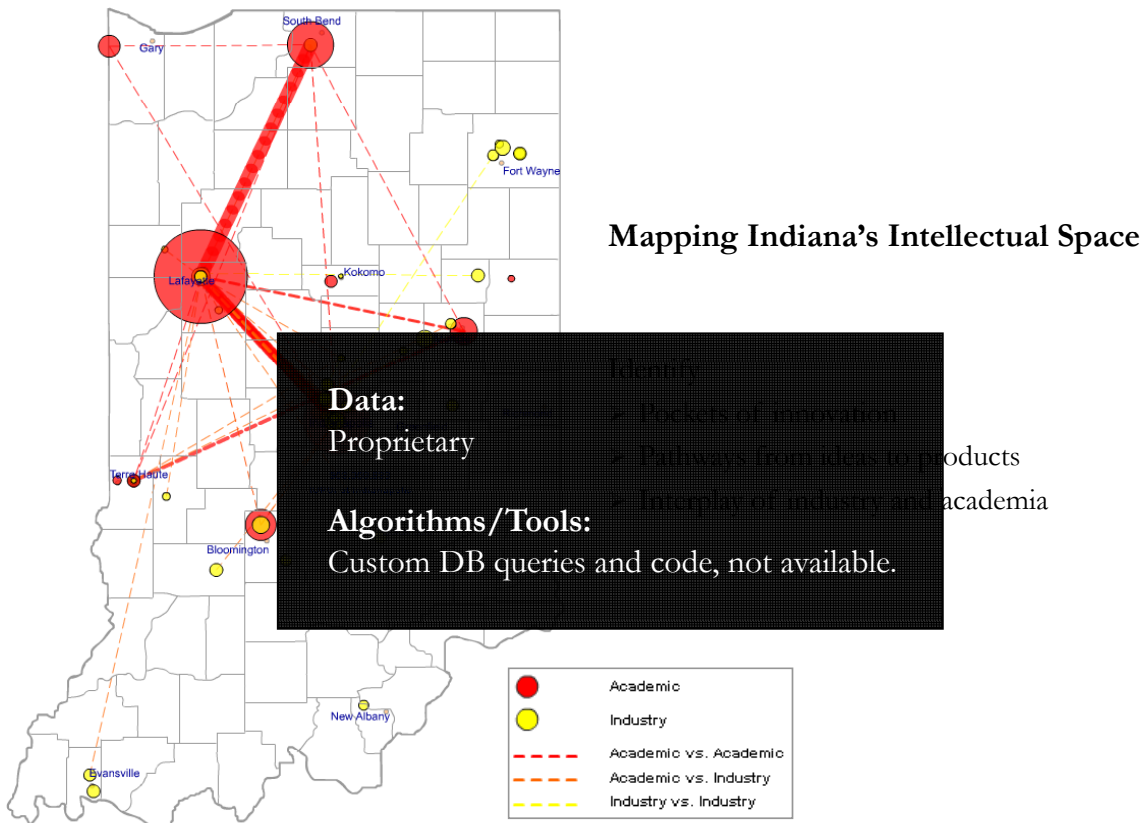
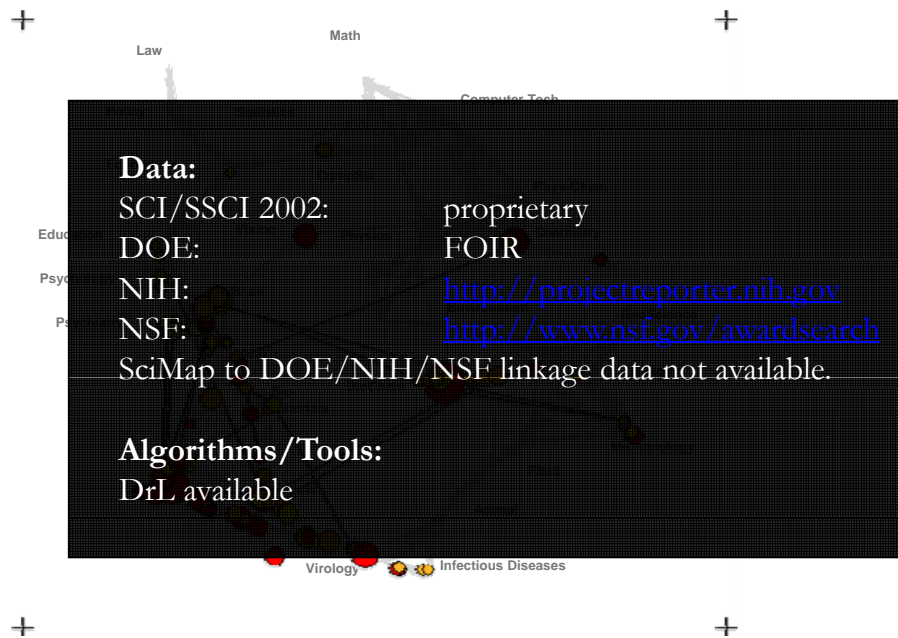
### Funding Patterns of the National Institutes of Health (NIH)



# Science map applications: Identifying core competency

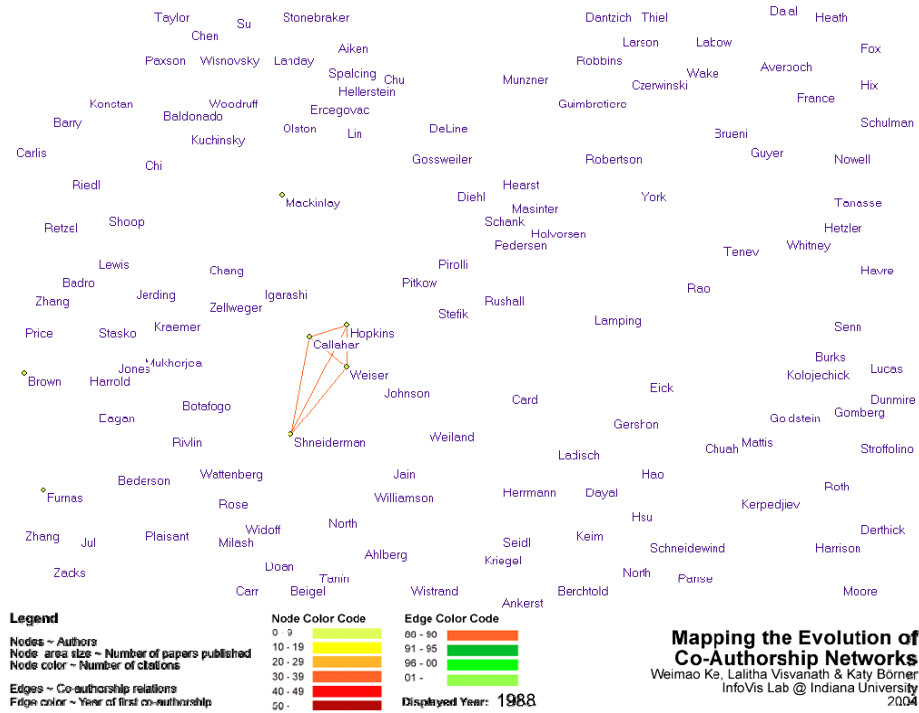
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

## Funding Patterns of the National Institutes of Health (NIH)



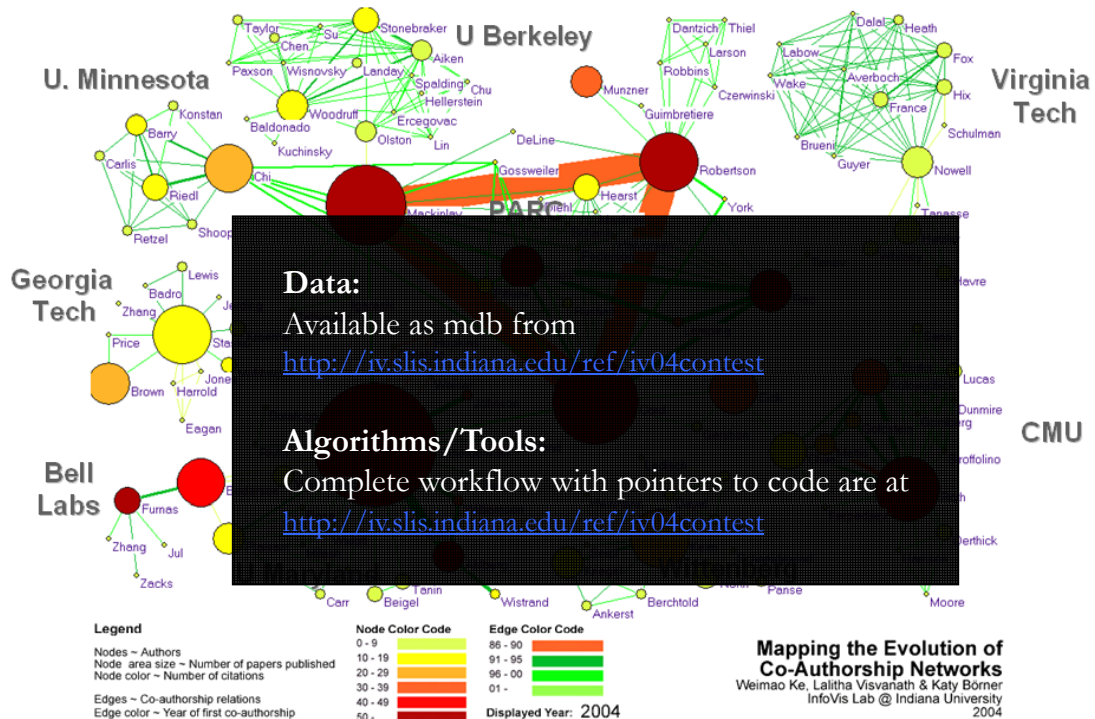
# Mapping the Evolution of Co-Authorship Networks

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.

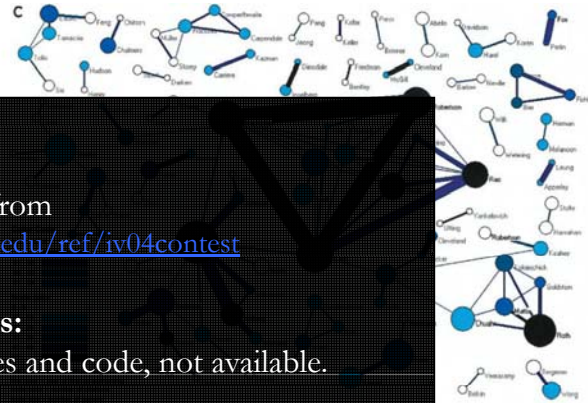


# Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

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?

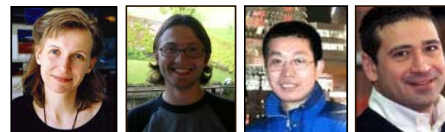


## Contributions:

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

**Data:**  
Available as mdb from <http://iv.slis.indiana.edu/ref/iv04contest>

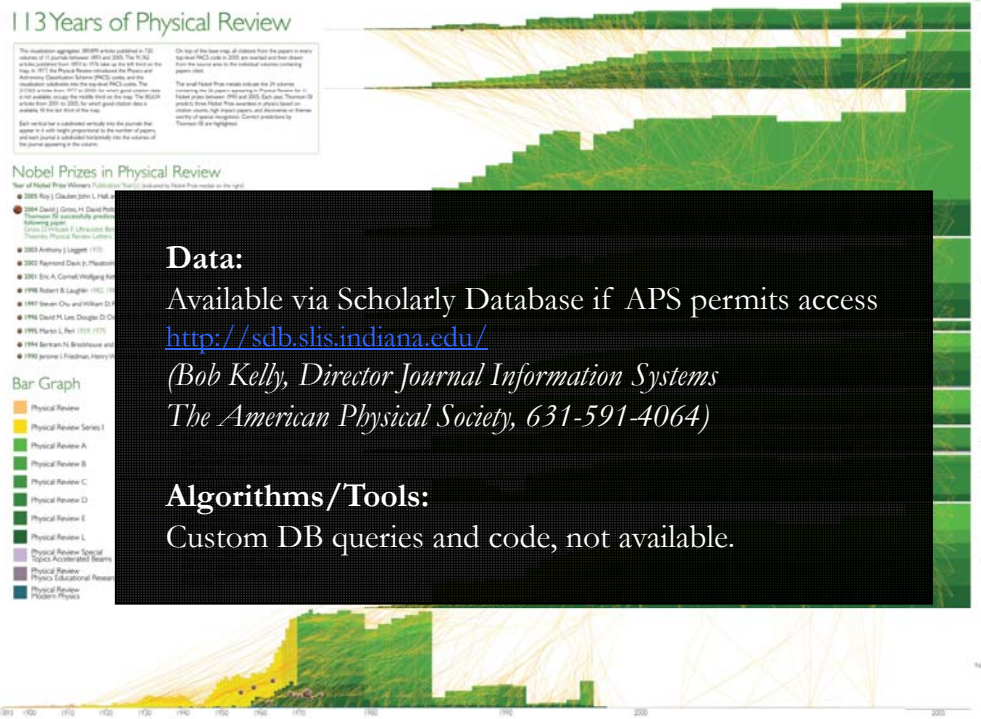
**Algorithms/Tools:**  
Custom DB queries and code, not available.



# 113 Years of Physical Review

[http://scimaps.org/dev/map\\_detail.php?map\\_id=171](http://scimaps.org/dev/map_detail.php?map_id=171)

Bruce W. Herr II and Russell Dabon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)



**Data:**  
Available via Scholarly Database if APS permits access  
<http://sdb.slis.indiana.edu/>  
(Bob Kelly, Director Journal Information Systems  
The American Physical Society, 631-591-4064)

**Algorithms/Tools:**  
Custom DB queries and code, not available.





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

Börner, Katy, Penumarty, 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 institutions in order to produce high quality data and quality research?
3. Does the Internet change patterns, i.e., more produced at geographic instructions?

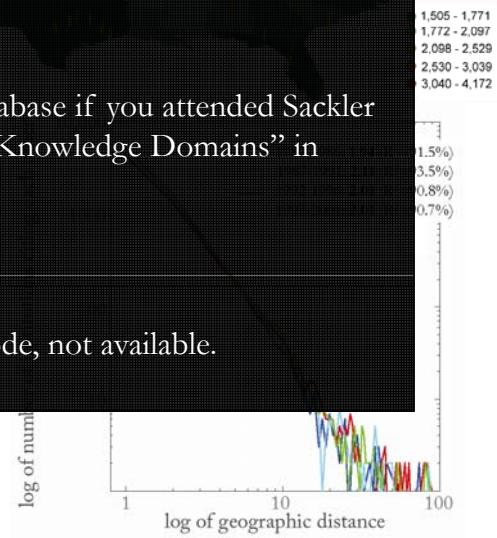


**Data:**  
 Available via Scholarly Database if you attended Sackler Colloquium on "Mapping Knowledge Domains" in May 2003  
<http://sdb.slis.indiana.edu/>

**Algorithms/Tools:**  
 Custom DB queries and code, not available.

## Contributions:

- Answer to Qs 1 - 2
- Answer to Qs 3 in 2006
- 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 Topic Bursts

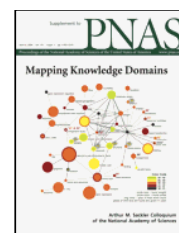
Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

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



**Data:**  
 Available via Scholarly Database if you attended Sackler Colloquium on "Mapping Knowledge Domains" in May 2003  
<http://sdb.slis.indiana.edu/>

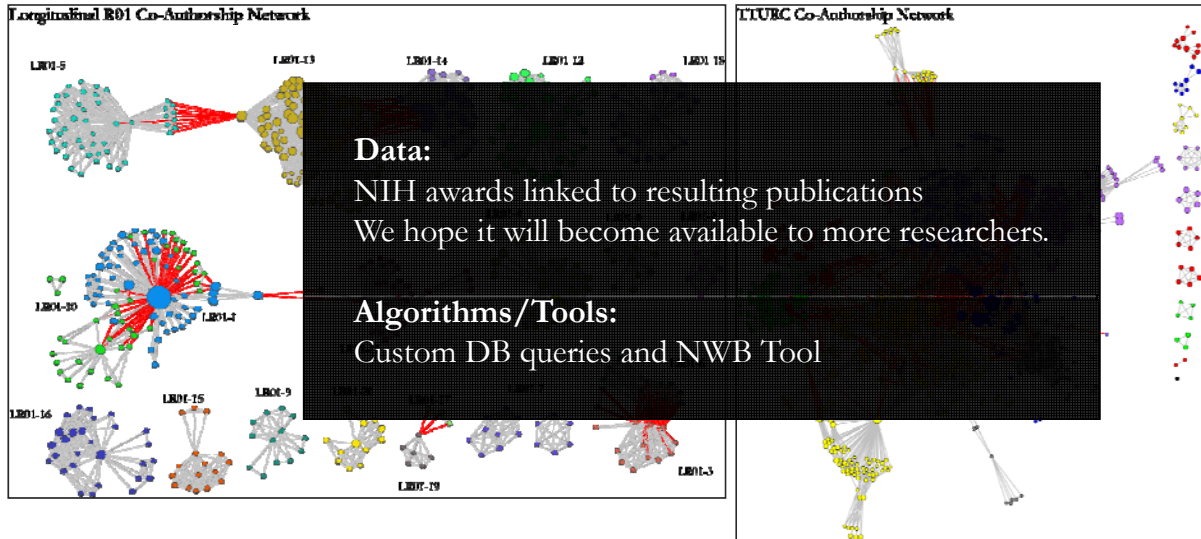
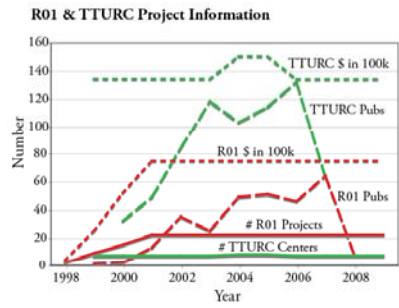
**Algorithms/Tools:**  
 Custom DB queries and code, not available.



# Mapping Transdisciplinary Tobacco Use Research Centers Publications

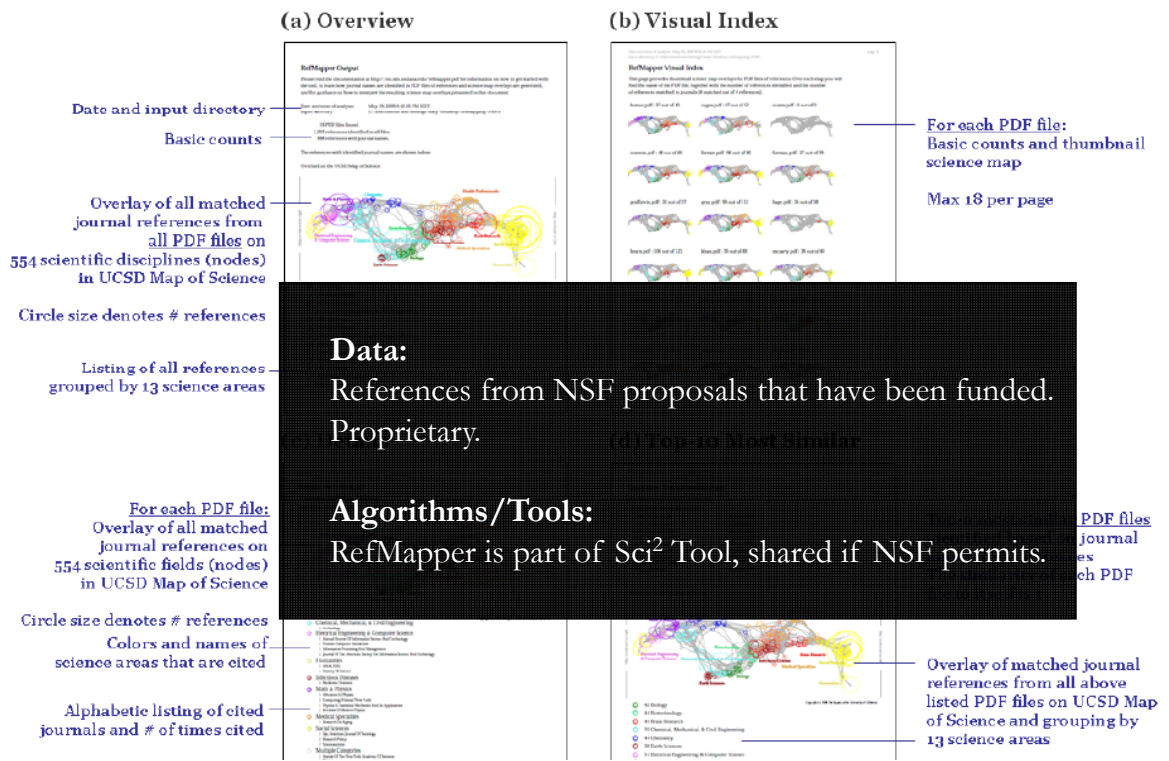
Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, *forthcoming*.



# Reference Mapper

Dubon & Börner, *forthcoming*.





## Overview

### Science of Science Studies

**Science of Science Cyberinfrastructure** (<http://sci.slis.indiana.edu>):

- **Scholarly Database (SDB)** (<http://sdb.slis.indiana.edu>) that provides free access to 23 million scholarly records
- **Sci<sup>2</sup> Tool** which reads SDB data and supports the identification of activity bursts, the extraction and display of co-author/inventor/investigator networks, and topical analysis, among others.

### Mapping Science Exhibit

Science of Science Cyberinfrastructure  
— P O R T A L —

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

**Introduction**  
E. O. Wilson writes in *Consilience: The Unity of Knowledge* (1998): "Features that distinguish science from pseudoscience are repeatability, economy, mensuration, heuristics, and consilience." Please see Börner's [recent presentation](#) at the *A Deeper Look at the Visualization of Scientific Discovery* NSF Workshop for a general introduction of the needs and the resources provided here.

**Needs Analysis**  
As part of the "TLS: Towards a Macroscopic for Science Policy Decision Making" NSF SBE-0738111 award, interviews with science policy makers are conducted to identify what science of science research results and tools might be most desirable and effective. So far, 30 formal, one-hour interviews have been conducted with science policy makers at university campus level, program officer level, and division director level for governmental, state, and private foundations. Data compilation will start in October 2008 and resulting report can be ordered by sending a request to Mark Price ([maaprice@indiana.edu](mailto:maaprice@indiana.edu)).

**Conceptualization of Science**  
A science of science requires a theoretically grounded and practically useful conceptualization of the structure and evolution of science. A special journal issue entitled "*Science of Science: Conceptualizations and Models of Science*" edited by [Katy Börner](#), Indiana University & [Andrea Schamhorst](#), Royal Netherlands Academy of Arts and Sciences invites contributions on this topic. It will be published in the *Journal of Informetrics* 3(1) in January 2009.

**Scholarly Database**  
The [Scholarly Database \(SDB\)](#) at Indiana University aims to serve researchers and practitioners interested in the analysis, modeling, and visualization of large-scale scholarly datasets. The database currently provides access to over 20 million papers, patents and grants. Resulting datasets can be downloaded in bulk. Register for free access at <https://sdb.slis.indiana.edu/>.

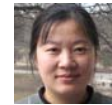
**Cyberinfrastructures**  
The Scientometrics filling of the [Network Workbench \(NWB\) Tool](#) provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and Bibtext files, or NSF awards can be read and diverse networks can be extracted and studied. Download [User Manual with focus on Scientometrics](#).

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



# Scholarly Database

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



Nianli Ma

“From Data Silos to Wind Chimes”

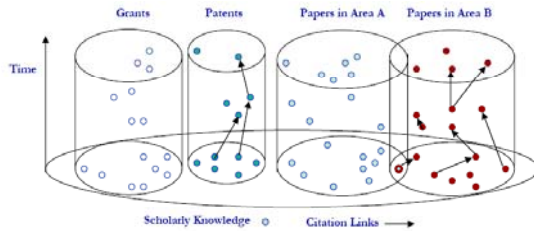
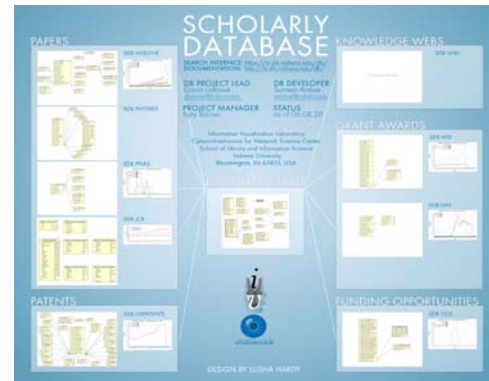


Figure 1: The interoperability and cross linkage problem. Many but not all of today's scholarly datasets, e.g., papers, patents, grants, are stored and made available so that 'vertical' citation linkages can be traversed. There are very few instances in which datasets of different origin and/or type are 'horizontally' interlinked.



- Create public databases that any scholar can use. Share the burden of data cleaning and federation.
- Interlink creators, data, software/tools, publications, patents, funding, etc.

La Rowe, Gavin, Ambre, Sumeet, Burgoon, John, Ke, Weimao and Börner, Katy. (2007) *The Scholarly Database and Its Utility for Scientometrics Research*. In *Proceedings of the 11th International Conference on Scientometrics and Informetrics, Madrid, Spain, June 25-27, 2007*, pp. 457-462. <http://ella.slis.indiana.edu/~katy/paper/07-issi-sdb.pdf>



## Scholarly Database: Web Interface

Anybody can register for free to search the about 23 million records and download results as data dumps.

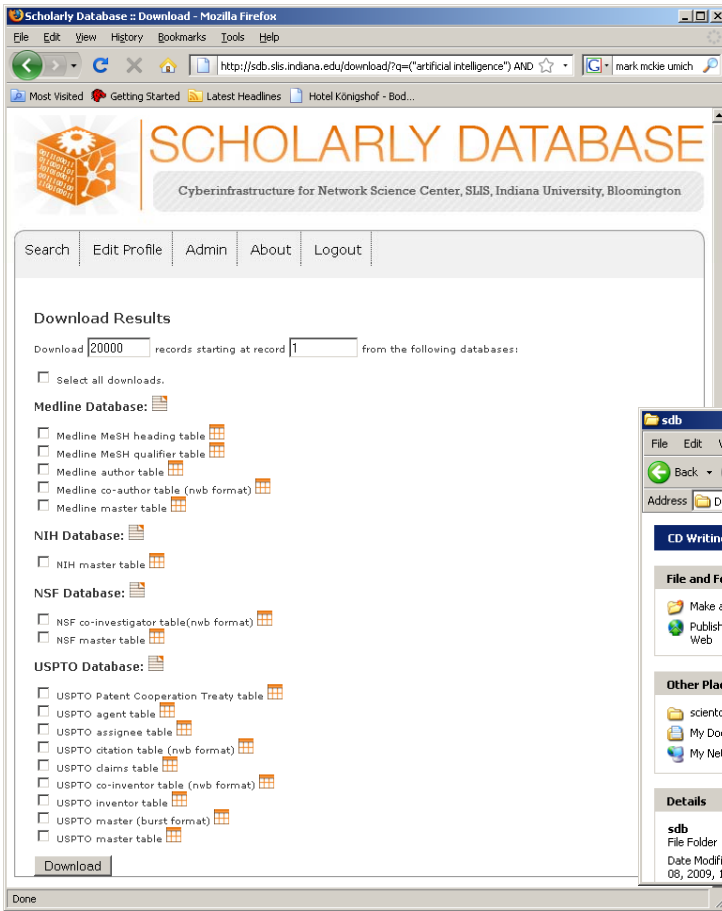
Currently the system has over 130 registered users from academia, industry, and government from over 60 institutions and four continents.

The screenshot shows the Scholarly Database web interface in a Mozilla Firefox browser. The search results page displays the following information:

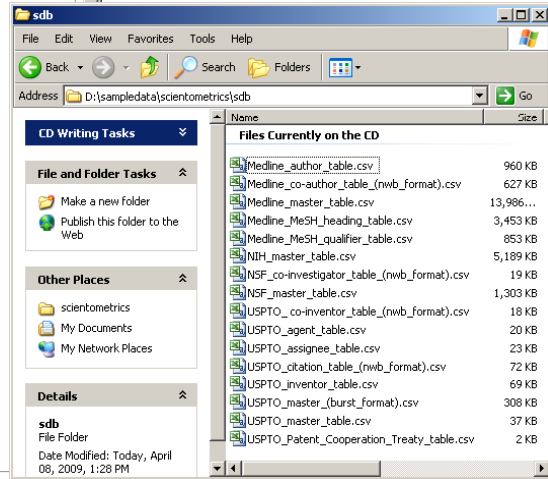
**Search Results:**

- Your search returned 13,233 results in 0.295 seconds.
- Total results per database: NIH: 2,103, Medline: 10,235, USPTO: 279, NSF: 614.
- Results 1 through 20.

Source	Authors/Creators	Year	Title	Score (out of 5.71)
Medline	LaCombe	1987	Artificial intelligence.	5.71
Medline	Schmitt	1989	Artificial intelligence: expert systems.	5.71
Medline	Schmitt	1990	[Artificial intelligence in dentistry]	5.71
Medline	Adlansing and Adlansing	2002	Artificial-intelligence-augmented systems.	5.60

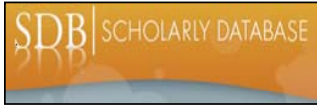


Since March 2009:  
Users can download networks:  
- Co-author  
- Co-investigator  
- Co-inventor  
- Patent citation  
and tables for  
burst analysis in NWB.



## SDB Demo

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

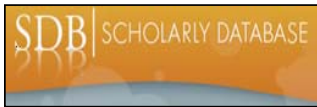


## Scholarly Database: # Records, Years Covered

Datasets available via the Scholarly Database (\* internally)

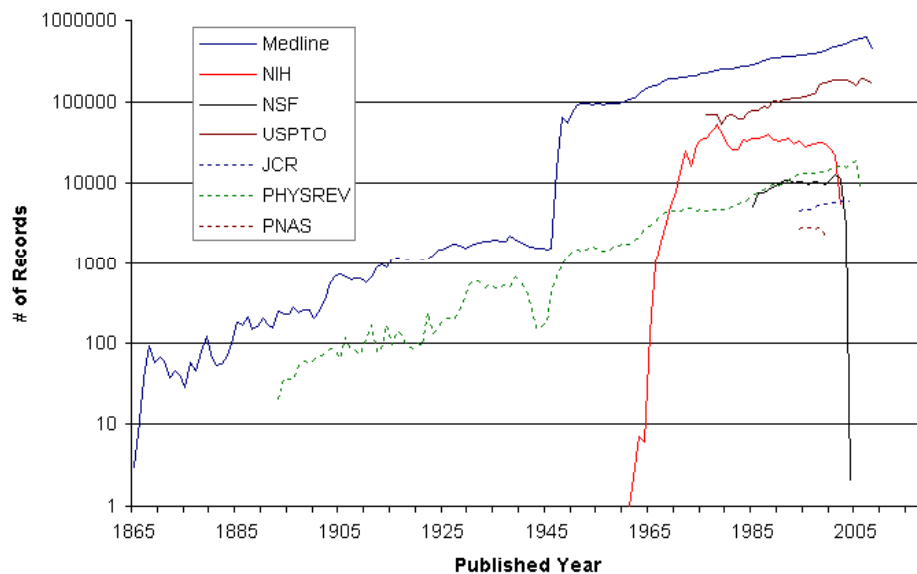
Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	17,764,826	1898-2008	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,875,694	1976-2008	Yes*	
NSF	174,835	1985-2002	Yes*	
NIH	1,043,804	1961-2002	Yes*	
<b>Total</b>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

Aim for comprehensive time, geospatial, and topic coverage.



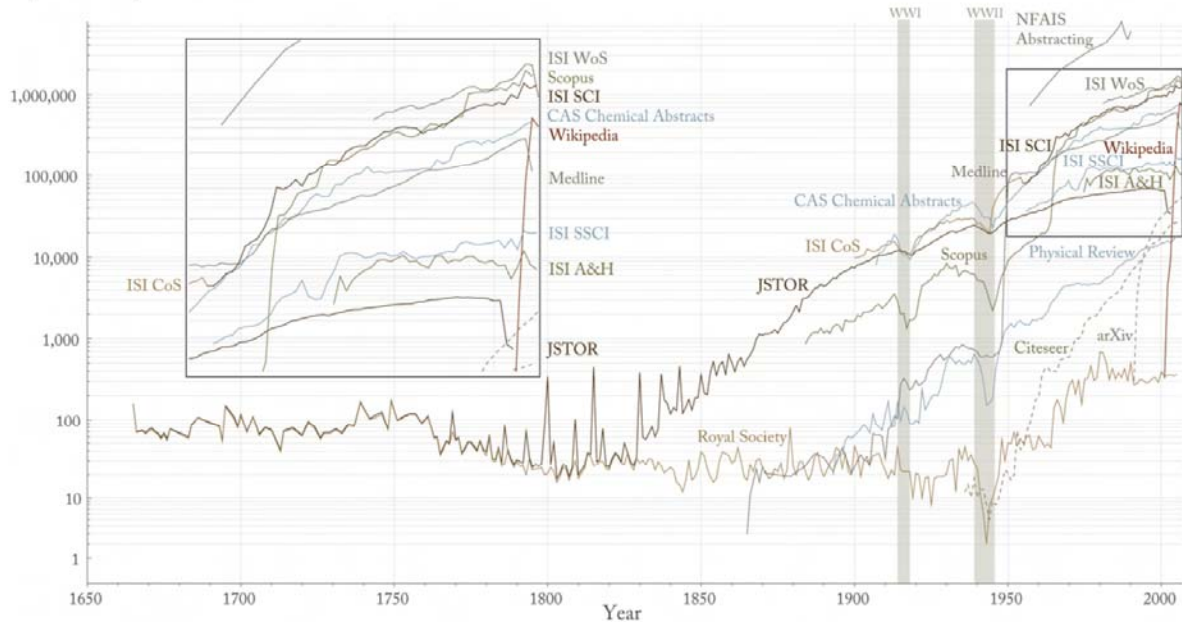
## Temporal and Geospatial Coverage

SDB Records in August 2009



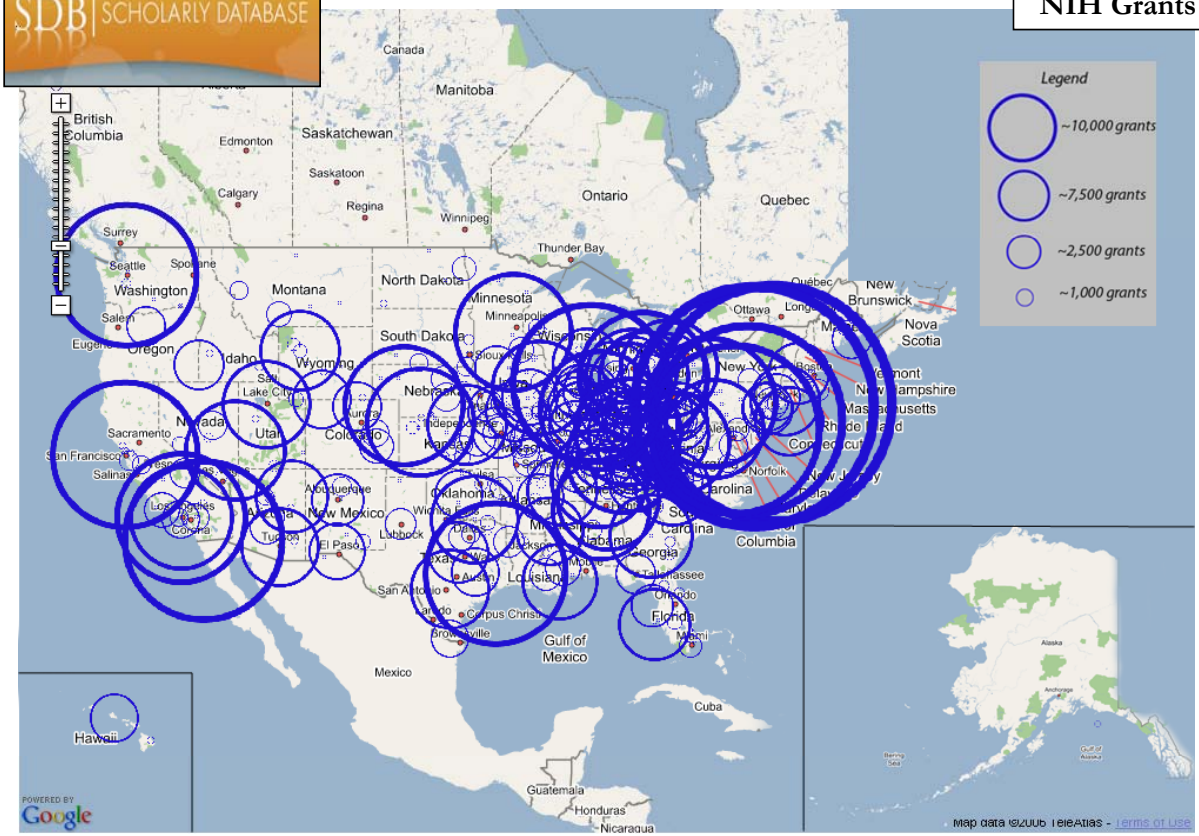
## Comparison with Major Publication Data commonly used in scientometric studies

Papers & Wikipedia Entries

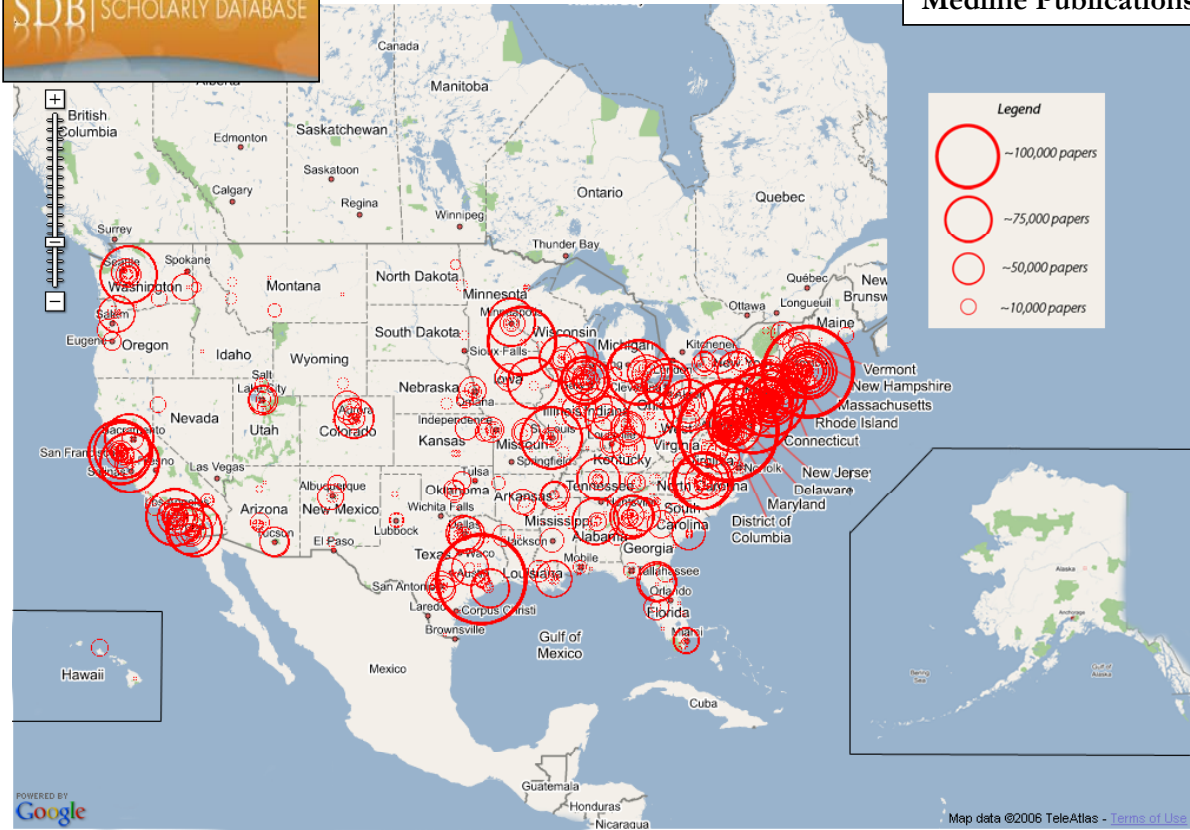


*'Atlas of Science: Guiding the Navigation and Management of Scholarly Knowledge', Part I: The Rise of Science and Technology. Chart showing the number of papers/wikipedia entries for different databases and publication years. Contact Katy Börner <katy@indiana.edu> or Elisba Hardy <ehardy@indiana.edu> for details.*

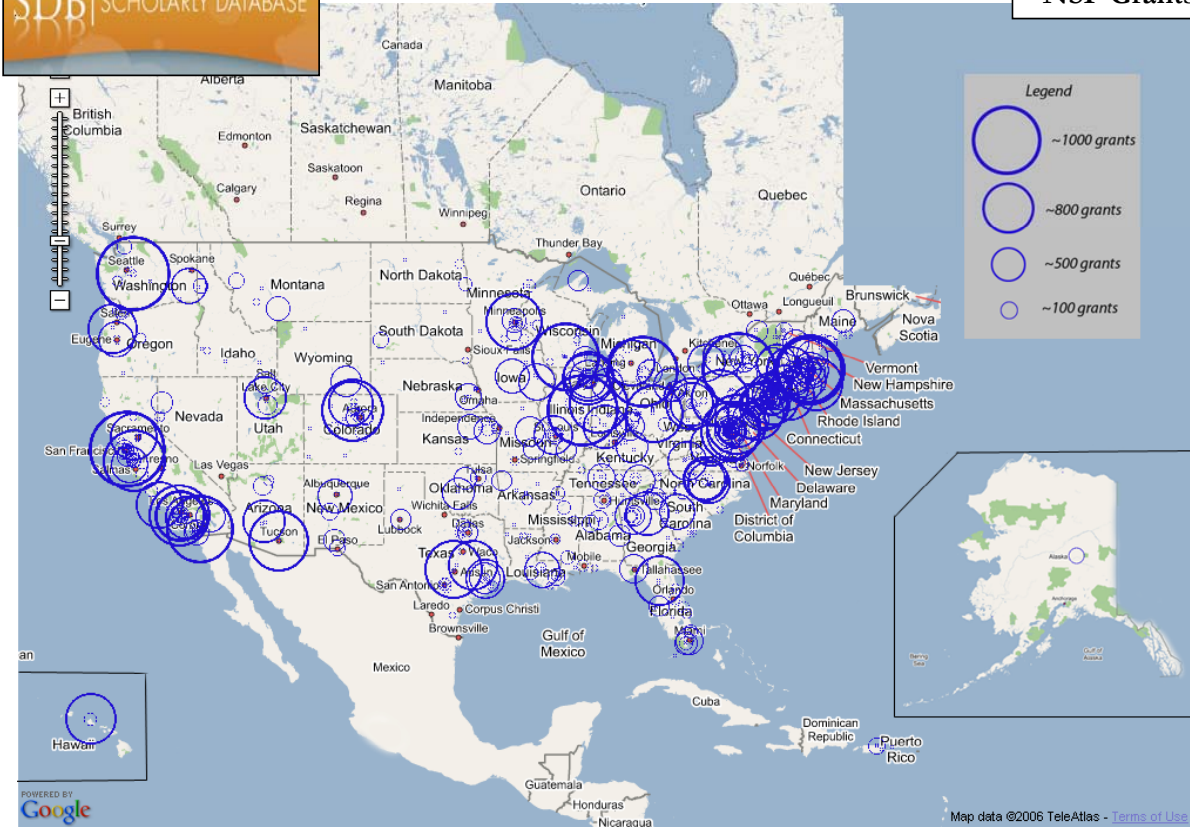
## NIH Grants



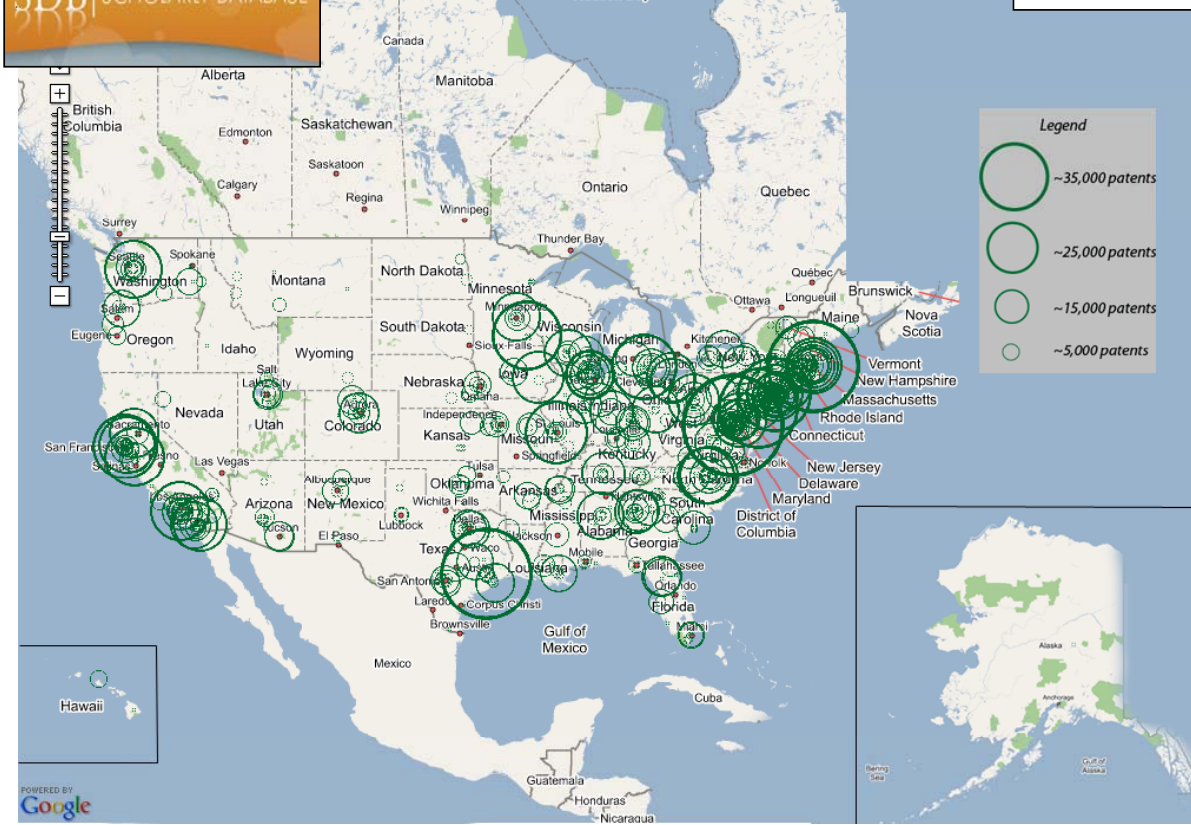
Medline Publications



NSF Grants





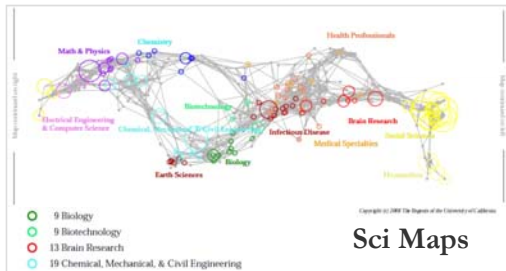


Sci<sup>2</sup> Tool

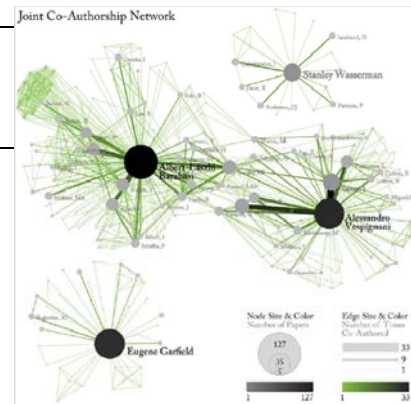
<http://sci.slis.indiana.edu>

“Open Code for S&T Assessment”

Branded OSGi/CIShell based tool with NWB plugins and many new plugins.



Sci Maps



GUESS Network Vis

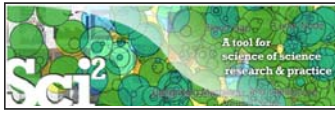
Geo Maps

Hierarchical Circular Visualization

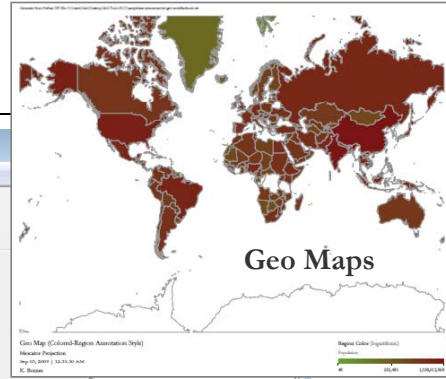
Horizontal Time Graphs



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, 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

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

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

The primary investigators are Katy Börner, Ingrid Isenhardt, 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>).

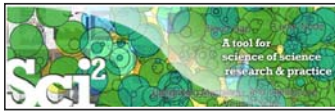
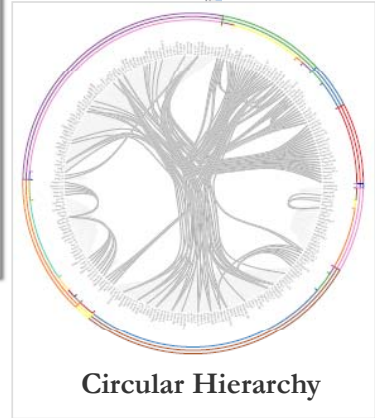
Please cite as follows:  
Sci² Team. (2009). Science of Science Tool. Ingrid Isenhardt, SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

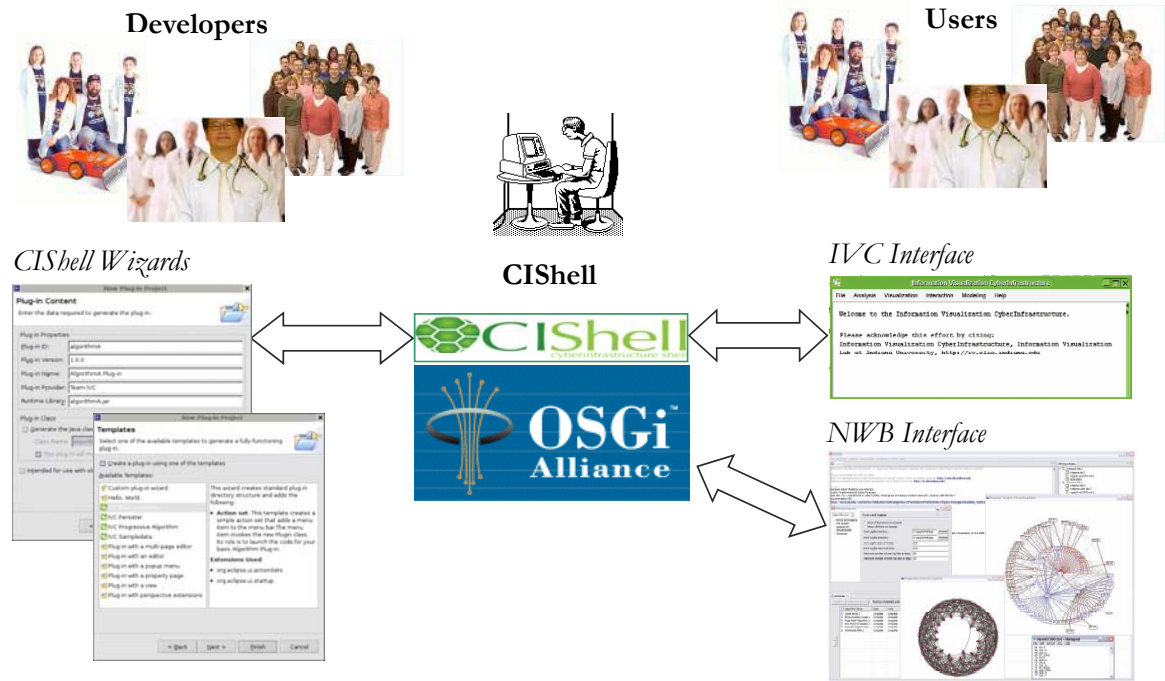
Remove From List  Remove completed

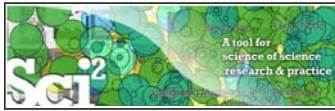
Algorithm Name	Date	Time	% Complete
Extract Co-Author Network	09/03/2009	00:15:20 AM	100%
Load and Clean ISI File	09/03/2009	00:15:05 AM	100%

- GUESS
- GnuPlot
- 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 annotations)
- Geo Map (region coloring annotations)
- Image Viewer
- RefMapper



# Serving Non-CS Algorithm Developers & Users





## Sci<sup>2</sup> Tool: Supported Data Formats

### Personal Bibliographies

- Bibtext (.bib)
- Endnote Export Format (.enw)

### Data Providers

- Web of Science by Thomson Scientific/Reuters (.isi)
- Scopus by Elsevier (.scopus)
- Google Scholar (access via *Publish or Perish* save as CSV, Bibtext, EndNote)
- Awards Search by National Science Foundation (.nsf)

### Scholarly Database (all text files are saved as .csv)

- Medline publications by National Library of Medicine
- NIH funding awards by the National Institutes of Health (NIH)
- NSF funding awards by the National Science Foundation (NSF)
- U.S. patents by the United States Patent and Trademark Office (USPTO)
- Medline papers – NIH Funding

### Network Formats

- NWB (.nwb)
- Pajek (.net)
- GraphML (.xml or .graphml)
- XGMML (.xml)

### Burst Analysis Format

- Burst (.burst)

### Other Formats

- CSV (.csv)
- Edgelist (.edge)
- Pajek (.mat)
- TreeML (.xml)

37



## NWB=Sci<sup>2</sup> Tool: Algorithms (July 1st, 2008)

See <https://nwb.slis.indiana.edu/community> and handout

### Preprocessing Edit

#### Remove Nodes

- [Extract Top Nodes](#)
- [Extract Nodes Above or Below Val](#)
- [Delete High Degree Nodes](#)
- [Delete Random Nodes](#)
- [Delete Isolates](#)

#### Remove Edges

- [Extract Top Edges](#)
- [Extract Edges Above or Below Val](#)
- [Remove Self Loops](#)
- [Trim By Degree<sup>2</sup>](#)
- [Pathfinder Network Scaling](#)

#### Sampling

- [Snowball Sampling \(n nodes\)](#)
- [Node Sampling](#)
- [Edge Sampling](#)

#### Transformations

- [Symmetrize](#)
- [Dichotomize](#)
- [Multipartite Joining](#)

### Modeling Edit

#### General

- [Random Graph](#)
- [Watts-Strogatz Small World](#)
- [Barabási-Albert Scale-Free](#)

#### Structured

- [CAN](#)
- [Chord](#)

#### Unstructured

- [Hypergrid](#)
- [PRU](#)

#### Other

- [TARL](#)
- [Discrete Network Dynamics](#)

### Analysis Edit

#### General Purpose

- [Network Analysis Toolkit<sup>2</sup>](#)

#### Unweighted & Undirected

- Based on degree/  
[Node Degree](#)  
[Node Distribution](#)

#### Based on clustering

- [k-Nearest Neighbor](#)
- [Watts Strogatz Clustering Coefficient](#)
- [Watts Strogatz Clustering Coefficient Over k](#)

#### Based on path

- [Diameter](#)
- [Average Shortest Path](#)
- [Shortest Path Distribution](#)
- [Node Betweenness Centrality](#)

#### Based on components

- [Connected Components](#)
- [Weak Component Clustering](#)

#### K-Core

- [Extract K-Core<sup>2</sup>](#)
- [Annotate K-Core<sup>2</sup>](#)

#### Unweighted & Directed

- Based on degree  
[Node Indegree](#)  
[Node Outdegree](#)  
[Indegree Distribution](#)  
[Outdegree Distribution](#)

#### Based on local graph structure

- [k-Nearest Neighbor](#)
- [Single Node In-Out Degree Correlations<sup>2</sup>](#)

#### Unnamed Category?

- [Page Rank](#)

#### Based on local graph structure #2

- [Dyad Reciprocity<sup>2</sup>](#)
- [Arc Reciprocity<sup>2</sup>](#)

### tion Edit

#### Tools

- [GUESS](#)
- [GnuPlot<sup>2</sup>](#)

#### Predefined Positions Layout

- [DrL \(VxOrd\)](#)
- [Pre-defined Positions \(prefuse beta\)<sup>2</sup>](#)

#### Move

- [Circular](#)

#### Tree Layouts

- [Radial Tree \(prefuse alpha\)](#)
- [Radial Tree with Annotations \(prefuse beta\)<sup>2</sup>](#)
- [Tree Map](#)
- [Tree View](#)
- [Balloon Graph \(prefuse alpha\)<sup>2</sup>](#)

#### Network Layouts

- [Force Directed with Annotation \(prefuse beta\)](#)
- [Kamada-Kawai \(JUNG\)](#)
- [Fruchterman-Reingold \(JUNG\)](#)
- [Fruchterman-Reingold with Annotation \(prefuse beta\)](#)
- [Spring \(JUNG\)](#)
- [Small World \(prefuse alpha\)](#)

#### Other Layouts

- [Parallel Coordinates \(demo\)<sup>2</sup>](#)
- [LaNet \(k-Core Decomposition\)](#)

### etrics Edit

#### Extract Network From Table

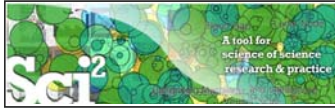
- [Extract Co-Authorship Network](#)
- [Extract Co-Occurrence Network From Table<sup>2</sup>](#)
- [Extract Directed Network From Table<sup>2</sup>](#)

#### Extract Network From Another Network

- [Extract Bibliographic Coupling Similarity Network](#)
- [Extract Co-Citation Similarity Network<sup>2</sup>](#)

#### Cleaning

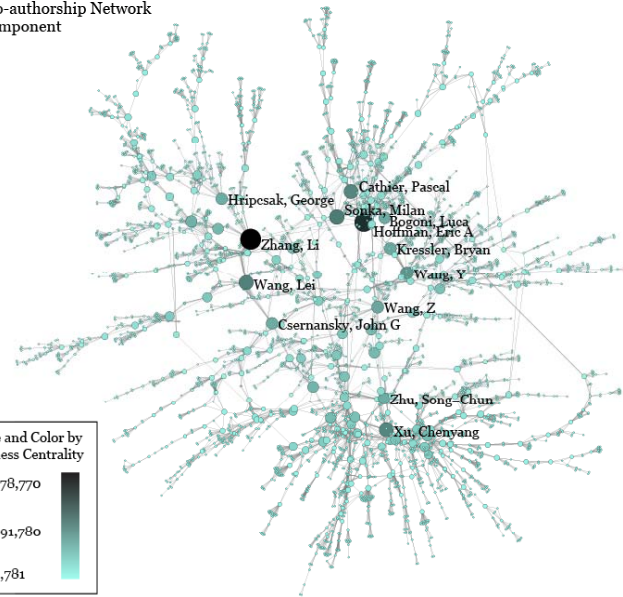
- [Remove ISI Duplicate Records](#)



## NWB=Sci<sup>2</sup> Tool: Output Formats

- NWB tool can be used for data conversion. Supported output formats comprise:
  - CSV (.csv)
  - NWB (.nwb)
  - Pajek (.net)
  - Pajek (.mat)
  - GraphML (.xml or .graphml)
  - XGMML (.xml)
- GUESS
  - Supports export of images into common image file formats.
- Horizontal Bar Graphs
- saves out raster and ps files.

Medline Co-authorship Network  
Largest Component



## Exemplary Analyses and Visualizations

### Individual Level

- A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
- B. Loading NSF grant data for network science researchers at Indiana U

### Institution Level

- C. Indiana U, C

### Scientific Field Level

- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDI

**Will be presented in hands-on Workshop on Thursday Sept 3, 2009, 1-5pm**

**Together with guidance on how to design workflows using 100+ algorithms and how to dissect and design effective visualizations.**

**Bonus: Create your custom tool.**

co-PI

bursting

# Sci<sup>2</sup> Tool Demo

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

NSF - Award Search - Awardee Information - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.nsf.gov/awardsearch/piSearch.do?SearchType=piSearch&page=1&QueryText=&PIFirstName=&PILastName= nsf awards search

Most Visited Getting Started Latest I headlines

Search What's Hot What's New Featured For You Notifications

Sources... Create Wiki page for... create Highlighting Hide Behavior Anatomy Chemicals Diseases Genes Physiology Living Being New Other

### Search Results

[Back](#)

Results are sorted by award date, with the most recent awards at the top. Click on a column heading to re-sort the results. The up/down arrows at the right of each column title control whether the sort is ascending or descending. To view the abstract, click on the award number or title. Click on the data in other columns to perform a new search with that parameter.

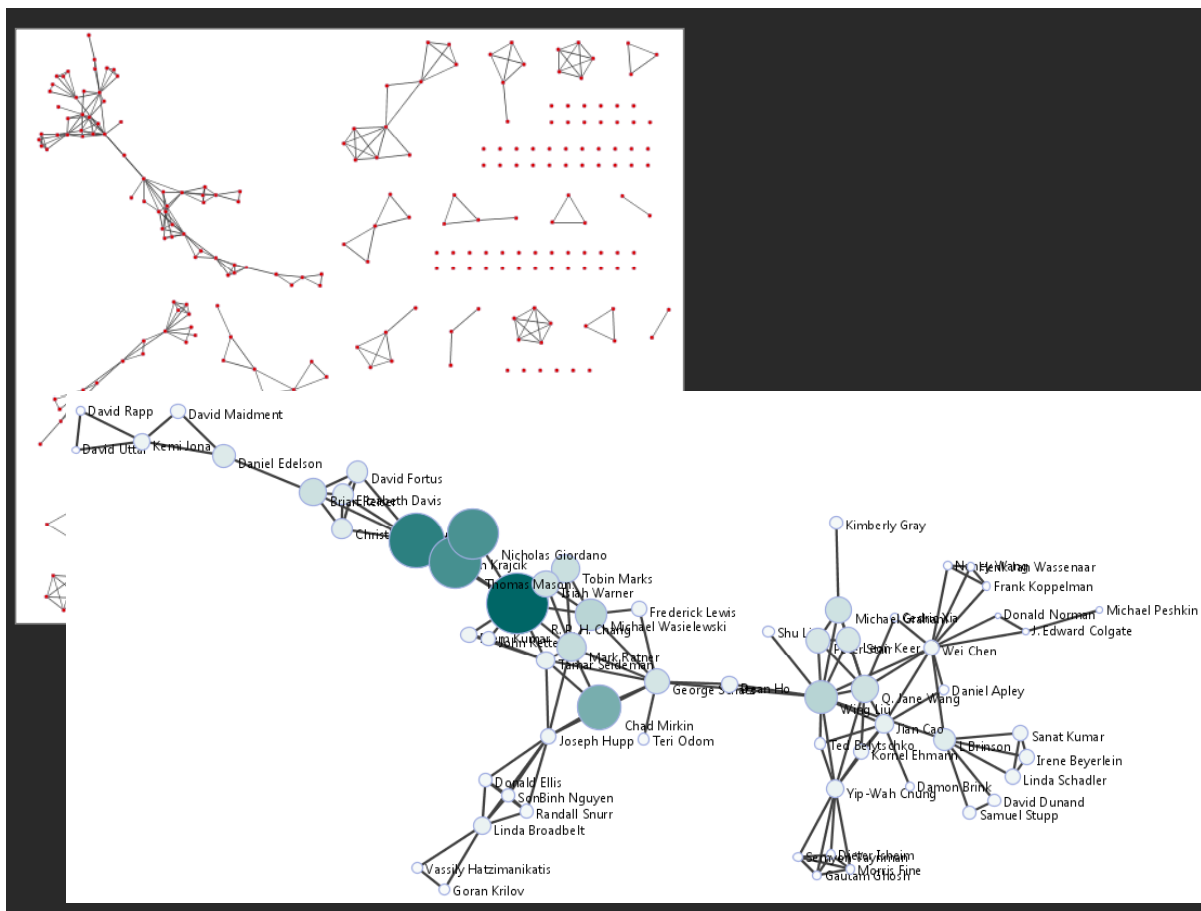
[Refine Search](#)

**344 awards found, displaying 1 to 50.**  
[First/Prev] 1, 2, 3, 4, 5, 6, 7 [Next/Last]

Award Number	Title	NSF Organization	Program(s)	Start Date	Principal Investigator	State	Organization
<a href="#">0938075</a>	<a href="#">Developing Guidelines for Using Digital Media Visualization Resources to Support Student Inquiry in Online Laboratory Investigations</a>	DUE	NATIONAL SMETE DIGITAL LIBRARY	01/01/2010	<a href="#">Jona, Kemi</a>	IL	<a href="#">Northwestern University</a>
<a href="#">0843252</a>	<a href="#">Causal Supports for Early Word Learning</a>	BCS	DEVELOP& LEARNING SCIENCES/CR1	09/01/2009	<a href="#">Booth, Amy</a>	IL	<a href="#">Northwestern University</a>
<a href="#">0845063</a>	<a href="#">CAREER: Unlocking the Synthetic Potential of N-Allylhydrazones</a>	CHE	METHODOLOGY	09/01/2009	<a href="#">Thomson, Regan</a>	IL	<a href="#">Northwestern University</a>

Done

<http://www.nsf.gov/awardsearch>



## Outlook

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 (Sc<sup>2</sup> 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 corpuses 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 or macroscopes will expand.



## Overview

### Science of Science Studies

Science of Science Cyberinfrastructure (<http://sci.slis.indiana.edu>):

- **Scholarly Database (SDB)** (<http://sdb.slis.indiana.edu>) that provides free access to 23 million scholarly records
- **Sci<sup>2</sup> Tool** which reads SDB data and supports the identification of activity bursts, the extraction and display of co-author/inventor/investigator networks, and topical analysis, among others.

### Mapping Science Exhibit

## Mapping Science Exhibit – 10 Iterations in 10 years

<http://scimaps.org>



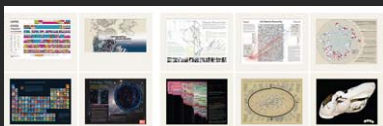
### The Power of Maps (2005)



### Science Maps for Economic Decision Makers (2008)



### The Power of Reference Systems (2006)



### 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)

### The Power of Forecasts (2007)



### 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  
- Wallenberg Hall, Stanford University, CA  
- Center of Advanced European Studies and Research, Bonn, Germany  
- Science Train, Germany.

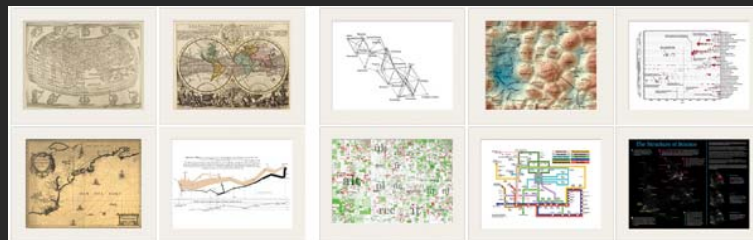




Debut of 5<sup>th</sup> Iteration of Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, <http://mediax.stanford.edu>, <http://scaleindependentthought.typepad.com/photos/scimaps>

# The Power of Maps

## Four Early Maps of Our World VERSUS Six Early Maps of Science



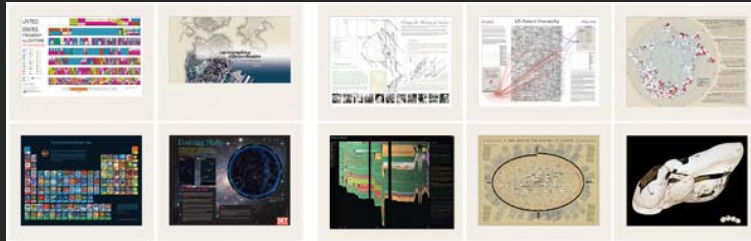
*(1st Iteration of Places & Spaces Exhibit - 2005)*





# The Power of Reference Systems

## Four Existing Reference Systems VERSUS Six Potential Reference Systems of Science

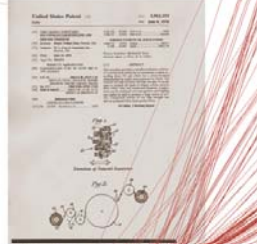


*(2<sup>nd</sup> Iteration of Places & Spaces Exhibit - 2006)*

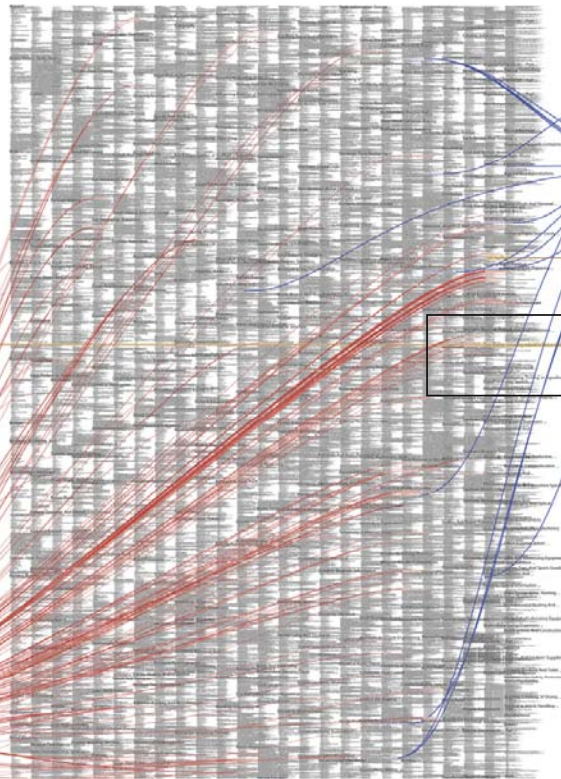
### Impact

The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that organizes them by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 165,523 categories in a hierarchy that can get as deep as 13 levels. We display the first three levels (13,329 categories) at right in what might be considered a virtual map of inventions. Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (enlarged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical connective implants, that mention Goretex as prior art.



### US Patent Hierarchy



### Prior Art



New patents often build on older ideas from many categories. Here, blue lines originate in distinct different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a "taxonomy validator" that helps people decide whether categories are good ones.

Categories can be redefined or combined, and sometimes need to be split when they become too large, a constant problem shared by many classifications systems in this information-rich century. But how can we determine exactly where to split a category in two, for example—if there are hundreds or thousands of elements in it?

The Taxonomy Validator measures a "distance to prototype" (how far each element is from an idealized "prototype" element for each bucket. This can be based on statistics, computational comparisons of words, or even human judgements. A simple bar chart can then show how good a category is. A good category has lots of small bars; a generally ragged category is one that might need scrutiny or reorganization; while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than can fit into memory as words, focusing people on just the right issues, and providing a vastly broader background to support more informed judgements.



## Impact

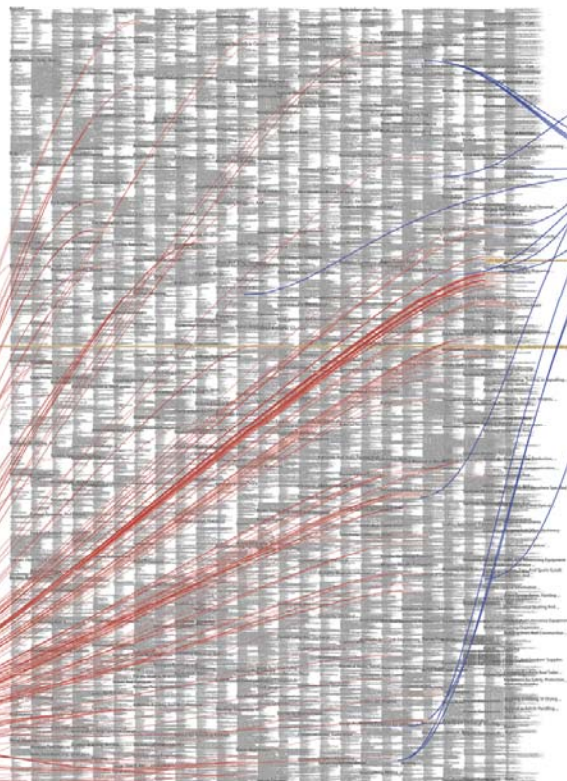
The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,523 categories in a hierarchy that can get as deep as 15 levels. We display the first three levels (13,529 categories) at right in what might be considered a visual map of inventions.

Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

The patent on Gore-tex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Gore-tex as prior art.



## US Patent Hierarchy



## Prior Art



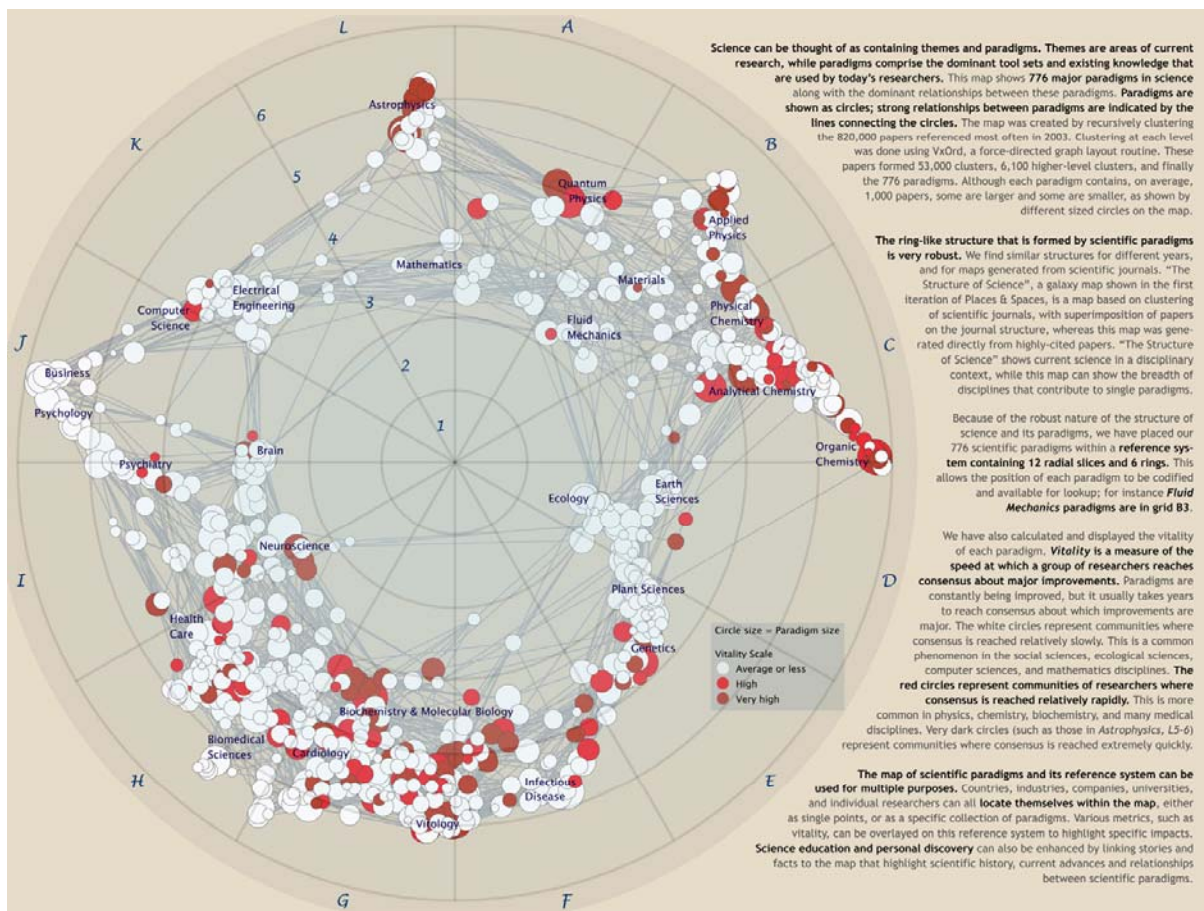
New patents often build on older ideas from many categories. Here, blue lines originate in sixteen different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a "Taxonomy Validator" that helps people decide whether categories are good ones.

Categories can be redefined or combined, and sometimes need to be split when they become too large; a constant problem shared by many classifications systems in this information rich century. But how can we determine exactly where to split a category in two, for example—if there are hundreds or thousands of elements in it?

The Taxonomy Validator measures a "distance to prototype" how far each element is from an idealized "prototype" element for each bucket. This can be based on statistics, computational comparisons of words, or even human judgement. A simple bar chart can then show how good a category is. A good category has lots of small bars; a generally rigged category is one that might need accuracy or reorganization, while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than can fit its memory as words, focusing people on just the right issues, and providing a vastly broader background to support more informed judgments.

	Synthetic Resins or Natural Rubber
	Ion-exchange Polymer or Process of Preparation
	Process of Regenerating Membrane or Process of Preparing
	Previously Formed Solid Ion-exchange Polymer Admixed With Membrane
	Polymer Characterized By Defined Size or Shape Other than Bead
	Chemically Treated Solid Polymer
	Solid Polymer Derived From Ethylenically Unsaturated Reactant
	Solid Polymer Derived From At Least One 1,2-epoxy Containing
	Solid Polymer Derived From Aldehyde or Derivative
	From Ethylenically Unsaturated Reactant Only
	From Aldehyde or Derivative
	Process of Treating Scrap or Waste Product Containing At Least One
	Process of Treating Scrap or Waste Product Containing At Least One
	Treating Rubber (or Rubberlike Materials) or Polymer Derived From
	Treating Polymer Derived From A Monomer Containing Only One
	Treating Polymer Derived From Hydrocarbon Monomers Only
	Treating Polysiloxane
	Treating Polyester
	Treating With Alcohol
	Treating Polyurethane, Polyurea (excluding Urea-formaldehyde)
	Treating With Alcohol or Amine
	Treating Polycarbonamide
	Cellular Products or Processes of Preparing or Treating
	Cellular Product Derived From Two or More Solid Polymers or From
	At Least One Polymer Is Derived From Reactant Containing Two
	At Least One Polymer Is Derived From An Aldehyde or Derivative
	At Least One Polymer Is Derived From A $-n=c=x$ Reactant Where



# The Power of Forecasts

## Four Existing Forecasts VERSUS Six Potential Science 'Weather' Forecasts



*(3<sup>rd</sup> Iteration of Places & Spaces Exhibit - 2007)*

### 113 Years of Physical Review

This visualization aggregates 38,000 articles published in 720 volumes of *Physical Review* between 1893 and 2005. The 7,732 articles published from 1893 to 1976 take up the left portion of the map. In 1977, the *Physical Review* introduced the *Physics and Astronomy Classification Scheme* (PACS) codes and the visualization subdivides into the top-level PACS codes. The 20,000 articles from 1977 to 2005 for which good PACS codes are available occupy the middle third of the map. The 8,624 articles from 2006 to 2008, for which good PACS codes are available, fill the last third of the map.

Each article link is color-coded according to the PACS code that appears in it with height proportional to the number of papers, and each journal is color-coded according to the volume of the journal appearing in the column.

On top of the base map, all citations from the papers in every top-level PACS code in 2005 are overlaid and then drawn from the source view to the individual volumes containing papers cited.

The small Nobel Prize models indicate the 24 volumes spanning the papers appearing in *Physical Review* by Nobel Prize between 1901 and 2003. Each year Thomson ISI predicts new Nobel Prize awardees or places based on citations, grants, high impact papers, and discoveries or the number of journal citations. Current predictions by Thomson ISI are highlighted.

#### Nobel Prizes in Physical Review

Year of Nobel Prize Winner's Publication Year(s) (indicated by Nobel Prize models on the right)

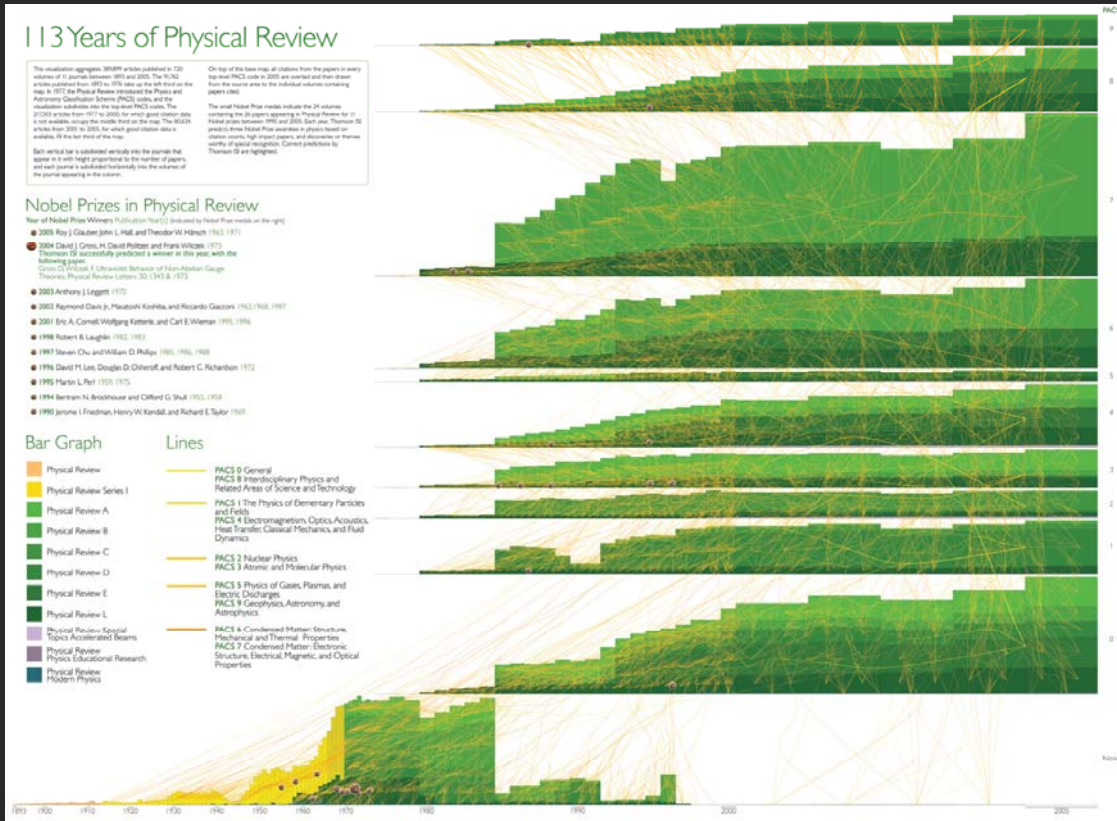
- 2005 Roy J. Glauber, John L. Hall, and Theodor W. Hänsch (1963, 1971)
- 2004 David J. Gross, H. David Politzer, and Frank Wilczek (1973)
- Thomson ISI successfully predicted a winner in this year, with the following paper:  
Gross, D.J., Politzer, H. Ultraviolet Behavior of Non-Abelian Gauge Theories, *Physical Review Letters* 30, 1303 (1973)
- 2003 Anthony J. Leggett (1972)
- 2002 Raymond Davis Jr., Masatoshi Koshiba, and Riccardo Giacconi (1963, 1968, 1987)
- 2001 Eric A. Cornell, Wolfgang Ketterle, and Carl E. Wieman (1995, 1996)
- 1998 Robert B. Laughlin (1982, 1987)
- 1997 Steven Chu, and William D. Phillips (1982, 1988, 1989)
- 1996 David H. Lee, Douglas D. Osheroff, and Robert C. Richardson (1972)
- 1995 Martin L. Perl (1975, 1976)
- 1994 Brian N. Brockhouse and Clifford G. Shull (1953, 1955)
- 1993 Jerome I. Friedman, Henry W. Kendall, and Richard E. Taylor (1969)

#### Bar Graph

- Physical Review
- Physical Review Series I
- Physical Review A
- Physical Review B
- Physical Review C
- Physical Review D
- Physical Review E
- Physical Review L
- Physical Review: Special Topics Accelerated Beams
- Physical Review: Physics Educational Research
- Physical Review: Modern Physics

#### Lines

- PACS 0 General
- PACS 8 Interdisciplinary Physics and Related Areas of Science and Technology
- PACS 1 The Physics of Elementary Particles and Fields
- PACS 4 Electromagnetism, Optics, Acoustics, Heat, Transport, Classical Mechanics, and Fluid Dynamics
- PACS 2 Nuclear Physics
- PACS 3 Atomic and Molecular Physics
- PACS 5 Physics of Gases, Plasmas, and Liquids
- PACS 9 Geophysics, Astronomy, and Astrophysics
- PACS 6 Condensed Matter: Structure, Mechanical and Thermal Properties
- PACS 7 Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties



*114 Years of Physical Review - Bruce W. Herr II, Russell Dubon, Katy Borner, Elisba Hardy, Shashikant Penumarthy - 2007*

# MAPS OF SCIENCE

A visualization of 7.2 million scholarly documents  
appearing in over 16,000 journals, proceedings or symposia  
between Jan, 2001 and Dec, 2005

This map of science was constructed by sorting more than 7.2 million journals into disciplines, disciplines represented as nodes, and sets of individual articles that cite one another link the lines between disciplines are pairs of disciplines that share a common literature. A three-dimensional map was used to determine the position of each discipline on the surface of a sphere based on the linkage between disciplines. The smallest branch links like rubber bands connecting to other nodes in the same cluster. Each of disciplines without links need to end up on different sides of the map.

The spherical map, which is not shown here, was converted to a two-dimensional projection (the same one used to draw the continents of the earth on a two-dimensional map) to give the large map shown below. This projection shows a perspective of the entire map in which all nodes that the disciplines tend to cite are along the middle of the map. If this were a map of the earth it would be like a single-colored map showing the equator. There are no disciplines at the top (north pole) or the bottom (south pole). Multiple projections also exist for disciplines, but need to be larger than the full circle to compare to the right side and bottom that the map is most important. In this map, the most important link is on the right connect with the computer science group on the left in new connections (see).

The six map projections shown at the bottom are images of what one would see if looking down to the north pole of the map, at six different locations. When viewed this way, the map looks like a wheel with an inner ring and outer ring. This wheel of nodes connects every closely every closely the two-dimensional maps we have previously produced.

**Forecasting Large Trends in Science**

Calculations were performed using the large volume quantities of disciplines (fields) to determine a set of three nodes likely to exist along in the change in the structure of science over time. Correlation coefficients between fields were calculated for each individual year (2001-2005). A single regression analysis was conducted to see if there were significant changes in these correlation coefficients from year to year.

If the structure of science shows future is more toward stability, we would expect some relations between neighboring fields to increase, and connections between distant fields to decrease. We found the opposite, suggesting that the underlying structure is unstable and likely to change dramatically over the next decade.

Six clusters, representing how the structure is likely to change, are provided below. Maps with white arrows represent increases of distant fields, and likely to be added close to each other in the future. Maps with dark arrows represent fields that are currently close to each other, but likely to become more dispersed. We expect that large changes of science will show changes in relative correspondences to these disciplines. Labels of the figure suggest, while the physical colors will toggle and draw close to the medical fields.

**Physics and Chemistry**, indicated by the light green color above the top center position. The connection with the largest growth rate (17%) are shown in black arrows, and will draw a close dispersion of the medical fields compared to the current location.

**Aeronautical Chemical Mechanical & Civil Engineering**, indicated by the light green color above the top center position. The connection with the largest growth rate (17%) are shown in black arrows, and will draw a close dispersion of the medical fields compared to the current location.

**Infectious Diseases**, indicated by the dark red color above the top center position. The connection with the largest growth rate (17%) are shown in black arrows, and will draw a close dispersion of the medical fields compared to the current location.

**Medical Specialties**, indicated by the dark red color above the top center position. The connection with the largest growth rate (17%) are shown in black arrows, and will draw a close dispersion of the medical fields compared to the current location.

**Health Professionals**, indicated by the dark red color above the top center position. The connection with the largest growth rate (17%) are shown in black arrows, and will draw a close dispersion of the medical fields compared to the current location.

**Social Sciences**, indicated by the yellow color above the top center position. The connection with the largest growth rate (17%) are shown in black arrows, and will draw a close dispersion of the medical fields compared to the current location.

Source: University of California, San Diego Knowledge Mapping Laboratory, Color Design: © Regents of the University of California. The underlying data come from two sources: Thomson ISI and Scopus. Mapping methodology and description text by Dick Ravans, President, iScholar Strategies, Inc., and Anne Peacock, Senior National Laboratories, Graphics & Geography by Ethan Walker and Mike Potts. Special acknowledgments to Kay Bowen, Dr. Bill, Dr. Stephen Hay, Lou Simon, and Henry Stahl. © 2007 by Dick Ravans, all rights reserved.

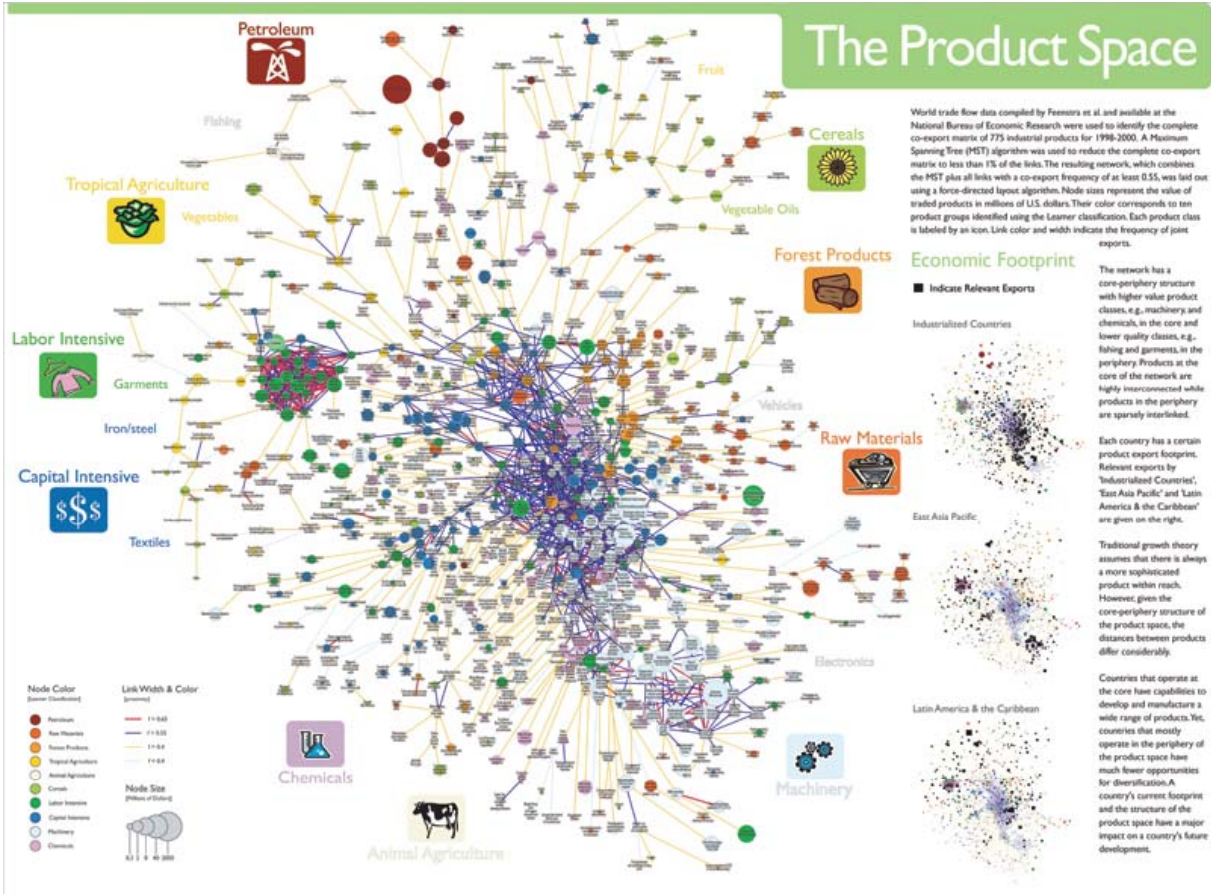
*Maps of Science: Forecasting Large Trends in Science - Richard Klavans, Kevin Boyack - 2007*

## Science Maps for Economic Decision Making

### Four Existing Maps VERSUS Six Science Maps

*(4<sup>th</sup> Iteration of Places & Spaces Exhibit - 2008)*

# The Product Space



## Happiness Depends on Various Factors

Social scientists are starting to include relative happiness with hard data on economic status, health, and other factors as they assess quality of life. They rely on surveys of 'subjective well-being'—how good people feel about their lives. A world map of one 'happiness index' shows many, but not all, wealthy northern countries faring well. Residents of sub-Saharan Africa and the former Soviet Union, meanwhile, report particularly low levels of contentment.

Any attempt to measure happiness will fall short—each life is a series of joys, struggles, and sorrows, and satisfaction can depend as much on outlook as on circumstances. Averages obscure the happy moments in struggling nations, as well as people who suffer from poor health, poverty, or discrimination in countries that rank high. Still, happiness indices can help researchers move beyond simple economics as they track progress—or backsliding—over time.

**MEASURING THE INTANGIBLE**  
 The map is derived from the New Economics Foundation's 2008 'Happy Planet Index,' which drew on over 100 surveys of subjective well-being. Its 'satisfaction with life scale'—a happiness index—ranks the relative happiness of nations, from a high of 273 (Denmark and Switzerland) to a low of 100 (Burundi).

**Happiness Index**  
 Very Happy  
 Happy  
 Content  
 Unhappy  
 No data  
 SOURCE: WFP, A. JON



**HEALTH**  
 Japan boasts the world's longest life expectancy—one measure of overall health. Swaziland, at the other end of the scale, is plagued by poverty, disease, and violence. Disparities in access to health care divide many countries into haves and have-nots.



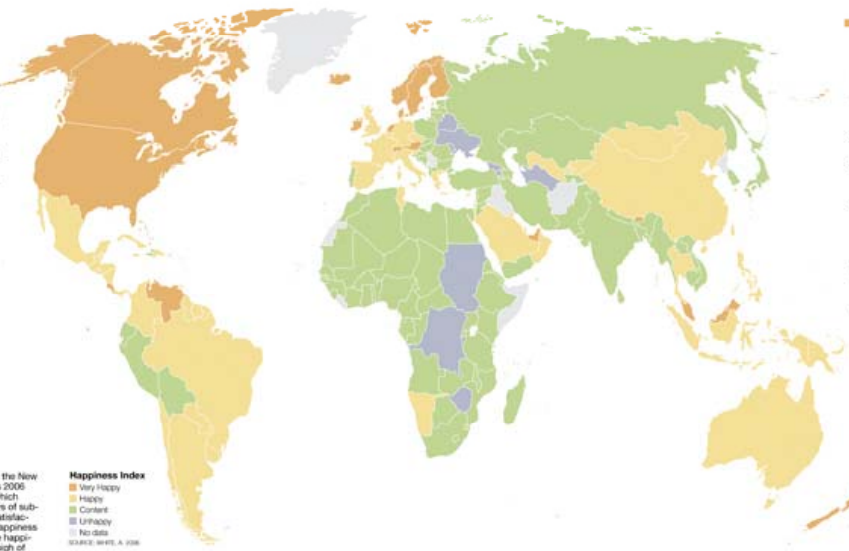
**WEALTH**  
 Money still can't buy love, or happiness, and wealthier people aren't always more content. Still, tiny Luxembourg, which takes top rank in per capita Gross Domestic Product (GDP), also rates a 253 on the happiness index. Real poverty means real misery, a fate shared by billions.



**EDUCATION**  
 Residents of Australia can expect to spend more time in school—an average of almost 21 years—than citizens of any other country. But only a basic education is needed to see a significant jump in overall happiness. Around the world, hundreds of millions lack even that.

"It's time we admitted there's more to life than money."

—David Cameron, U.K. leader of the opposition, 2010



- RANKING THE WORLD'S HAPPIEST PLACES**  
 Northern Europe, North America, and several wealthy countries make the list, but so do many less prosperous island nations.
- 1 DENMARK, SWITZERLAND
  - 2 AUSTRIA, ICELAND
  - 3 BAHAMAS, FINLAND, SWEDEN
  - 4 BHUTAN, BRUNEI, CANADA, IRELAND, LUXEMBOURG
  - 5 COSTA RICA, MALTA, NETHERLANDS
  - 6 ANTIGUA AND BARBUDA, MALAYSIA, NEW ZEALAND, NORWAY, SEYCHELLES, ST. KITTS AND NEVIS, UNITED ARAB EMIRATES, UNITED STATES, VANUATU, VENEZUELA

**DEFINING WELL-BEING**  
 By comparing the happiness index to data from the UN, the CIA, and other sources, a U.K. psychologist determined that good health and health care, enough money for fundamental needs, and access to basic education are the most important factors for subjective well-being. European countries top all three measures.

# Science Maps for Science Policy Making

## Four Existing Maps VERSUS Six Science Maps



(5<sup>th</sup> Iteration of Places & Spaces Exhibit - 2009)

### LEGEND

- Physics
- Chemistry
- Biology
- Earth & Planetary Sciences
- Social Sciences
- Humanities

— Connection

## CLICKSTREAM MAP OF SCIENCE

This is the first map created from large-scale, world-wide, scholarly usage data. It visualizes the collective flow of scientists' movements from one journal to another in their online navigation behavior.

The MESUR project ([www.mesur.org](http://www.mesur.org)) collected a database of nearly 1 billion user requests recorded by the web portals of some of the world's most significant publishers, aggregators and large university consortia, among them Thomson Scientific (Wiley & Science, Elsevier (Scopus), JSTOR, Inspec, University of Texas (4 campuses), Health Institutions), and California State University (23 campuses). All usage logs captured by the MESUR project contain session identifiers that identify the individual clickstreams of individual scientists navigating from one article to the next.

Pairs of journals are connected when they have a high probability of being followed by each other in same clickstreams. The circles represent individual journals. A line between two circles indicates that they are strongly connected in either direction. The colors indicate the scientific domain a journal belongs to according to their Dewey (journal) and JCR classification codes that were mapped into the Getty Research Center's Arts and Architecture Thesaurus (AAT) to allow disambiguation at various levels of detail. The size of circles corresponds to the strength (degree centrality) of a journal's connections in the map. The map is structured by the Fruchterman-Reingold Algorithm that treats connections like springs; connected journals are drawn together, but they are not allowed to get too close.

The map is derived from usage data and therefore also reflects the actions of those who read the literature but never publish themselves, e.g. practitioners and laypersons. As a result, practitioner-driven domains such as nursing, social work and tourism studies are prominently featured. The natural sciences on the other extreme and humanities groups, as well distinct clusters that are connected via various specific interdisciplinary "spillover" nodes. Most domains are highly interdisciplinary, but this is more so the case for the social sciences and humanities. Geography, mathematics and computer science are not represented as one specific cluster, but spread-out through the map.

Like citation maps, this map is based upon a particular sample of the scientific community, albeit one that includes non-publishing scientists and practitioners and a much greater sample of publications. From MESUR's database of 1 billion user events, we created a matrix of 6 million connections between approximately 100,000 journals. From that matrix we selected only 50,000 connections with the highest number of observations, ranging from approximately 40,000 to 170 observations. This subset of connections pertained to the 2,307 most used journals. The procedure may produce specific biases which require investigation. This map should therefore not be considered as a final map of scientific activity, but as a hypothesis for the feasibility of tracking scientific activity from usage data. We hope this methodology will provide unique insights into the real-time structure of scientific activity as it can be derived from scholarly clickstream data.

When we cut the AAT taxonomy at the top level, only two distributions remain: natural sciences (blue nodes) vs. the social sciences and humanities (yellow nodes). Some journals along the spines of the wheel have classifications (colors) that do not correspond to their location in the map. The rationale either that journal in question is highly interdisciplinary, and/or has been assigned a classification that does not correspond to how scientists actually use the particular journal.

**DATA** 03/01/06 - 02/01/07

- 356,000,000 user requests
- 6,700,000 connections from raw data
- 97,532 records in raw data
- 50,000 top connections for map (> 170)
- 2,307 journals for map

**mesur**

More information on this map can be found in Bollen, J., Van de Sompel, H., Hagberg, A., Goltencourt, L., Chute, R., Rodriguez, M. and Balakireva, L. (2008) Clickstream Data Yields High-Resolution Maps of Science. PLoS ONE 4(3): e4003. doi:10.1371/journal.pone.0040033 (Peer-reviewed online)

Design layout by: Jeremy D. Chason

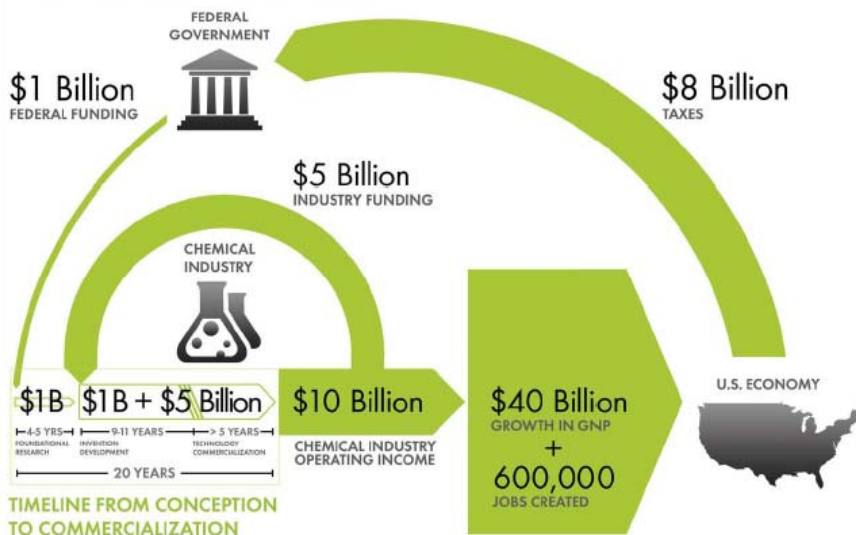
*A Clickstream Map of Science – Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriguez, Lyudmila Balakireva - 2008*



# Chemical Research & Development Powers the U.S. Innovation Engine

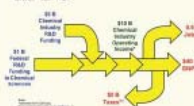
Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences

## INVESTMENT IN CHEMICAL SCIENCE R&D



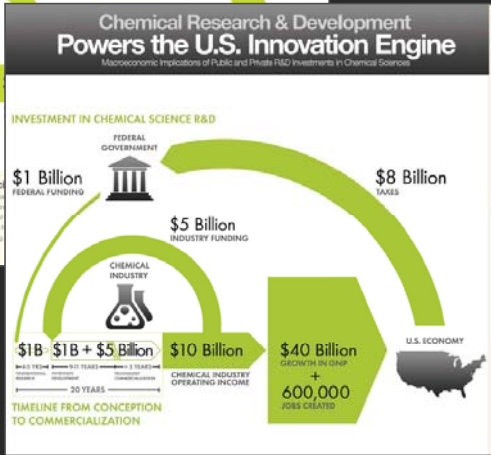
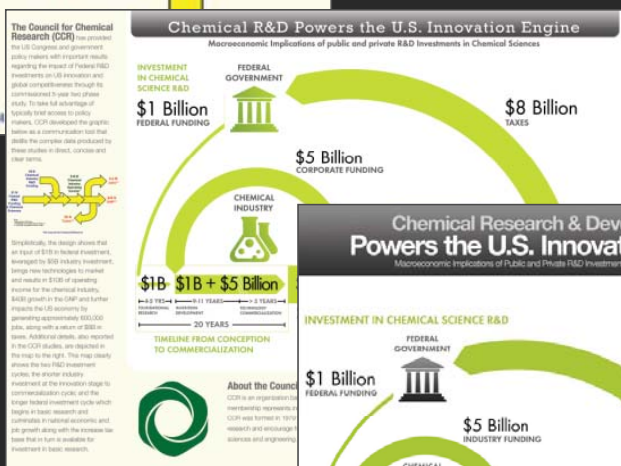
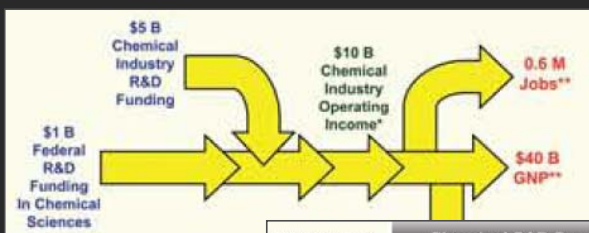
## The Council for Chemical Research (CCR)

has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise and clear terms.



The design shows that an input of \$1B in federal investment, leveraged by \$5B industry investment, brings new technologies to market and results in \$10B of operating income for the chemical industry, \$40B growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of \$8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles; the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase tax base that in turn is available for investment in basic research.

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



The Council for Chemical Research (CCR) has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise and clear terms.

The design shows that an input of \$1B in federal investment, leveraged by \$5B industry investment, brings new technologies to market and results in \$10B of operating income for the chemical industry, \$40B growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of \$8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles; the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase tax base that in turn is available for investment in basic research.

# Additional Elements of the Exhibit

## Illuminated Diagram Display

## Hands-on Science Maps for Kids

## Worldprocessor Globes

### Illuminated Diagram Display

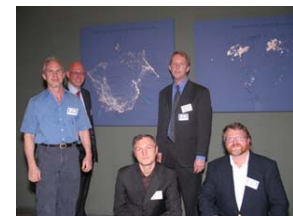
W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007)  
*Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007.*

#### Questions:

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

#### Contributions:

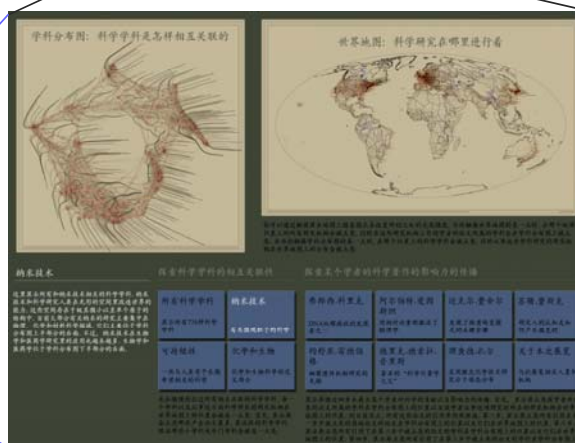
- Interactive, high resolution interface to access and make sense of data about scholarly activity.

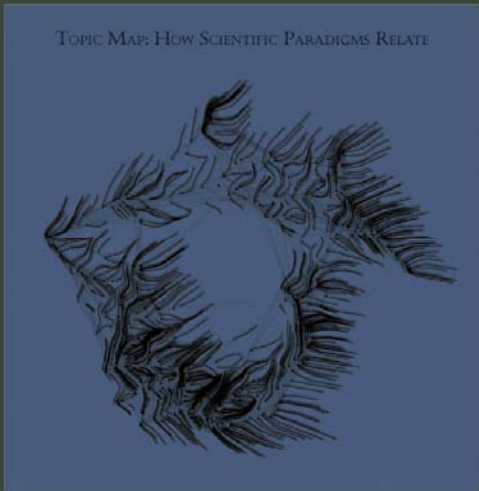


Large-scale, high resolution prints illuminated via projector or screen.



Interactive touch panel.





You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

### Nanotechnology

This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

#### All Topics

Sweep through all 376 scientific paradigms

#### Nanotechnology

Science on the tiny scale of molecules

#### Francis H. C. CRICK

Co-discovered DNA's double helix

#### Albert EINSTEIN

Revitalized physics with Relativity theories

#### Michael E. FISHER

Models critical phase transitions of matter

#### Susan T. FISKE

Connects perception and stereotypes

#### Sustainability

The science behind our long-term hopes

#### Biology & Chemistry

The interface between these two vital fields

#### Joshua LEDERBERG

Pioneer in bacterial genetic mechanisms

#### Derek J. de Solla PRICE

Known as the "Father of Scientometrics"

#### Richard N. ZARE

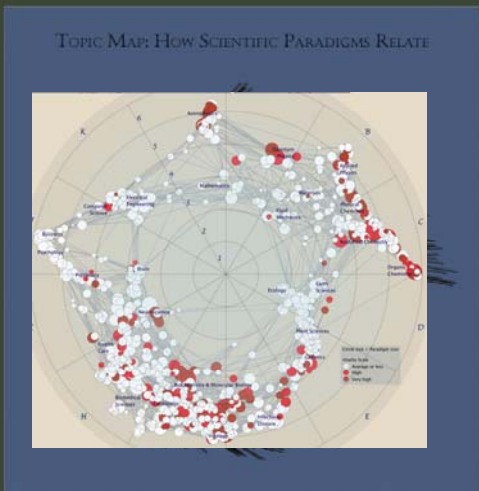
Uses laser chemistry in molecular dynamics

#### About this display

People & organizations that helped create it

We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to far more topics than did the original work. The third snapshot lights science that cites the second, and the fourth lights science that cites the third.



You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

### Nanotechnology

This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

#### All Topics

Sweep through all 376 scientific paradigms

#### Nanotechnology

Science on the tiny scale of molecules

#### Francis H. C. CRICK

Co-discovered DNA's double helix

#### Albert EINSTEIN

Revitalized physics with Relativity theories

#### Michael E. FISHER

Models critical phase transitions of matter

#### Susan T. FISKE

Connects perception and stereotypes

#### Sustainability

The science behind our long-term hopes

#### Biology & Chemistry

The interface between these two vital fields

#### Joshua LEDERBERG

Pioneer in bacterial genetic mechanisms

#### Derek J. de Solla PRICE

Known as the "Father of Scientometrics"

#### Richard N. ZARE

Uses laser chemistry in molecular dynamics

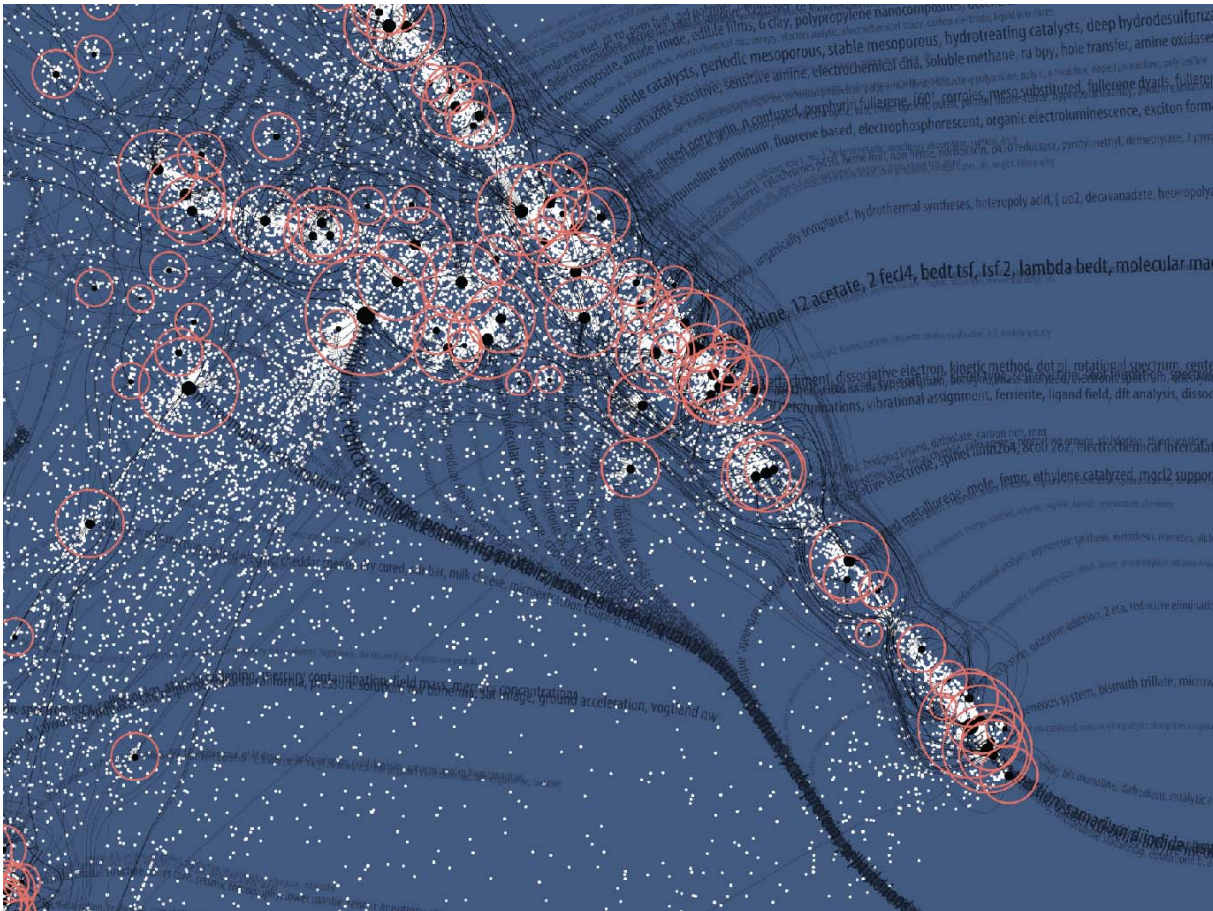
#### About this display

People & organizations that helped create it

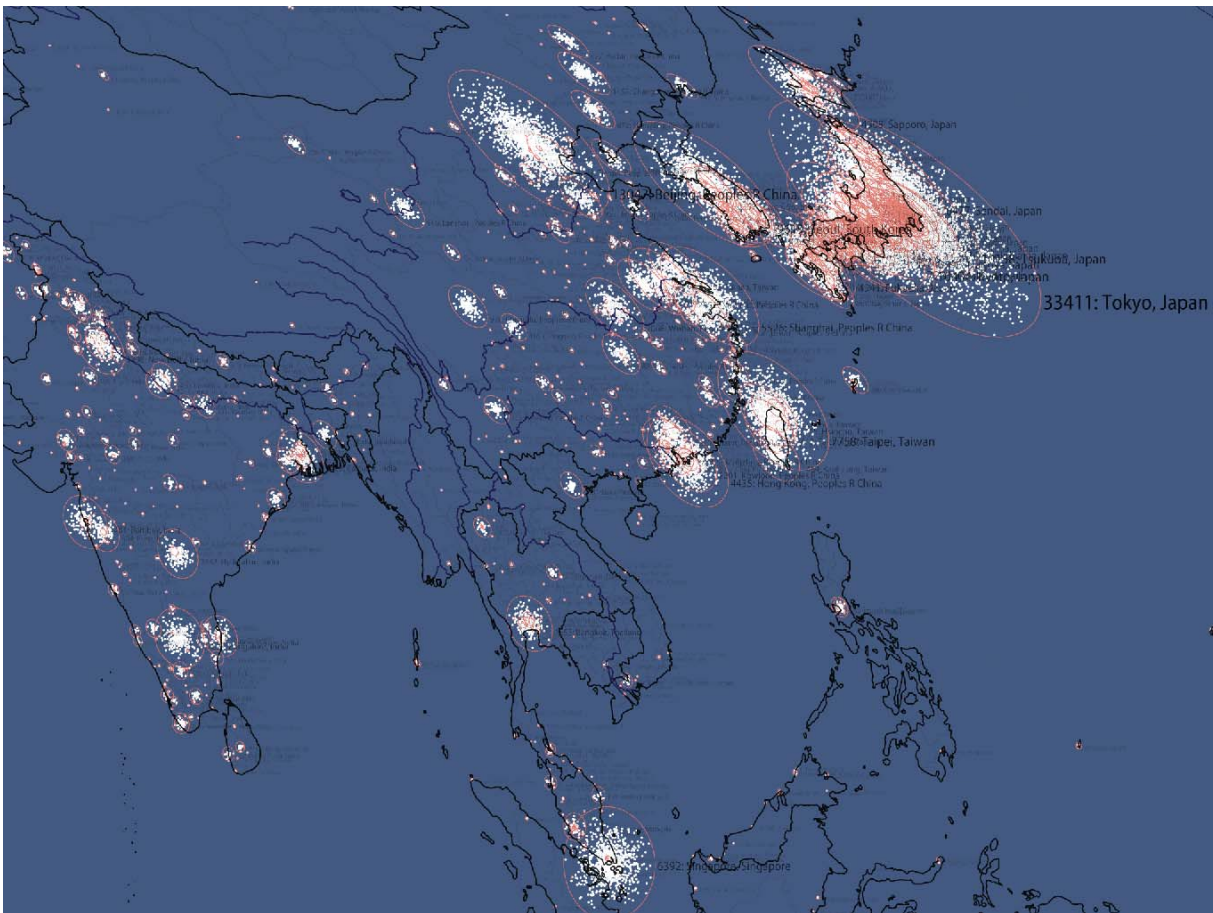
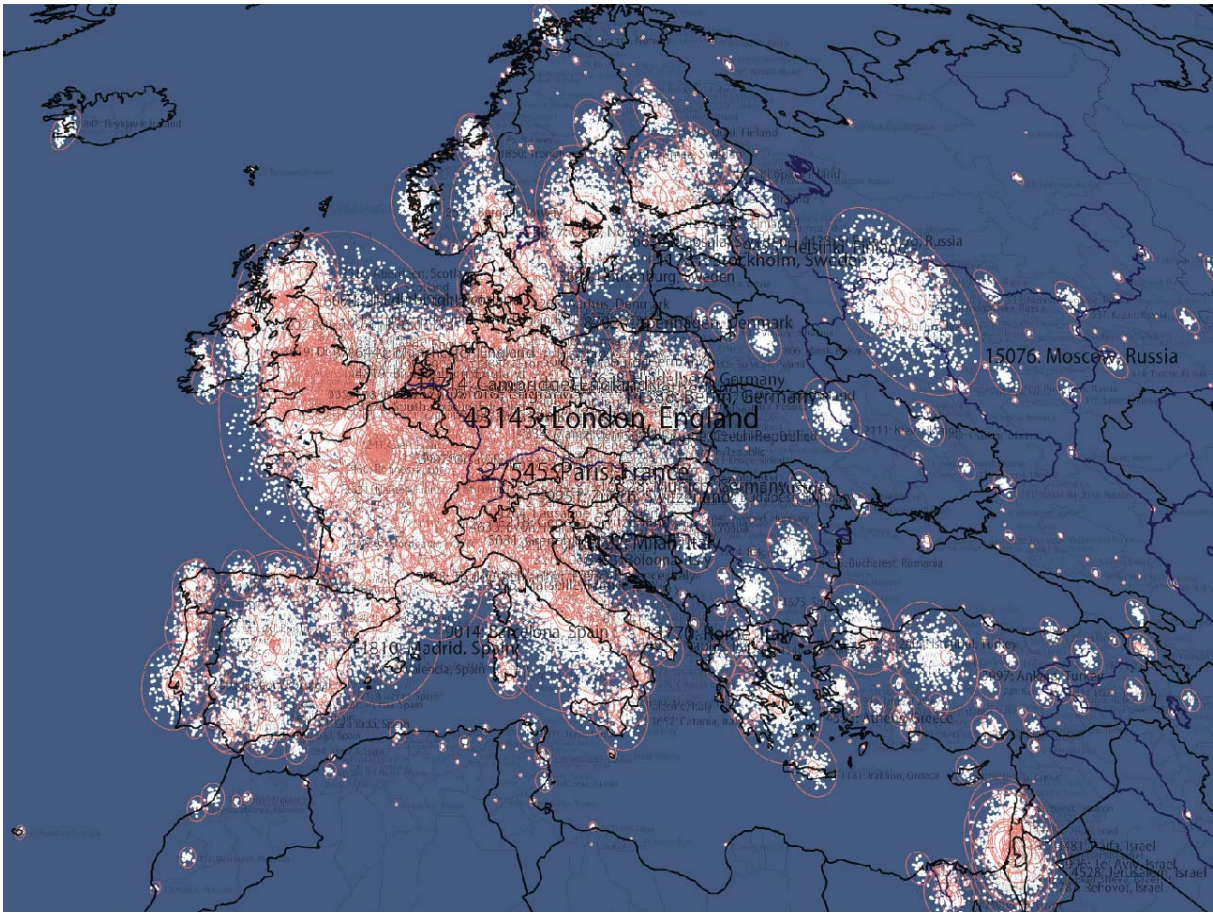
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to far more topics than did the original work. The third snapshot lights science that cites the second, and the fourth lights science that cites the third.

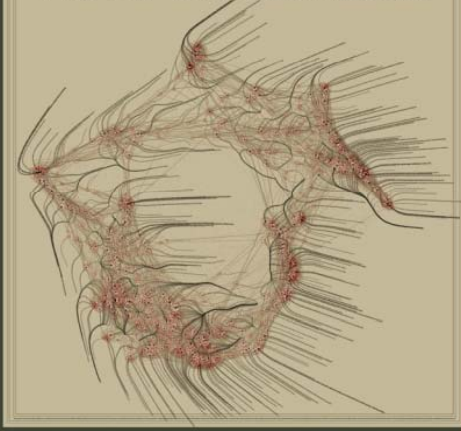
# TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE



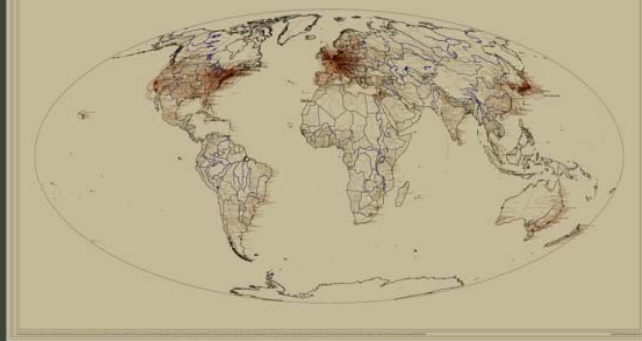




学科分布图：科学学科是怎样相互关联的



世界地图：科学研究在哪里进行着



你可以通过触摸屏在地图上随意指点来改变所到之处的光亮强度。当你触摸世界地图的某一点时，在那个地理位置上的所有研究机构会被点亮。同时在这些研究机构工作的学者的论文所属的学科会在学科分布图上被点亮。而当你触摸学科分布图的某一点时，在那个位置上的科学学科会被点亮，同时从事这些学科研究的研究机构在世界地图上的分布会被点亮。

纳米技术

这里显示所有和纳米技术相关的科学学科。纳米技术和科学研究人在无形的空间里改造世界的的能力。这些空间存在于极其微小以至单个原子的结构中。目前大部分有关纳米的研究主要集中在物理、化学和材料科学领域。它们主要位于学科分布图上半部分的右面。不过，纳米技术在生物学和医药学研究里的应用也越来越多。生物学和医药学位于学科分布图下半部分的右面。



探索科学学科的相互关联性

所有科学学科 显示所有776种科学学科	纳米技术 有关微观粒子的科学
可持续性 一些与人类寄予长期希望相关的科学	化学和生物 化学和生物科学的交叉部分

光标缓慢的扫过所有相互关联的科学学科，每一个学科以及从事这方面科学研究的研究机构在世界地图上的位置会被逐一点亮。首先，显示屏会点亮那些产出论文最多、最活跃的科学学科，然后那些小学科或冷门学科会被逐一点亮。

探索某个学者的科学著作的影响力的传播

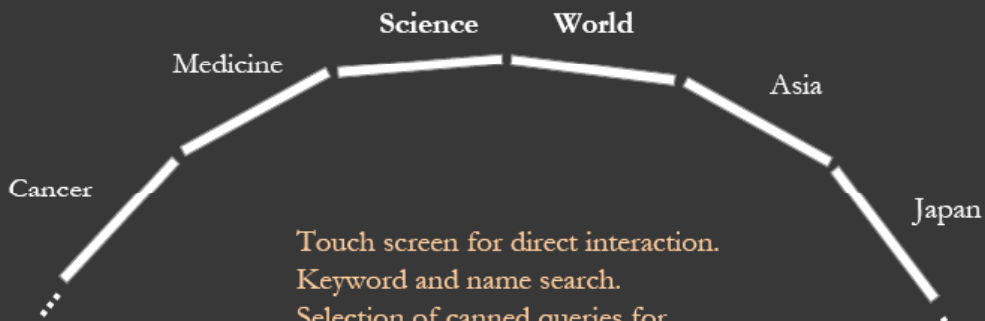
弗郎西·科里克 DNA双螺旋结构的发现者之一	阿尔伯特·爱因斯坦 用相对论重新激活了物理学	迈克尔·费舍尔 发现了物质转变模式的关键步骤	苏珊·费斯克 研究人的认知是如何产生偏见的
约舒亚·雷德伯格 细菌遗传机制研究先驱	德里克·德索拉·普里斯 著名的“科学计量学之父”	理查德·扎尔 采用激光化学技术研究分子动态分布	关于本次展览 与此展览相关人员和机构

显示屏通过四步来展示某个学者对科学的贡献以及影响力的传播。首先，显示屏点亮该学者所发表的论文所属的学科在学科分布图上的位置以及该学者从事这项研究时的研究机构在世界地图上的位置。到目前为止，所有这些论文的引用率仍然很高。第二步，显示屏点亮所有引用在第一步中被点亮的原始论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第三步，显示屏点亮所有引用了在第二步中被点亮的论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第四步，显示屏点亮所有引用了在第三步中被点亮的论文的论文在学科分布图上的位置以及它们在世界地图上的位置。

Re-implementation of Illuminated Diagram Software

by Advanced Visualization Lab, Indiana University

Drives unlimited number of ID screens.



- Touch screen for direct interaction.  
 Keyword and name search.  
 Selection of canned queries for
- interdisciplinary research areas
  - famous people
  - activity patterns, e.g., bursts, trends, etc.





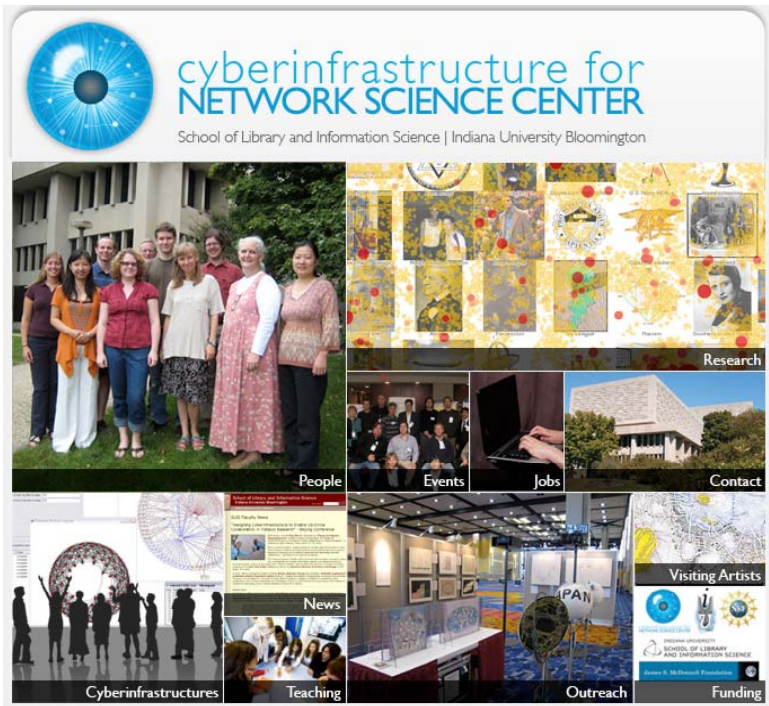
Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months, 12 coaches, 300 m long. Opening was on April 23<sup>rd</sup>, 2009 by German Chancellor Merkel, <http://www.expedition-zukunft.de>

This is the only mockup in this slide show.

Everything else is available today.







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