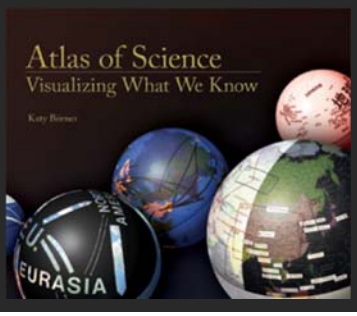
Mining, Mapping, and Accelerating Scholarly Networks

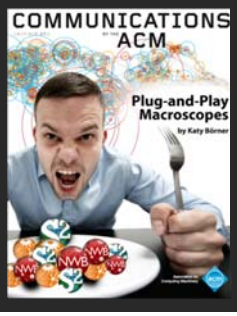
Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

With special thanks to the members at the Cyberinfrastructure for Network Science Center; the Sci2, NWB, and EpiC teams; and the VIVO Collaboration

Shared Horizons: Data, Biomedicine, and the Digital Humanities
Maryland Institute for Technology.

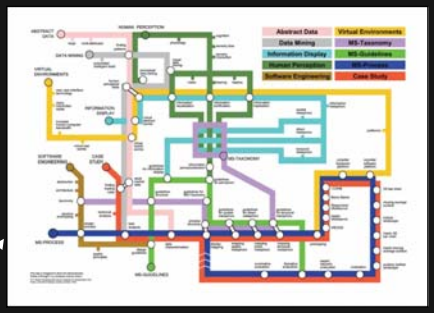
April 11, 2013





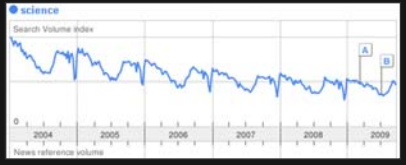
Terra bytes of data

Descriptive & Predictive Models



Find your way

Find collaborators, friends



Identify trends

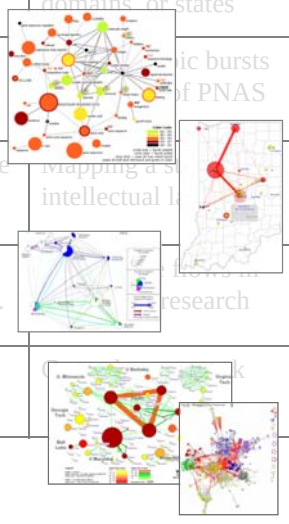
Type of Analysis vs. Level of Analysis

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

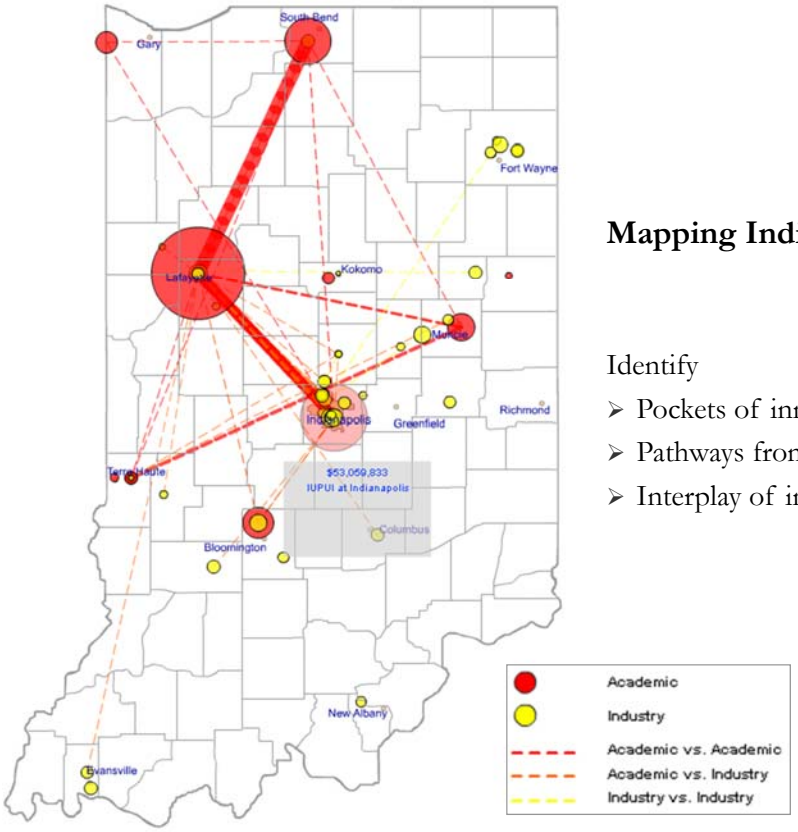
cyberinfrastructure for NETWORK SCIENCE CENTER
School of Library and Information Science | Indiana University Bloomington

Type of Analysis vs. Level of Analysis

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



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



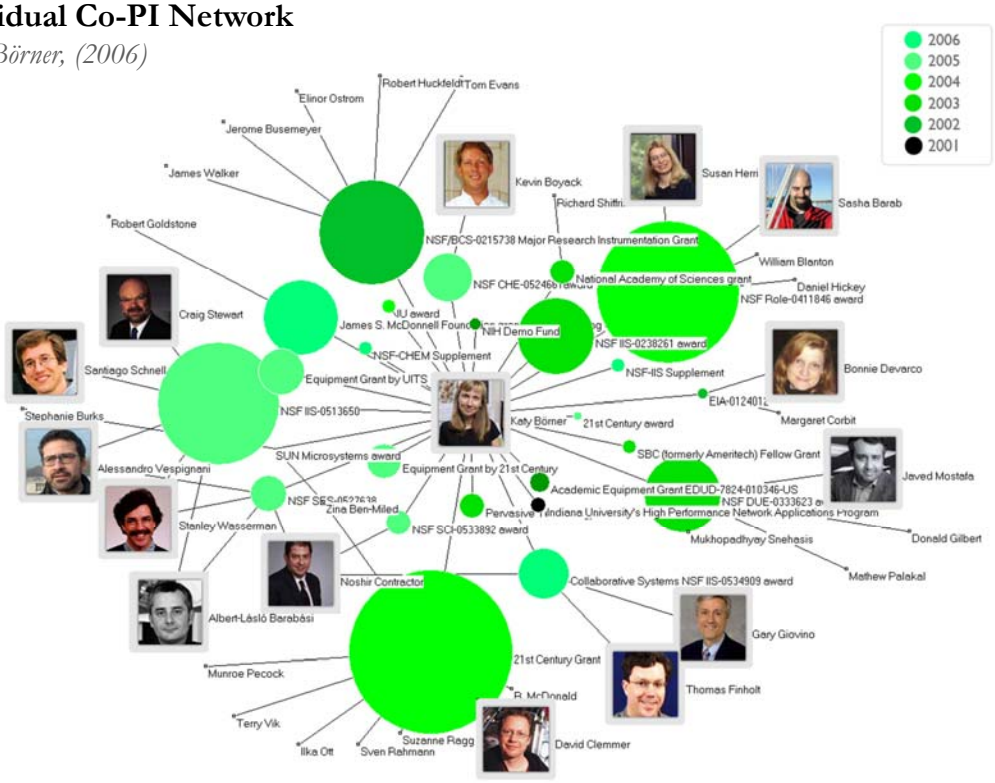
Mapping Indiana's Intellectual Space

Identify

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

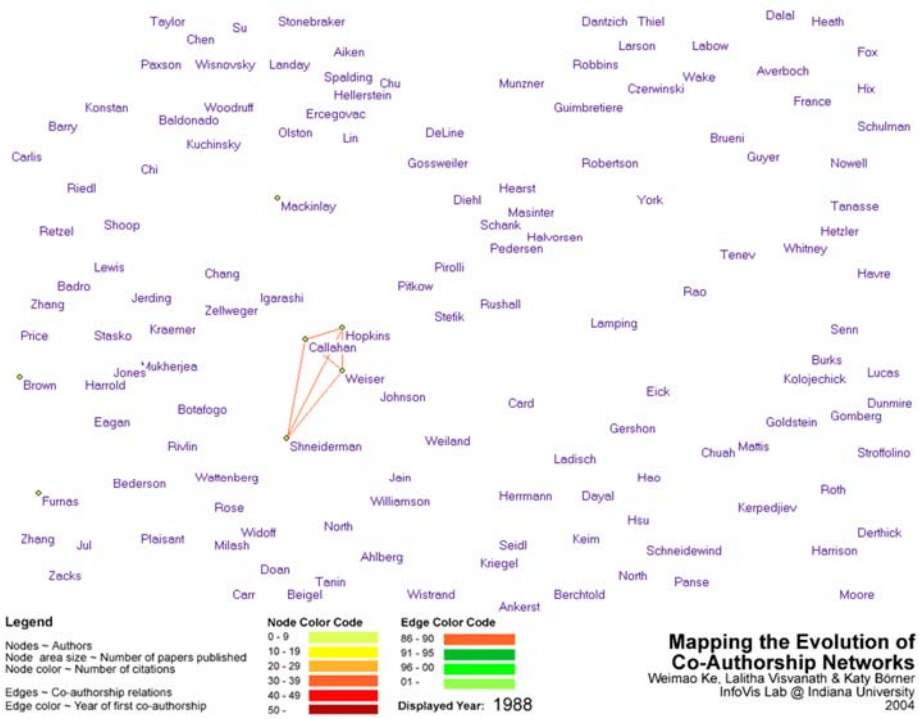
Individual Co-PI Network

Ke & Börner, (2006)



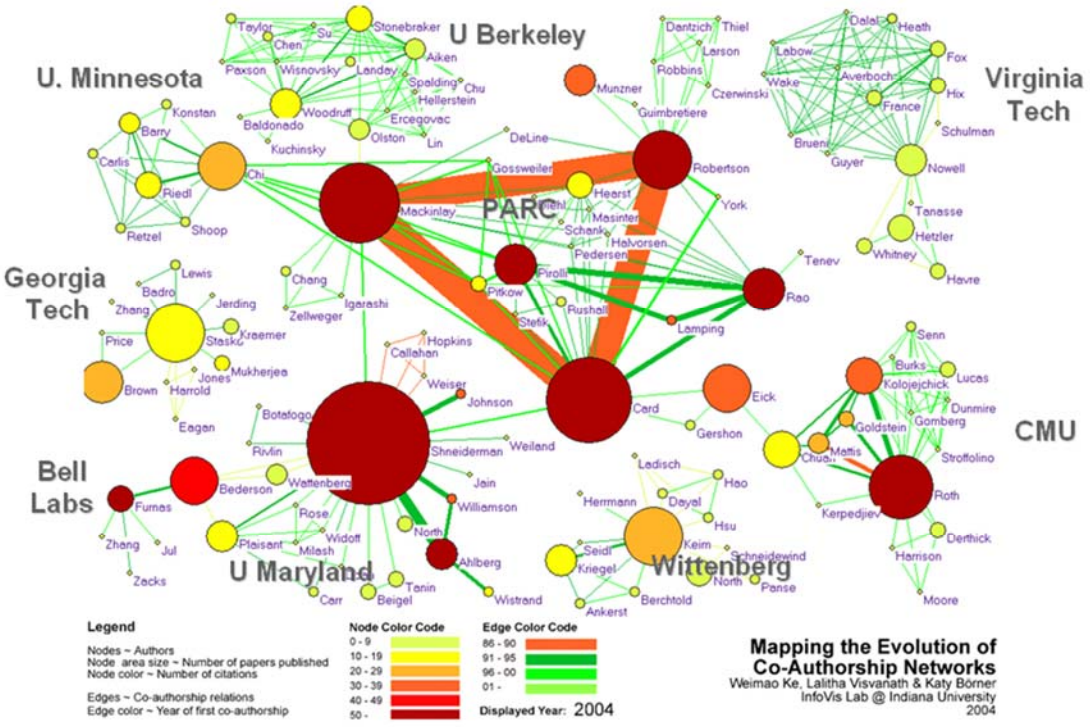
Mapping the Evolution of Co-Authorship Networks

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



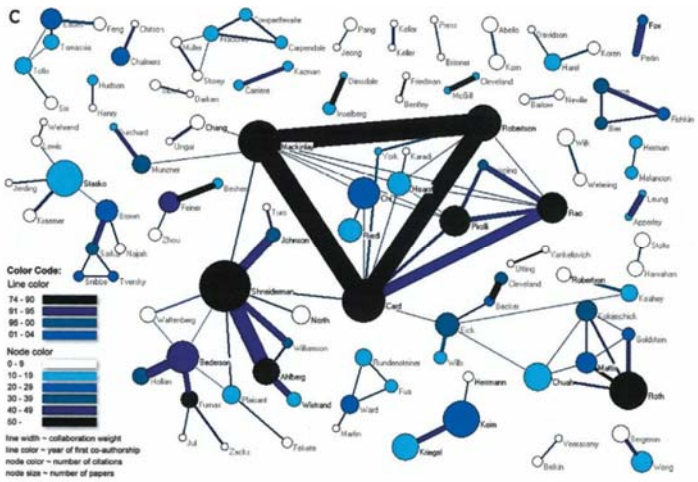
Mapping the Evolution of Co-Authorship Networks

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:

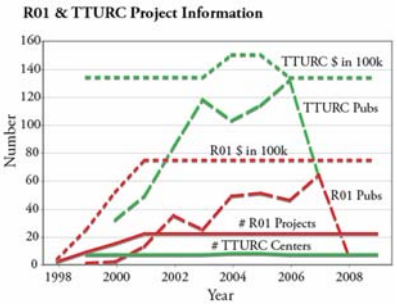
- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size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

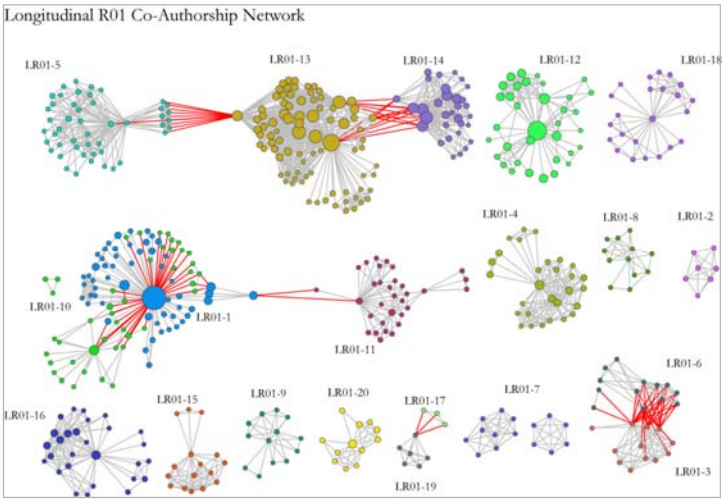
Mapping Transdisciplinary Tobacco Use Research Centers Publications

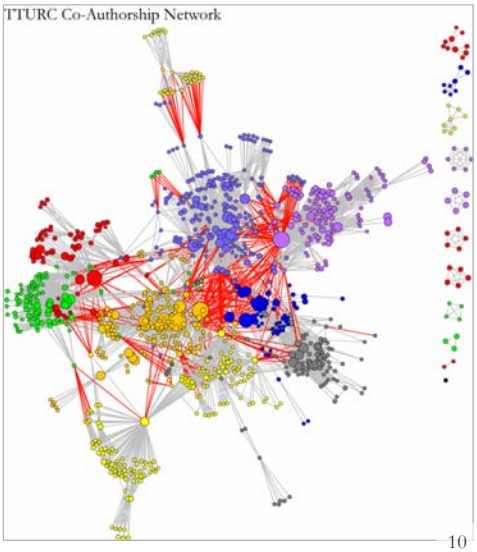
Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812



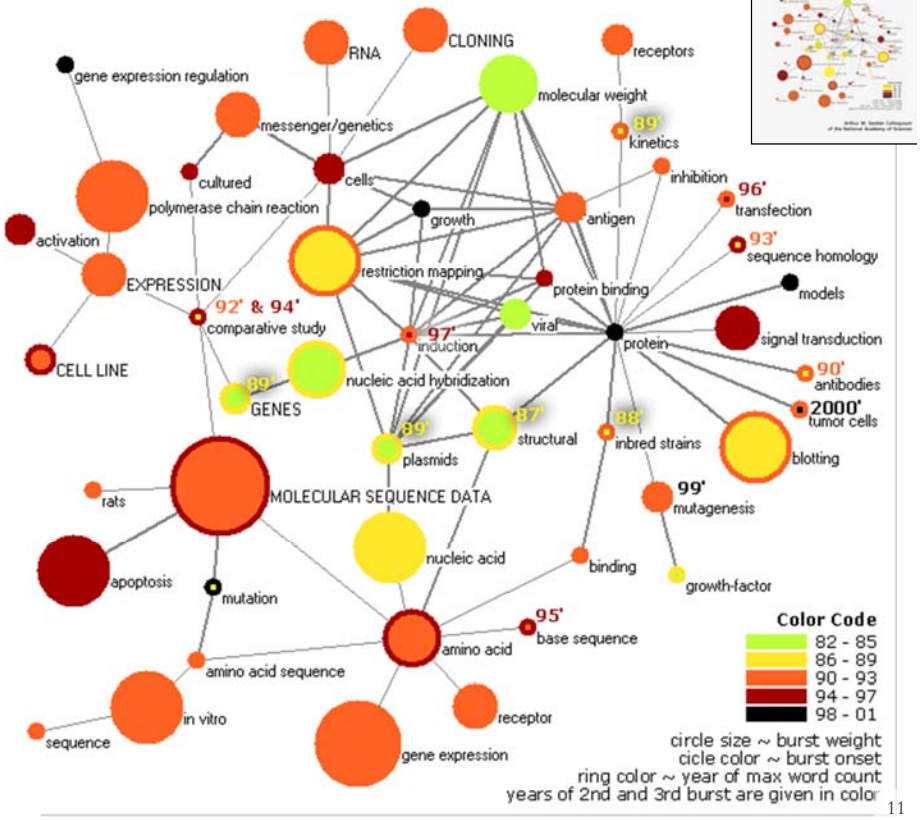




Mapping Topic Bursts

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.



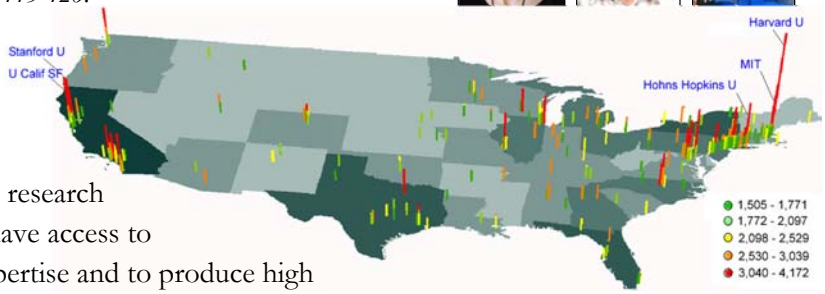
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

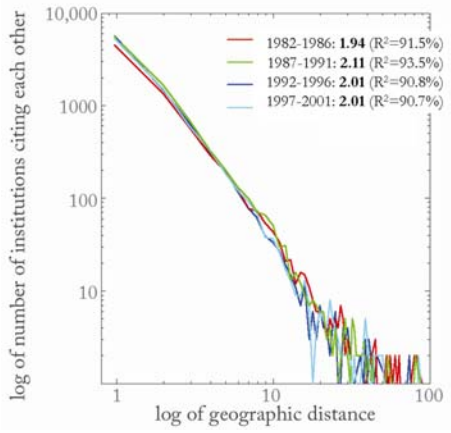


Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

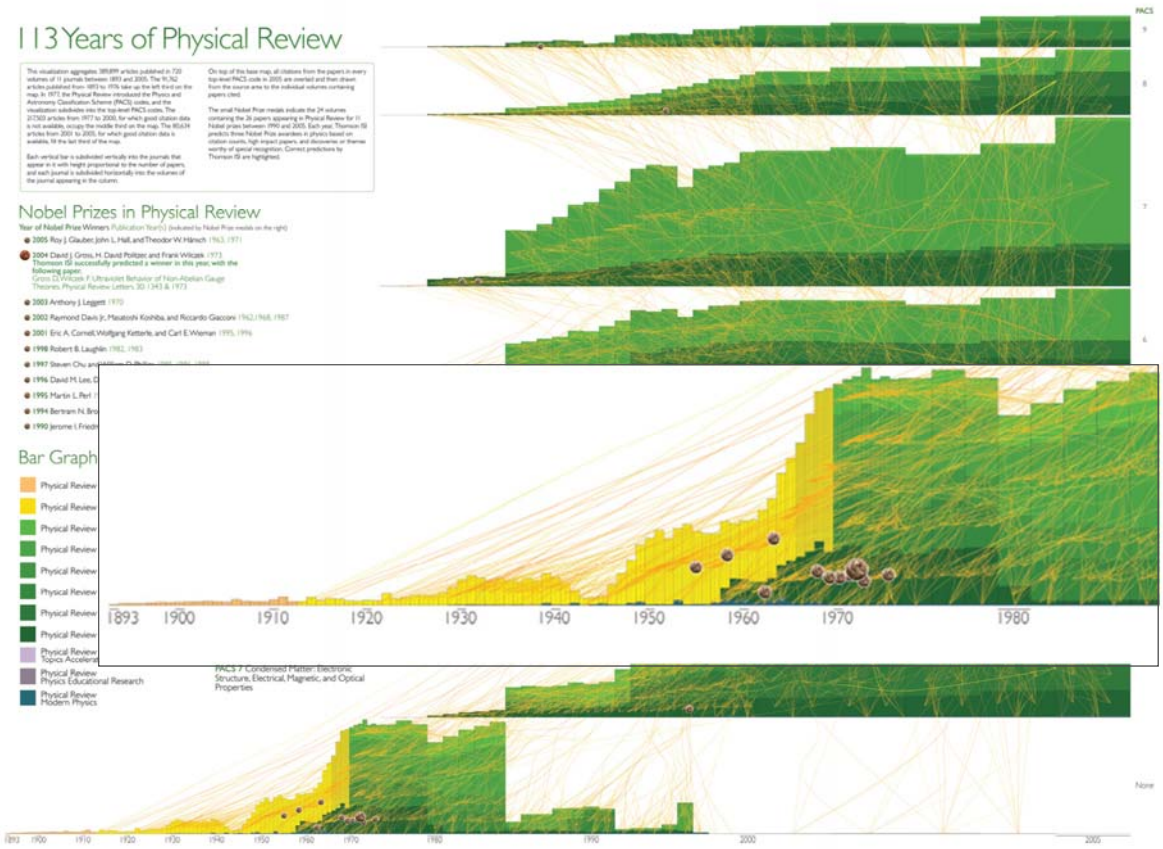
1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links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





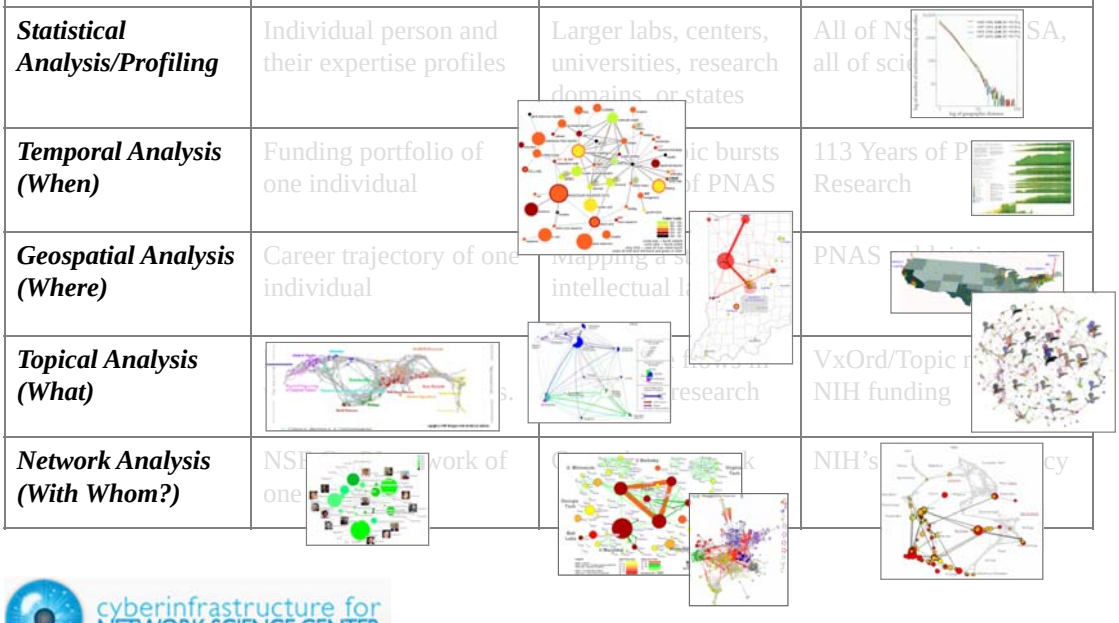
Contributions:

- 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.



Type of Analysis vs. Level of Analysis

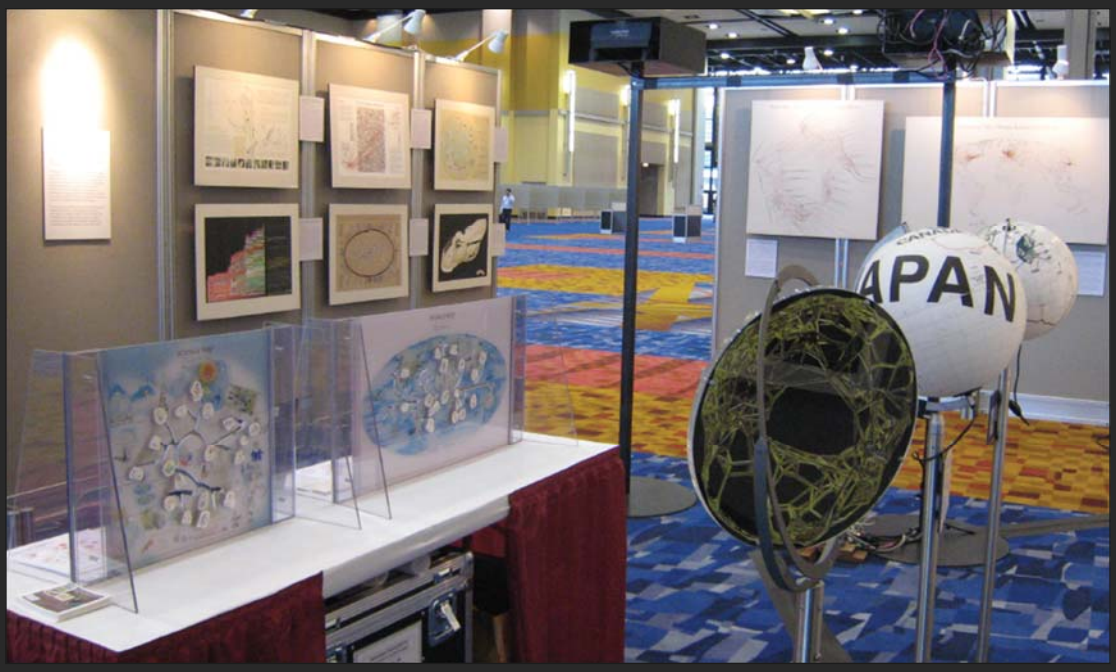
	Micro/Individual (1-100 records)	*Meso/Local (101–10,000 records)*	*Macro/Global (10,000 < records)*
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NS SA, all of sci
Temporal Analysis (When)	Funding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



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

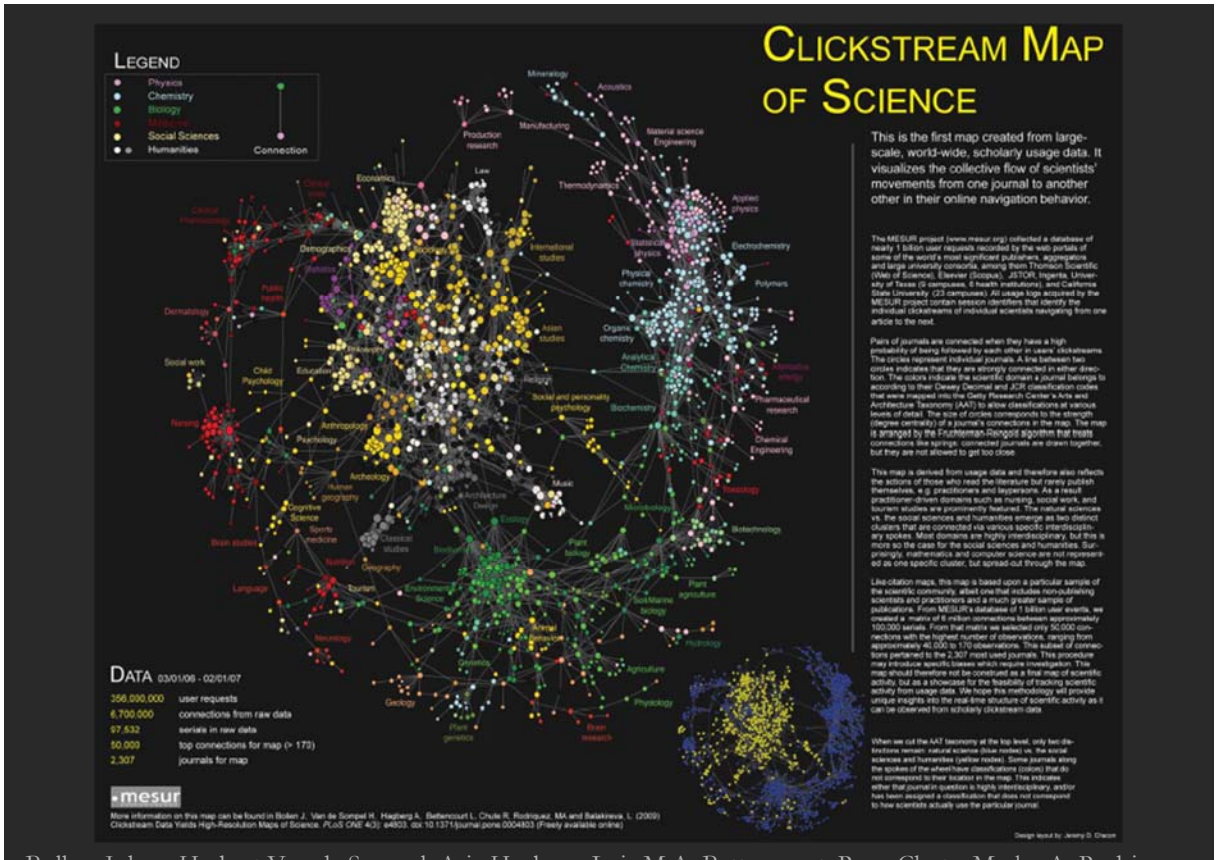
Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org/

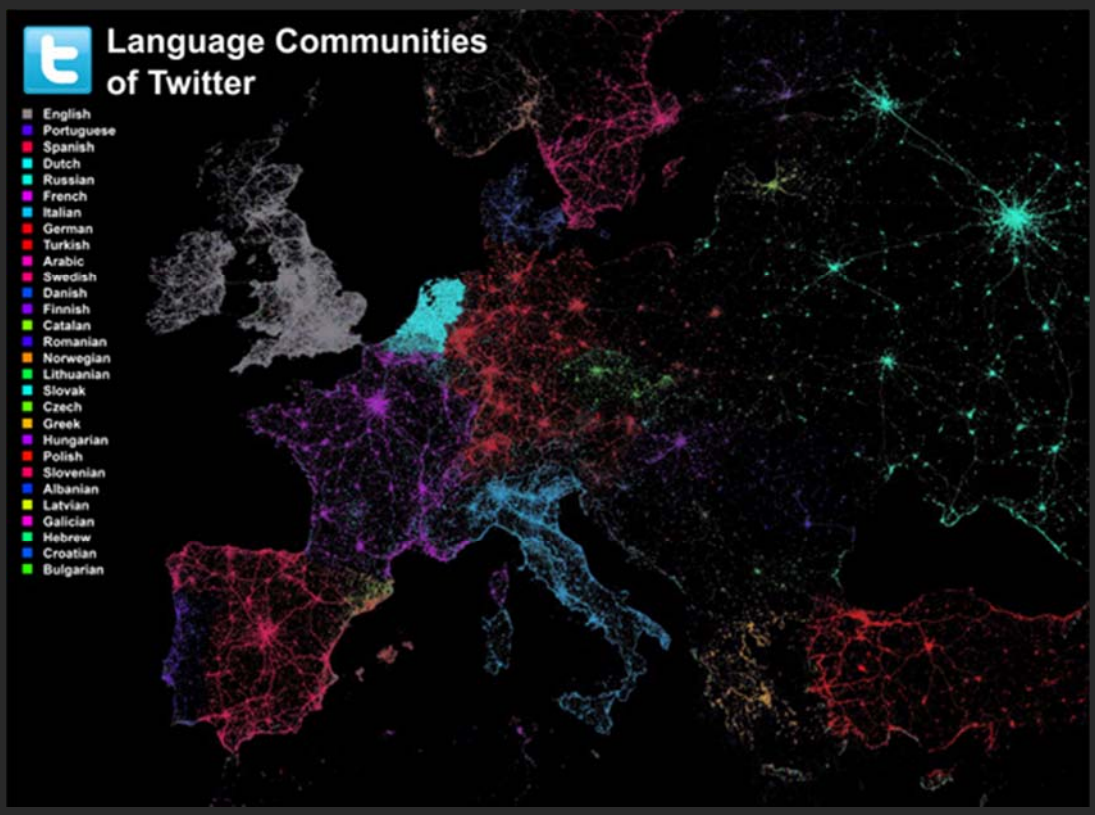


Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, http://mediax.stanford.edu, http://scaleindependentthought.typepad.com/photos/scimaps

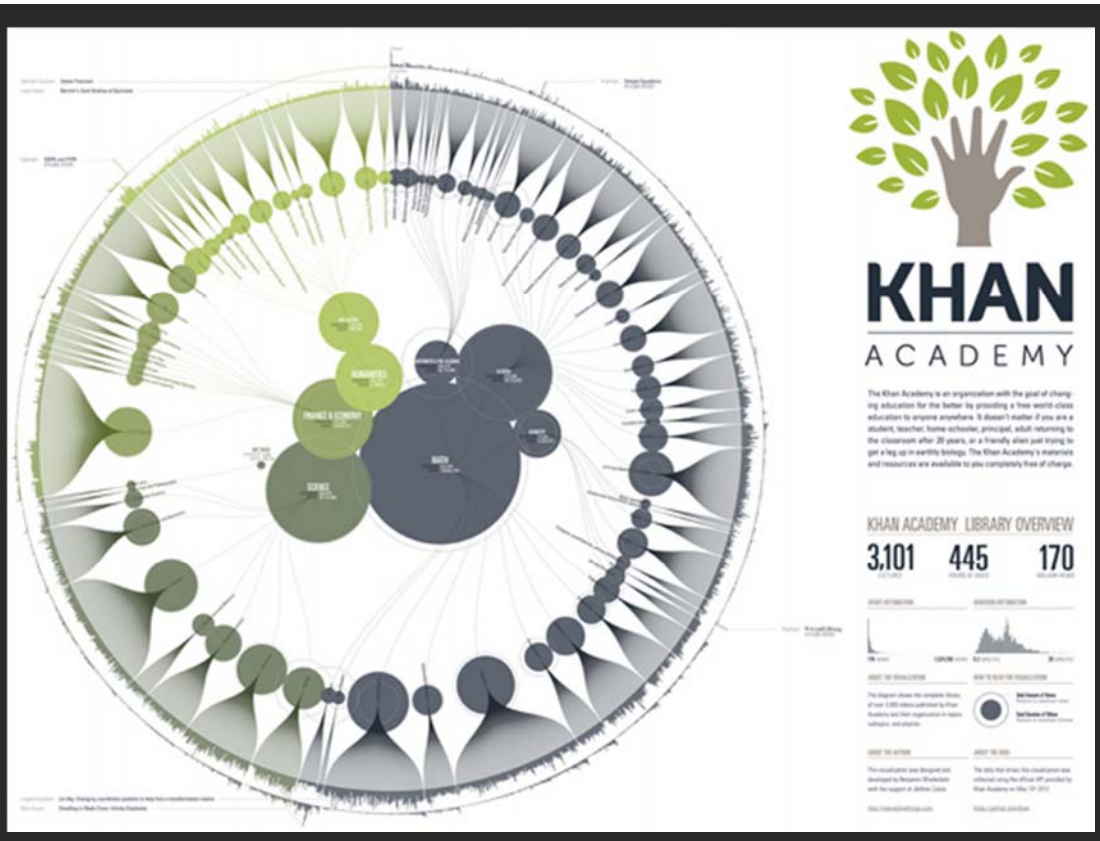
Places & Spaces *Digital Display in North Carolina State's brand new Immersion Theater*



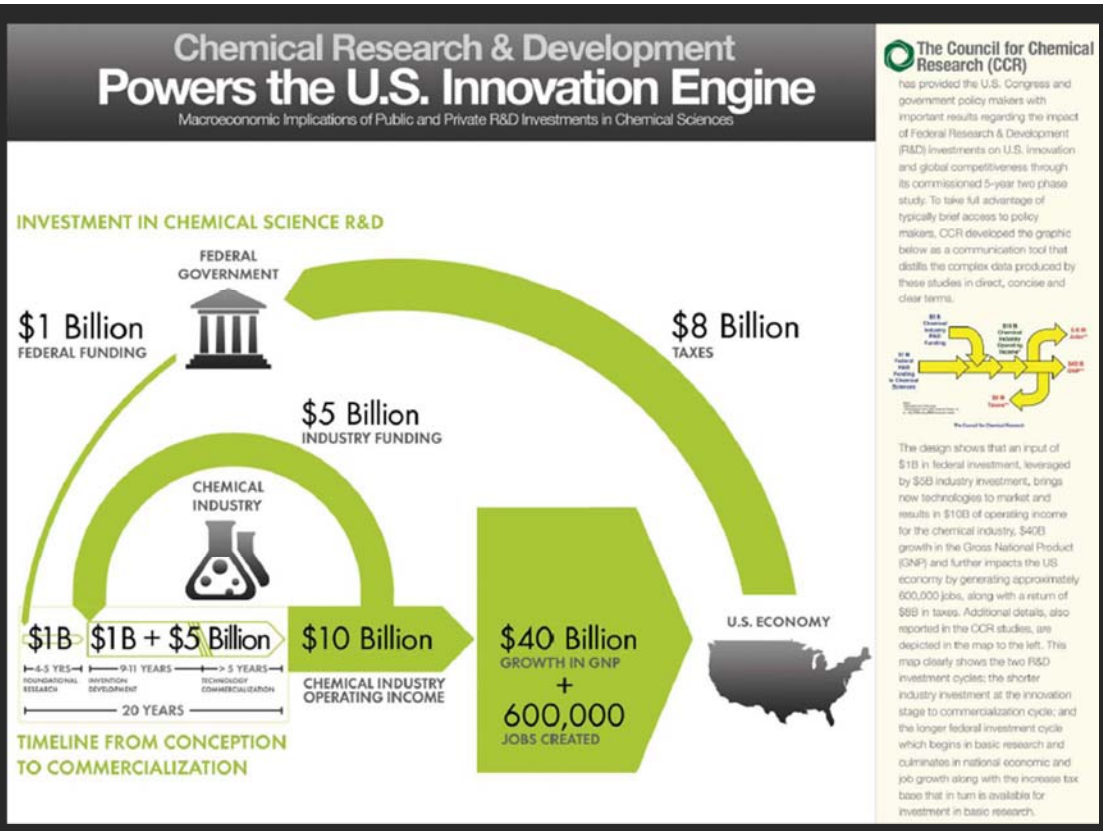
Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.



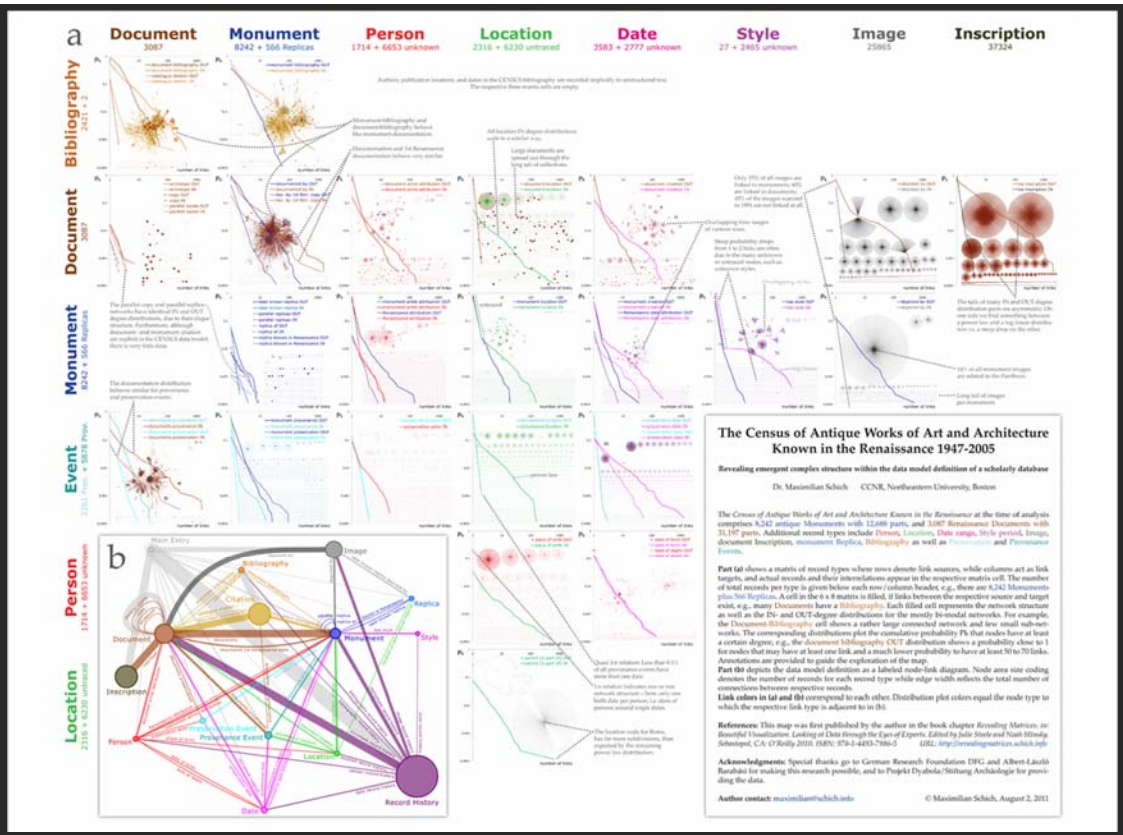
Eric Fischer. 2012. Language Communities of Twitter.



Benjamin Wiederkehr and Jérôme Cukier. 2012. Khan Academy Library Overview.

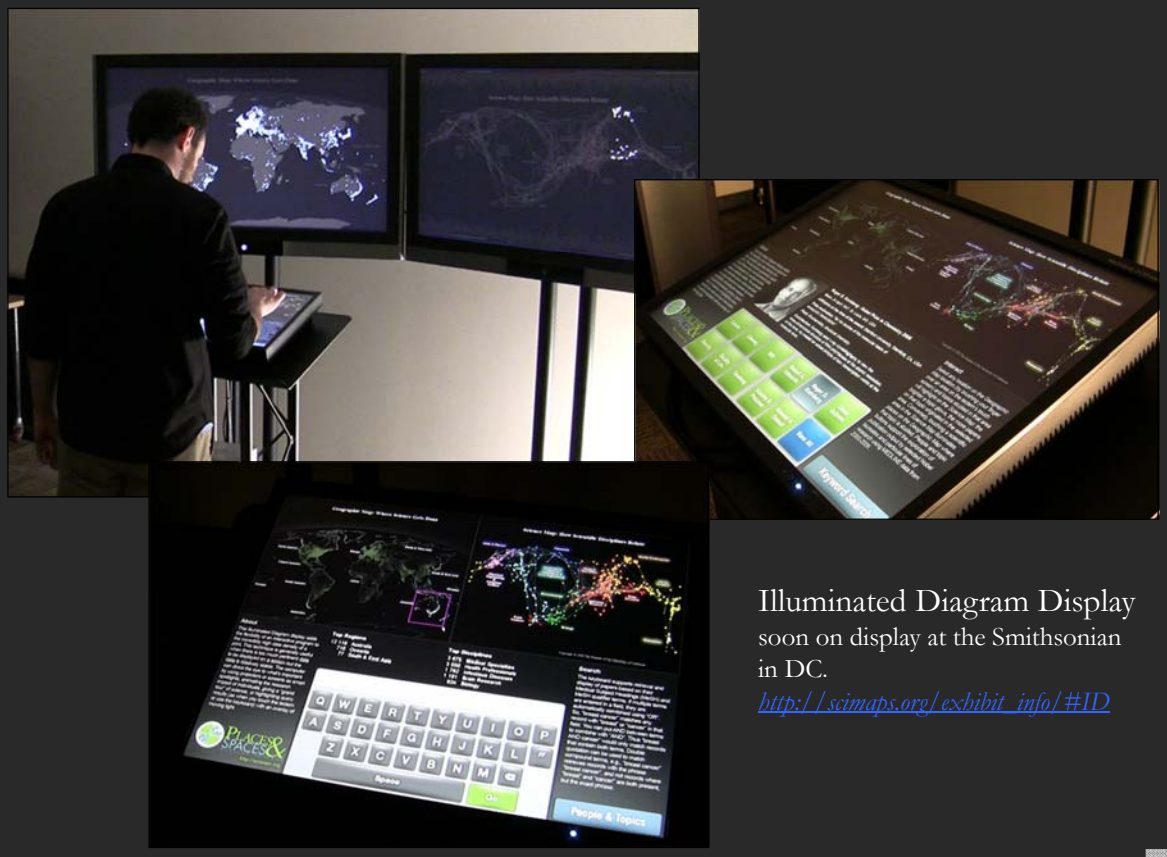


Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research.

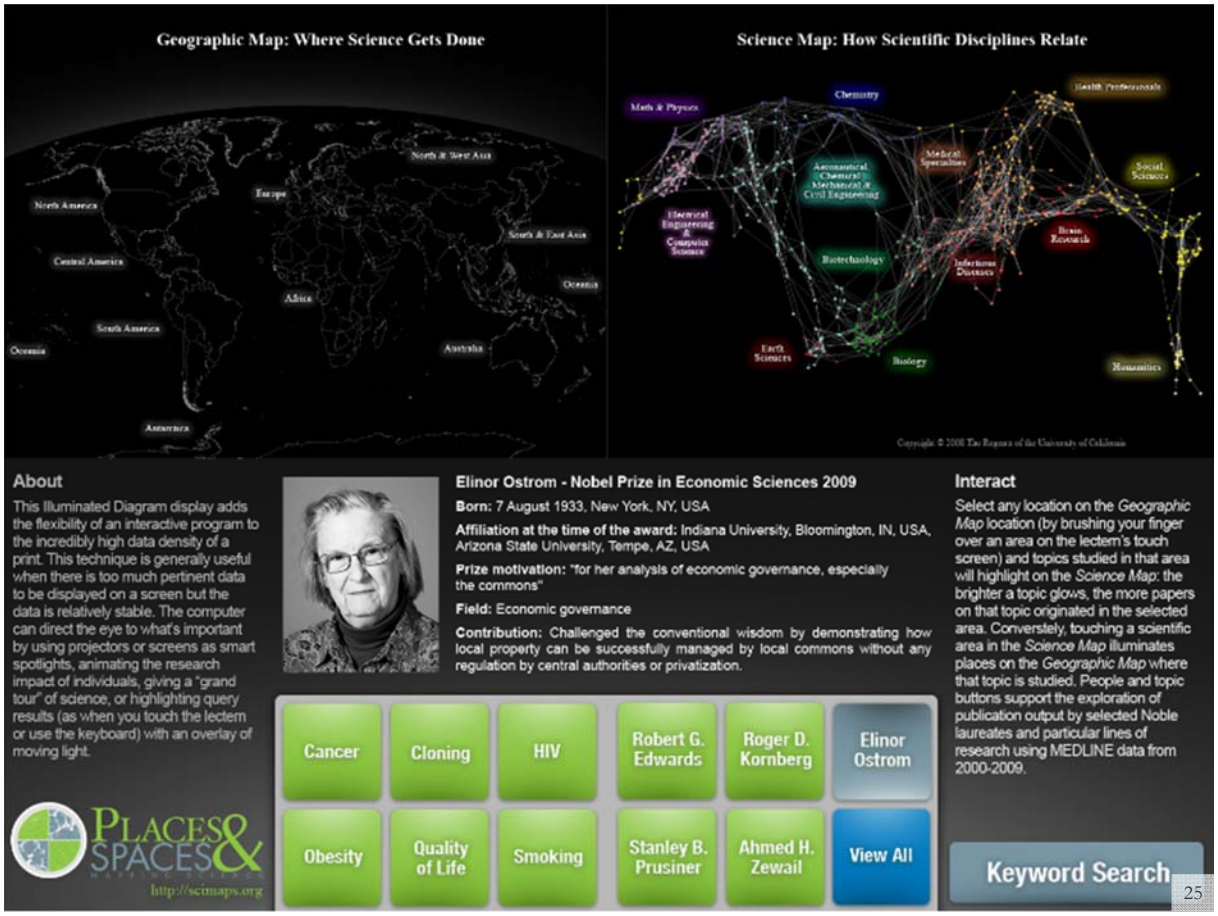


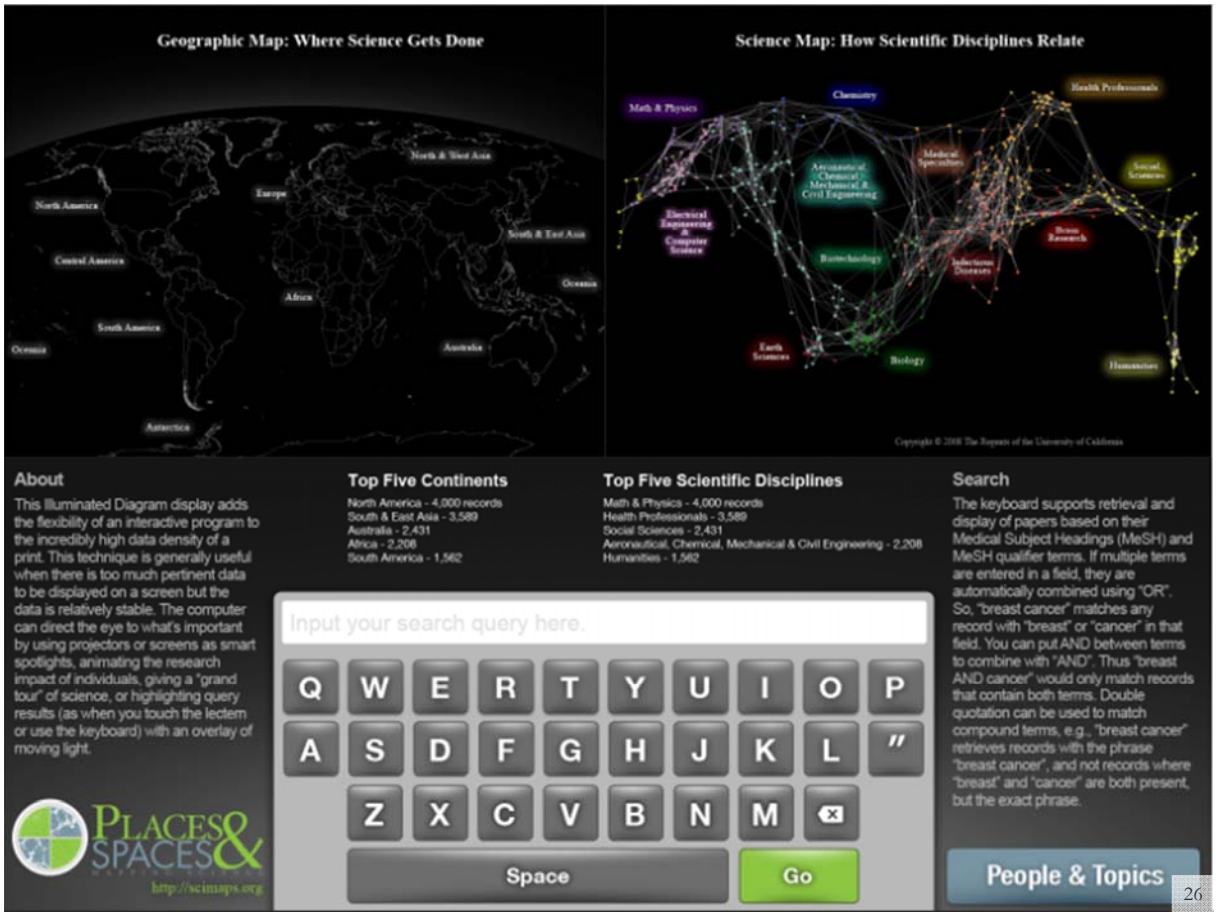
Maximilian Schich. 2011. The Census of Antique Works of Art and Architecture Known in the Renaissance, 1947-2005.

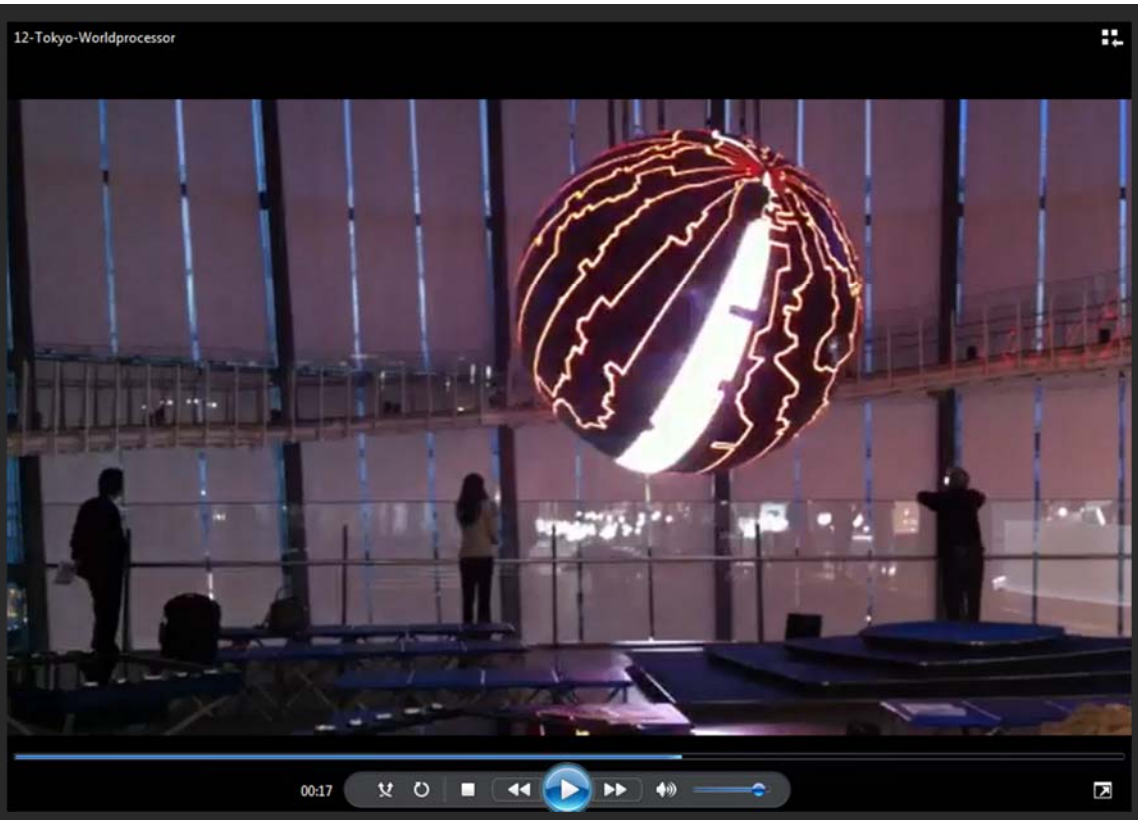
Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months 12 coaches, 300 m long
Opening was on April 23rd, 2009 by German Chancellor Merkel
http://www.expedition-zukunft.de



Illuminated Diagram Display
soon on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID


Geographic Map: Where Science Gets Done
North America
Europe
North & West Asia
South & East Asia
Central America
Africa
Oceania
South America
Australia
Antarctica
Science Map: How Scientific Disciplines Relate
Math & Physics
Chemistry
Health Professionals
Medical Specialties
Social Sciences
Aeronautical, Chemical, Mechanical & Civil Engineering
Electrical Engineering & Computer Science
Brain Research
Biotechnology
Infectious Diseases
Earth Sciences
Biology
Humanities
Copyright © 2008 The Regents of the University of California
About
This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.
PLACES & SPACES
MAPPING SCIENCE
http://scimaps.org
Elinor Ostrom - Nobel Prize in Economic Sciences 2009
Born: 7 August 1933, New York, NY, USA
Affiliation at the time of the award: Indiana University, Bloomington, IN, USA, Arizona State University, Tempe, AZ, USA
Prize motivation: "for her analysis of economic governance, especially the commons"
Field: Economic governance
Contribution: Challenged the conventional wisdom by demonstrating how local property can be successfully managed by local commons without any regulation by central authorities or privatization.
Cancer
Cloning
HIV
Robert G. Edwards
Roger D. Kornberg
Elinor Ostrom
Obesity
Quality of Life
Smoking
Stanley B. Prusiner
Ahmed H. Zewail
View All
Interact
Select any location on the Geographic Map location (by brushing your finger over an area on the lectern's touch screen) and topics studied in that area will highlight on the Science Map: the brighter a topic glows, the more papers on that topic originated in the selected area. Conversely, touching a scientific area in the Science Map illuminates places on the Geographic Map where that topic is studied. People and topic buttons support the exploration of publication output by selected Noble laureates and particular lines of research using MEDLINE data from 2000-2009.
Keyword Search


Geographic Map: Where Science Gets Done
North America
Europe
North & West Asia
South & East Asia
Central America
Africa
Oceania
South America
Australia
Antarctica
Science Map: How Scientific Disciplines Relate
Math & Physics
Chemistry
Health Professionals
Medical Specialties
Social Sciences
Aeronautical, Chemical, Mechanical & Civil Engineering
Electrical Engineering & Computer Science
Brain Research
Biotechnology
Infectious Diseases
Earth Sciences
Biology
Humanities
Copyright © 2008 The Regents of the University of California
About
This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.
PLACES & SPACES
MAPPING SCIENCE
http://scimaps.org
Top Five Continents
North America - 4,000 records
South & East Asia - 3,589
Australia - 2,431
Africa - 2,208
South America - 1,562
Top Five Scientific Disciplines
Math & Physics - 4,000 records
Health Professionals - 3,589
Social Sciences - 2,431
Aeronautical, Chemical, Mechanical & Civil Engineering - 2,208
Humanities - 1,562
Input your search query here.
Q W E R T Y U I O P
A S D F G H J K L "
Z X C V B N M
Space
Go
Search
The keyboard supports retrieval and display of papers based on their Medical Subject Headings (MeSH) and MeSH qualifier terms. If multiple terms are entered in a field, they are automatically combined using "OR". So, "breast cancer" matches any record with "breast" or "cancer" in that field. You can put AND between terms to combine with "AND". Thus "breast AND cancer" would only match records that contain both terms. Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer", and not records where "breast" and "cancer" are both present, but the exact phrase.
People & Topics



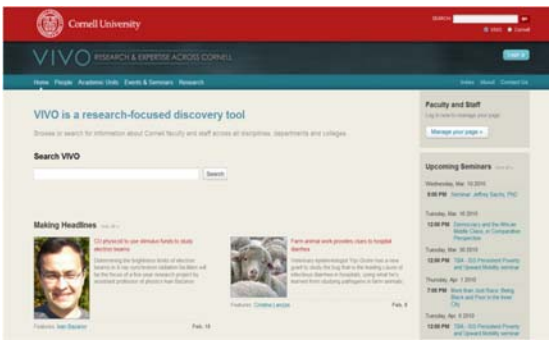
Ingo Gunther's Worldprocessor globe design now on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan

VIVO International Researcher Network

VIVO
Enabling a National Network of Scientists

VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

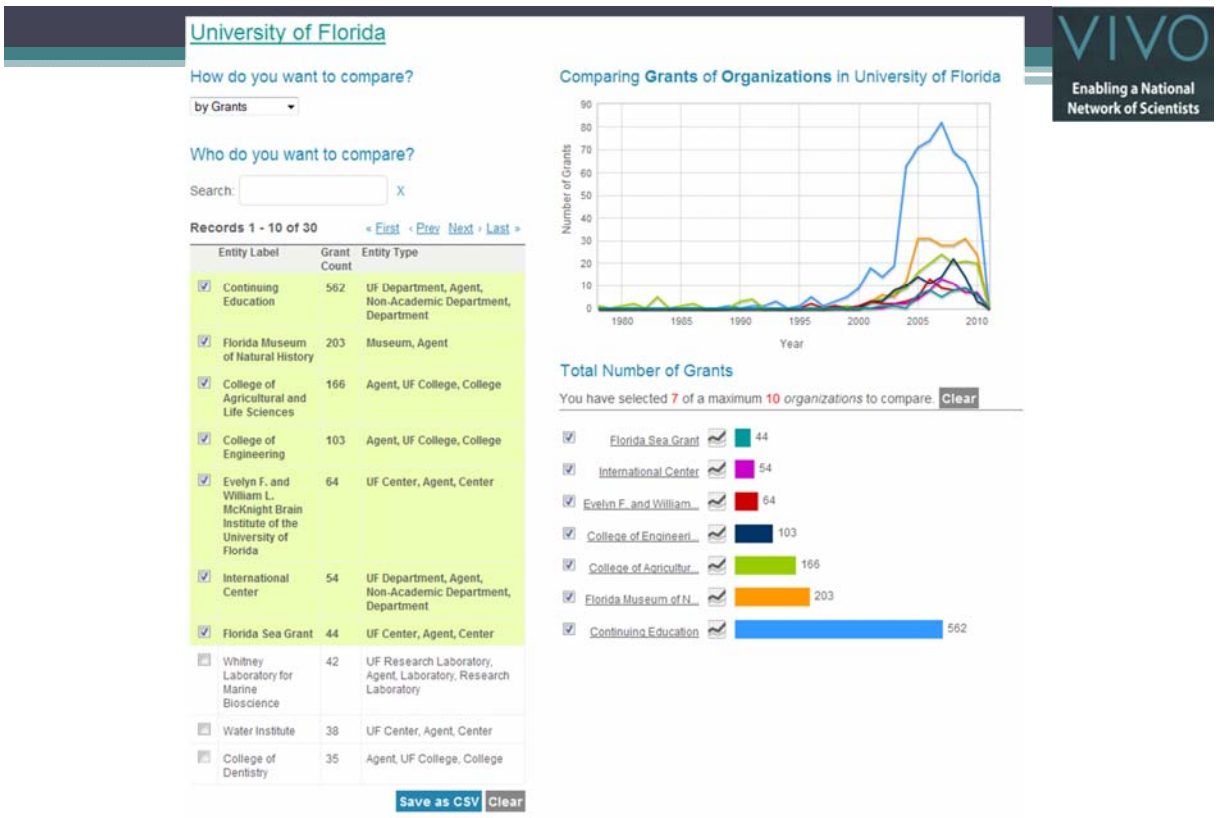
- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.



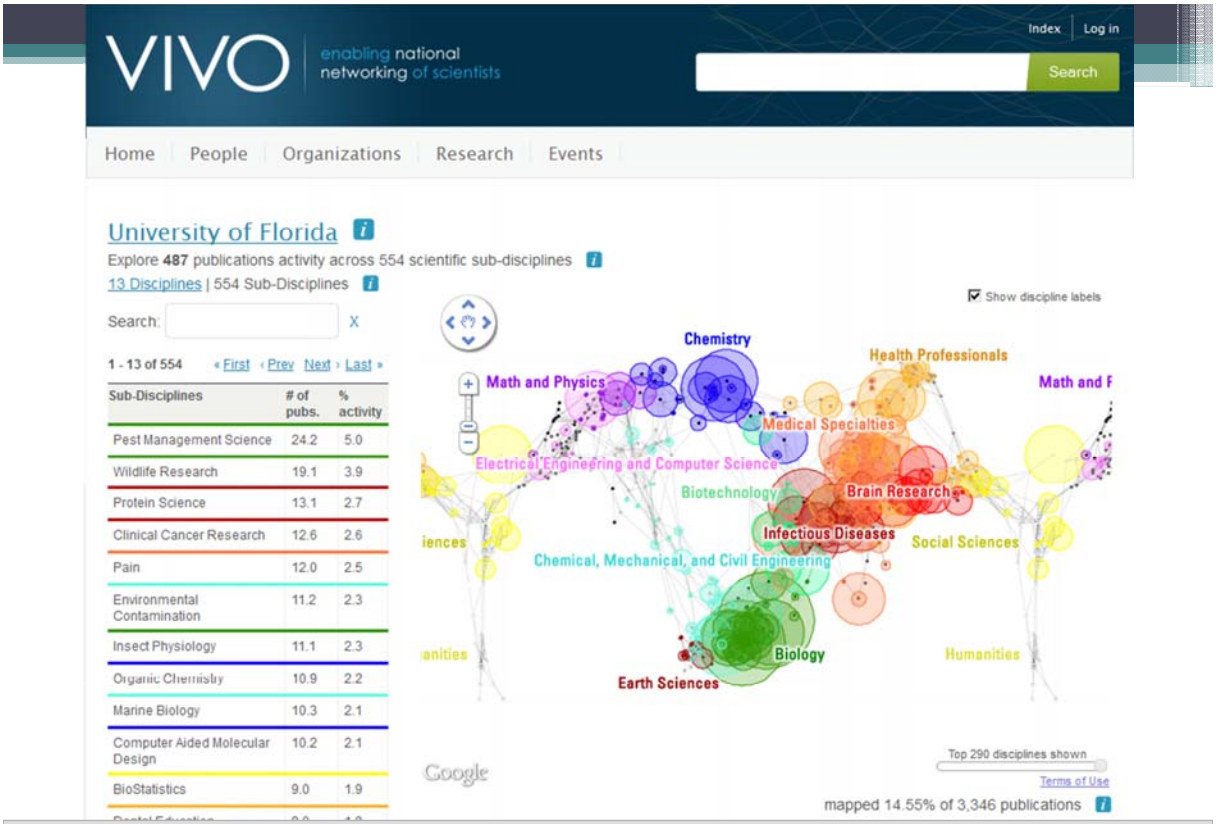
Funded by $12 million NIH award.

Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University**: Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaeleen Trimarchi. **Washington University, St. Louis**: Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klaz, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.

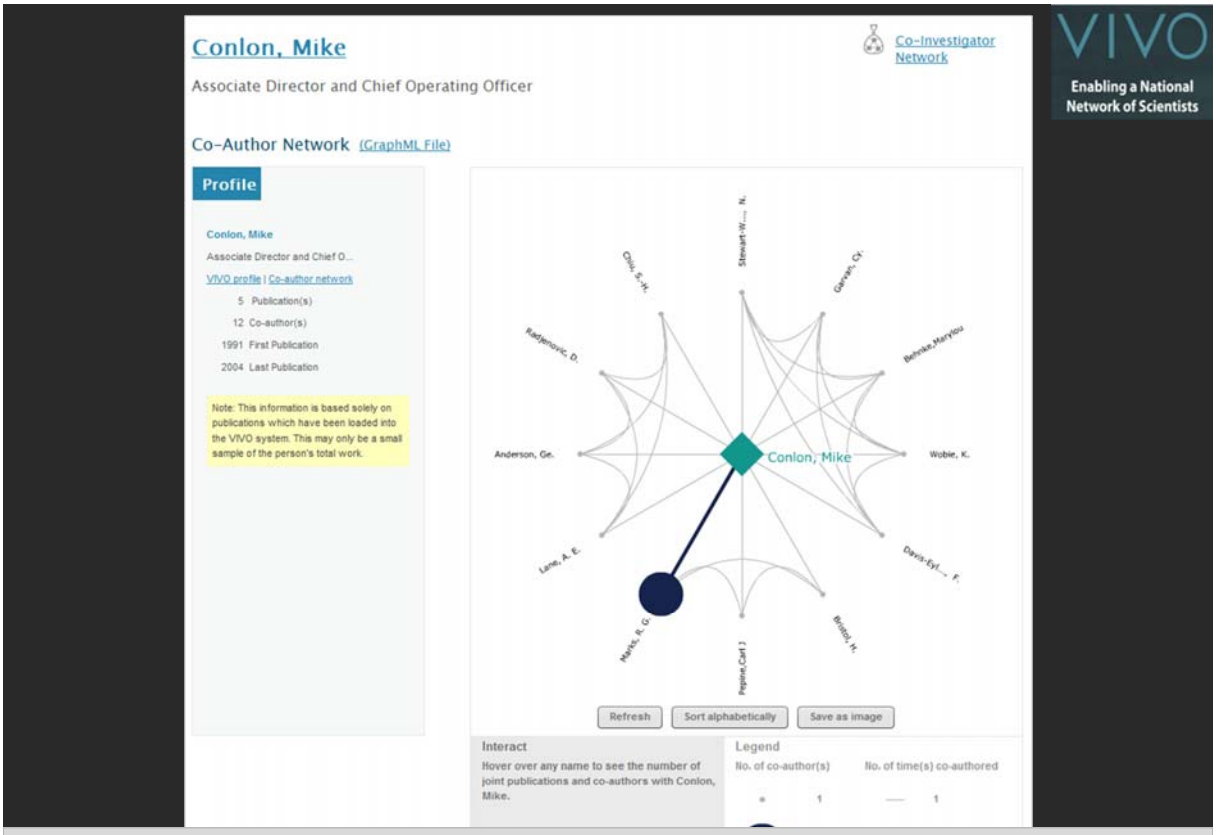
UF Clinical and Translational Science Institute
UNIVERSITY of FLORIDA



Temporal Analysis (When) Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level



Topical Analysis (What) Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)



Network Analysis (With Whom?) Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?

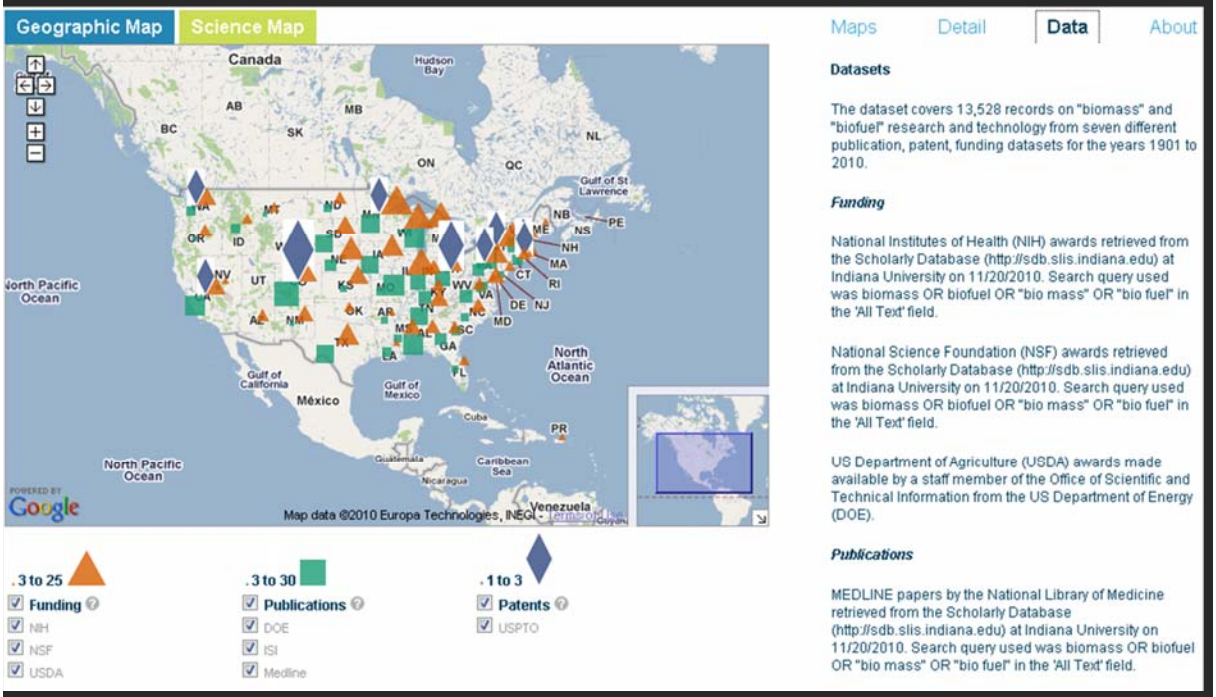
Geospatial Analysis (Where) Where is what science performed by whom? Science is global and needs to be studied globally.

VIVO
Enabling a National Network of Scientists

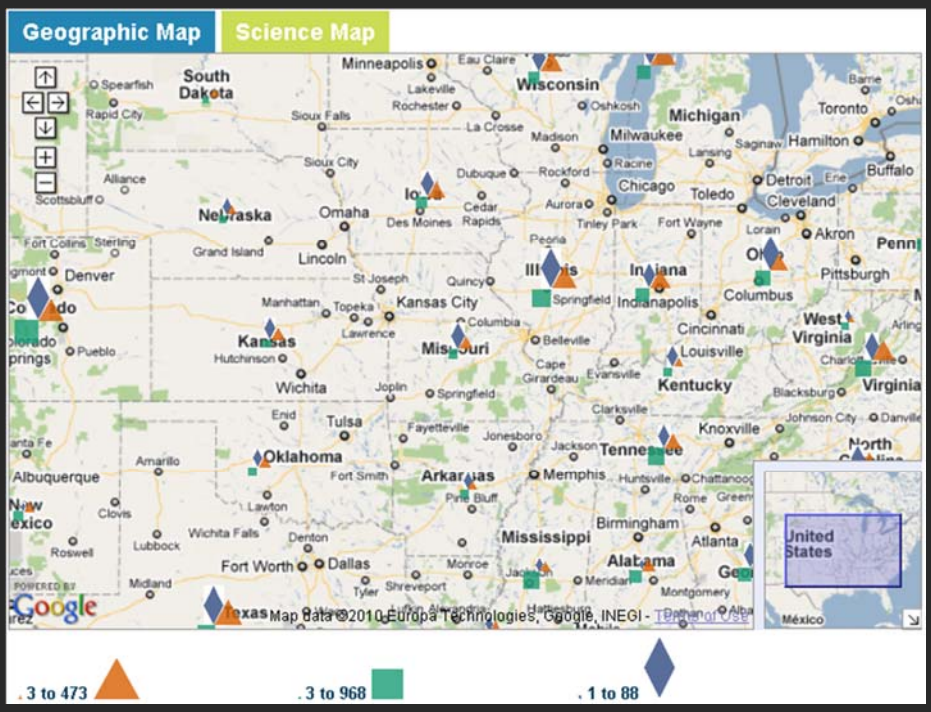
VIVO On-The-Go

Overview, Interactivity, Details on Demand
come to commonly used devices and environments

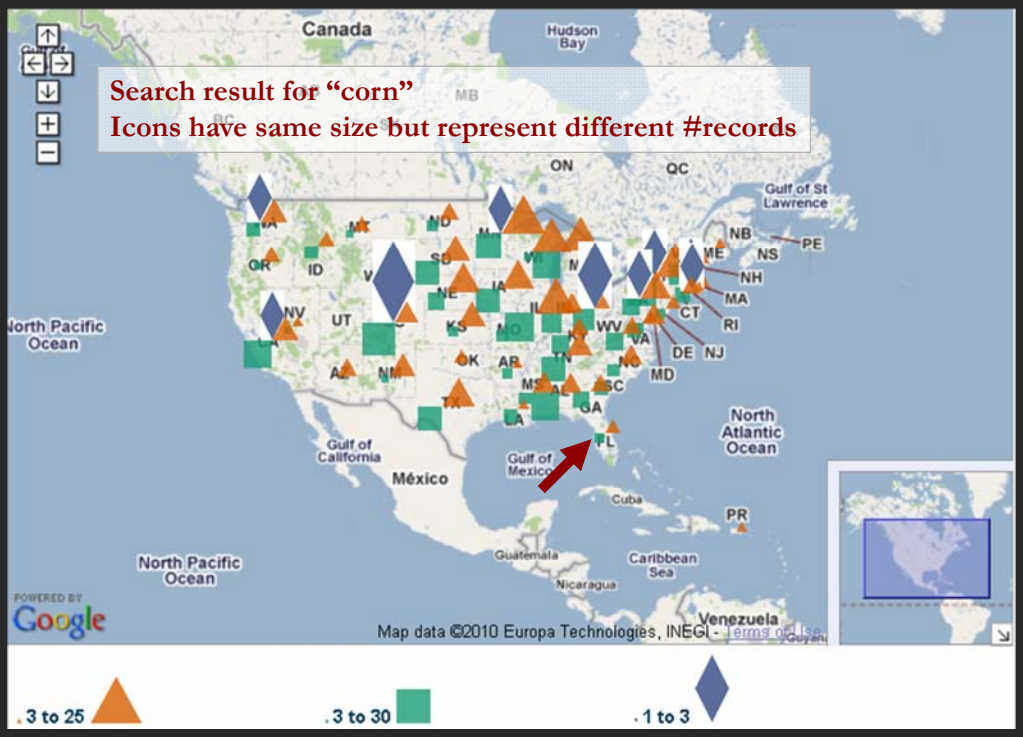
Online Interactive Maps for Sustainability Research and NIH

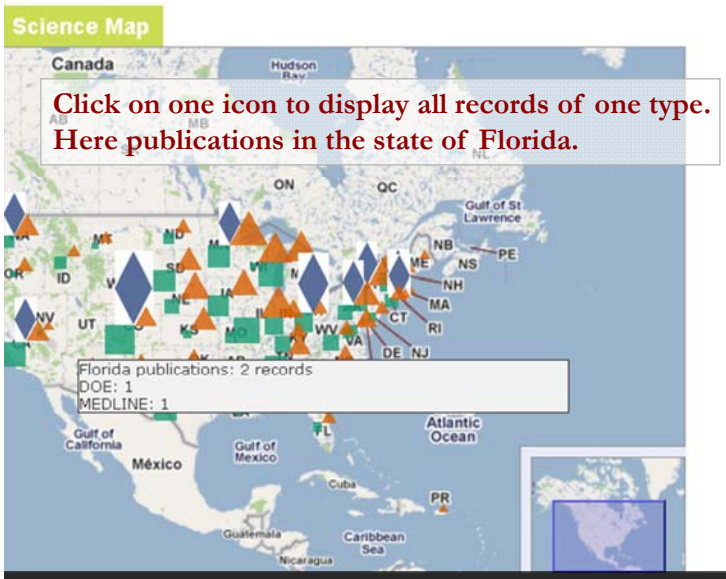


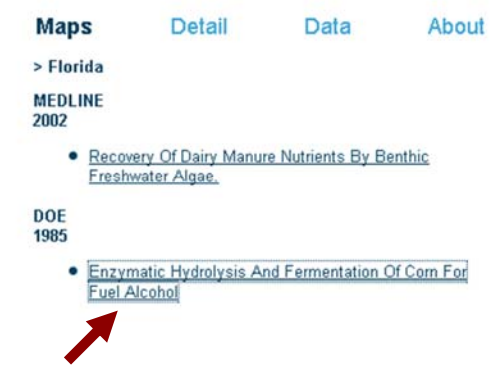
The geographic map at **state level**.

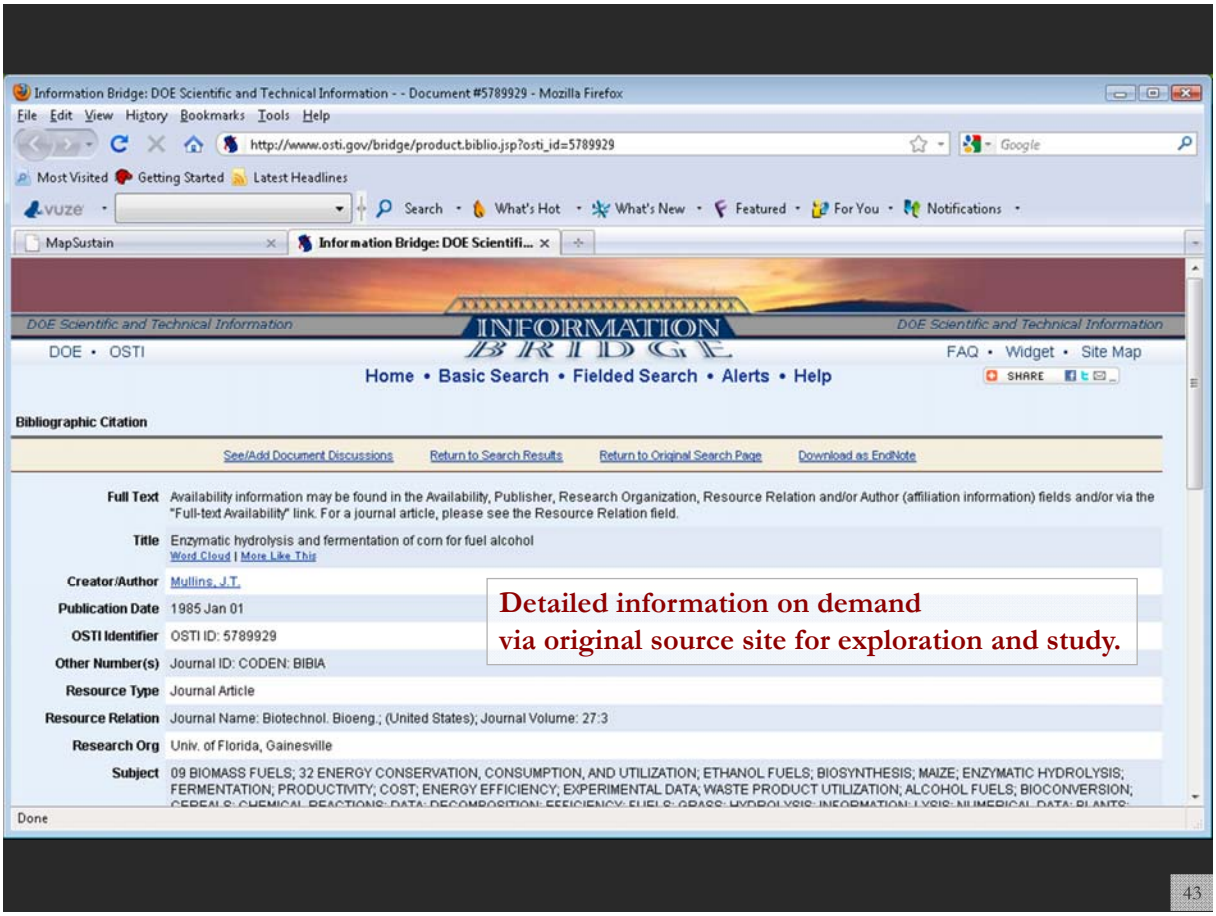


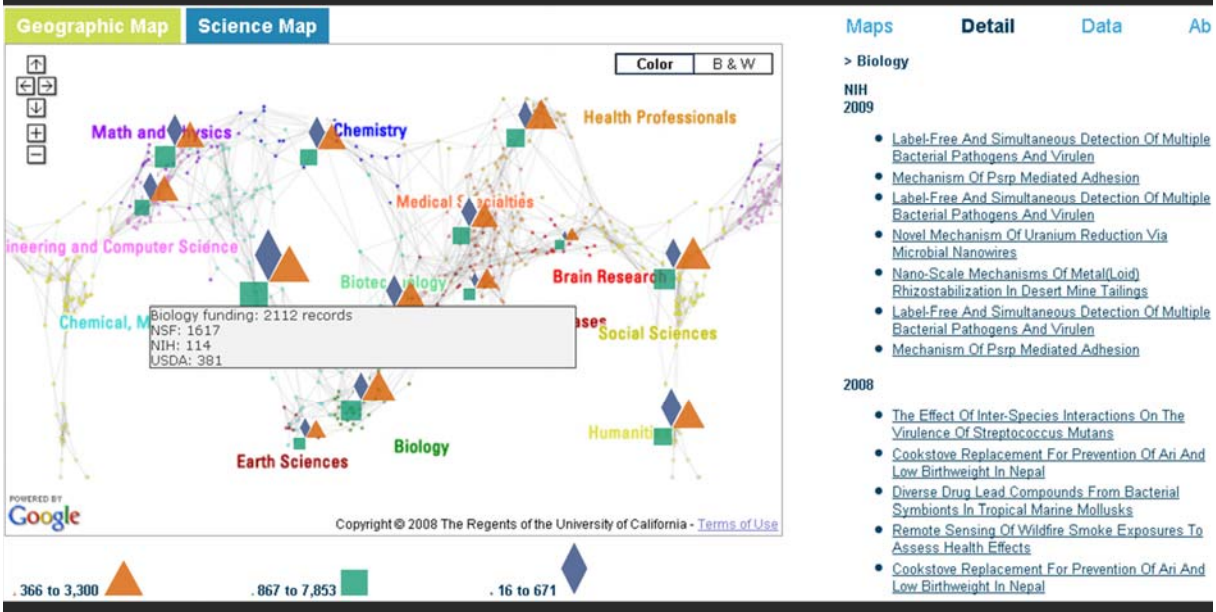
The geographic map at **city level**.


Search result for "corn"
Icons have same size but represent different #records
Canada
Hudson Bay
North Pacific Ocean
México
Gulf of California
Gulf of Mexico
North Atlantic Ocean
Cuba
PR
Guatemala
Caribbean Sea
Nicaragua
Venezuela
Map data ©2010 Europa Technologies, INEGI
.3 to 25
.3 to 30
.1 to 3

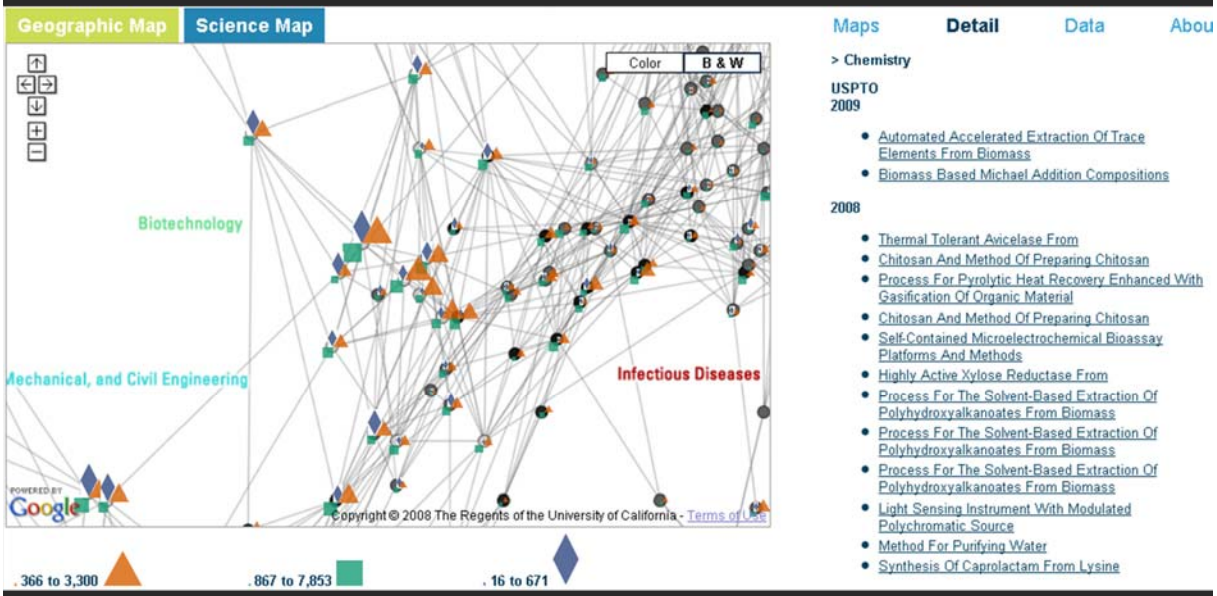

Science Map
Click on one icon to display all records of one type.
Here publications in the state of Florida.
Florida publications: 2 records
DOE: 1
MEDLINE: 1


Maps
Detail
Data
About
> Florida
MEDLINE
2002
Recovery Of Dairy Manure Nutrients By Benthic Freshwater Algae.
DOE
1985
Enzymatic Hydrolysis And Fermentation Of Corn For Fuel Alcohol

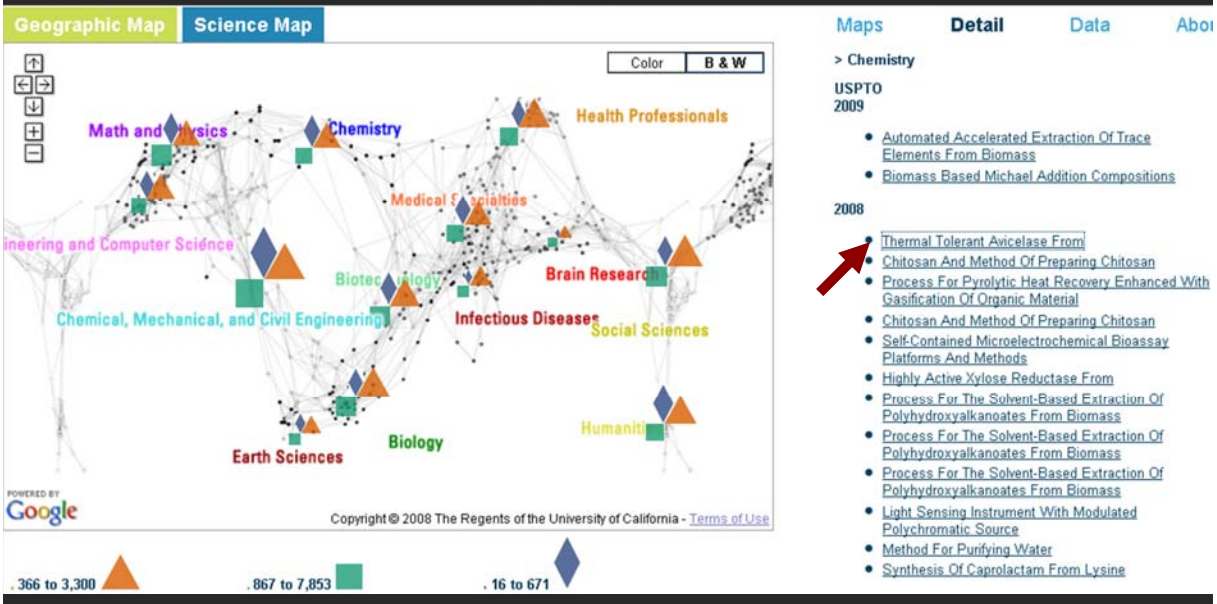


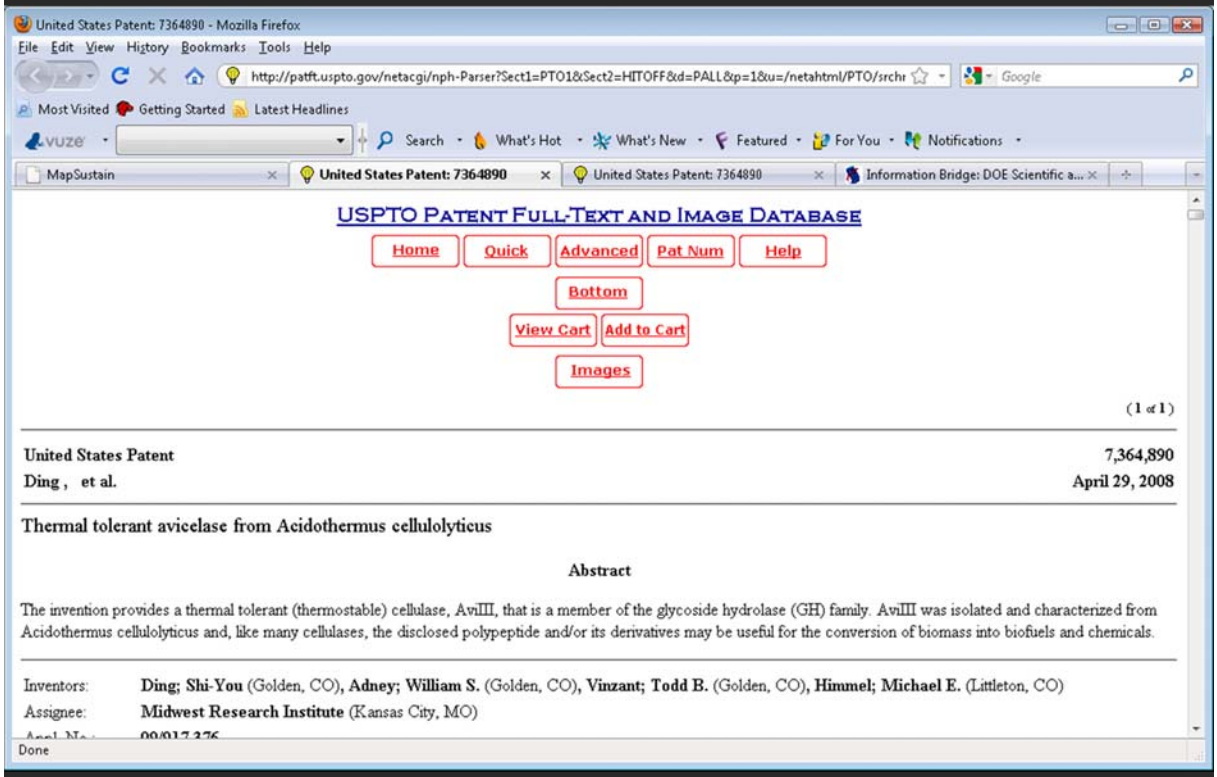


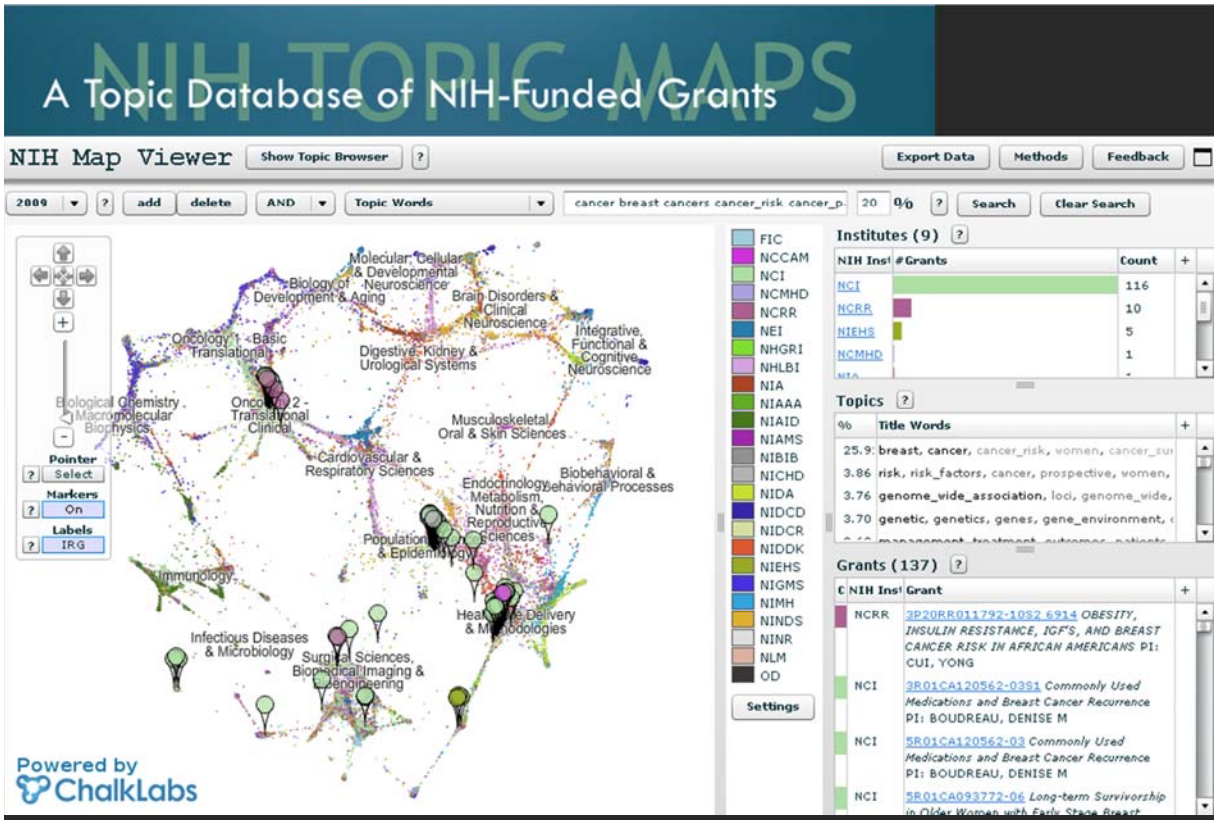
The science map at 13 top-level scientific **disciplines level**.



The science map at 554 **sub-disciplines level**.

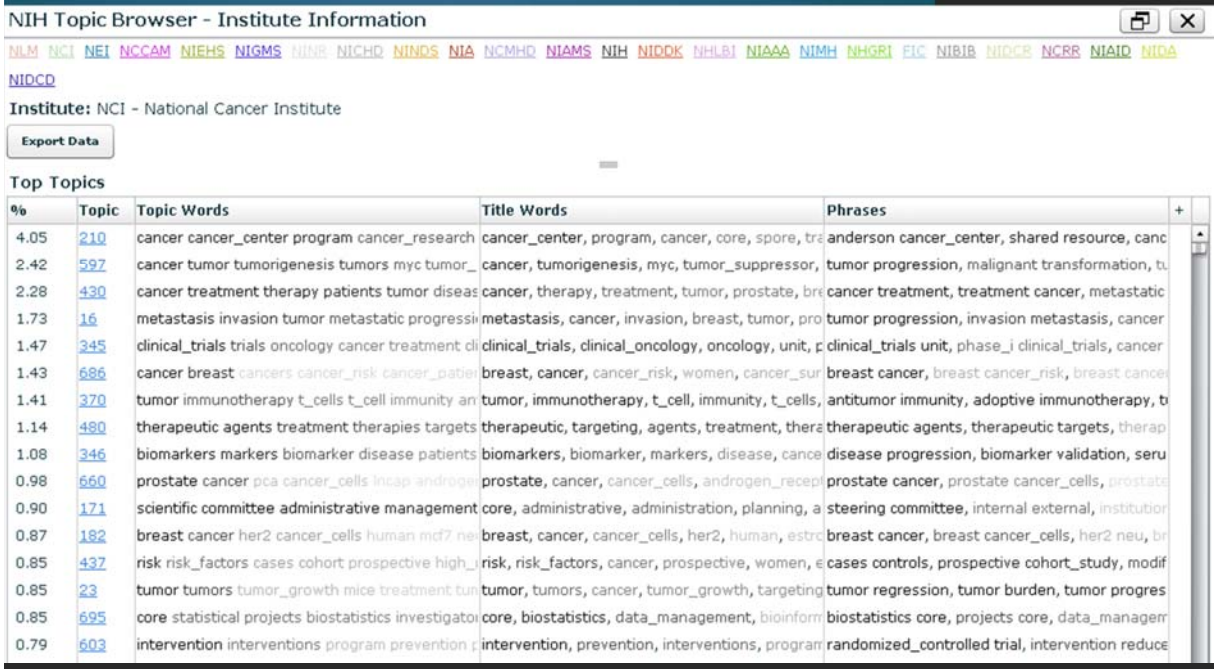






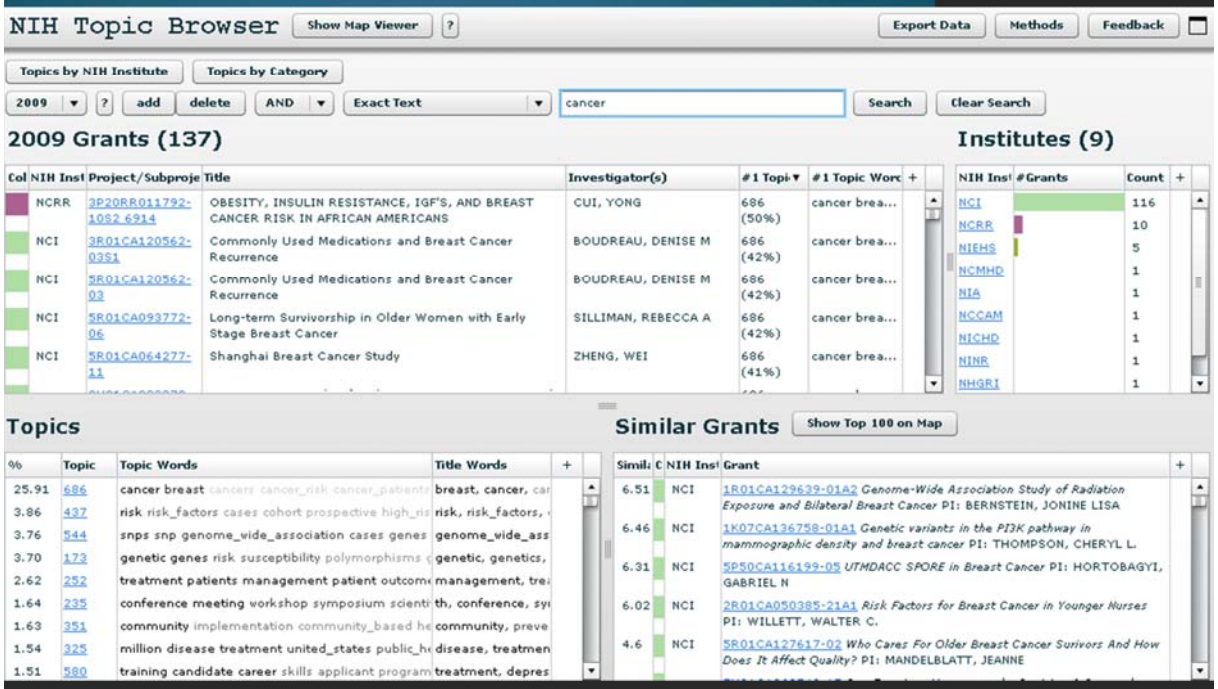
https://app.nihmaps.org

A Topic Database of NIH-Funded Grants



https://app.nihmaps.org

A Topic Database of NIH-Funded Grants

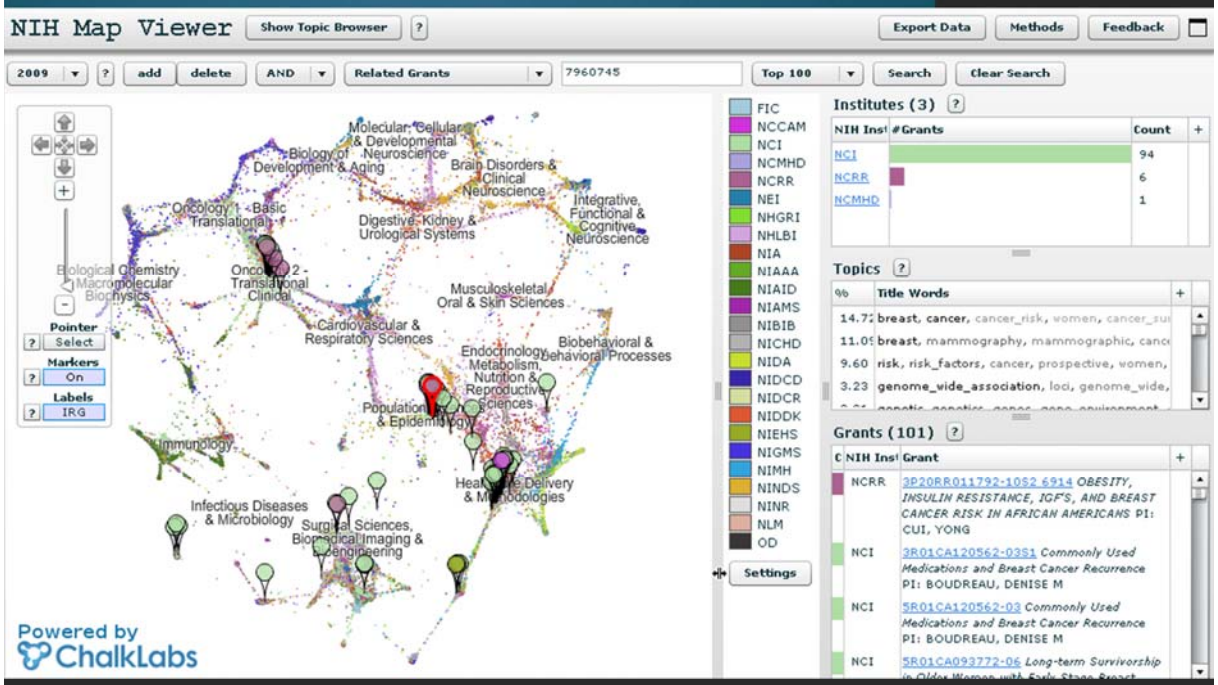


https://app.nihmaps.org

A Topic Database of NIH-Funded Grants

https://app.nihmaps.org

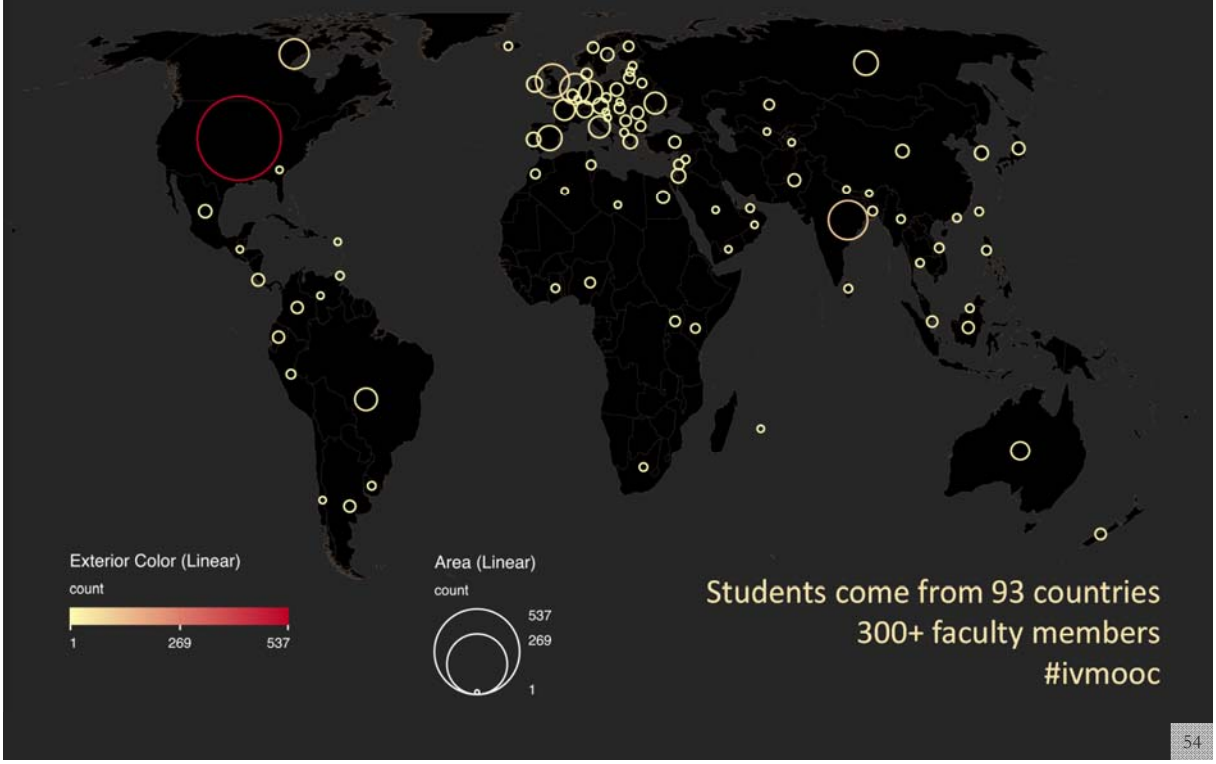
A Topic Database of NIH-Funded Grants

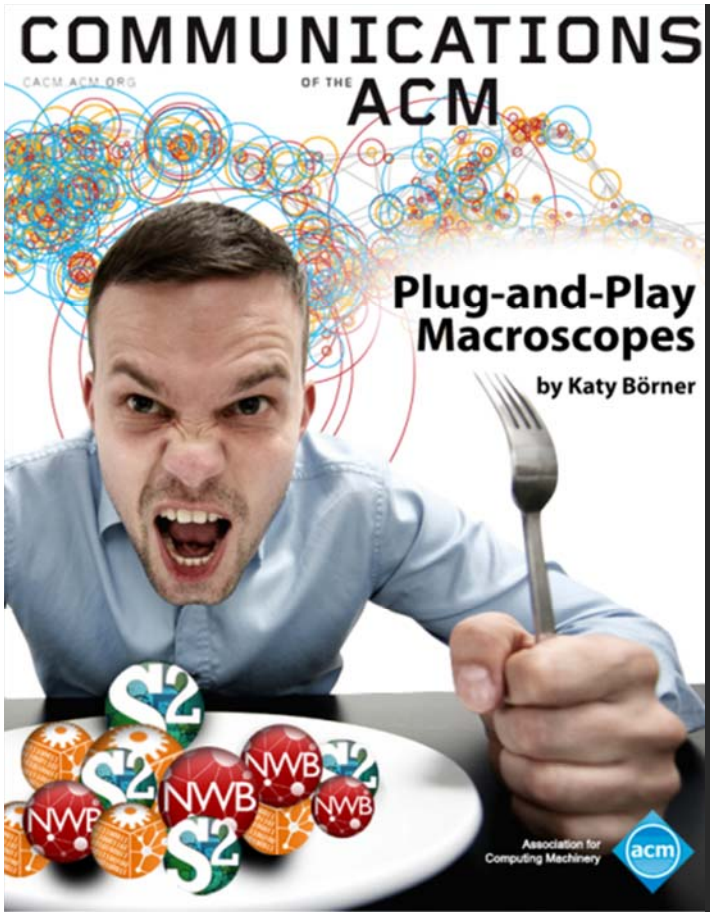


https://app.nihmaps.org

The Information Visualization MOOC

The Information Visualization MOOC
ivmooc.cns.iu.edu



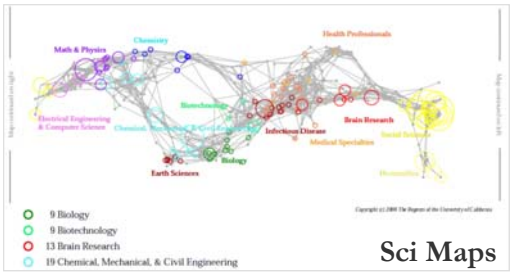


Börner, Katy. (2011). Plug-and-Play Macroscopes. *Communications of the ACM, 54(3), 60-69.*

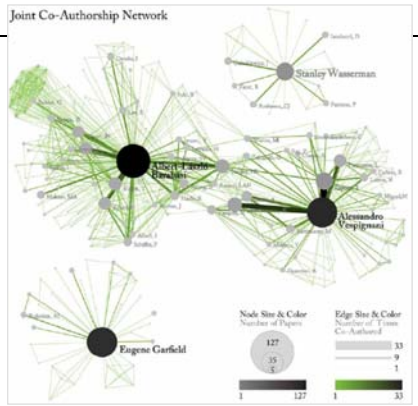
Video and paper are at http://www.scivee.tv/node/27704

Sci² Tool – "Open Code for S&T Assessment"

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



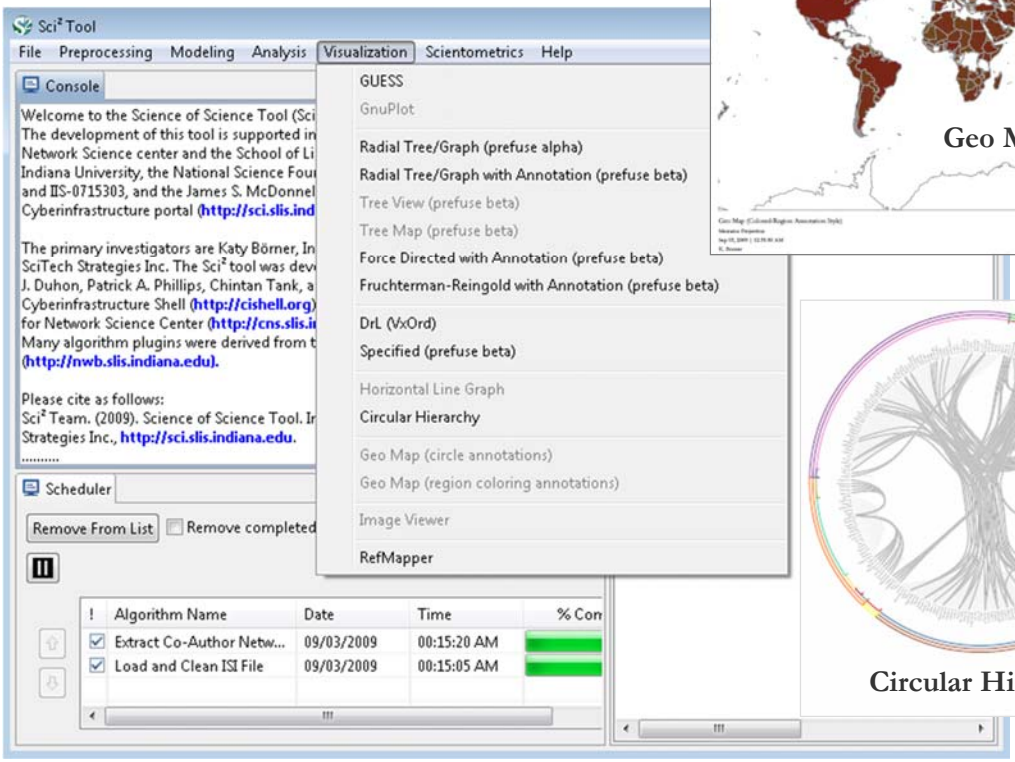
Sci Maps

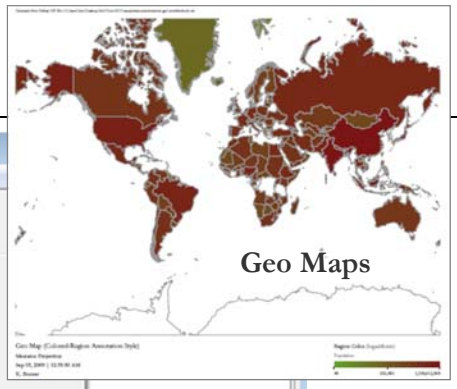


GUESS Network Vis

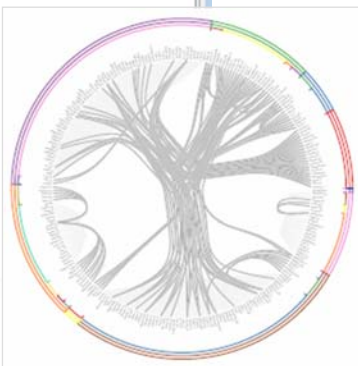
Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Duhon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of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

Sci² Tool Vis cont.





Geo Maps



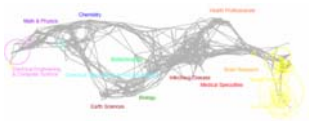
Circular Hierarchy

Mapping NEH awards and MEDLINE publications, 1980-2009

NEH Grants:
41,258 grants of 47,197 started between 1980 and 2009, encompassing 3.21 billion of the 3.77 billion dollars awarded.

Geo-coding by zipcode:
36,512 of 41,258 grants encompassing 3.13 billion of a potential 3.21 billion were geocoded to 3,510 distinct locations.

Science-coding by topic:
37,132 of 41,258 grants encompassing 2.09 billion of a potential 3.21 billion were mapped to 42 distinct subdisciplines.



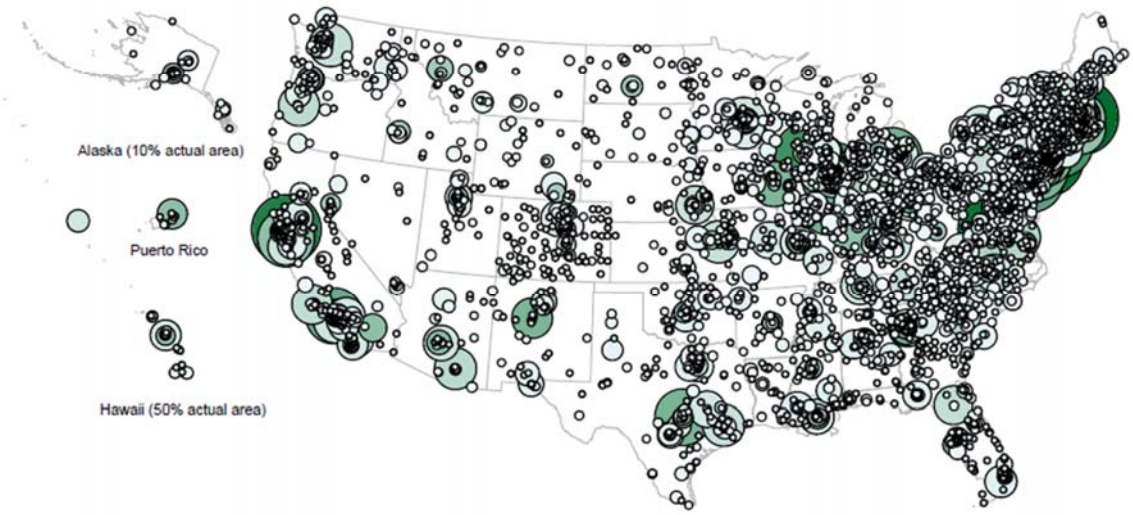
MEDLINE publications:
12.95 million papers were published between 1980 and 2009.

Geo-coding:
Not possible with the data we have.

Science-coding by journal:
11.62 million of 12.95 million papers were science located (89.7%). Science located 5,941 out of 14,561 journals (40.8%) to 415 distinct subdisciplines.

Geospatial Visualization (Proportional Symbol Map)

NEH Grants by ZIP Code (1980-2009)
Apr 09, 2013 | 01:09:37 PM EDT



Legend
Interior Color (Linear)
award_amount
0 20,700,000 41,400,000

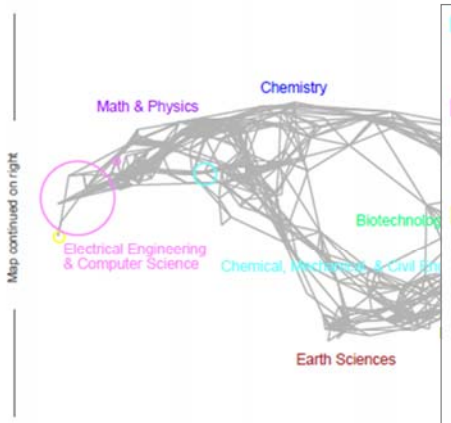
Area (Linear)
count
490
246
1

How to Read this Map
This *proportional symbol map* shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

Topical Visualization: UCSD Map of Science

NEH grants, mapped based on categories



Legend
Circle area: Dollars Awarded
Unclassified = 0
Minimum = 8,471
Maximum = 509,399,136
Color: Discipline

Area
216,889,824
113,027,584
9,216,623

Chemical, Mechanical, & Civil Engineering
13,469,200 Mechanical Design Engineering

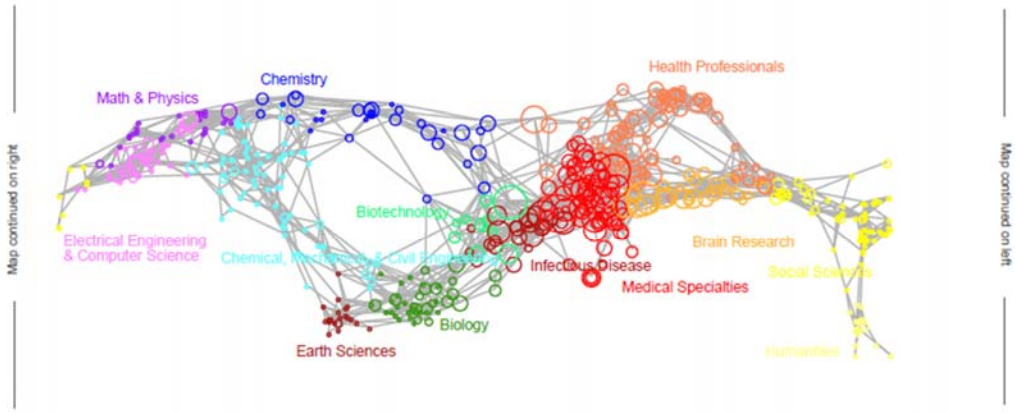
Electrical Engineering & Computer Science
153,044,896 Library Science; Infomation Retrieval
724,505 Logic

Humanities
509,399,136 American History
101,708,568 Art History
87,938,056 Asian Studies
15,956,450 Biblical Literature
33,668,248 Classics
347,050 Critical Studies
36,538,632 Cross Disciplinary Stu
238,170,928 English Literature
8,471 Ethics
4,126,550 German Studies
38,912,520 Hispanic Studies
2,234,555 Italian Studies
36,181,320 Linguistics
33,062,848 Literary Criticism
56,233,540 Medieval History
65,690,320 Modern Language
98,065,344 Music & Theatre
58,949,420 Philosophy Psychology
42,320,944 Science History
905,530 Semiotics
52,800,752 Social History
100,976,304 Socio-Cultural Anthropology

Social Sciences
2,593,480 Child Development
1,852,100 Communication Res
1,482,022 Economics
58,847,100 Education
53,223,792 Ethnology
79,414,408 Higher Education
71,120 Human Resource Mana
15,179,400 International Develc
3,008,390 International Econor
10,232,764 Law
4,693,470 Political Geography
28,897,260 Political Science
832,157 Public Administration
21,255,080 Regional Studies
1,967,490 Rural Studies
15,774,390 Sociology
4,333,450 Urban Studies

Topical Visualization: UCSD Map of Science

Medline papers, mapped based on journal names



2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.

Legend

Circle area: Fractional Journal Count
Unclassified = 0
Minimum = 0
Maximum = 344,175
Color: Discipline

Area

2,168,898.25
1,130,275.88
92,166.23

How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

References

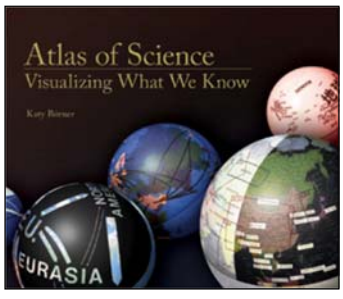
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255.
http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

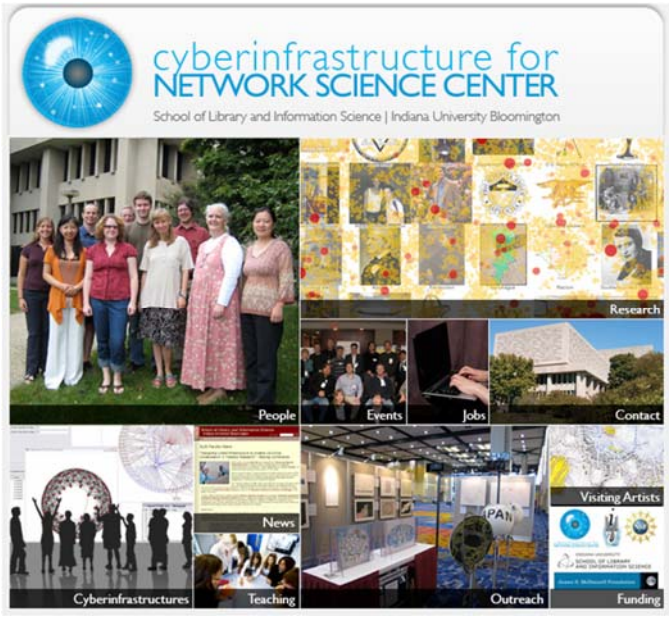
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).
http://www.pnas.org/content/vol101/suppl_1/

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., Volume 41, Chapter 12, pp. 537-607.
http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

Börner, Katy (2010) **Atlas of Science**. MIT Press.
http://scimaps.org/atlas

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.





All papers, maps, tools, talks, press are linked from http://cns.iu.edu

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience