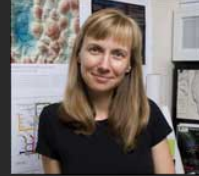


The Scholarly Database and Its Utility for Scientometrics Research



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

Co-Authors are
Gavin LaRowe, Sumeet Ambre, John Burgoon & Weimao Ke
Cyberinfrastructure for Network Science Center

*ISI 2007, Madrid, Spain
June 26, 2007*



This Talk has Three Parts:

1. Motivation
2. Scholarly Database
3. Discussion and Outlook

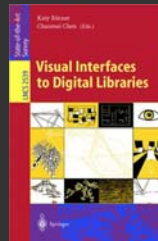


This Talk has Three Parts:

1. Motivation
2. Scholarly Database
3. Discussion and Outlook



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.
- 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).
- Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (in press). **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.
- **Places & Spaces: Mapping Science** exhibit, see also <http://scimaps.org>.



places & spaces &

Places & Spaces: Mapping Science
An exhibition created to demonstrate the power of maps to understand, navigate, and manage not only physical places, but also abstract information spaces.

Home Browse Maps Compare & Contrast Maps Schedule Connect

Home

Exhibit Purpose and Goals

The Places & Spaces: Mapping Science exhibit has been created to demonstrate the power of maps. An initial theme of this exhibit is to compare and contrast first maps of our entire planet with the first maps of all of science as we know it.

Check out the [schedule of physical showings](#) and come see with your own eyes the extent to which maps can be employed to help make sense of the flood of information we are confronted with and how domain maps can be used to locate complex and beautiful information.

"Places & Spaces: Mapping Science" on display at the New York Hall of Science, Dec. 9, 2006 - Feb. 25, 2007.

Places & Spaces at the [NYPL Science, Industry, and Business Library](#) (Madison/34th), New York, April 3rd - August 31st, 2006.

ORDER MAPS

ORDER DVD HERE!

Places & Spaces: Mapping Science
a science exhibit that introduces people to maps of sciences, their makers and users.
<http://scimaps.org>

Exhibit Curators:
Dr. Katy Börner & Julie Smith, Indiana University

Places & Spaces: Mapping Science exhibit, see also <http://scimaps.org>.

Chart toppers

An exhibition explores the diverse ways of putting data on the map.

From the simple pie charts and bar graphs to the complex network diagrams and interactive maps, the exhibition features a wide range of data visualization techniques. The charts are arranged in a way that allows visitors to explore the history and evolution of data visualization, from the early days of hand-drawn maps to the modern era of digital data analysis.

2006 GALLERY

BRILLIANT DISPLAY

From a jewel-like bird, rather than any diamond, to the delicately poetic swirls generated inside aircraft engines, the pursuit of knowledge turns up its fair share of beauty. This issue, *Nature* wraps up the year with an arresting series of images from 2006. We've divided them into the art of the natural world, planet-scapes both domestic and extraterrestrial, and the splendour of modern technology. Just because something enhances our knowledge doesn't mean it can't also be bewitching.

Research and art by **Evans Munk**.

AMAP OF SCIENCE
A map of science that maps relationships among scientific paradigms, as published in the journal *Nature*. This map was constructed by sorting roughly 6,000 published papers into 776 different scientific paradigms (shown as pale circular nodes) based on how often the papers were cited together by authors of their papers. Information, *Ethicalics*, an association, founded by *man*, *scientists*.

SEED
SCIENTIFIC METHOD: F
AMONG SCIENTIFIC PA

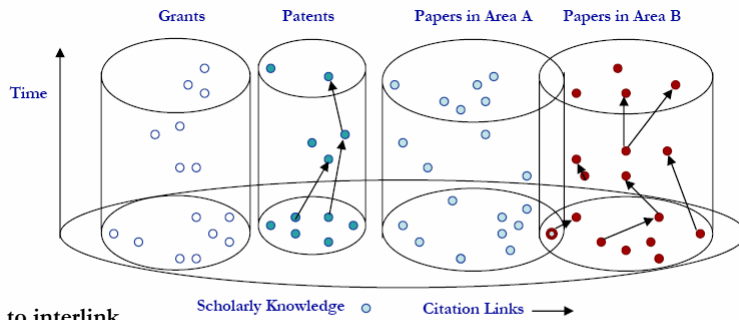
Mapping Science
[Map, science, network]

<http://scimaps.org>

This Talk has Three Parts:

1. Motivation
2. Scholarly Database
3. Discussion and Outlook

Challenges - Interlink Input & Publication/Patent Citation Output

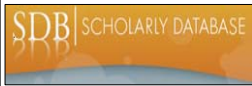


Need to interlink

- Grants and papers/patents.
- Grants/papers/patents and their PIs/authors/inventors, etc.

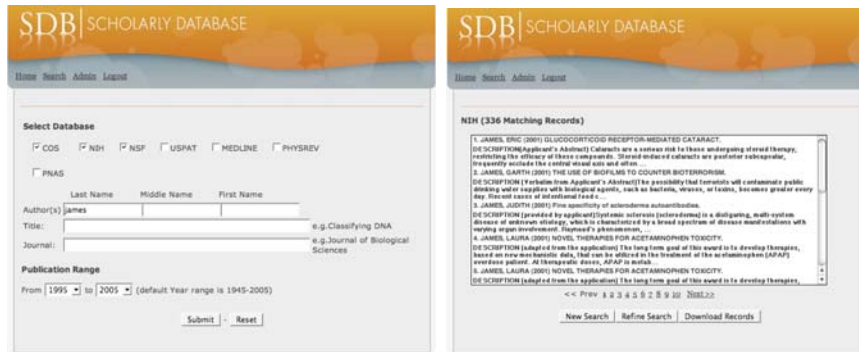
Use resulting networks to

- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.

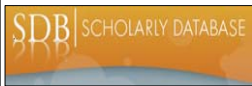


Scholarly Database: Web Interface

Search across publications, patents, grants.
Download records and/or (evolving) co-author, paper-citation networks.



Register for free access at <https://sdb.slis.indiana.edu>.

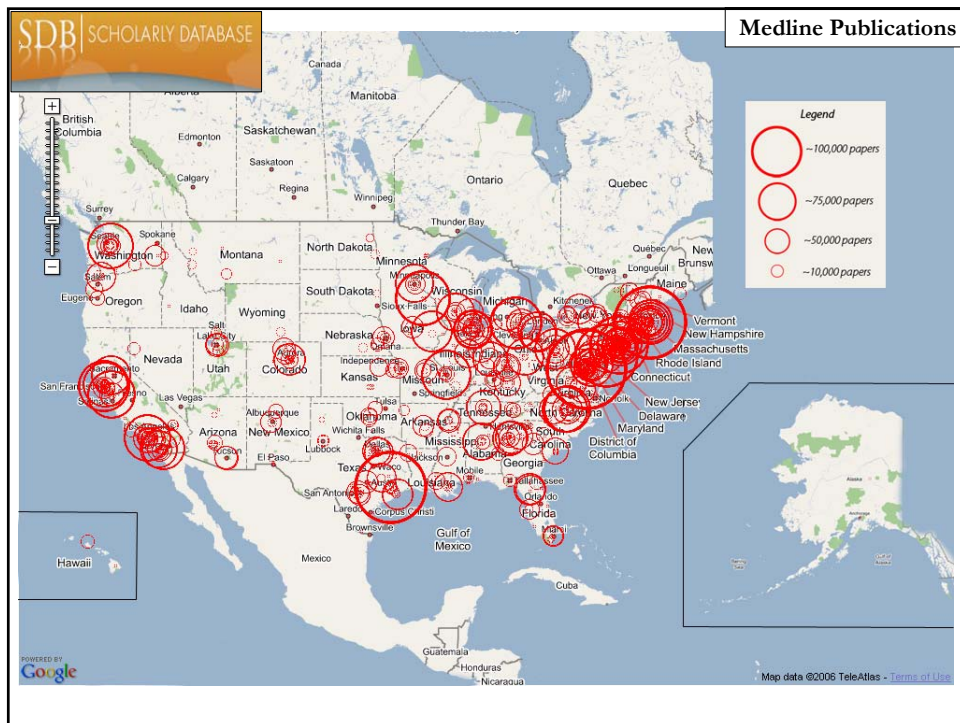
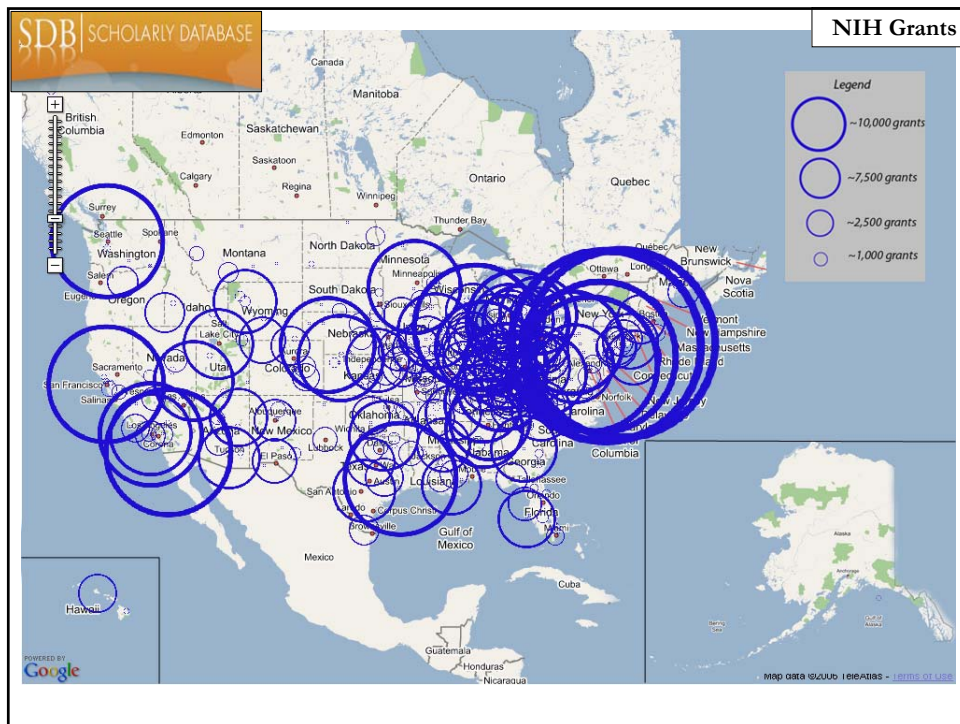


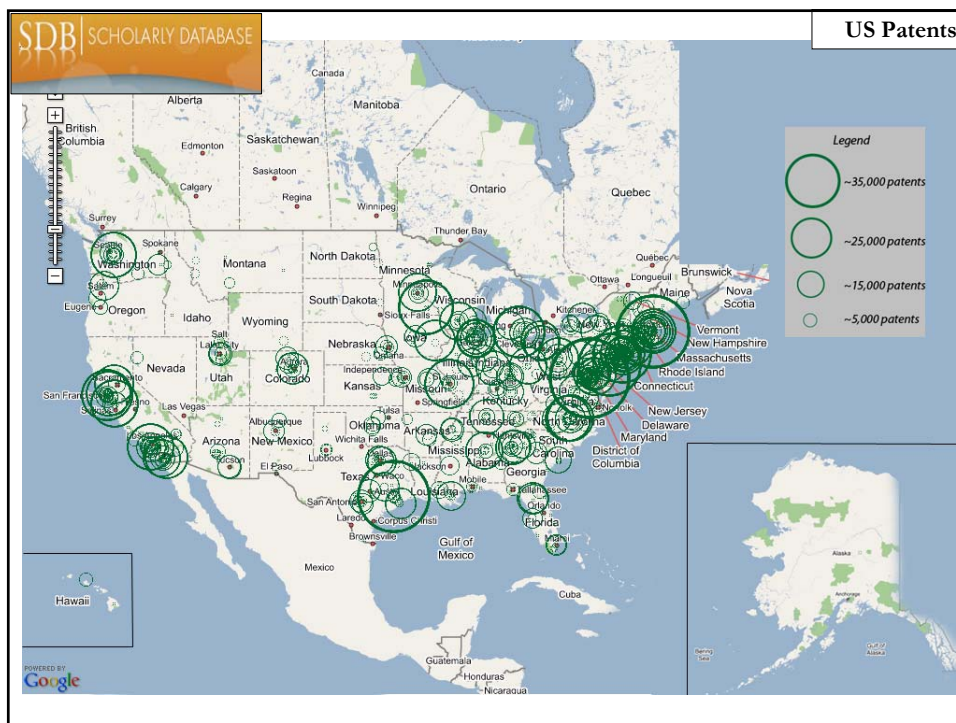
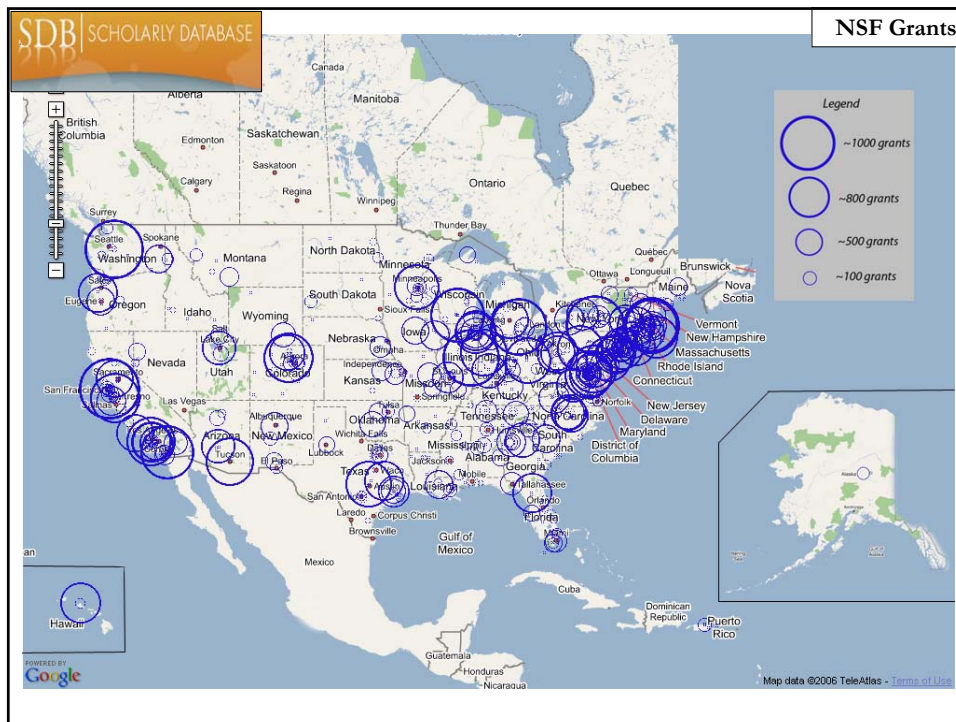
Scholarly Database: # Records & Years Covered

Datasets available via the Scholarly Database (* future feature)

Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	13,149,741	1965-2005	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,179,930	1976-2004	Yes*	
NSF	174,835	1985-2003	Yes*	
NIH	1,043,804	1972-2002	Yes*	
Total	18,021,560	1893-2006	4	3

Aim for comprehensive time, geospatial, and topic coverage.





This Talk has Three Parts:

1. Motivation
2. Scholarly Database
3. Discussion and Outlook

Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES
An Oracle database provides access to publications, papers, grants and grant opportunities. The database is continuously and automatically updated from the InfoVis website.

COMPUTING RESOURCES
The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Center, consisting of three Sun X86 servers with 12 GB RAM, 100 GB hard disk and 100 MB network card. A Sun 1000 server acts as a proxy and file server between the web browser and the database server.

SOFTWARE
An open source R2 framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. The algorithms, data analysis methods, flow control and the visualization software tools can be easily "plugged in" to the framework.

LEARNING MODULES
A set of associated learning modules aims to equip learners with a practical skill set in providing code, algorithms and diverse research techniques and methods. These modules are the product of an ongoing research project.

CAREER: *Visualizing Knowledge Domains*. NSF IIS-0238261 award (Katy Börner, \$440,000) Sept. 03-Aug. 08.
<http://in.slis.indiana.edu/>

NetworkWorkbench
A Workbench for Network Scientists



SEI: *Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research*. NSF IIS-0513650 award (Katy Börner, Albert-László Barabási, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Craig Stewart (Senior Personnel), \$1,120,926) Sept. 05 - Aug. 08.
<http://nwb.slis.indiana.edu>

The Future: Science (Forecast) Maps Online

There is a need to

- Understand the inner workings of science in order to support scientific progress and innovation.
- To monitor and support science as a dynamic system that changes quantitatively and qualitatively over time.
- To bridge between industry and academia to support the development of ideas into products.
- To communicate the results of science and technology to anybody for educational purposes but also to justify expenditures.

Only high quality data will lead to high quality maps and insights.

Hence, there is a need to harvest higher coverage (international and interdisciplinary) data, to clean data, and to interlink data types.

This can only be achieved in a collective effort similar to Wikipedia.

The Future: Science (Forecast) Maps Online

- Needed is a 'marketplace' of scholarly data, algorithms, services, expertise, publications.
- Maps of science can serve as a visual interface to these scholarly data types.
- Maps can be served in a Google Maps like fashion with wikipedia functionality, i.e., anybody can add data, comment on data, rate resources similar to Flickr or YouTube.

Incentive Design

- Many people will like to see themselves on the map providing incentives for them to help add and clean data.
- If funding agencies use these maps to see the impact of scholars as well as their work context then another major incentive is created.
- If industry uses these maps to enlist needed technologies and available funding opportunities then a 3rd incentive for being on the map is created.

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



(2nd Iteration of Places & Spaces Exhibit - 2006)

The End.