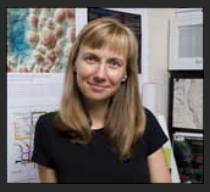
The Scholarly Database and Its Utility for Scientometrics Research



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

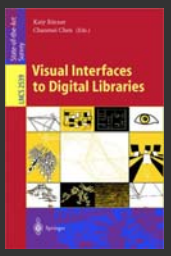
This Talk has Three Parts:

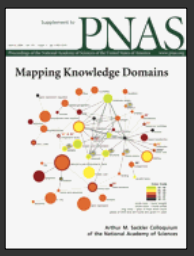
1. Motivation
2. Scholarly Database
3. Discussion and Outlook

This Talk has Three Parts:

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

Computational Scientometrics:
Studying Science by Scientific Means



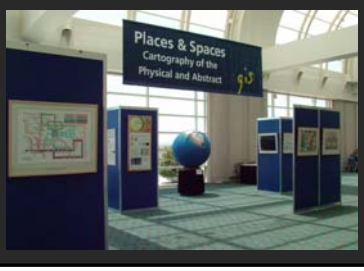


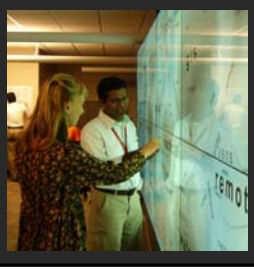
- *Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003).* ***Visualizing Knowledge Domains.*** *In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255.*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

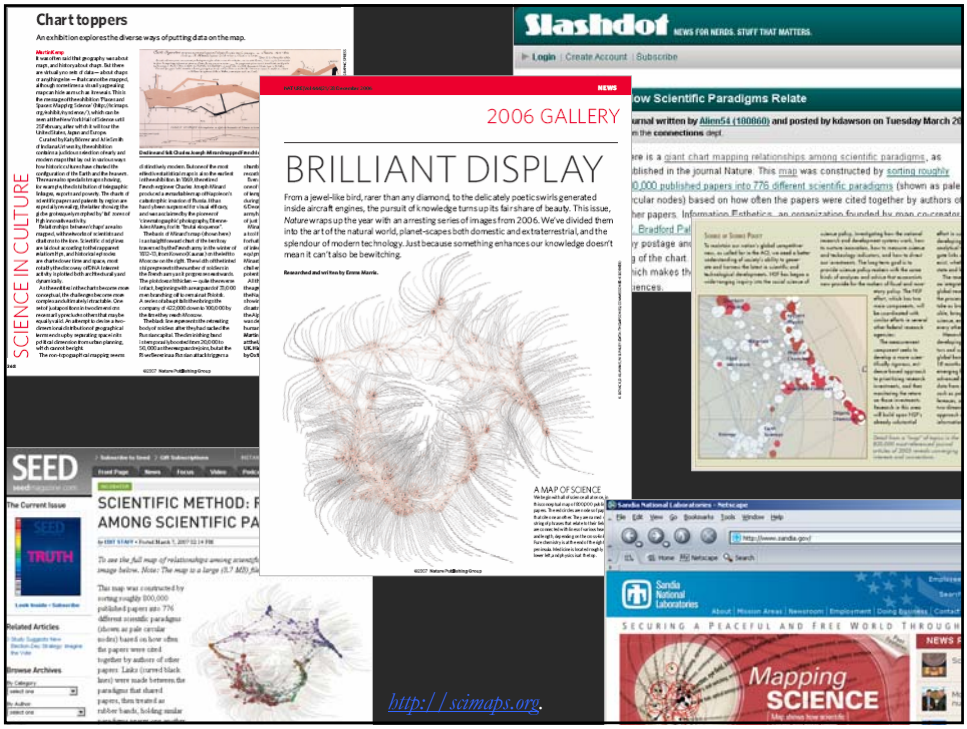
Places & Spaces: Mapping Science
a science exhibit that introduces people to maps of sciences, their makers and users.
http://scimaps.org

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





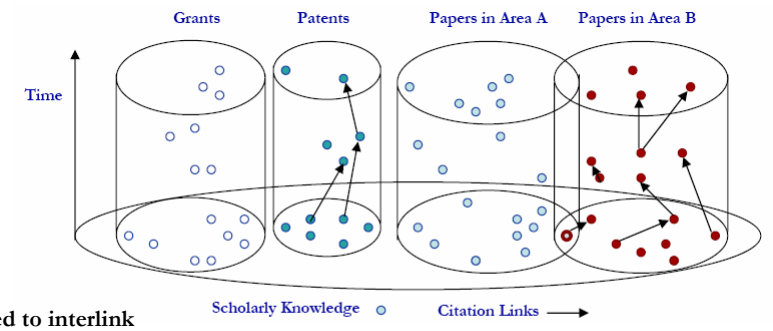
Places & Spaces: Mapping Science *exhibit, see also http://scimaps.org.*



This Talk has Three Parts:

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

Challenges - Interlink $ Input & Publication/Patent Citation Output

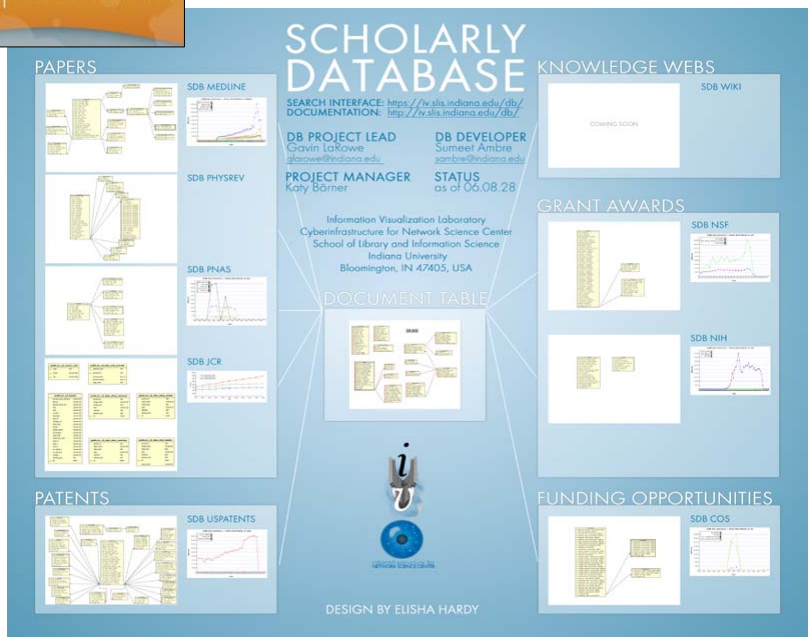


Need to interlink

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

Use resulting networks to

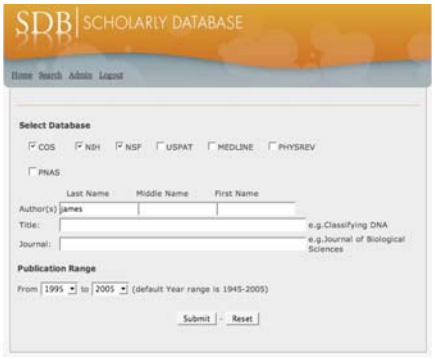
- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.



Scholarly Database: Web Interface

Search across publications, patents, grants.

Download records and/or (evolving) co-author, paper-citation networks.



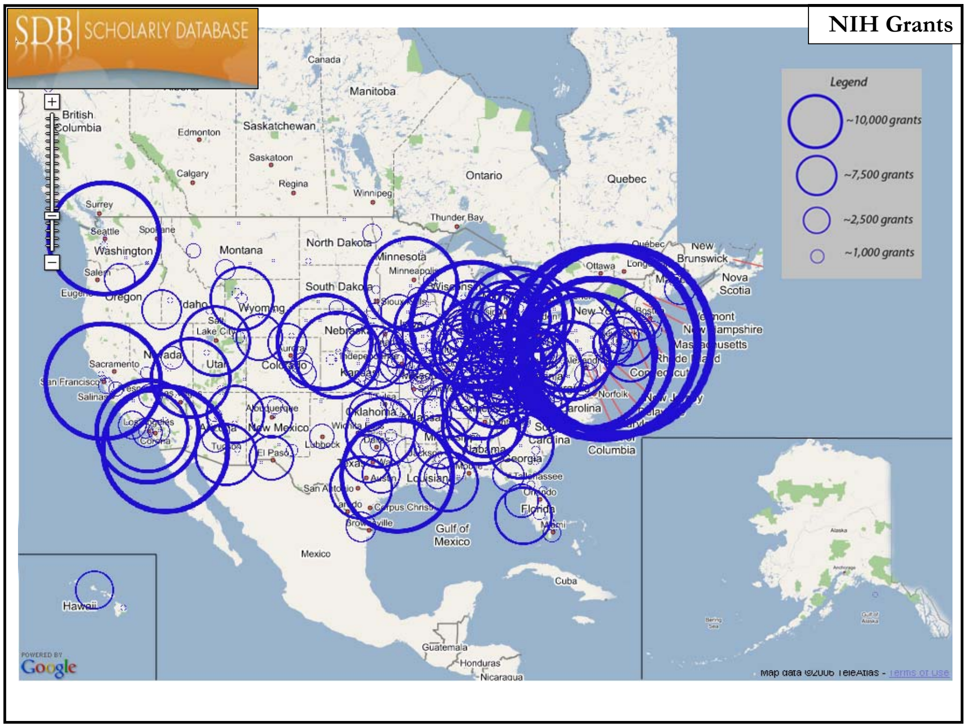
Register for free access at https://sdb.slis.indiana.edu.

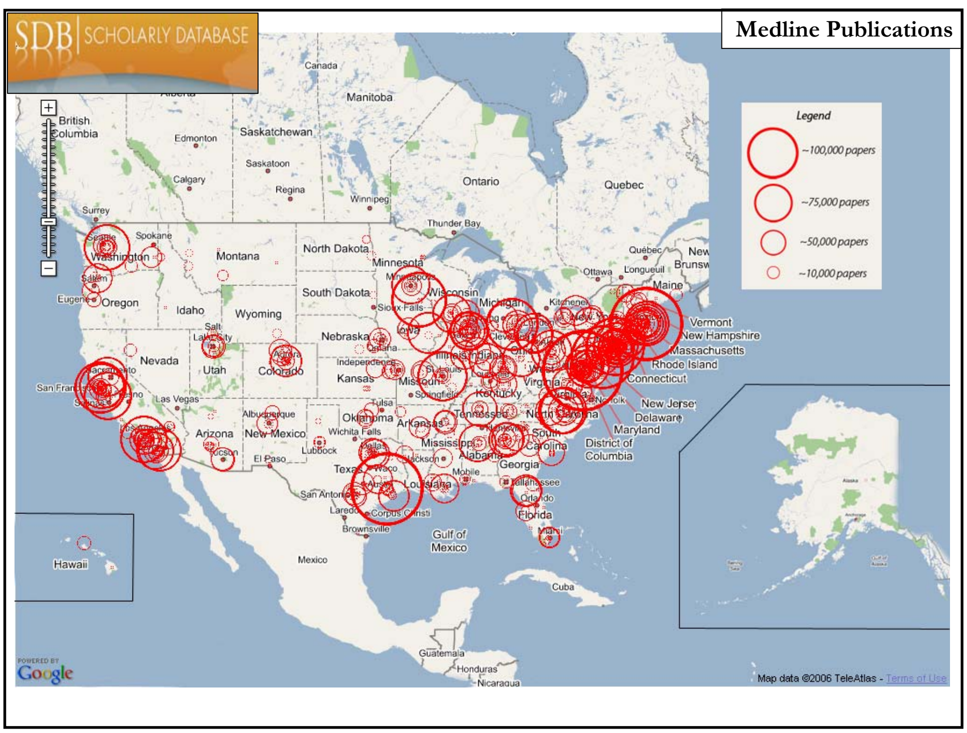
Scholarly Database: # Records & Years Covered

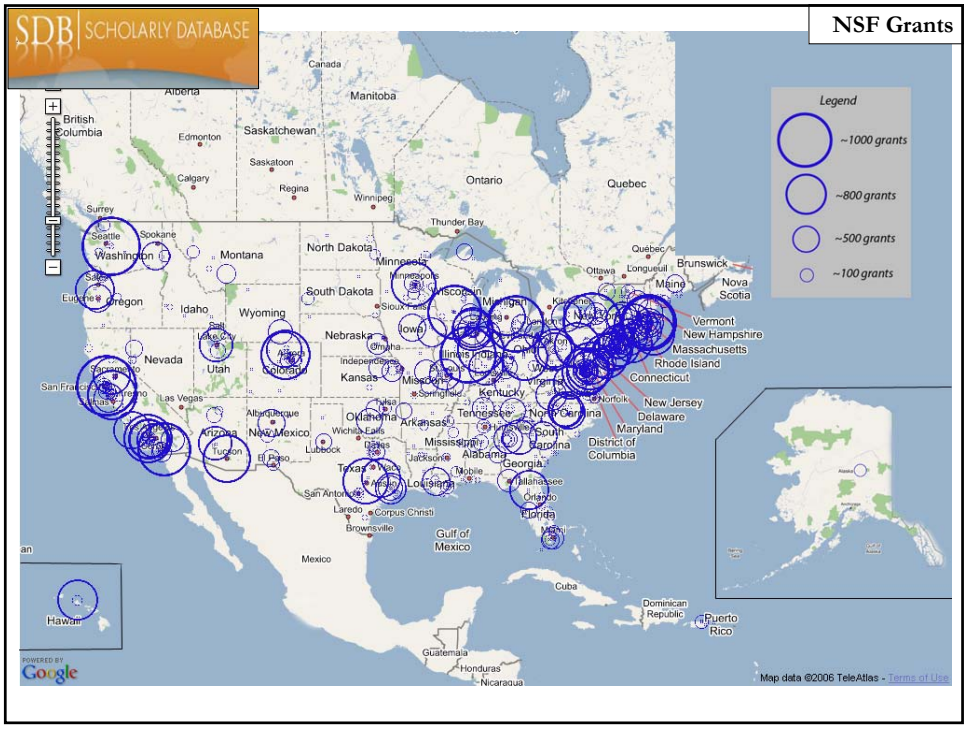
Datasets available via the Scholarly Database (* future feature)

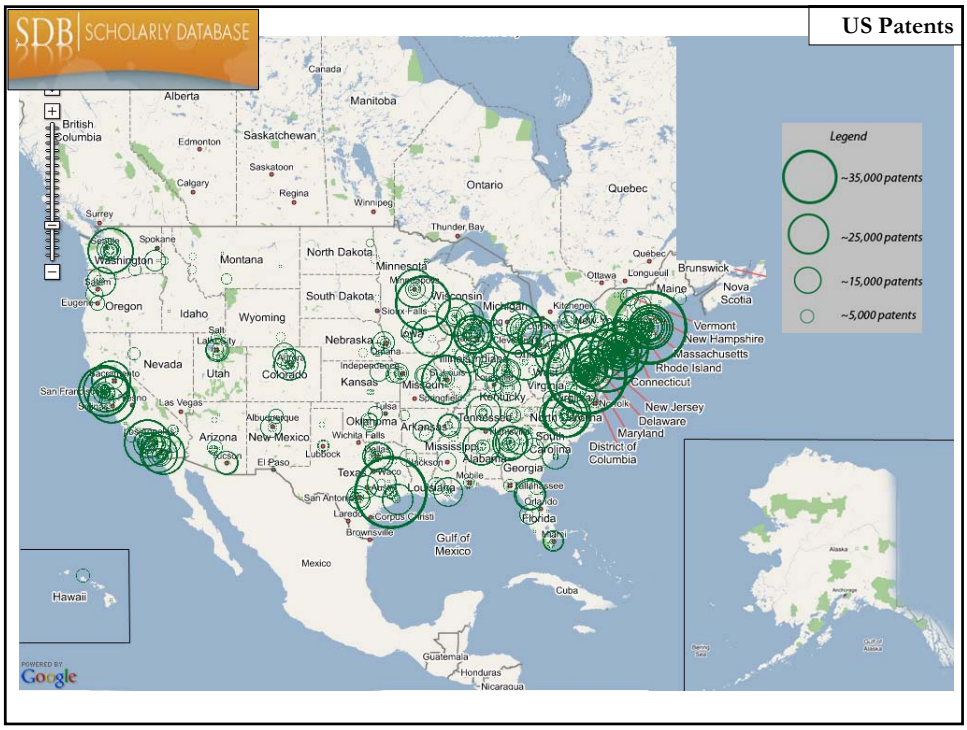
Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	13,149,741	1965-2005	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,179,930	1976-2004	Yes*	
NSF	174,835	1985-2003	Yes*	
NIH	1,043,804	1972-2002	Yes*	
Total	**18,021,560**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

Aim for comprehensive time, geospatial, and topic coverage.


SDB | SCHOLARLY DATABASE
NIH Grants
Legend
~10,000 grants
~7,500 grants
~2,500 grants
~1,000 grants
Map data ©2006 TeleAtlas

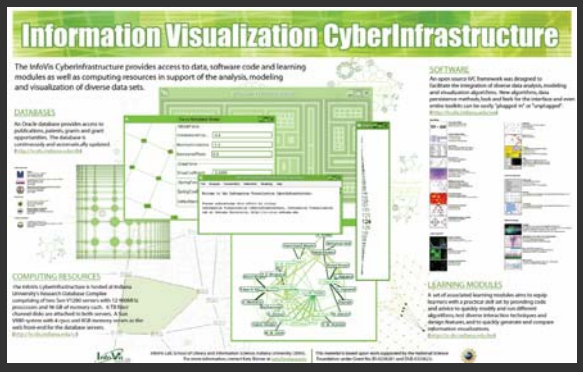

SDB | SCHOLARLY DATABASE
Medline Publications
Legend
~100,000 papers
~75,000 papers
~50,000 papers
~10,000 papers
Map data ©2006 TeleAtlas - Terms of Use


NSF Grants
SDB | SCHOLARLY DATABASE
Legend
~1000 grants
~800 grants
~500 grants
~100 grants
Canada
Alberta
Manitoba
British Columbia
Edmonton
Saskatchewan
Saskatoon
Calgary
Regina
Ontario
Quebec
Winnipeg
Thunder Bay
Surrey
Seattle
Spokane
Washington
Montana
North Dakota
Minnesota
Ottawa
Longueuil
Brunswick
Maine
Nova Scotia
Salem
Eugene
Oregon
Idaho
South Dakota
Wisconsin
Michigan
Wyoming
Salt Lake City
Nebraska
Iowa
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
Nevada
Utah
Colorado
Omaha
Kansas
Missouri
Kentucky
San Francisco
Las Vegas
Albuquerque
Oklahoma
Arkansas
Tennessee
New Jersey
Delaware
Maryland
Arizona
New Mexico
Wichita Falls
Lubbock
Mississippi
Alabama
South Carolina
District of Columbia
Tucson
El Paso
Texas
Georgia
Mobile
Tallahassee
Louisiana
San Antonio
Orlando
Florida
Laredo
Corpus Christi
Brownsville
Gulf of Mexico
Mexico
Cuba
Dominican Republic
Puerto Rico
Hawaii
Guatemala
Honduras
Nicaragua
Google
Map data ©2006 TeleAtlas - Terms of Use


US Patents
SDB | SCHOLARLY DATABASE
Legend
~35,000 patents
~25,000 patents
~15,000 patents
~5,000 patents
Canada
Alberta
Manitoba
British Columbia
Edmonton
Saskatchewan
Saskatoon
Calgary
Regina
Ontario
Quebec
Winnipeg
Thunder Bay
Surrey
Seattle
Spokane
Washington
Montana
North Dakota
Minnesota
Ottawa
Longueuil
Brunswick
Maine
Nova Scotia
Salem
Eugene
Oregon
Idaho
South Dakota
Wisconsin
Michigan
Wyoming
Salt Lake City
Nebraska
Iowa
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
Nevada
Utah
Colorado
Omaha
Kansas
Missouri
Kentucky
San Francisco
Las Vegas
Albuquerque
Oklahoma
Arkansas
Tennessee
New Jersey
Delaware
Maryland
Arizona
New Mexico
Wichita Falls
Lubbock
Mississippi
Alabama
South Carolina
District of Columbia
El Paso
Texas
Georgia
Mobile
Tallahassee
Louisiana
San Antonio
Orlando
Florida
Laredo
Corpus Christi
Brownsville
Gulf of Mexico
Mexico
Cuba
Hawaii
Guatemala
Honduras
Nicaragua
Google

This Talk has Three Parts:

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



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

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-Laszlo Barabasi, 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 an 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

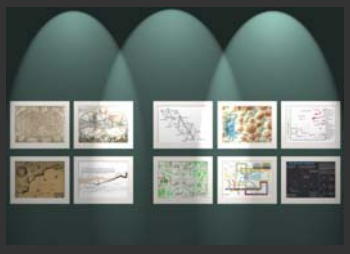
- Needed is a 'marketplace' of scholarly data, algorithms, services, expertise, publications.
- Maps of science can serve as a visual interface to these scholarly data types.
- Maps can be served in a Google Maps like fashion with wikipedia functionality, i.e., anybody can add data, comment on data,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

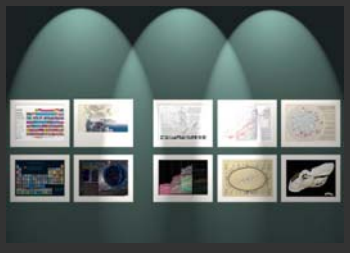
Four Early Maps of Our World
VERSUS
Six Early Maps of Science



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

The Power of Reference Systems

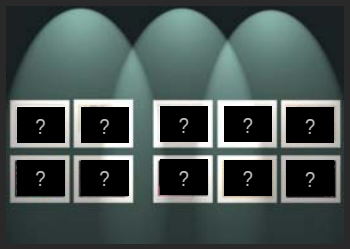
Four Existing Reference Systems
VERSUS
Six Potential Reference Systems of Science



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

The Power of Forecasts

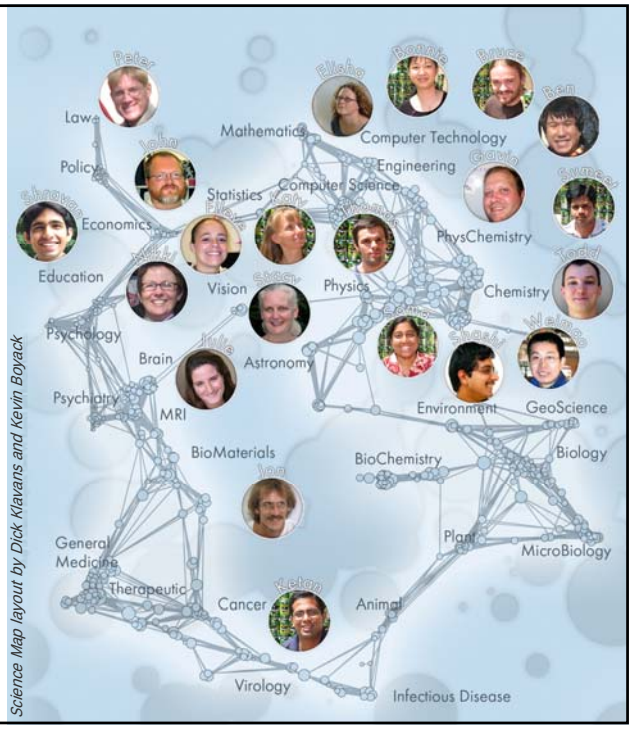
Four Existing Forecasts
VERSUS
Six Potential Science 'Weather' Forecasts



(3rd Iteration of Places & Spaces Exhibit - 2007)

I would like to thank all my colleagues and collaborators.

If not otherwise indicated, this work was conducted at the Information Visualization Laboratory and the Cyberinfrastructure for Network Science Center at Indiana University.



The End.