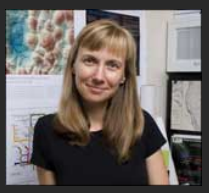
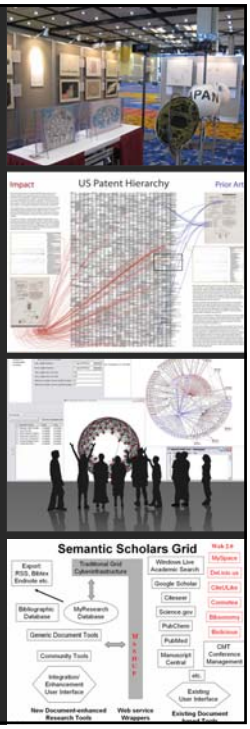
The Science of Science

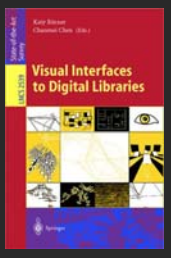


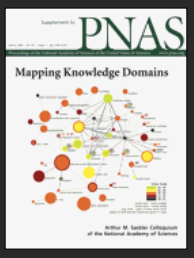
Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

Practical Semantic Astronomy Workshop at Caltech, Pasadena, California
February 20th, 2008



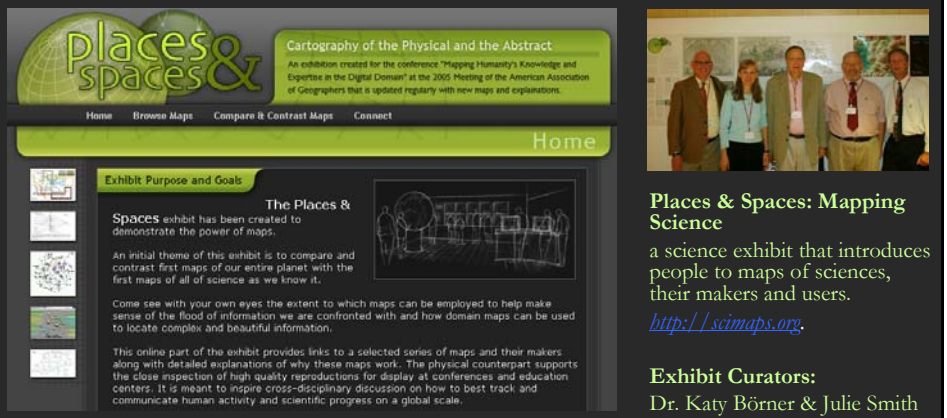
Computational Scientometrics: Studying Science by Scientific Means





- *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/
- *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

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



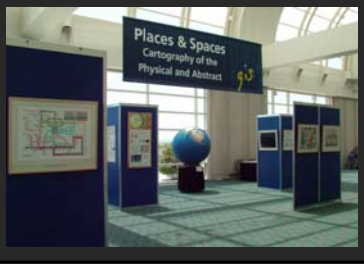
Places & Spaces: Mapping Science

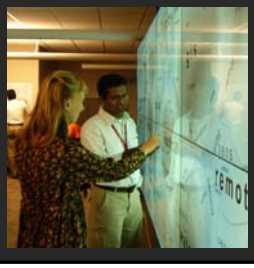
a science exhibit that introduces people to maps of sciences, their makers and users.

http://scimaps.org.

Exhibit Curators:

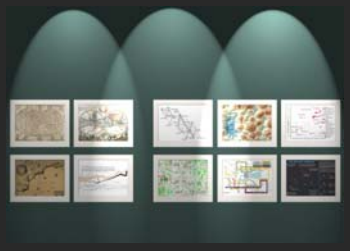
Dr. Katy Börner & Julie Smith



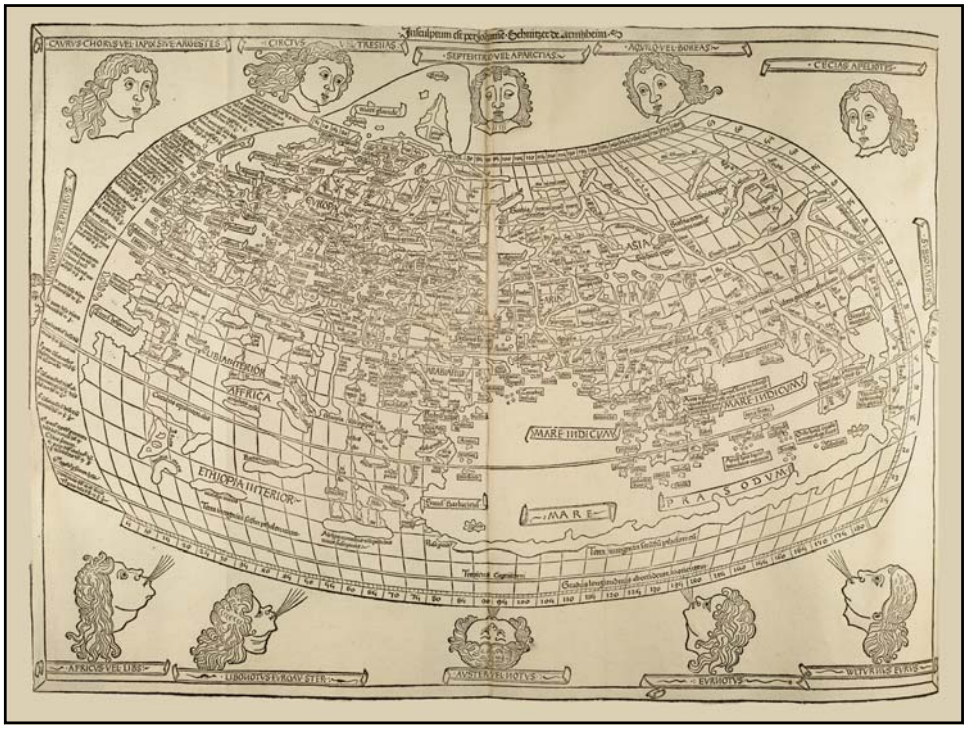


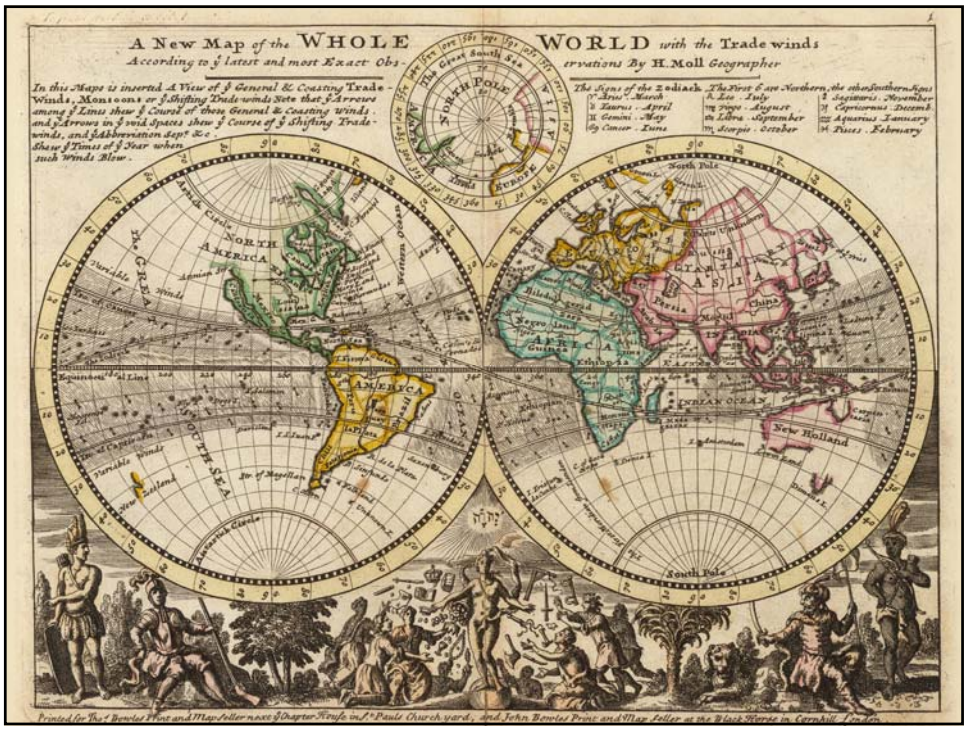
The Power of Maps

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



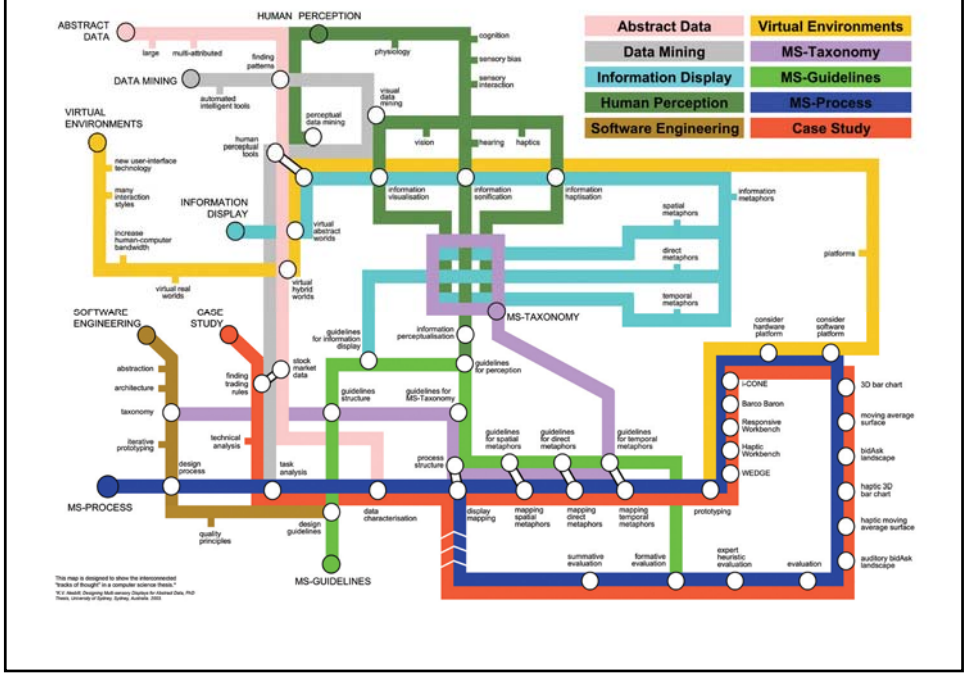
(1st Iteration of Places & Spaces Exhibit - 2005)

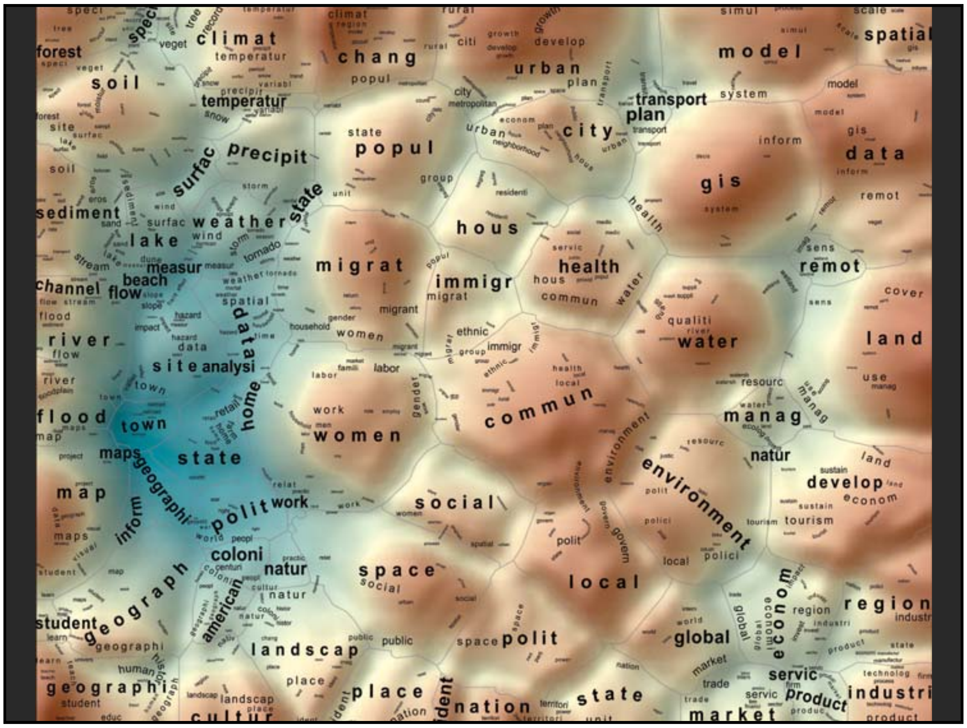

CAVRVS CHORVS VEL IAPIX SIVE ARGESTES
CIRCIVS VEL THRESIAS
SEPTENTRIO VEL APARCTIAS
AQVILO VEL BOREAS
CAECIAS APELIOTES
AFRICA
ETHIOPIA INTERIOR
ASIA
MARE INDICVM
MARE
SCYTHIA
AFRICVS VEL LIBS
LIBONOTVS EVRONOTVS
AVSTER VEL NOTVS
EVRONOTVS
ZEPHYRVS

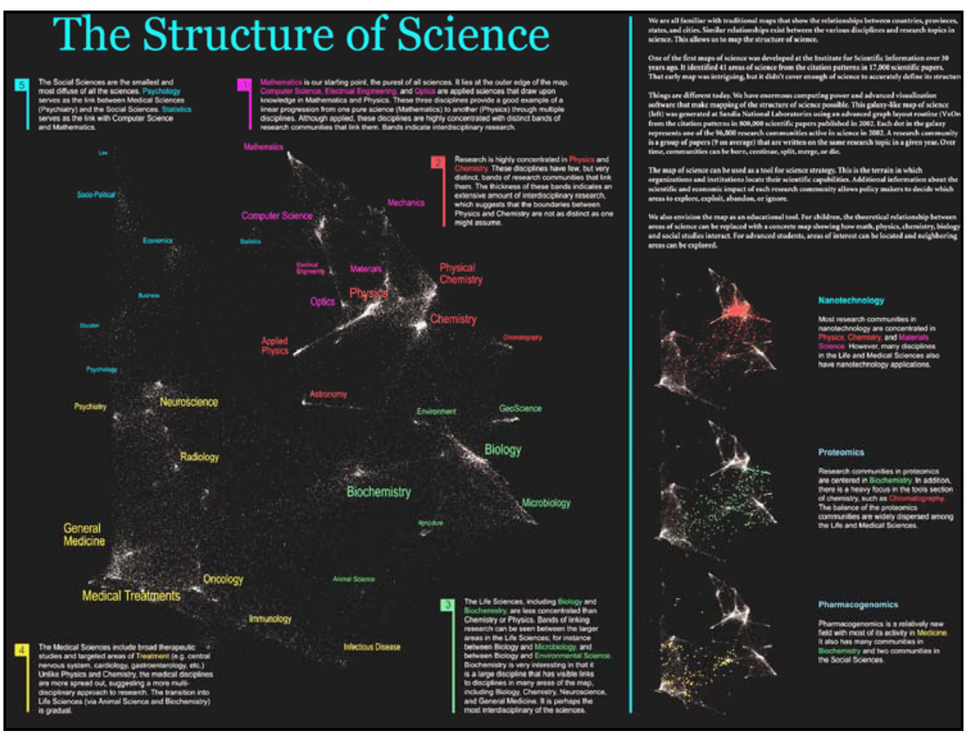

A New Map of the WHOLE WORLD with the Trade winds
According to ye latest and most Exact Obs-
ervations By H. Moll Geographer
In this Mapp is inserted A View of ye General & Coasting Trade-Winds, Monsoons or ye Shifting Trade-winds Note that ye Arrowes among ye Lines shew ye Course of these General & Coasting Winds. and ye Arrows in ye void Spaces shew ye Course of ye Shifting Trade-winds, and ye Abbreviation Sept. &c. Shew ye Times of ye Year when such Winds Blow.
The Signs of the Zodiack
The First & are Northern, the other Southern Signs
NORTH POLE
The Great South Sea
NORTH AMERICA
SOUTH SEA
AMERICA
AFRICA
ASIA
INDIAN OCEAN
New Holland
North Pole
South Pole

How would a map of science look?

What metaphors would work best?

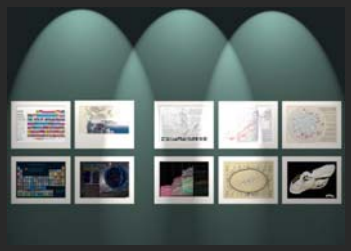



climat
chang
urban
model
spatial
forest
soil
temperatur
transport
plan
city
popul
precipit
data
gis
sediment
weather
hous
lake
migrat
health
remot
channel
beach
flow
measur
immigr
river
water
land
site analysi
flood
town
commun
manag
state
women
natur
develop
environment
map
polit
work
social
coloni
natur
space
local
region
student
geograph
landscap
polit
global
econom
servic
geographi
place
nation
state
market
product
industri

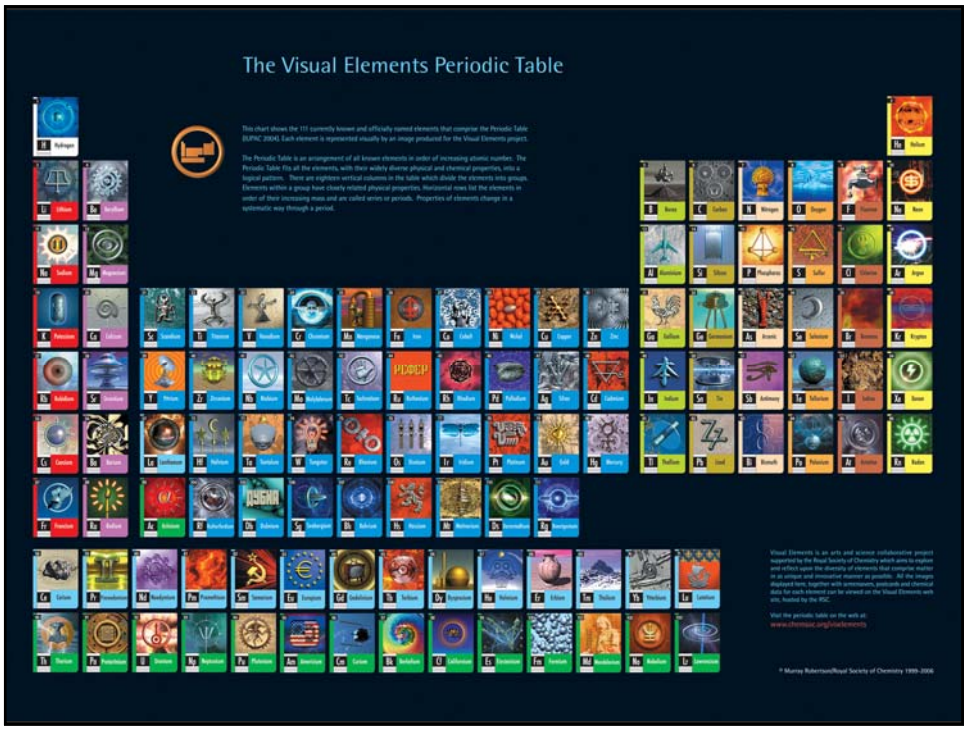

The Structure of Science
5
The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. Statistics serves as the link with Computer Science and Mathematics.
1
Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Computer Science, Electrical Engineering, and Optics are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.
2
Research is highly concentrated in Physics and Chemistry. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.
Law
Socio-Political
Economics
Statistics
Business
Psychology
Psychiatry
Mathematics
Computer Science
Mechanics
Electrical Engineering
Materials
Physical Chemistry
Optics
Physics
Chemistry
Applied Physics
Astronomy
Neuroscience
Radiology
Environment
GeoScience
Biology
Biochemistry
Microbiology
Agriculture
General Medicine
Oncology
Medical Treatments
Animal Science
Immunology
Infectious Disease
4
The Medical Sciences include broad therapeutic studies and targeted areas of Treatment (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.
3
The Life Sciences, including Biology and Biochemistry, are less concentrated than Chemistry or Physics. Bands of linking research can be seen between the larger areas in the Life Sciences; for instance between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.
One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.
Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd) from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.
The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.
We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest can be located and neighboring areas can be explored.
Nanotechnology
Most research communities in nanotechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.
Proteomics
Research communities in proteomics are centered in Biochemistry. In addition, there is a heavy focus in the tools section of chemistry, such as Chromatography. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.
Pharmacogenomics
Pharmacogenomics is a relatively new field with most of its activity in Medicine. It also has many communities in Biochemistry and two communities in the Social Sciences.

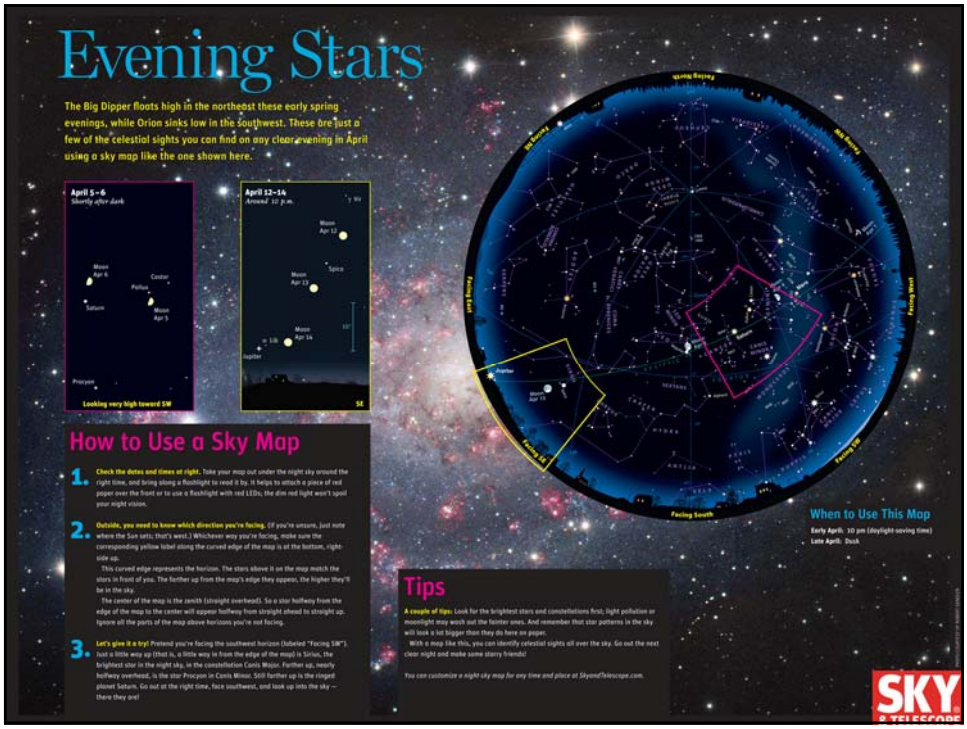
The Power of Reference Systems

Four Existing Reference Systems
VERSUS
Six Potential Reference Systems of Science



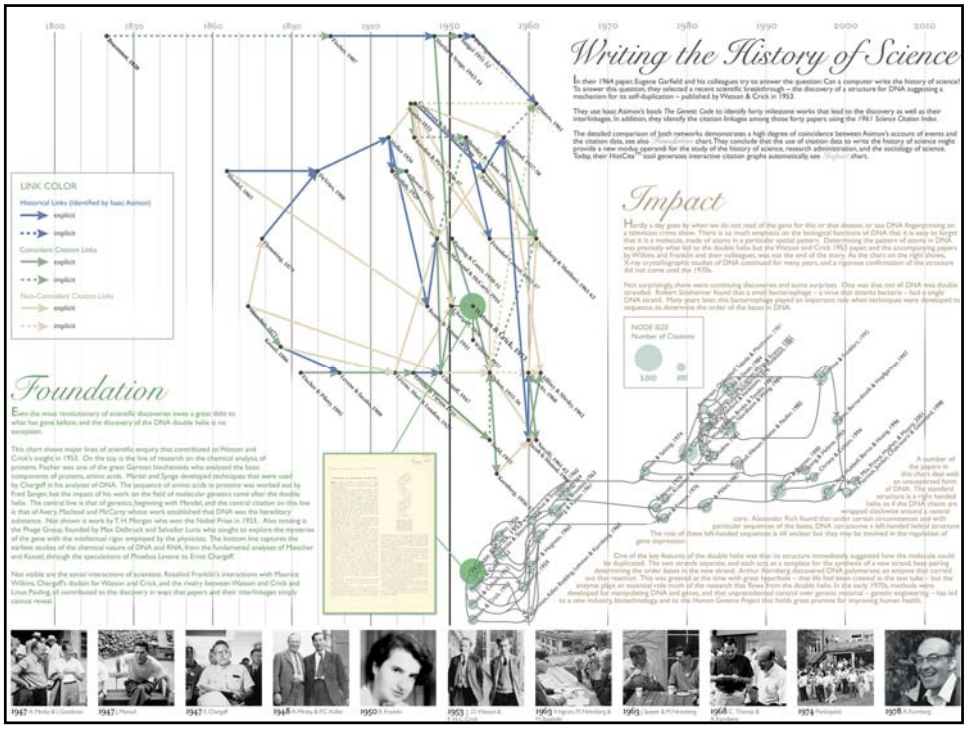
(2nd Iteration of Places & Spaces Exhibit - 2006)

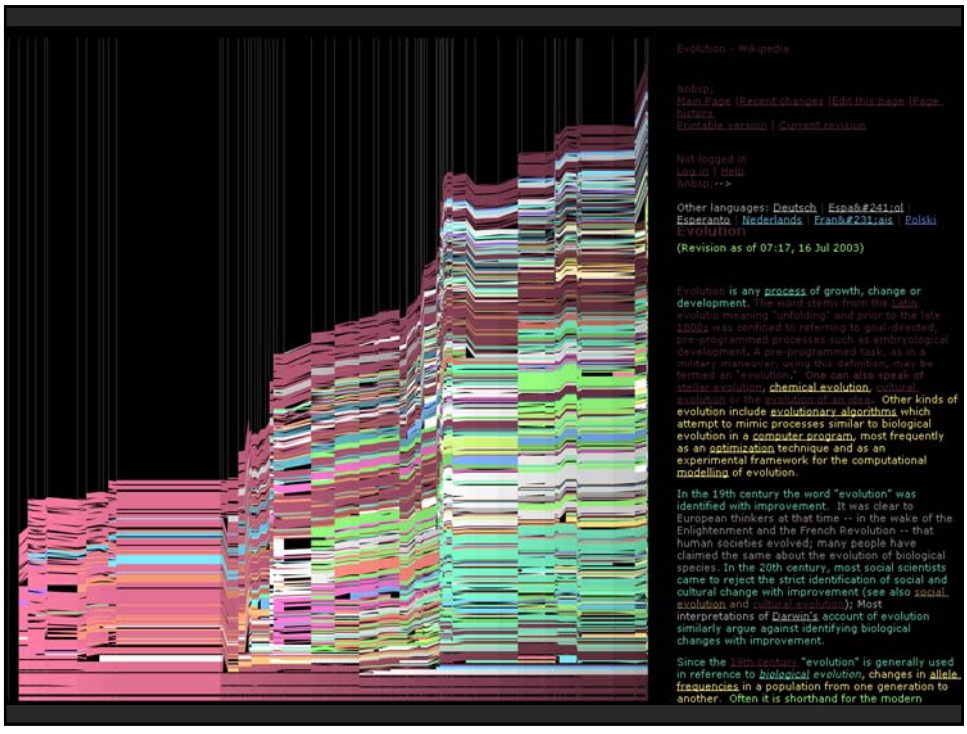


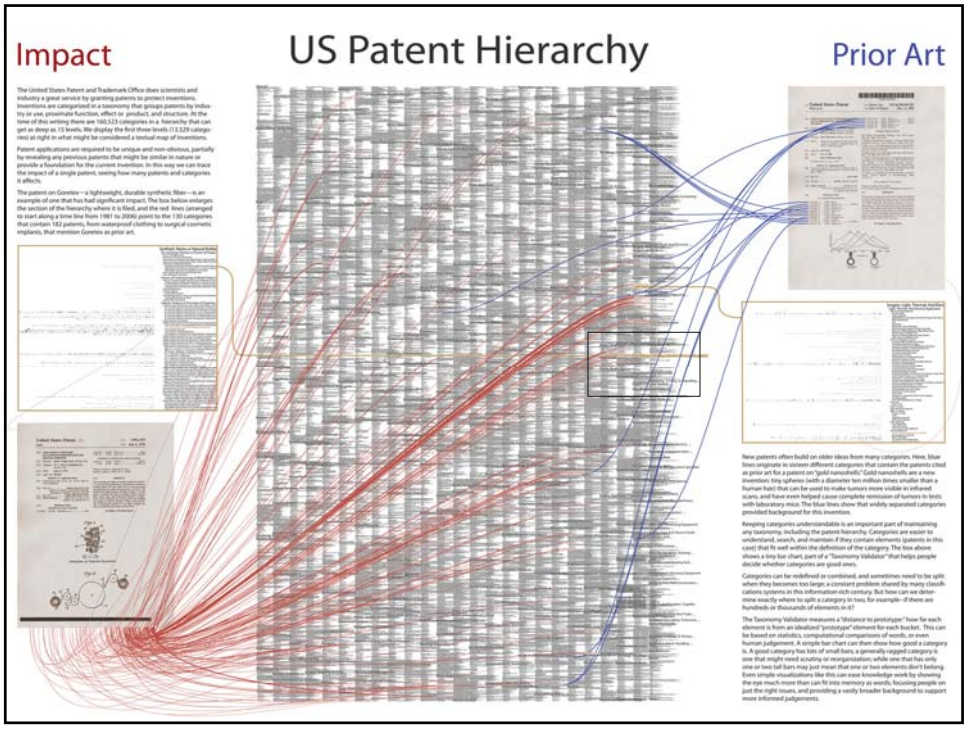


How would a reference system for all of science look?

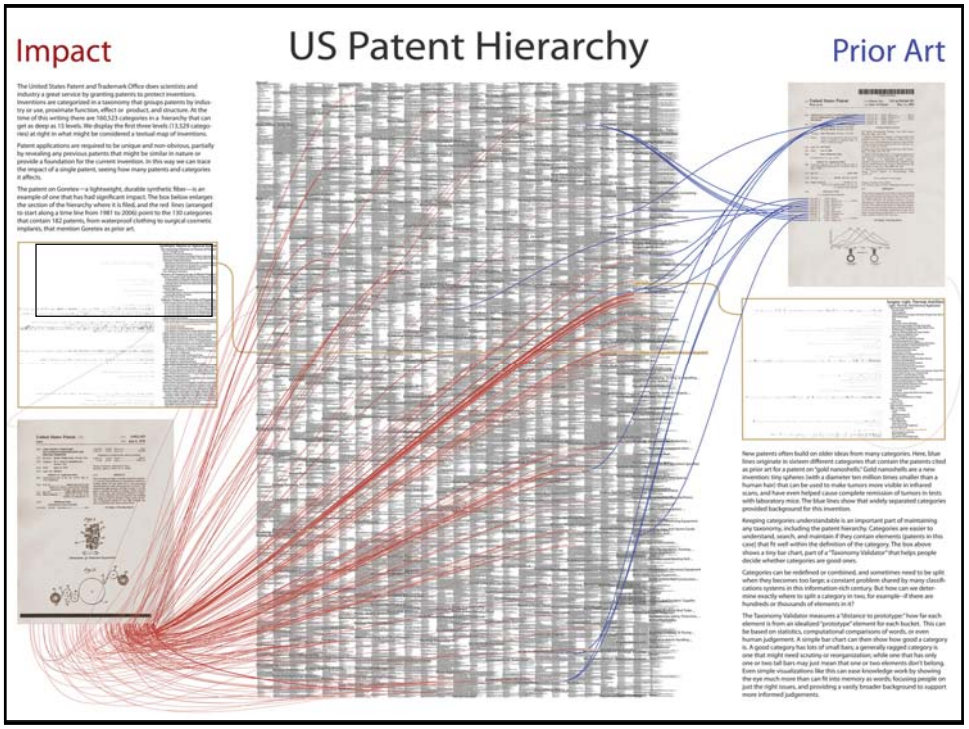
What dimensions would it have?

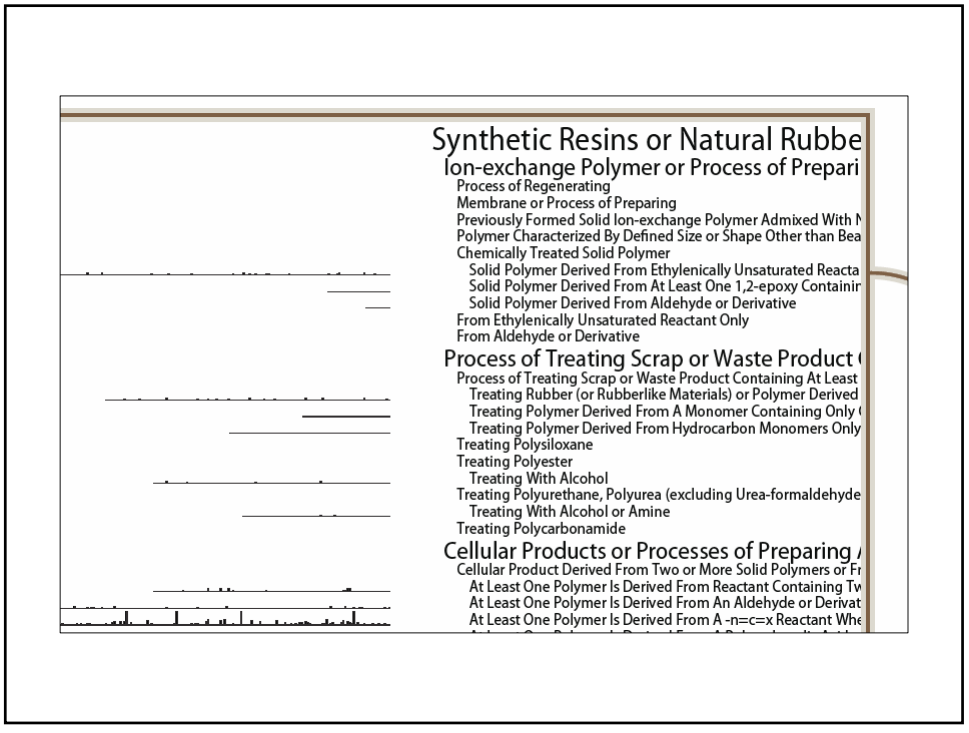

Writing the History of Science
Impact
Foundation
LINK COLOR
Historical Links (identified by Isaac Asimov)
explicit
implicit
Coincidors Citation Links
explicit
implicit
Non-Coincidors Citation Links
explicit
implicit
NODE SIZE
Number of Citations
1,000
100

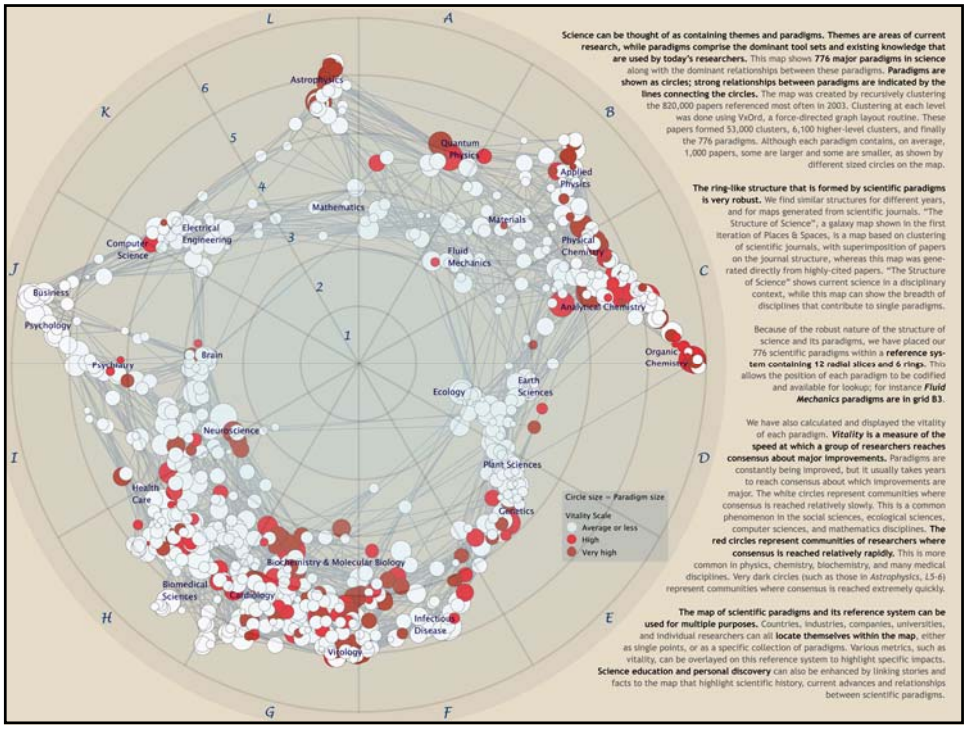

Evolution - Wikipedia
Main Page | Recent changes | Edit this page | Page history
Printable version | Current revision
Not logged in
Log in | Help
Other languages: Deutsch | Español | Esperanto | Nederlands | Français | Polski
Evolution
(Revision as of 07:17, 16 Jul 2003)
Evolution is any process of growth, change or development. The word stems from the Latin evolutio meaning "unfolding" and prior to the late 1800s was confined to referring to goal-directed, pre-programmed processes such as embryological development. A pre-programmed task, as in a military maneuver, using this definition, may be termed an "evolution." One can also speak of stellar evolution, chemical evolution, cultural evolution or the evolution of an idea. Other kinds of evolution include evolutionary algorithms which attempt to mimic processes similar to biological evolution in a computer program, most frequently as an optimization technique and as an experimental framework for the computational modelling of evolution.
In the 19th century the word "evolution" was identified with improvement. It was clear to European thinkers at that time -- in the wake of the Enlightenment and the French Revolution -- that human societies evolved; many people have claimed the same about the evolution of biological species. In the 20th century, most social scientists came to reject the strict identification of social and cultural change with improvement (see also social evolution and cultural evolution); Most interpretations of Darwin's account of evolution similarly argue against identifying biological changes with improvement.
Since the 19th century "evolution" is generally used in reference to biological evolution, changes in allele frequencies in a population from one generation to another. Often it is shorthand for the modern

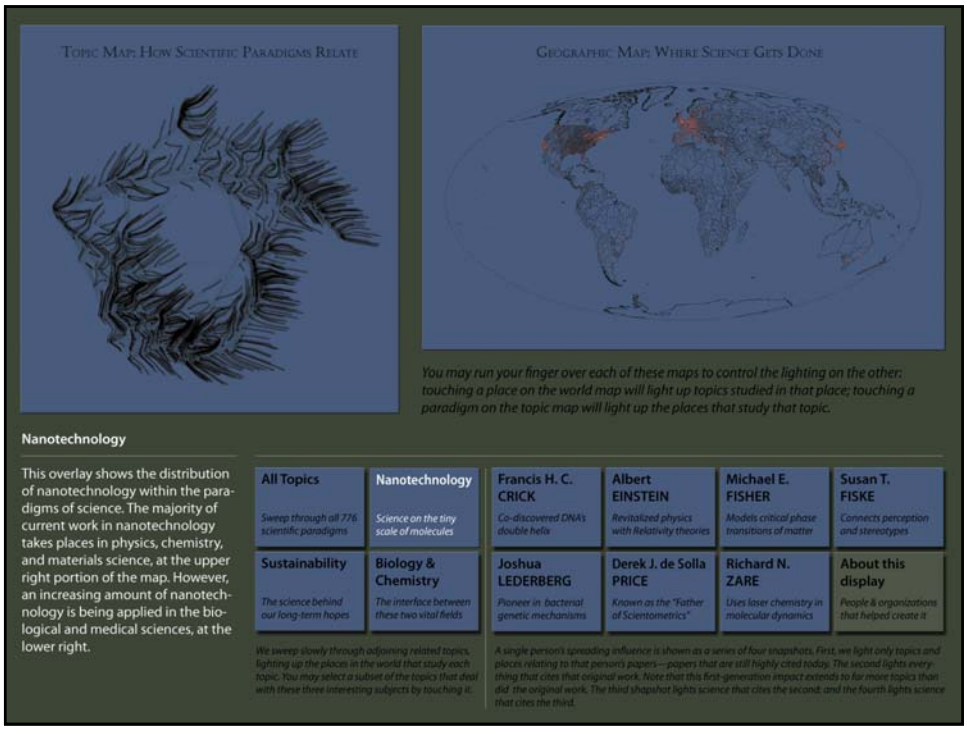

Impact
US Patent Hierarchy
Prior Art

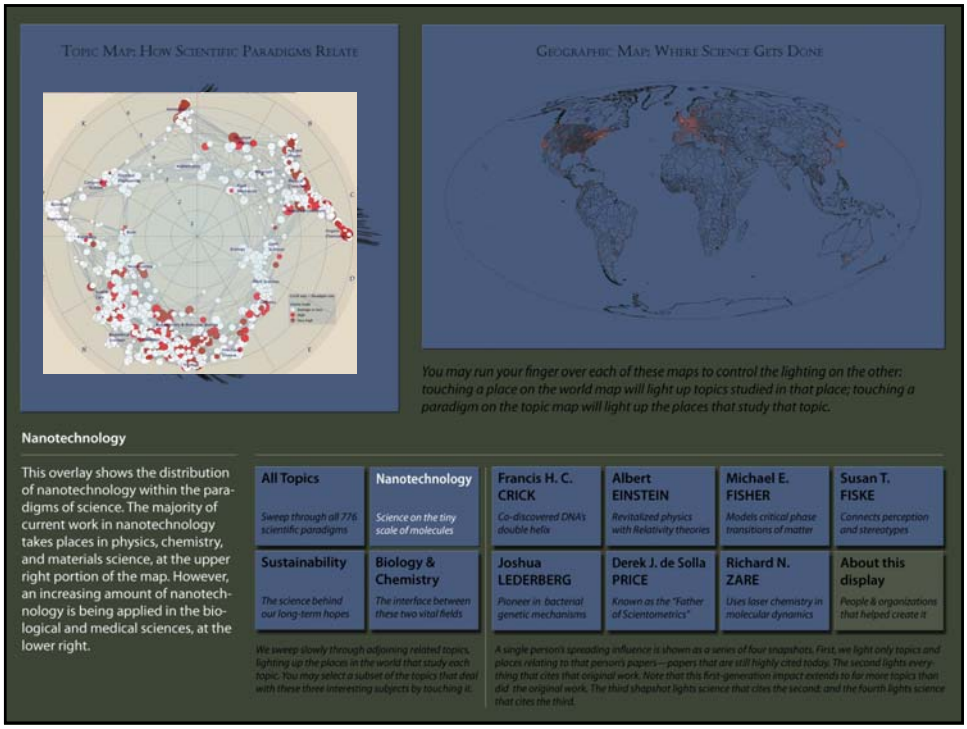

Extruding a solid polymer containing material to
Synthetic Resins or Natural ...
Coating Processes
Rotary Shafts, Gudgeons, ...
Data Processing: Structural ...
Synthetic Resins or Natural ...
Measuring, Testing
Data Processing: Speech
Data Processing: Financial
Jewelry, Symbolic
Threaded, Headed Fastener
Synthetic R
Data Processing: Artificial ...
Data Processing: Database ...
Synthetic Resins or Natural ...
Electrical Computer ... Arithmet
Amusement Devices
Electrical Computers And ...


Impact
US Patent Hierarchy
Prior Art

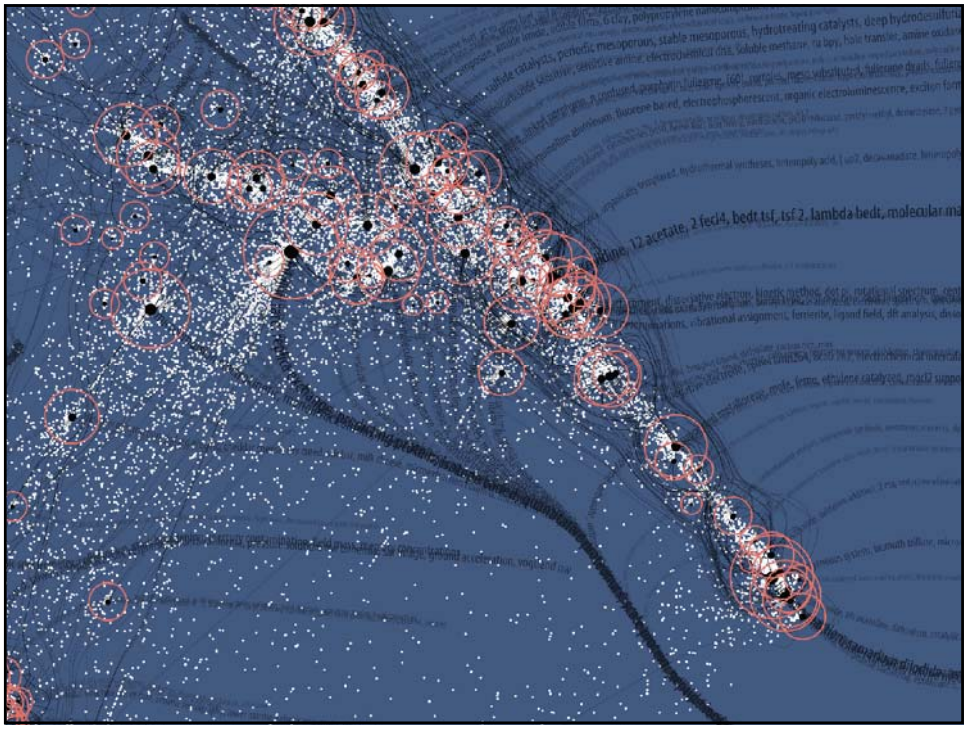

Synthetic Resins or Natural Rubbe
Ion-exchange Polymer or Process of Prepari
Process of Regenerating
Membrane or Process of Preparing
Previously Formed Solid Ion-exchange Polymer Admixed With
Polymer Characterized By Defined Size or Shape Other than Bea
Chemically Treated Solid Polymer
Solid Polymer Derived From Ethylenically Unsaturated Reacta
Solid Polymer Derived From At Least One 1,2-epoxy Containir
Solid Polymer Derived From Aldehyde or Derivative
From Ethylenically Unsaturated Reactant Only
From Aldehyde or Derivative
Process of Treating Scrap or Waste Product
Process of Treating Scrap or Waste Product Containing At Least
Treating Rubber (or Rubberlike Materials) or Polymer Derived
Treating Polymer Derived From A Monomer Containing Only
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
Cellular Product Derived From Two or More Solid Polymers or Fr
At Least One Polymer Is Derived From Reactant Containing Tw
At Least One Polymer Is Derived From An Aldehyde or Derivat
At Least One Polymer Is Derived From A -n=c=x Reactant Whe

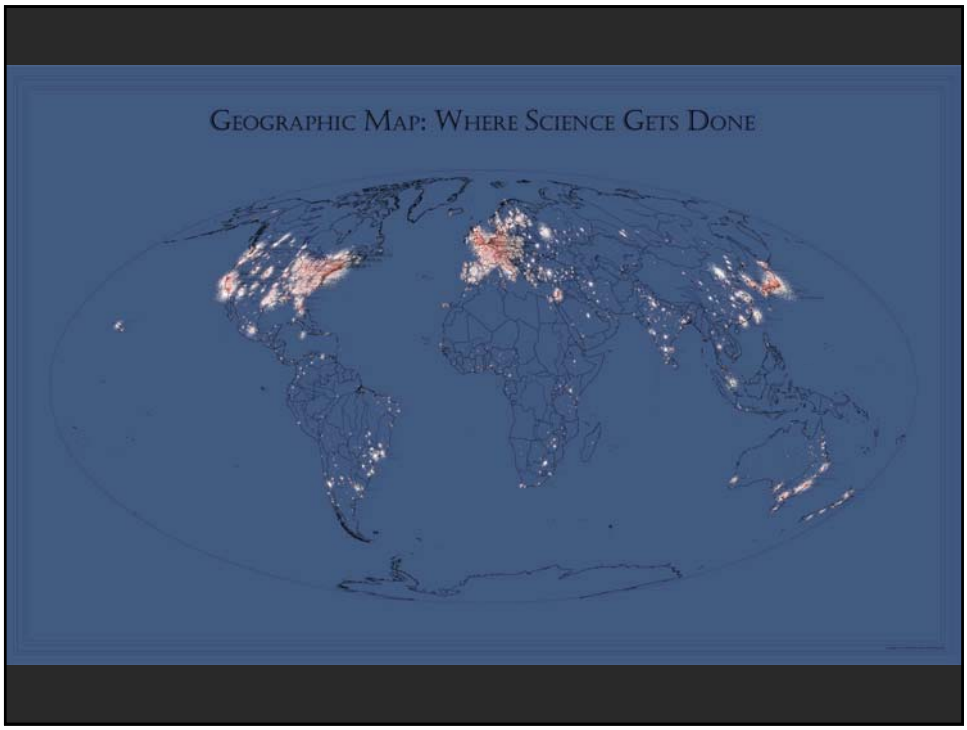

L
A
K
B
J
C
I
D
H
E
G
F
1
2
3
4
5
6
Astrophysics
Quantum Physics
Applied Physics
Mathematics
Materials
Electrical Engineering
Computer Science
Fluid Mechanics
Physical Chemistry
Business
Psychology
Analytical Chemistry
Brain
Psychiatry
Organic Chemistry
Ecology
Earth Sciences
Neuroscience
Plant Sciences
Health Care
Genetics
Circle size = Paradigm size
Vitality Scale
Average or less
High
Very high
Biochemistry & Molecular Biology
Biomedical Sciences
Cardiology
Infectious Disease
Virology
Science can be thought of as containing themes and paradigms. Themes are areas of current research, while paradigms comprise the dominant tool sets and existing knowledge that are used by today's researchers. This map shows 776 major paradigms in science along with the dominant relationships between these paradigms. Paradigms are shown as circles; strong relationships between paradigms are indicated by the lines connecting the circles. The map was created by recursively clustering the 820,000 papers referenced most often in 2003. Clustering at each level was done using VxOrd, a force-directed graph layout routine. These papers formed 53,000 clusters, 6,100 higher-level clusters, and finally the 776 paradigms. Although each paradigm contains, on average, 1,000 papers, some are larger and some are smaller, as shown by different sized circles on the map.
The ring-like structure that is formed by scientific paradigms is very robust. We find similar structures for different years, and for maps generated from scientific journals. "The Structure of Science", a galaxy map shown in the first iteration of Places & Spaces, is a map based on clustering of scientific journals, with superimposition of papers on the journal structure, whereas this map was generated directly from highly-cited papers. "The Structure of Science" shows current science in a disciplinary context, while this map can show the breadth of disciplines that contribute to single paradigms.
Because of the robust nature of the structure of science and its paradigms, we have placed our 776 scientific paradigms within a reference system containing 12 radial slices and 6 rings. This allows the position of each paradigm to be codified and available for lookup; for instance Fluid Mechanics paradigms are in grid B3.
We have also calculated and displayed the vitality of each paradigm. Vitality is a measure of the speed at which a group of researchers reaches consensus about major improvements. Paradigms are constantly being improved, but it usually takes years to reach consensus about which improvements are major. The white circles represent communities where consensus is reached relatively slowly. This is a common phenomenon in the social sciences, ecological sciences, computer sciences, and mathematics disciplines. The red circles represent communities of researchers where consensus is reached relatively rapidly. This is more common in physics, chemistry, biochemistry, and many medical disciplines. Very dark circles (such as those in Astrophysics, J.5-6) represent communities where consensus is reached extremely quickly.
The map of scientific paradigms and its reference system can be used for multiple purposes. Countries, industries, companies, universities, and individual researchers can all locate themselves within the map, either as single points, or as a specific collection of paradigms. Various metrics, such as vitality, can be overlayed on this reference system to highlight specific impacts. Science education and personal discovery can also be enhanced by linking stories and facts to the map that highlight scientific history, current advances and relationships between scientific paradigms.

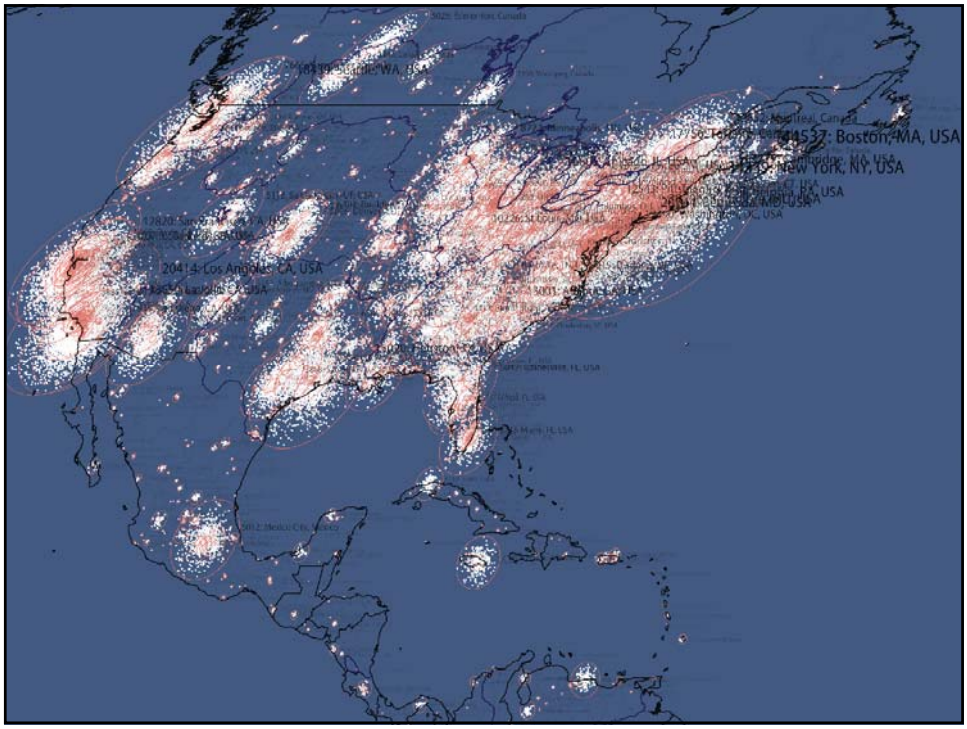

Topic Map: How Scientific Paradigms Relate
Geographic Map: Where Science Gets Done
You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the 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 776 scientific paradigms
Nanotechnology
Science on the tiny scale of molecules
Sustainability
The science behind our long-term hopes
Biology & Chemistry
The interface between these two vital fields
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.
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
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
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.

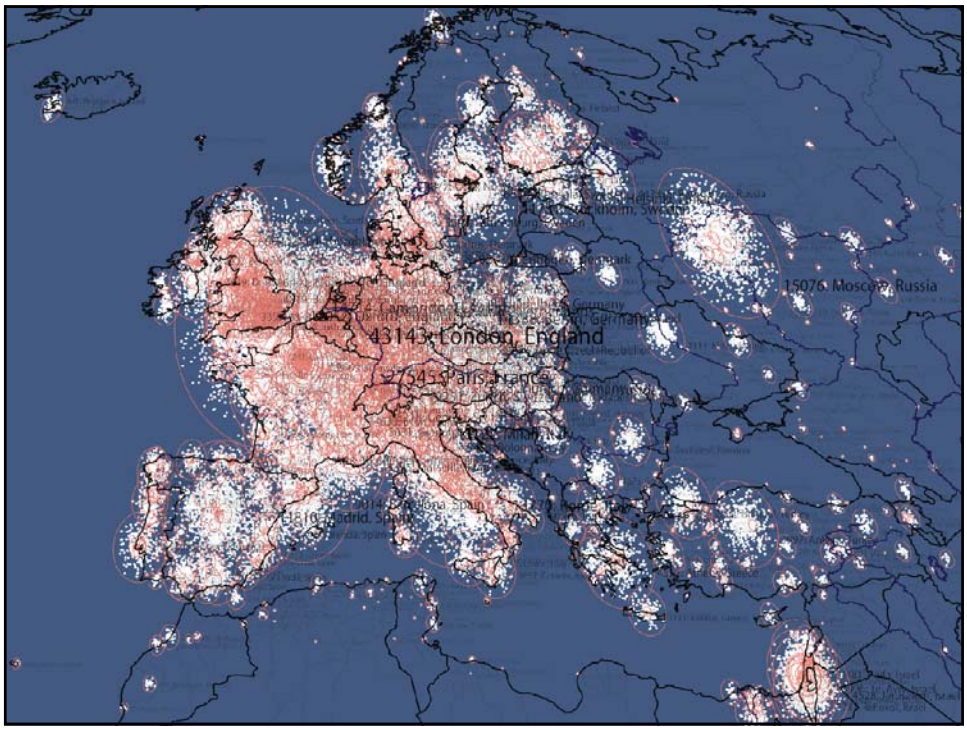

Topic Map: How Scientific Paradigms Relate
Geographic Map: Where Science Gets Done
You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the 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 776 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 shapshot lights science that cites the second; and the fourth lights science that cites the third.

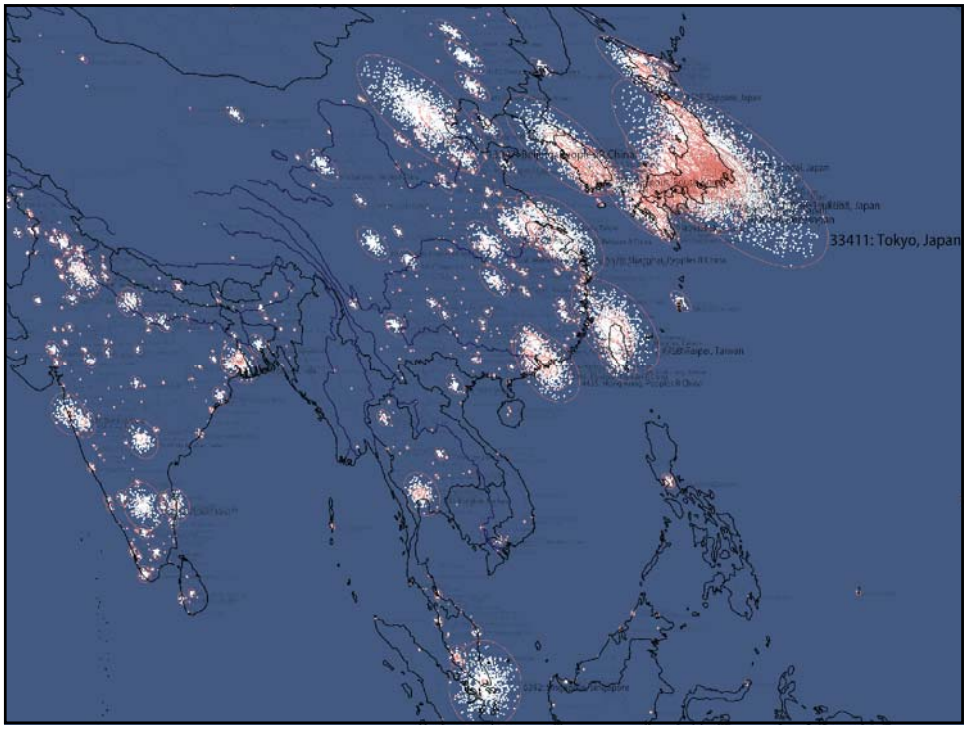

Topic Map: How Scientific Paradigms Relate




Geographic Map: Where Science Gets Done


20414 Los Angeles, CA, USA
New York, NY, USA

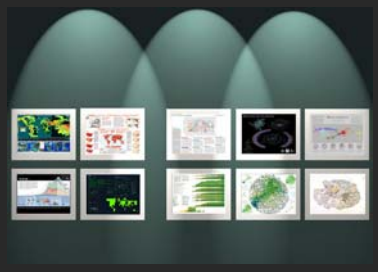

15076 Moscow, Russia
43143 London, England
27543 Paris, France



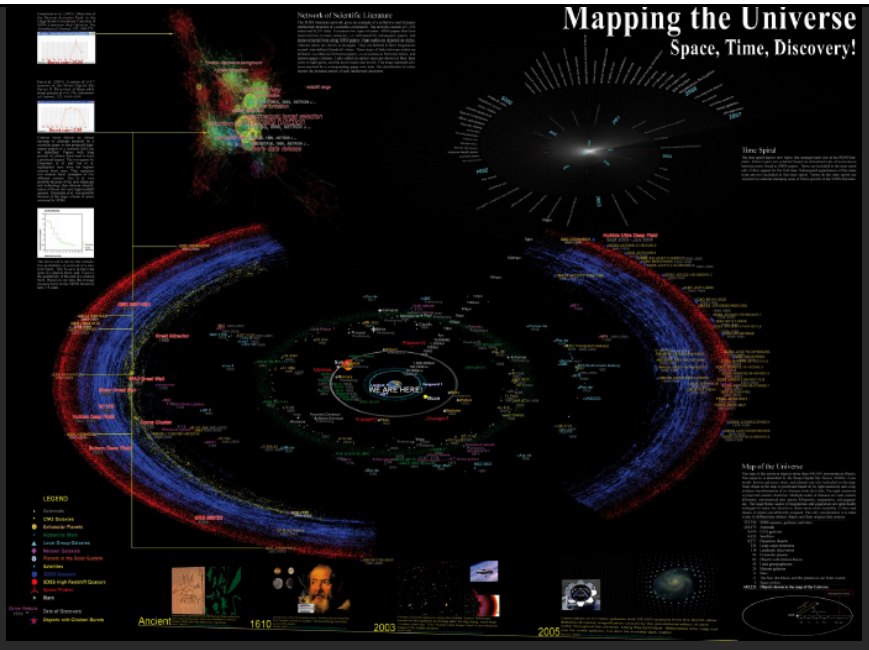
Illuminated Diagram Display
http://www.youtube.com/watch?v=bXABcOABG4E

The Power of Forecasts

Four Existing Forecasts
VERSUS
Six Potential Science 'Weather' Forecasts



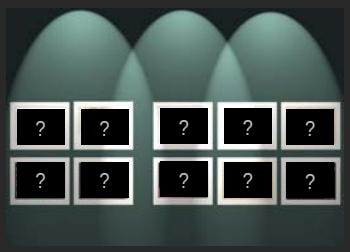
(3rd Iteration of Places & Spaces Exhibit - 2007)



Mapping the Universe: Space, Time, and Discovery!, by Chaomei Chen, Jian Zhang, Michael S. Vogeley, J. Richard Gott III, Mario Juric, and Lisa Kershner

Science Maps for Economic Decision Making

Four Existing Maps
VERSUS
Six Science Maps



(4th Iteration of Places & Spaces Exhibit - 2008)

A Potential Future: Science Maps in Action

KIDS first …

All maps of science are on sale via
http://scimaps.org/ordermaps/


Inventors & Inventions
ASTRONOMY
PHYSICS
ENGINEERING
COMPUTER SCIENCE
MATH
SOCIAL SCIENCE
PHYSICAL CHEMISTRY
BRAIN
MENTAL HEALTH
EARTH SCIENCES
ORGANIC THEORY
ECOLOGY
HEALTH CARE
BIOLOGY
THERAPEUTIC RESEARCH
INFECTIOUS DISEASE
CELL BIOLOGY
VIRUSES


Inventors
ASTRONOMY
PHYSICS
ENGINEERING
COMPUTER SCIENCE
MATH
SOCIAL SCIENCE
PHYSICAL CHEMISTRY
BRAIN
MENTAL HEALTH
EARTH SCIENCES
ORGANIC THEORY
ECOLOGY
HEALTH CARE
BIOLOGY
THERAPEUTIC RESEARCH
INFECTIOUS DISEASE
CELL BIOLOGY
VIRUSES

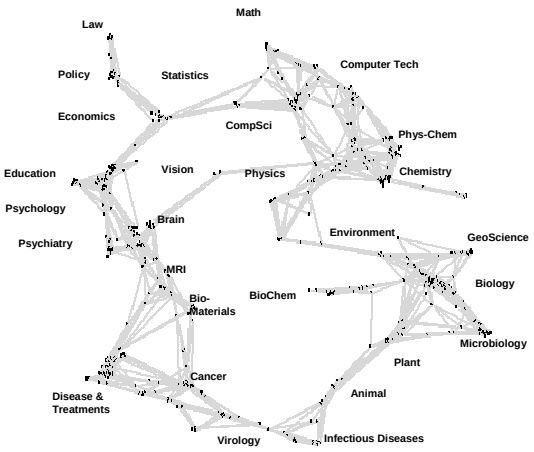

54
SOCCER

… my SPONSORS next …

Latest 'Base Map' of Science

Kevin W. Boyack & Richard Klavans, unpublished work.

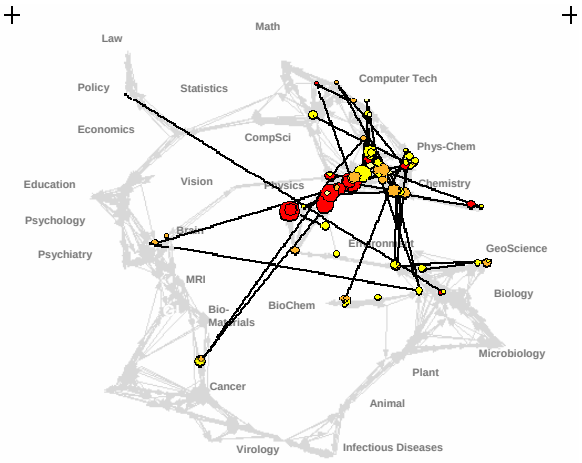
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

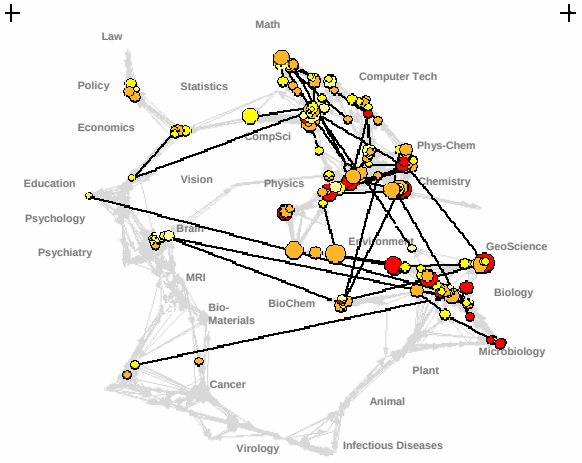
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

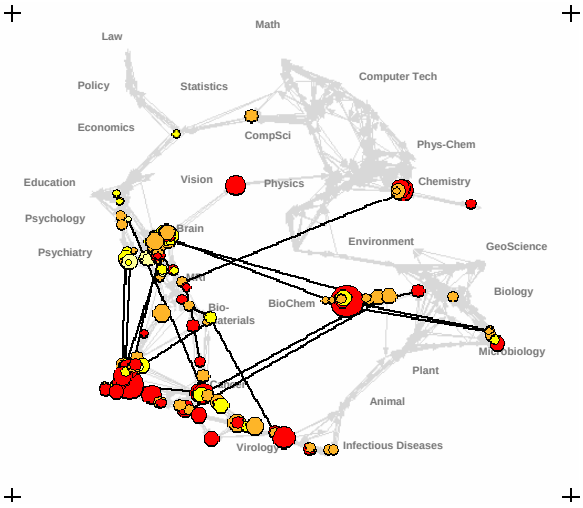
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

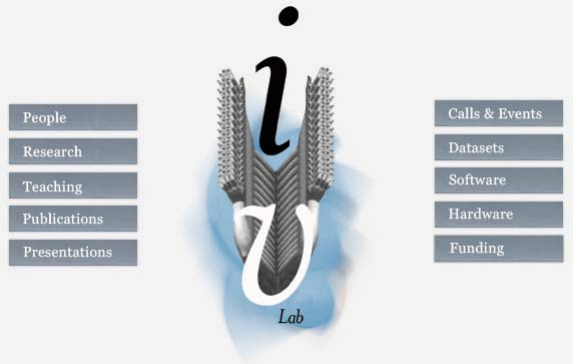
Funding Patterns of the National Institutes of Health (NIH)



… then SCIENTISTS …

Lab/Center Management System vs. Spacebook and MS Famulus

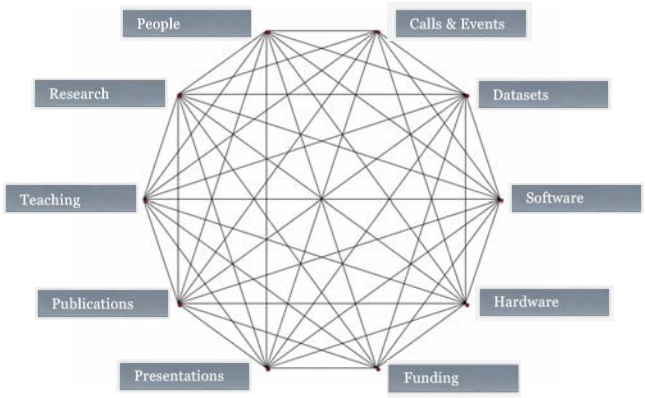
Designed to track, manage, and make use of data relevant for the daily operation of a medium size research team.



http://ivl.slis.indiana.edu

Data Entities and Interlinkages

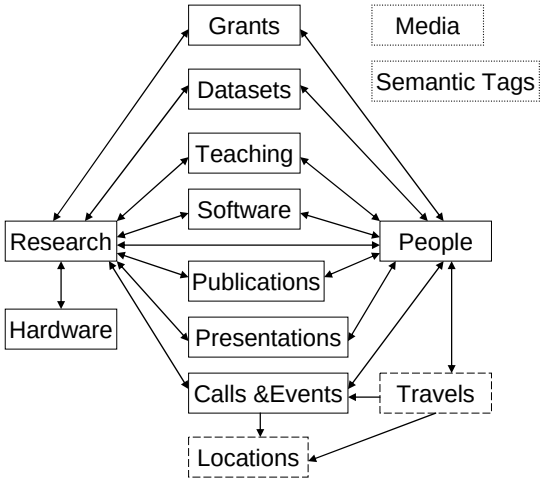
Designed for team leads, members, IT admins but also for external scholars and funding agencies.



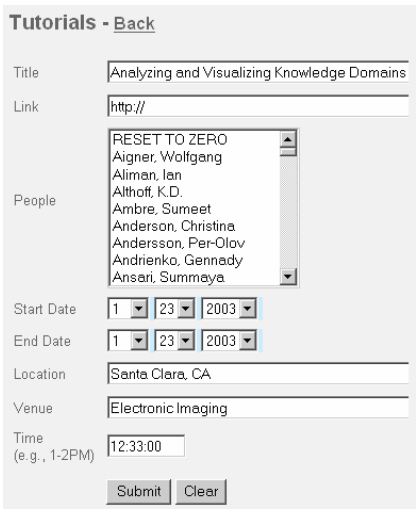
Not covered:
- Queries
- Workflows
- Protocols

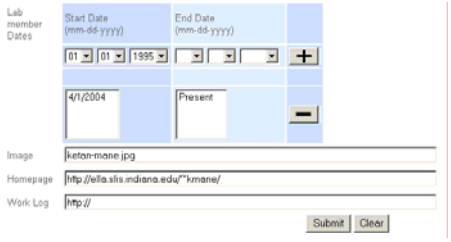
- Comments
- Bookmarks
- Ratings

Simplified representation of the IVL database schema

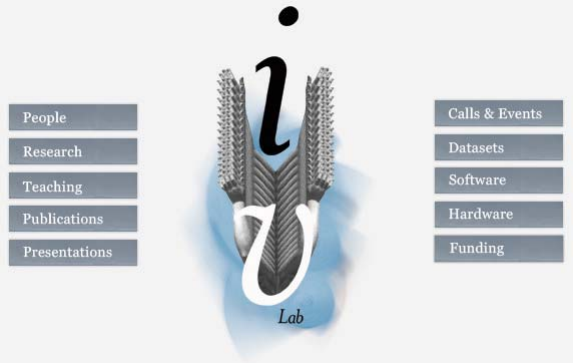


Data Entry



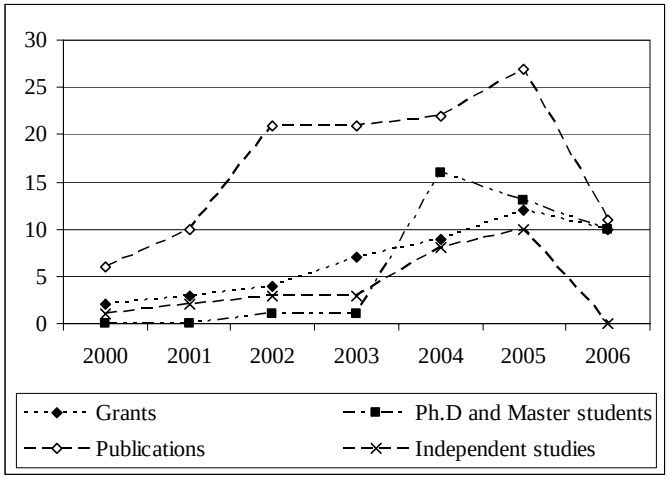


Demo

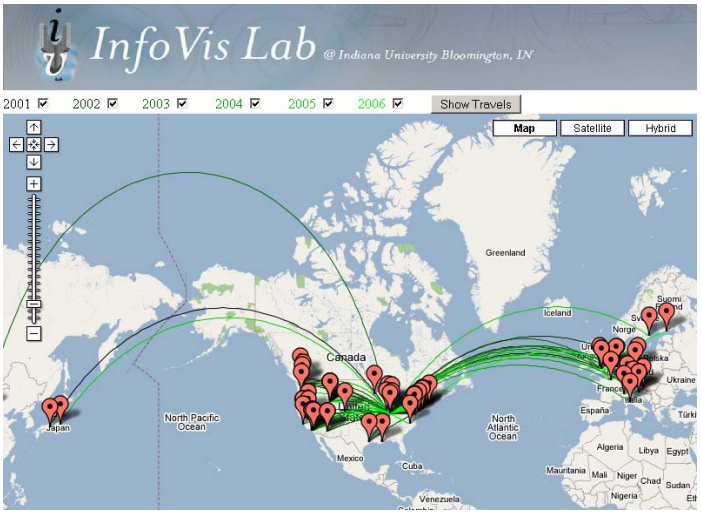


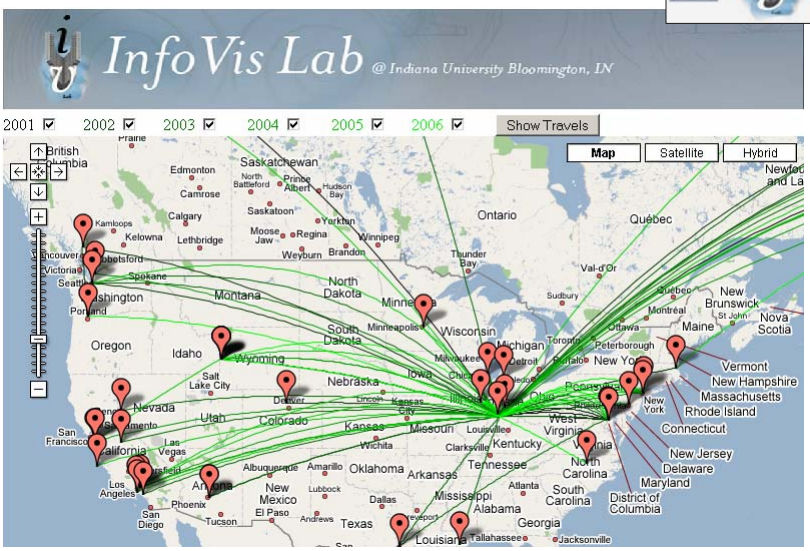
http://ivl.slis.indiana.edu

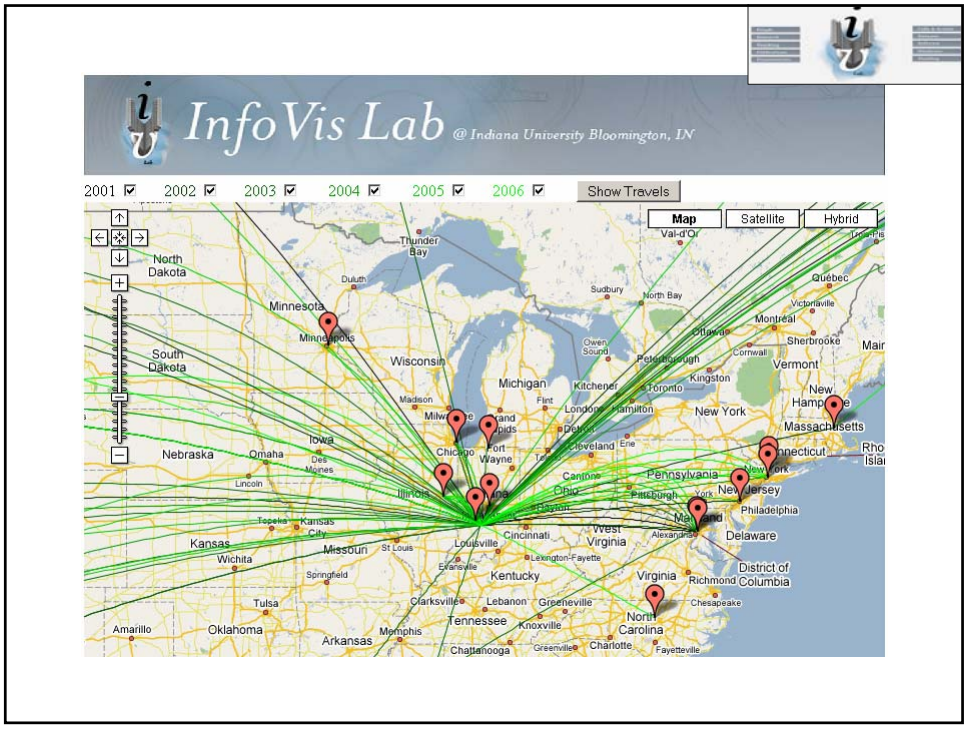
Time series analysis & visualization

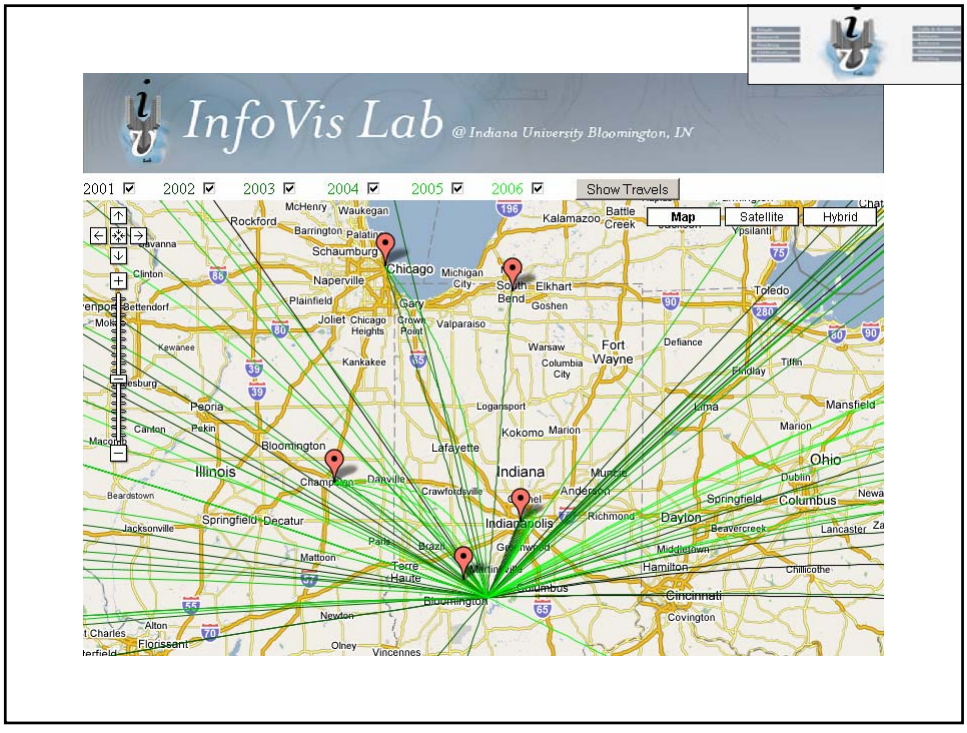


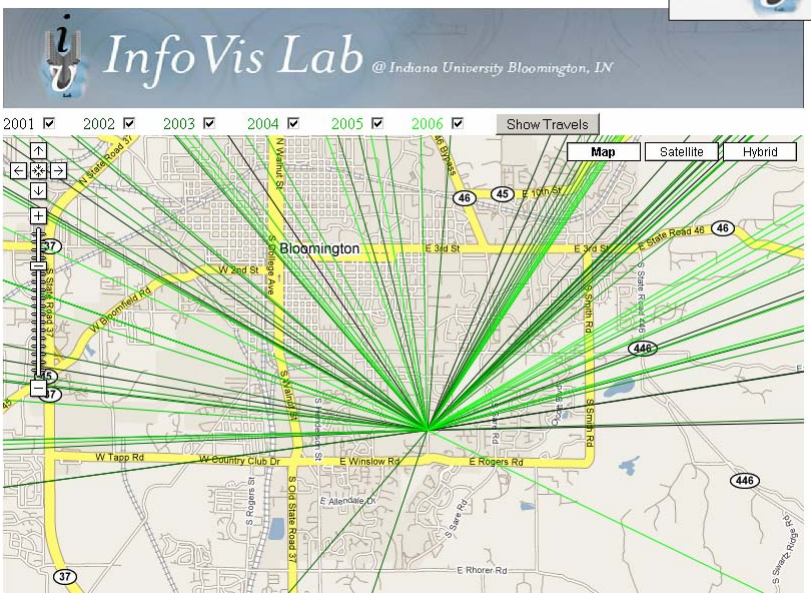
Katy's Travels in 2000-2006



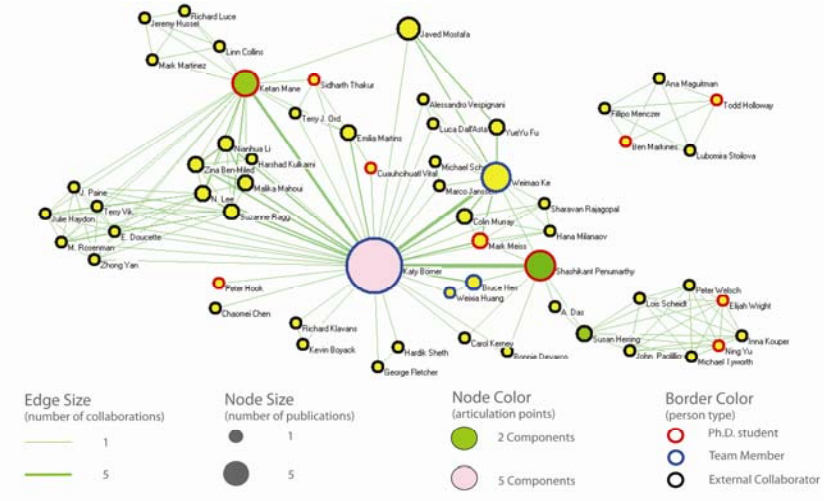


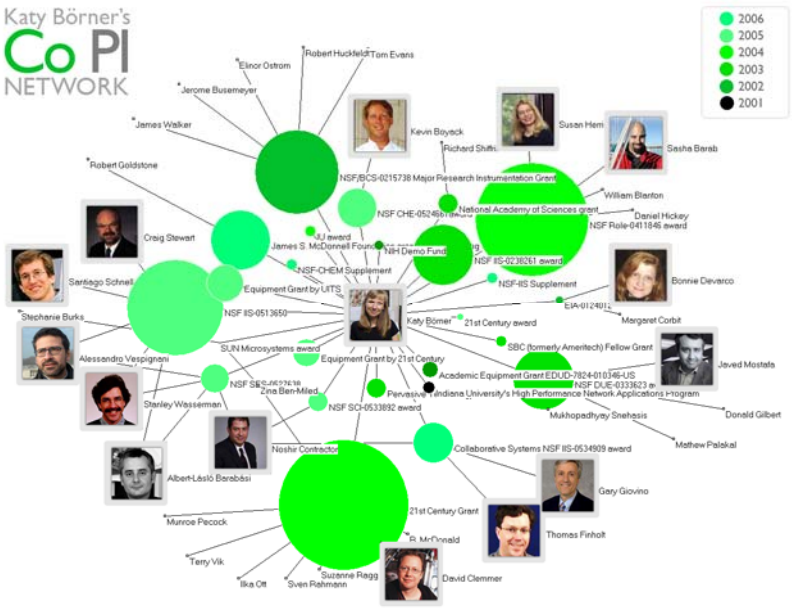

InfoVis Lab @ Indiana University Bloomington, IN
2001 2002 2003 2004 2005 2006
Show Travels
Map
Satellite
Hybrid


InfoVis Lab @ Indiana University Bloomington, IN
2001 2002 2003 2004 2005 2006
Show Travels
Map
Satellite
Hybrid



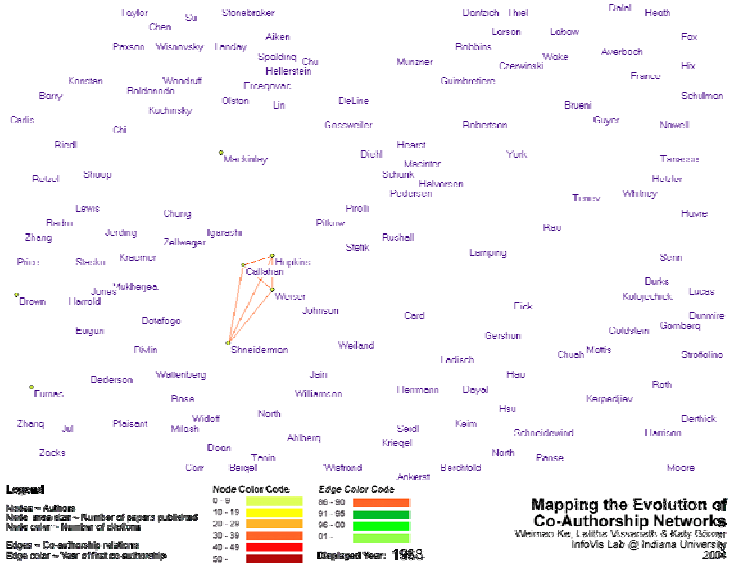
Co-authorship network for all publications of InfoVis lab group members in 2005 to 2006





Mapping the Evolution of Co-Authorship Networks

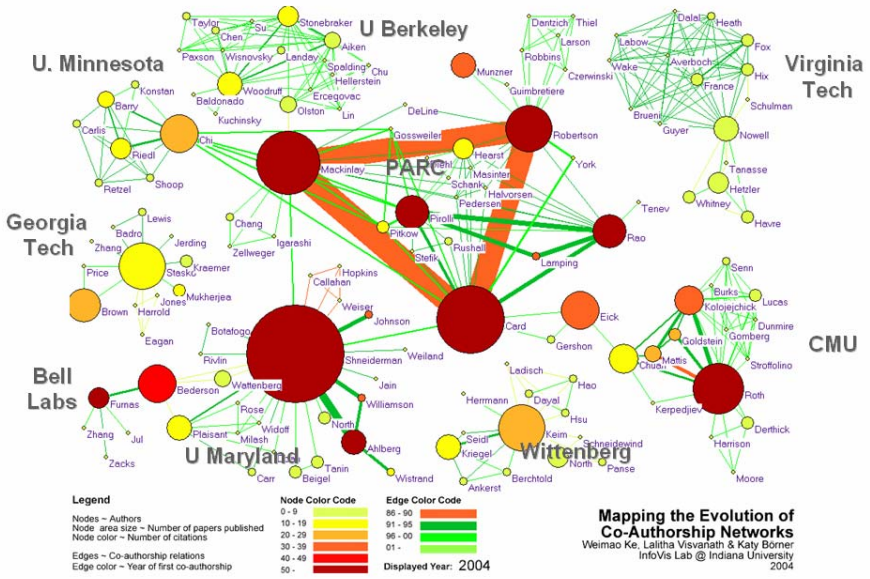
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



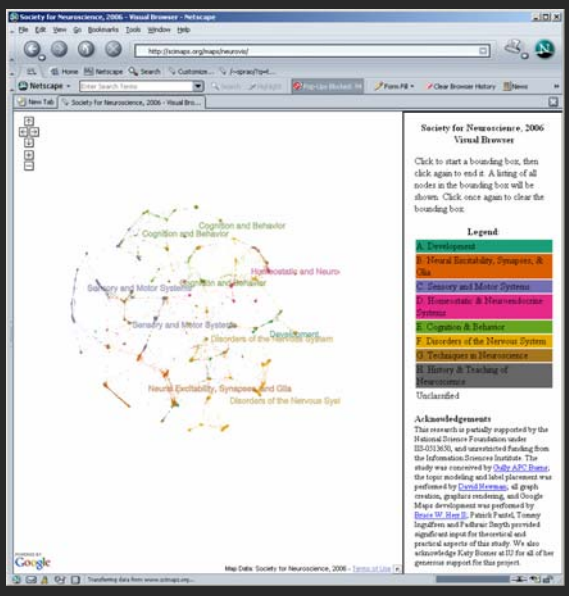
60

Mapping the Evolution of Co-Authorship Networks

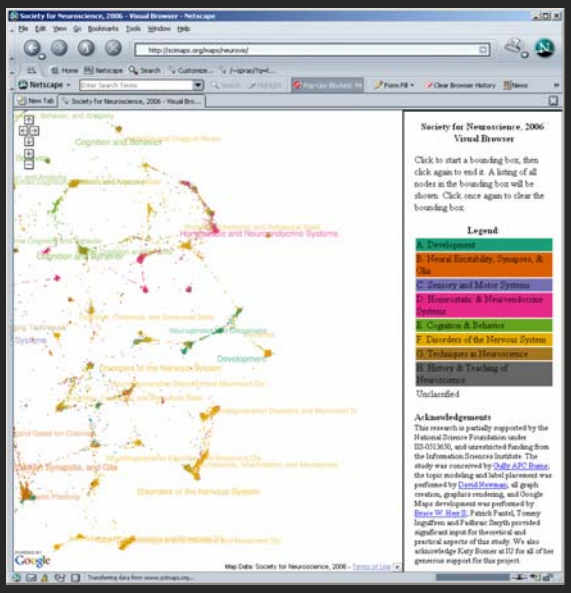
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest



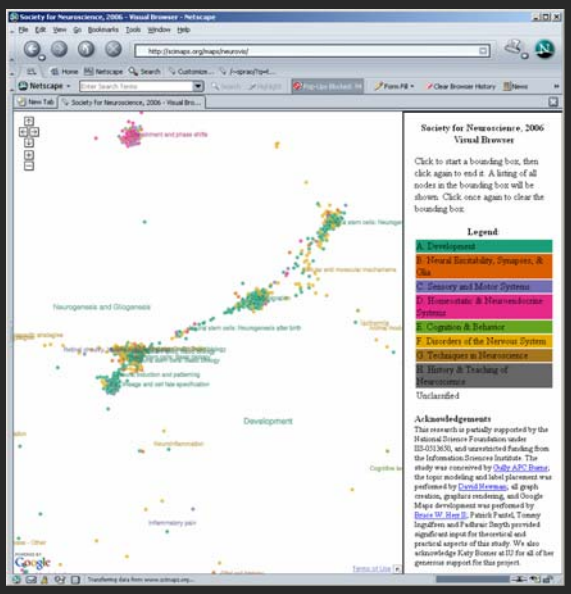
61



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/

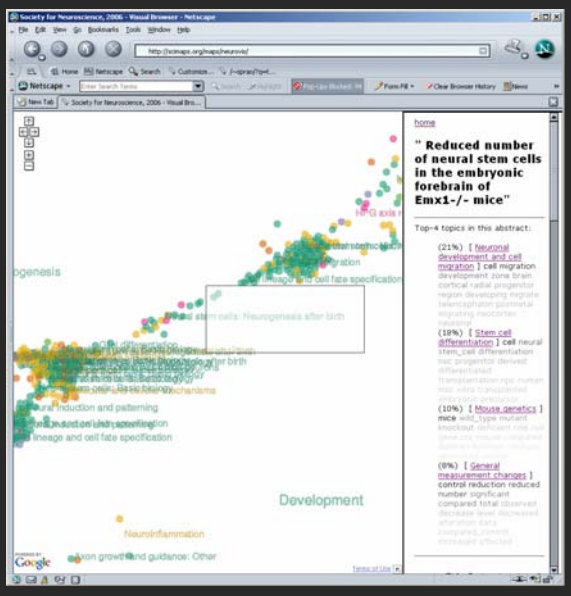


Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/

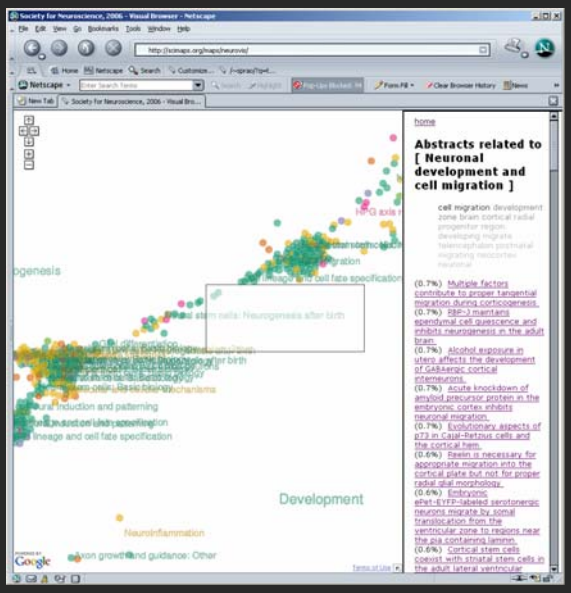


Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



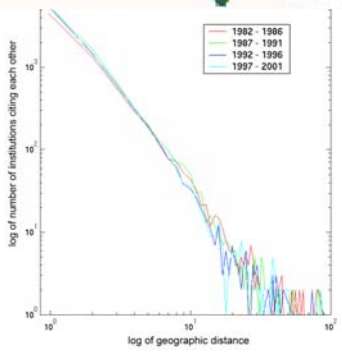
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/

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

Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?



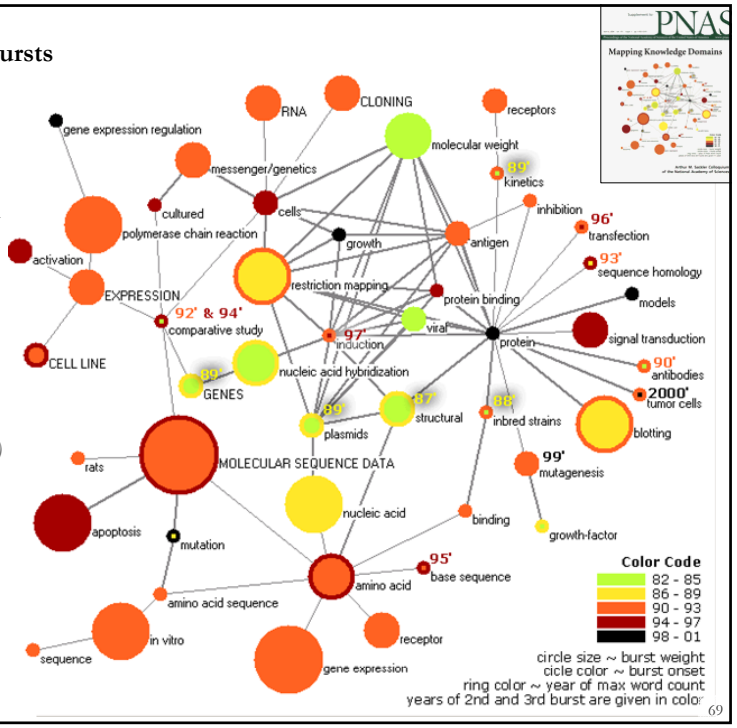
Contributions:

- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

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.



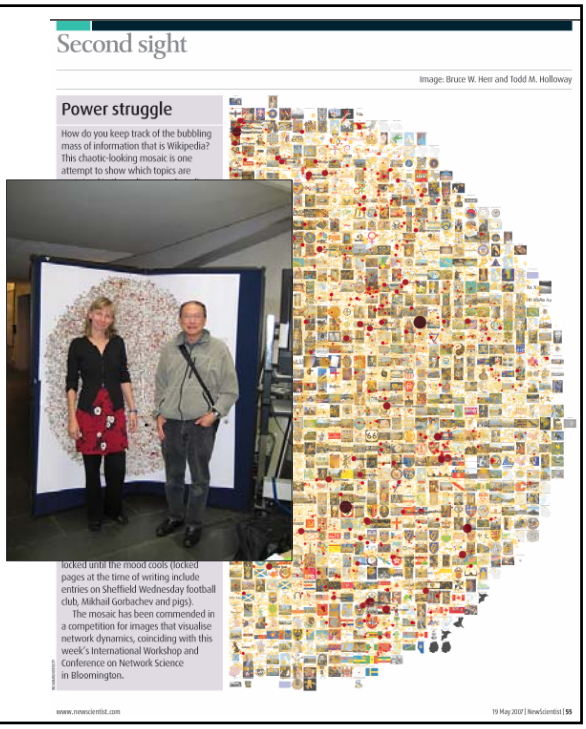
69

Wikipedian Activity

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

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



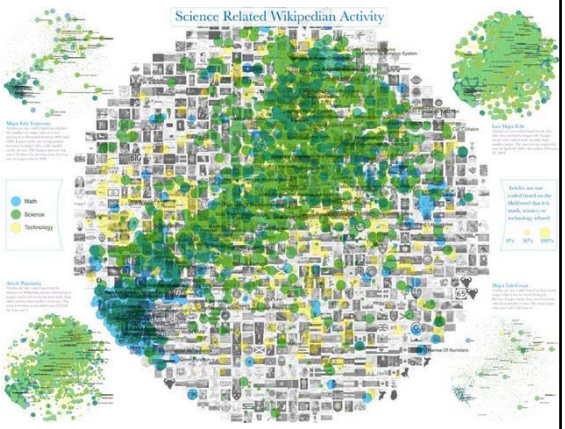
Science Related Wikipedian Activity

http://scimaps.org/dev/map_detail.php?map_id=165

Same base map.

Overlaid are 3,599 math (blue), 6,474 science (green), and 3,164 technology relevant articles (yellow).
All other articles are given in grey.

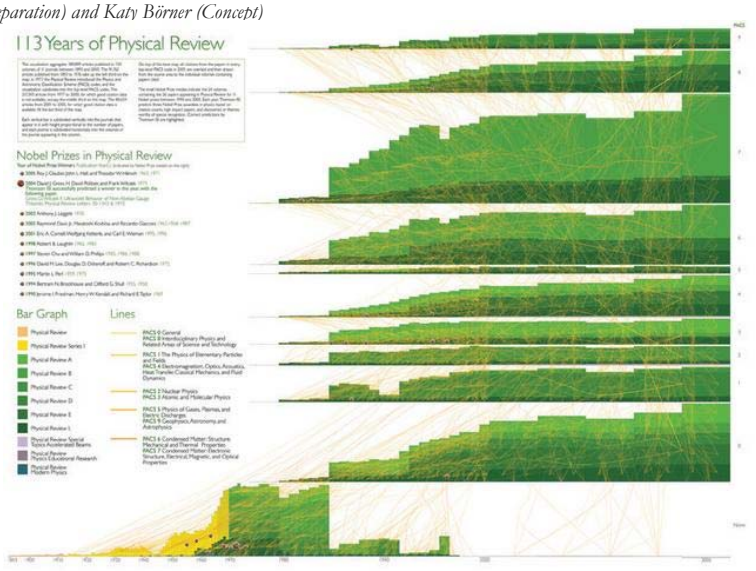
Corners show articles size coded according to
-article edit activity (top left),
- number of major edits (top right),
- number of bursts in edit activity (bottom, right)
- indegree (bottom left).



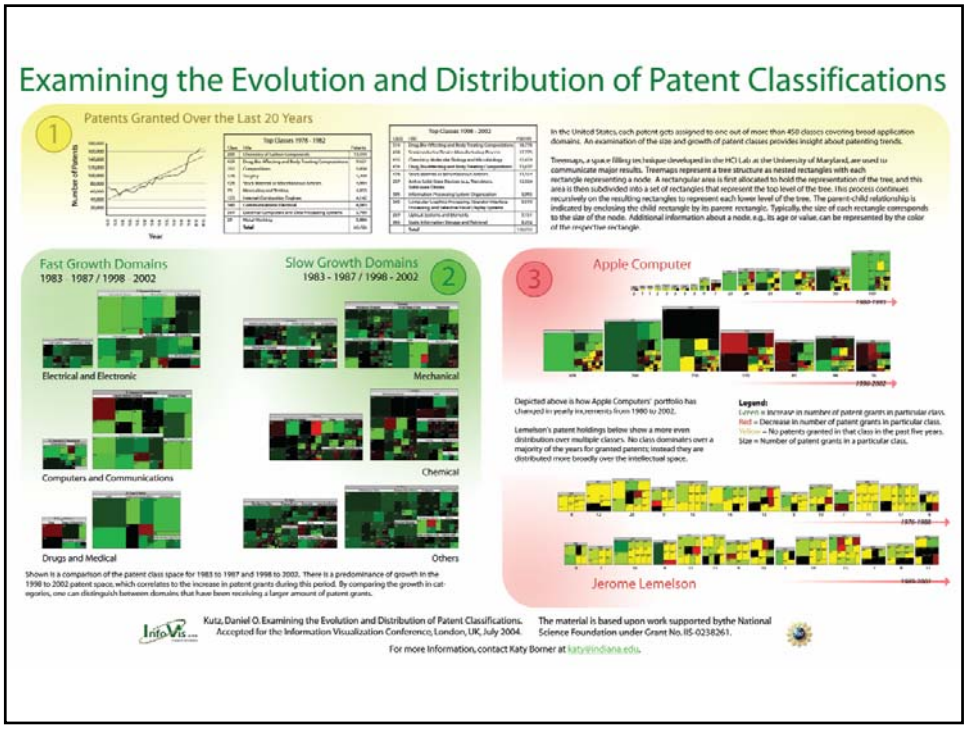
113 Years of Physical Review

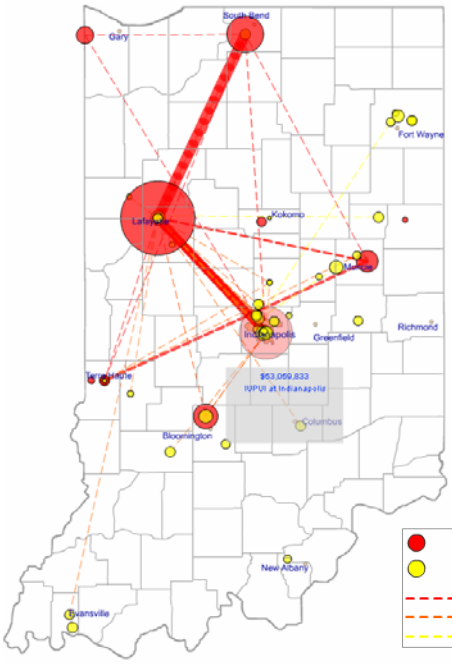
http://scimaps.org/dev/map_detail.php?map_id=171

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



… and INDUSTRY too.

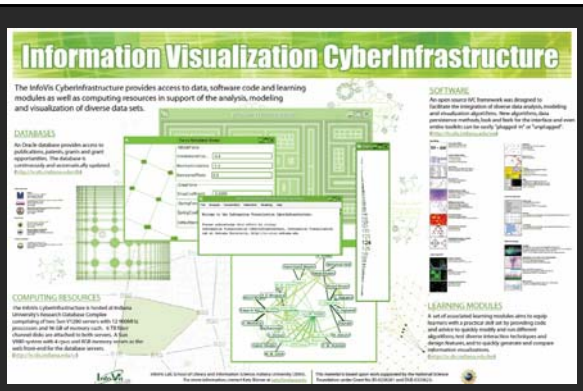




Mapping Indiana's Intellectual Space

Identify

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



CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, $451,000) Sept. 03-Aug. 08.
http://iv.slis.indiana.edu/

CIShell

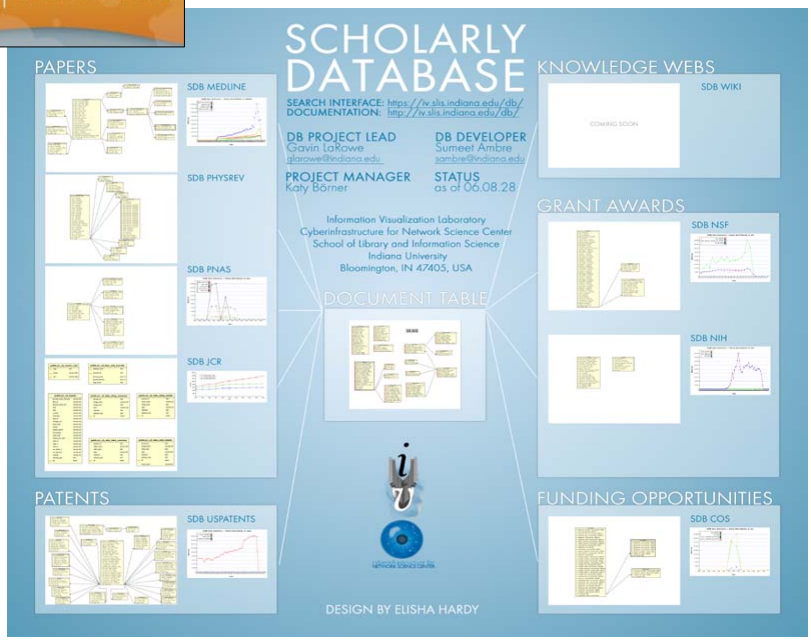
SDB | SCHOLARLY DATABASE

Scholarly Database
http://sdb.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-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), $1,120,926) Sept. 05 - Aug. 08. http://nwb.slis.indiana.edu

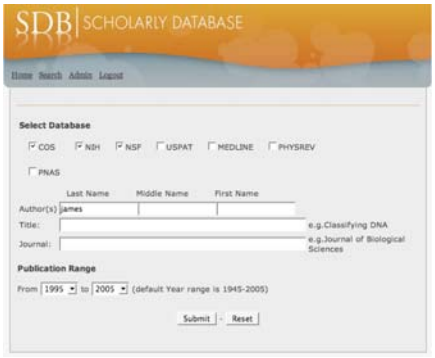
SDB | SCHOLARLY DATABASE



Scholarly Database: Web Interface

Search across publications, patents, grants.

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

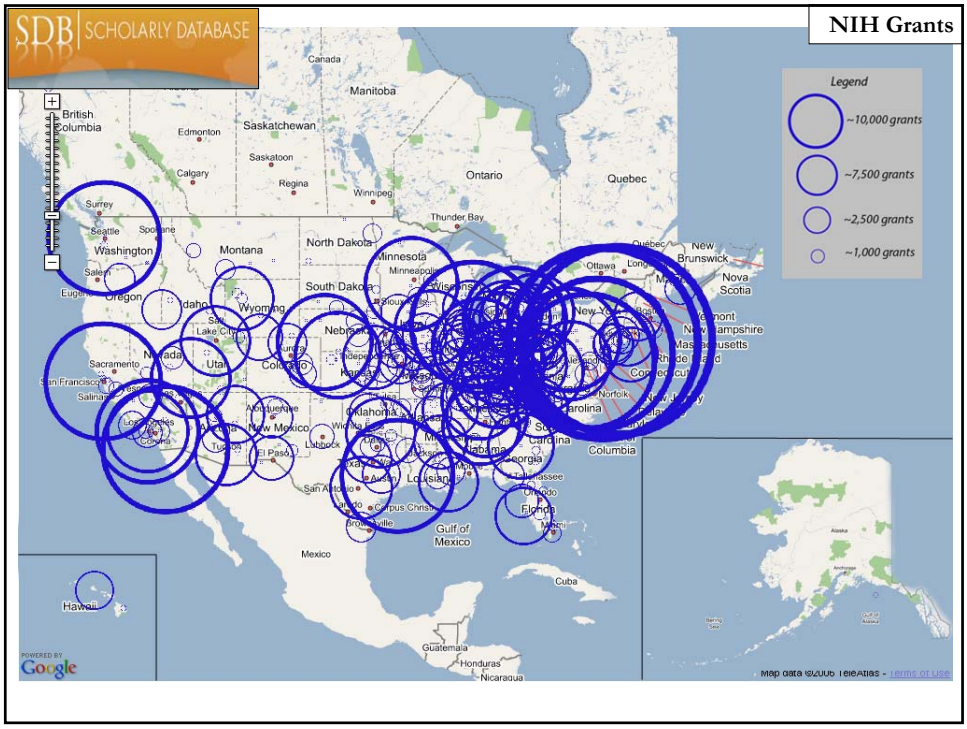


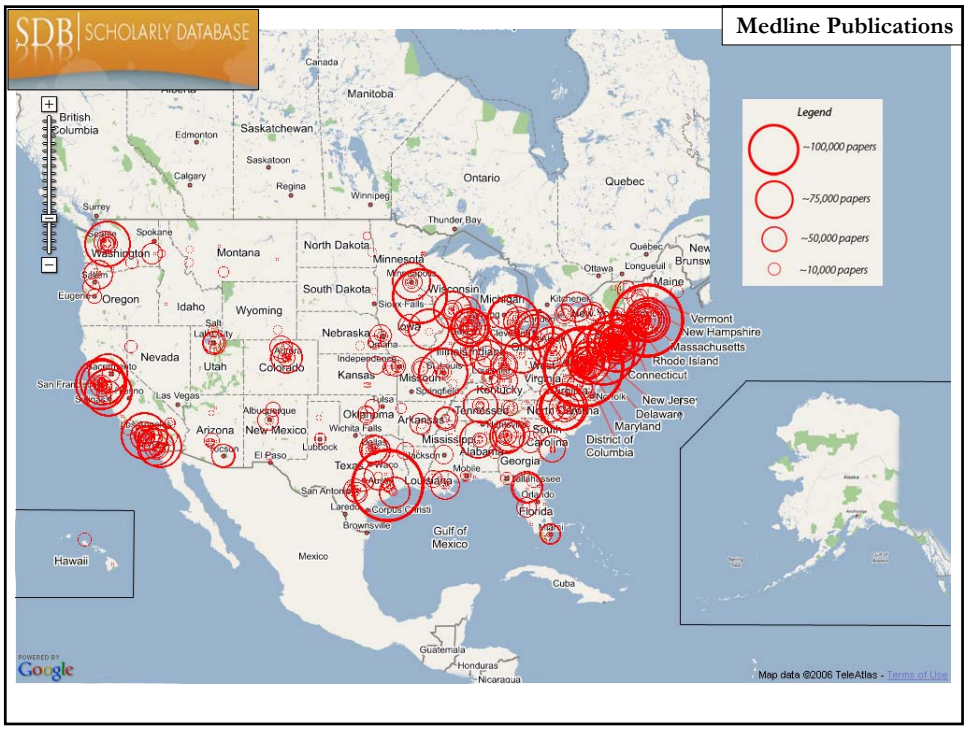
Scholarly Database: # Records & Years Covered

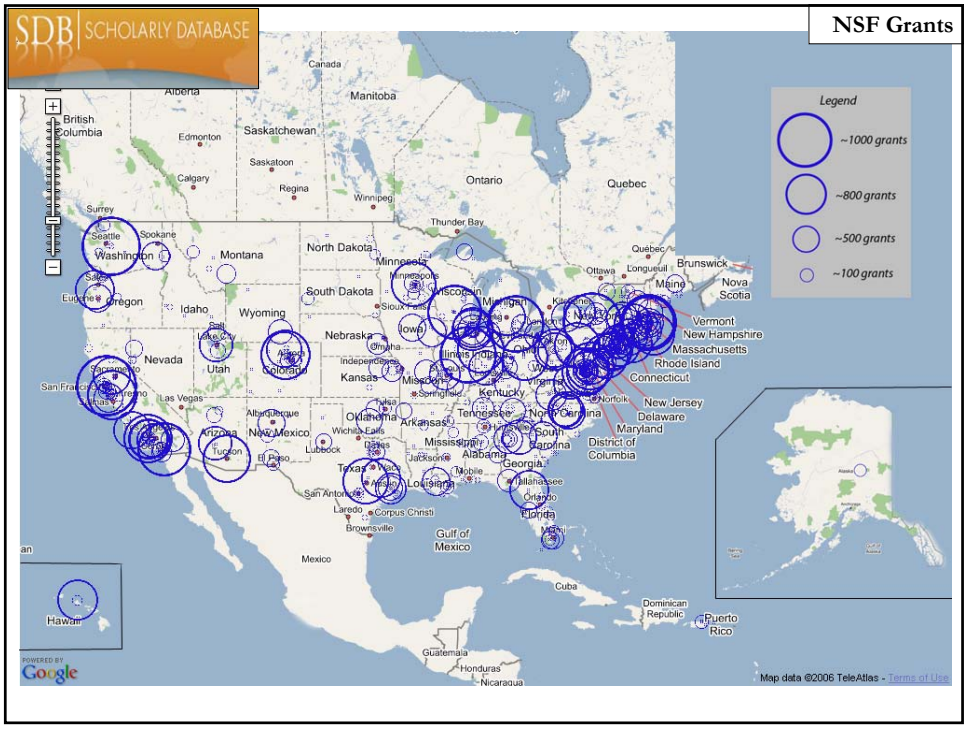
Datasets available via the Scholarly Database (* future feature)

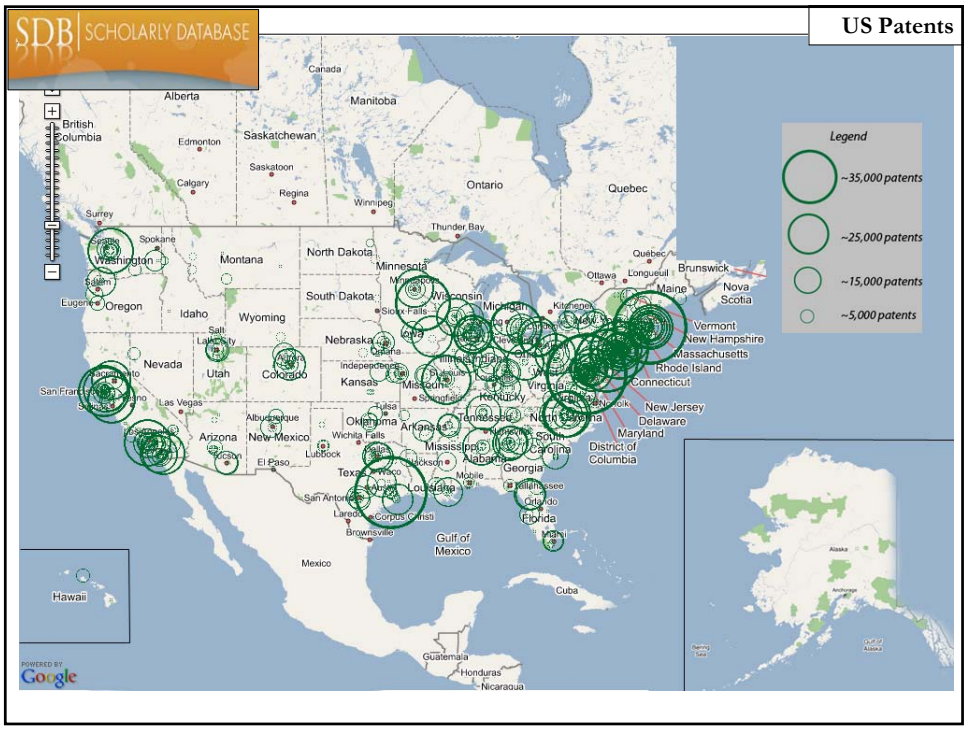
Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	13,149,741	1965-2005	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,179,930	1976-2004	Yes*	
NSF	174,835	1985-2003	Yes*	
NIH	1,043,804	1972-2002	Yes*	
Total	**18,021,560**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

Aim for comprehensive time, geospatial, and topic coverage.


NIH Grants
SDB | SCHOLARLY DATABASE
Legend
~10,000 grants
~7,500 grants
~2,500 grants
~1,000 grants
Canada
Manitoba
Saskatchewan
Ontario
Quebec
Montana
North Dakota
Minnesota
South Dakota
Wyoming
Nebraska
Oregon
Idaho
Washington
Colorado
Oklahoma
New Mexico
Texas
Louisiana
Florida
Georgia
Gulf of Mexico
Mexico
Cuba
Hawaii
Guatemala
Honduras
Nicaragua
Google


Medline Publications
SDB | SCHOLARLY DATABASE
Legend
~100,000 papers
~75,000 papers
~50,000 papers
~10,000 papers
Canada
Manitoba
Saskatchewan
Ontario
Quebec
Montana
North Dakota
Minnesota
South Dakota
Wyoming
Nebraska
Oregon
Idaho
Washington
Nevada
Utah
Colorado
Kansas
Oklahoma
Arizona
New Mexico
Texas
Louisiana
Florida
Georgia
Mississippi
Arkansas
Tennessee
Kentucky
Virginia
Maine
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
New Jersey
Delaware
Maryland
District of Columbia
Gulf of Mexico
Mexico
Cuba
Hawaii
Guatemala
Honduras
Nicaragua
Google
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
Oregon
Idaho
South Dakota
Wyoming
Nebraska
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
Nevada
Utah
Colorado
Kansas
Las Vegas
New Jersey
Delaware
Maryland
District of Columbia
Albuquerque
Oklahoma
Arkansas
Tennessee
Arizona
New Mexico
Wichita Falls
Lubbock
Mississippi
Alabama
Georgia
El Paso
Texas
Louisiana
San Antonio
Laredo
Corpus Christi
Brownsville
Florida
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
Oregon
Idaho
South Dakota
Wyoming
Nebraska
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
Nevada
Utah
Colorado
Kansas
Las Vegas
New Jersey
Delaware
Maryland
District of Columbia
Albuquerque
Oklahoma
Arkansas
Tennessee
Arizona
New Mexico
Wichita Falls
Lubbock
Mississippi
Alabama
Georgia
El Paso
Texas
Louisiana
San Antonio
Laredo
Corpus Christi
Brownsville
Florida
Gulf of Mexico
Mexico
Cuba
Hawaii
Guatemala
Honduras
Nicaragua
Google

Building Market Places not Cathedrals

- 'Software glue' has to interlink datasets and algorithms written in different languages using different data formats.
- The smaller the glue or 'CI Shell', the more likely it can be maintained.

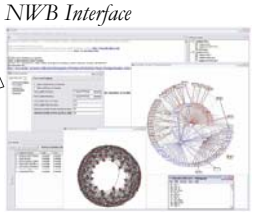
CIShell – Serving Non-CS Algorithm Developers & Users

CIShell Wizards

CIShell

IVC Interface

NWB Interface



CIShell – Build on OSGi Industry Standard

CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework.

OSGi (http://www.osgi.org) is

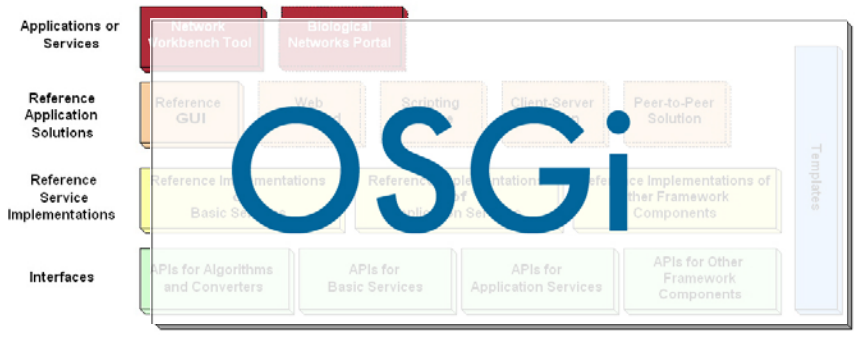
- A standardized, component oriented, computing environment for networked services.
- Successfully used in the industry from high-end servers to embedded mobile devices since 7 years.
- Alliance members include IBM (Eclipse), Sun, Intel, Oracle, Motorola, NEC and many others.
- Widely adopted in open source realm, especially since Eclipse 3.0 that uses OSGi R4 for its plugin model.

Advantages of Using OSGi

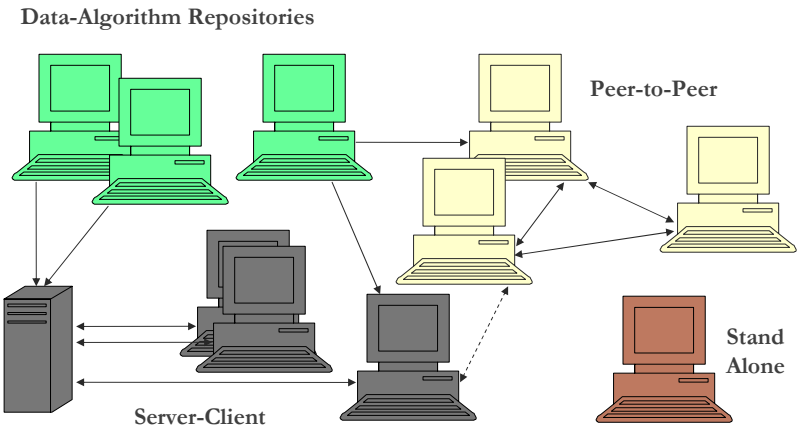
- Any CIShell algorithm is a service that can be used in any OSGi-framework based system.
- Using OSGi, running CIShells/tools can connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

Ideally, CIShell becomes a standard for creating OSGi Services for algorithms.

CIShell – Layer Cake



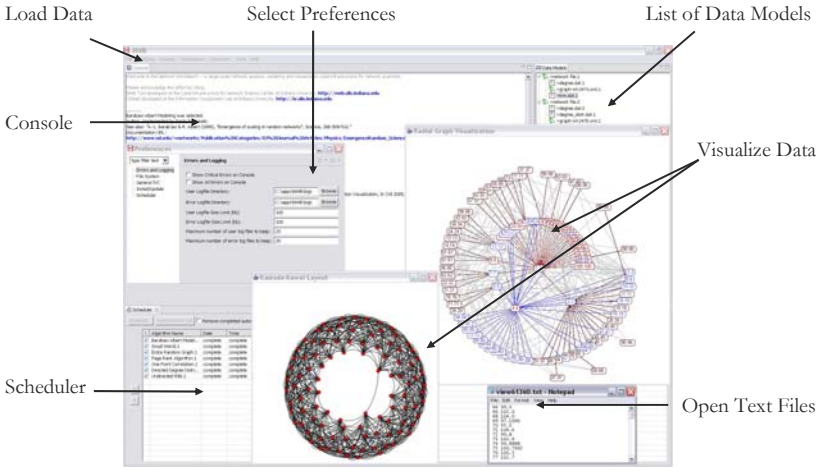
CIShell – Deployment



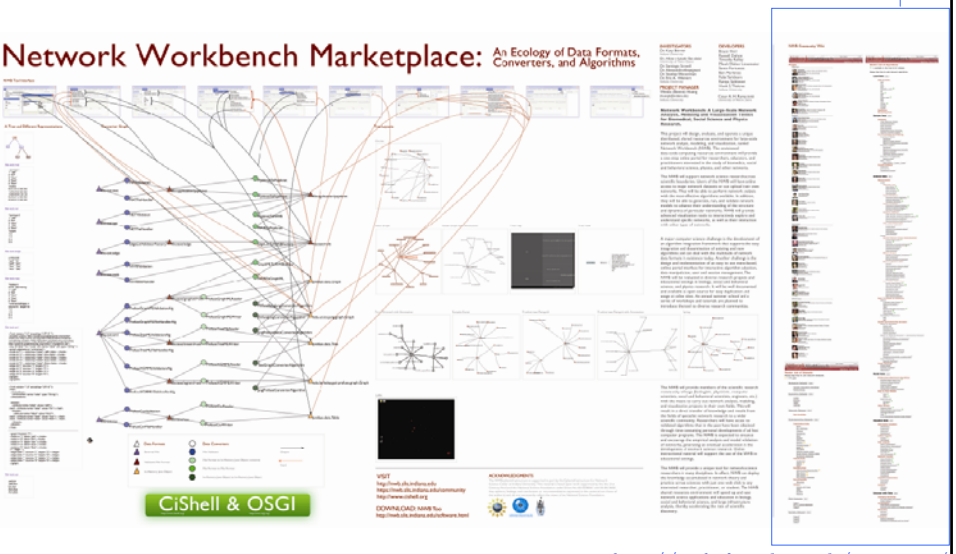
CIShell applications can be deployed as distributed data and algorithm repositories, stand alone applications, peer-to-peer architectures, and server-client architectures.

NWB Tool: Interface Elements

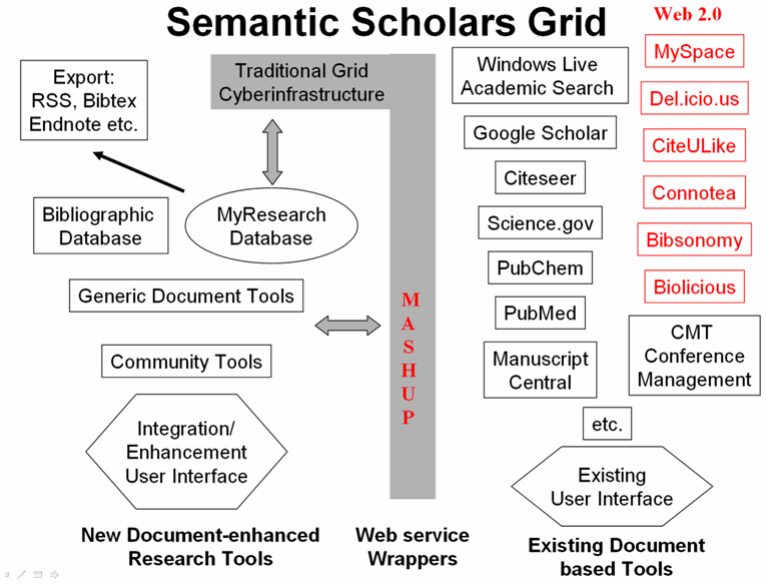
http://nwb.slis.indiana.edu



NWB
Community Wiki



https://nwb.slis.indiana.edu/community/



Slide with regards from Geoffrey Fox, IUB

The End.