

# Envisioning Big Data

CNS, SLIS, IU, Bloomington, IN

[katy@indiana.edu](mailto:katy@indiana.edu) | <http://cns.iu.edu>

Big Data Symposium

The Becton Engineering and Applied Science Center, New Haven, CT

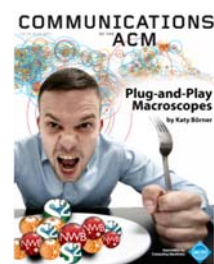
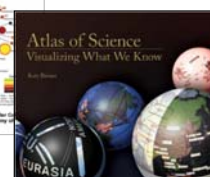
<http://www.cs.yale.edu/calendars/schedule.html>

October 26, 2012



## Overview

1. **Data mining and visualization research** that aims to increase our scientific understanding of the structure and dynamics of science and technology.
2. **Novel approaches and services** that improve information access, researcher networking, and research management.
3. **Data services and plug-and-play macroscope tools** that commoditize data mining and visualization.





Terra bytes of data

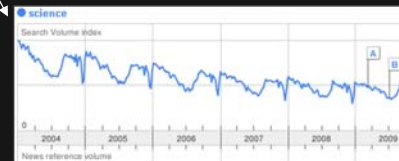
Descriptive & Predictive Models



Find your way



Find collaborators, friends



Identify trends



Terra bytes of data

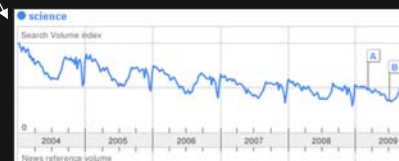
Plug-and-Play  
Macrosopes



Find your way



Find collaborators, friends



Identify trends

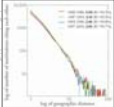
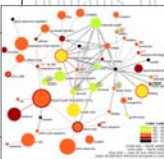



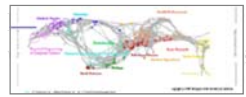
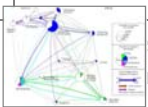
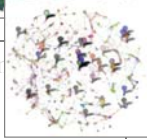

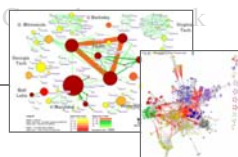
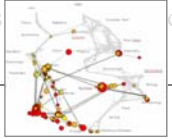
## Type of Analysis vs. Level of Analysis

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



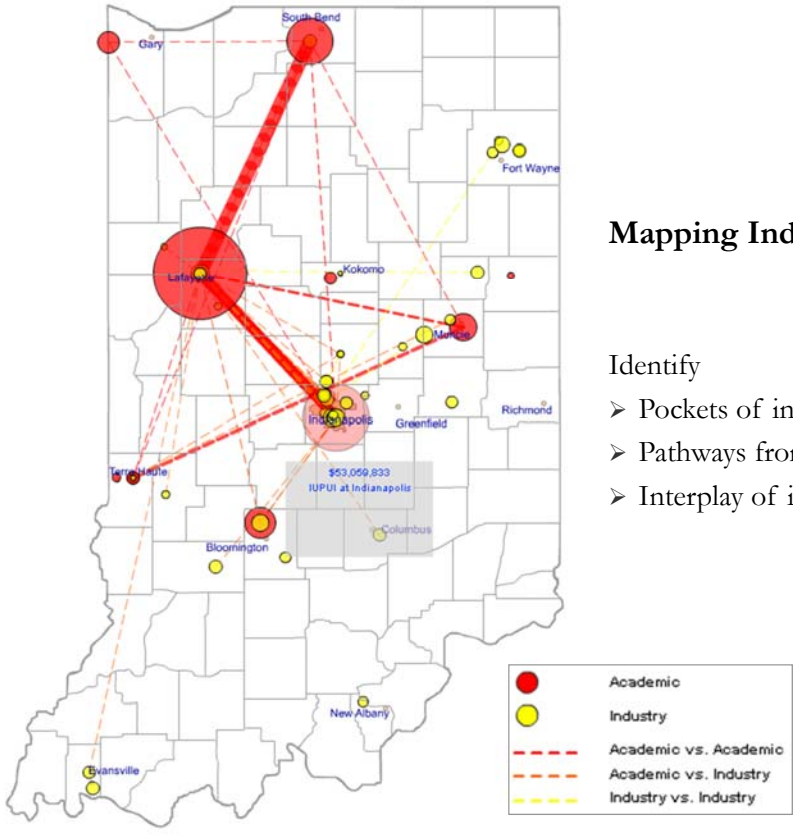
5

## Type of Analysis vs. Level of Analysis

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



6

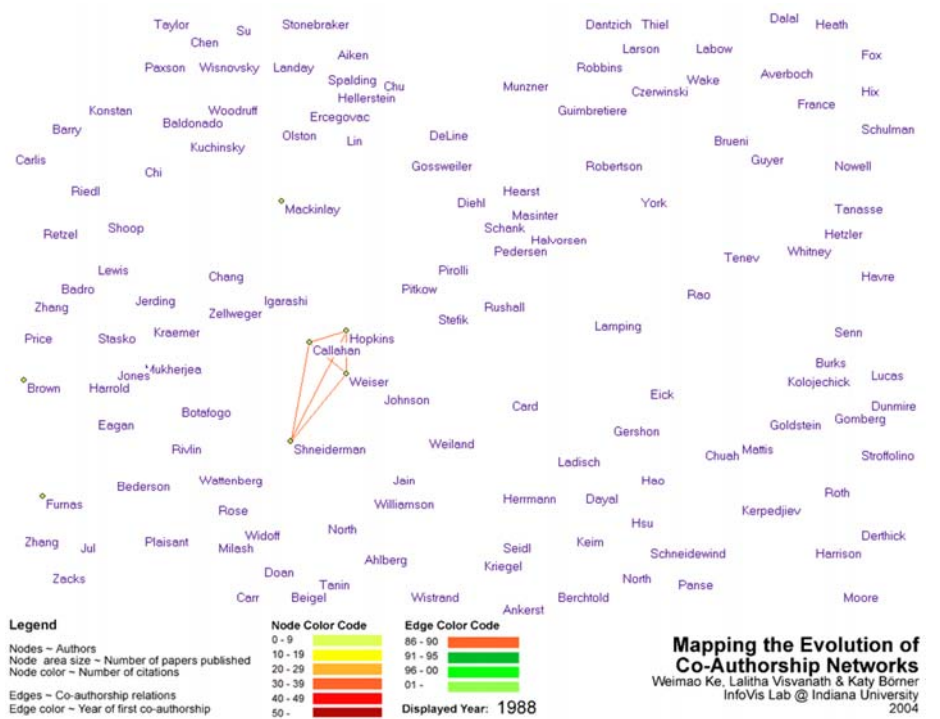


### Mapping Indiana's Intellectual Space

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

### Mapping the Evolution of Co-Authorship Networks

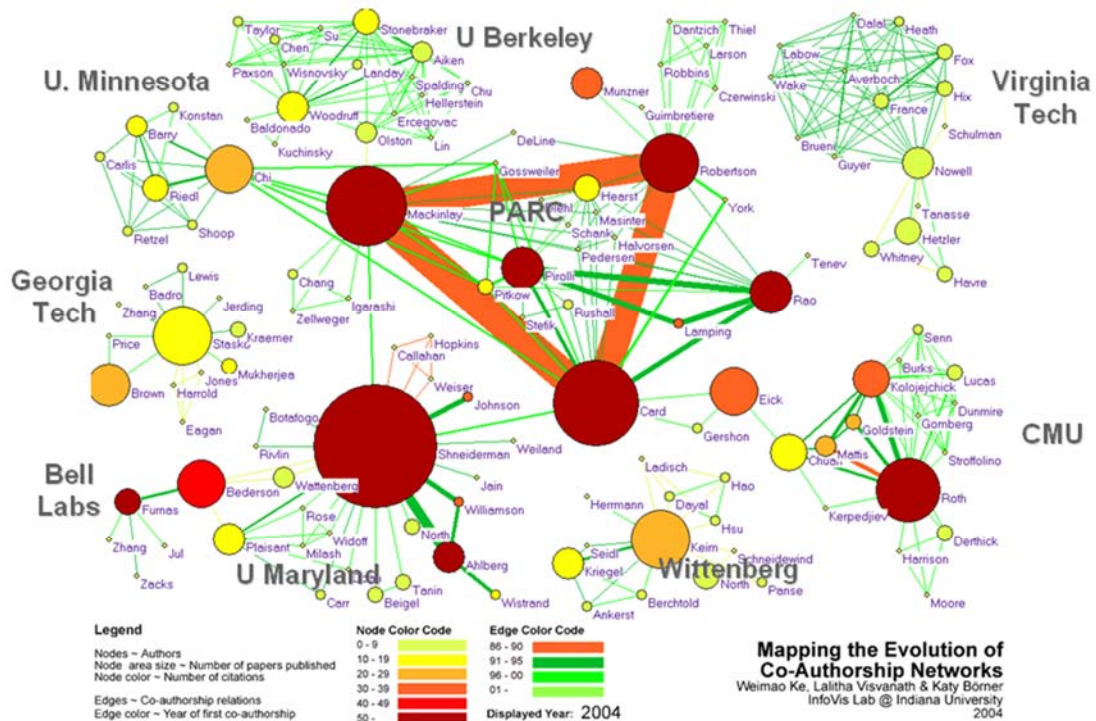
*Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.*



**Mapping the Evolution of Co-Authorship Networks**  
 Weimao Ke, Lalitha Visvanath & Katy Börner  
 InfoVis Lab @ Indiana University  
 2004

# Mapping the Evolution of Co-Authorship Networks

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



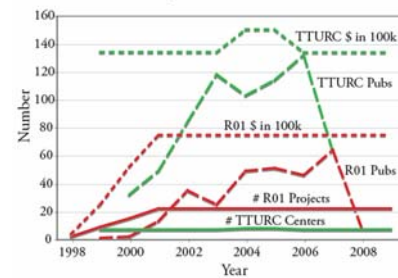
## Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

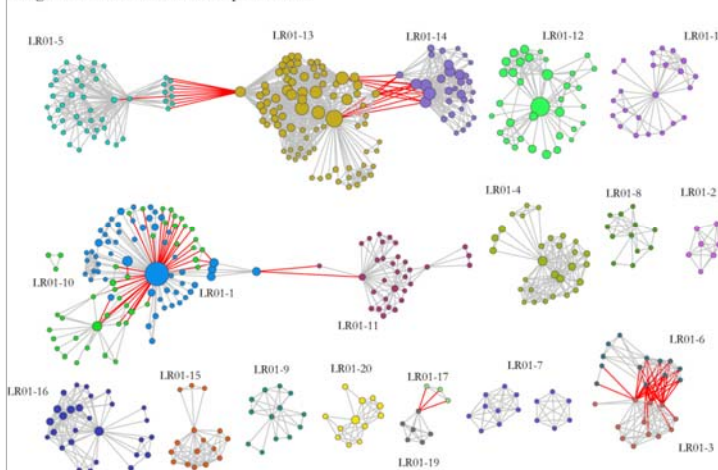
Zoss & Börner, forthcoming.

Supported by NIH/NCI Contract HHSN261200800812

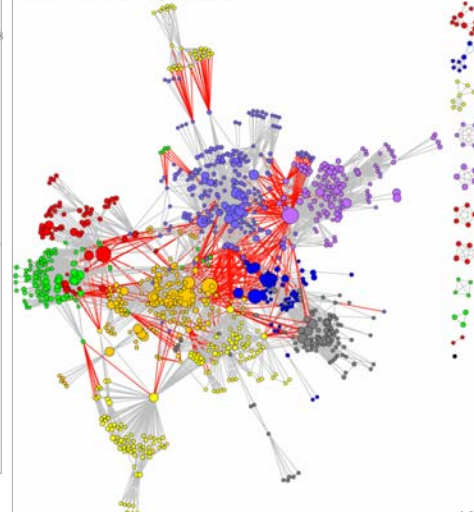
R01 & TTURC Project Information



Longitudinal R01 Co-Authorship Network

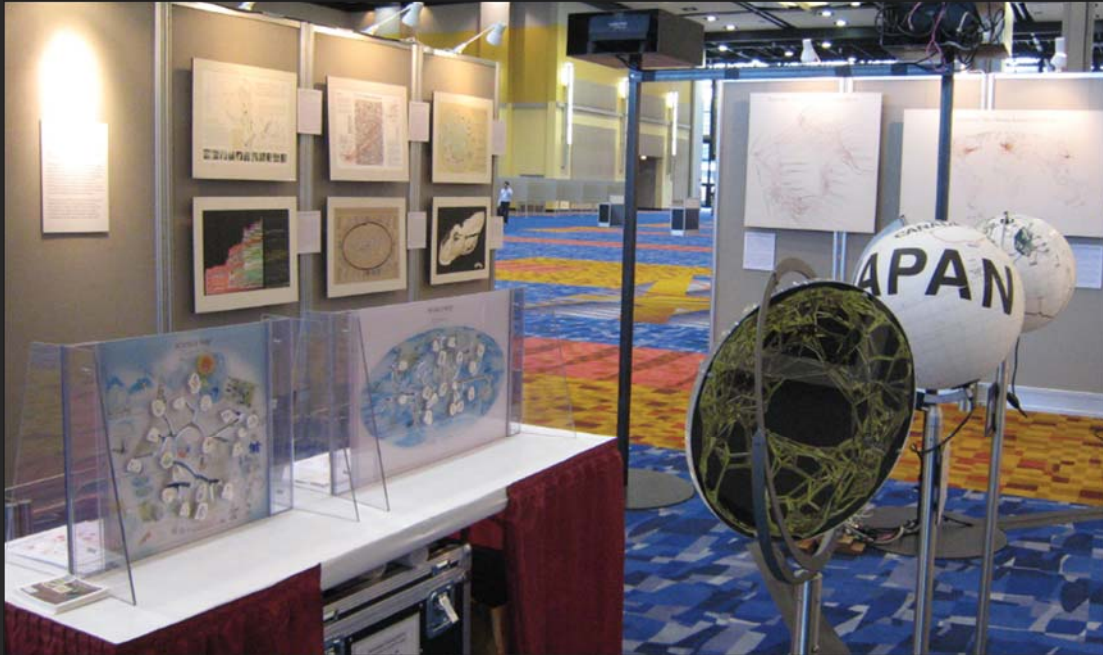


TTURC Co-Authorship Network





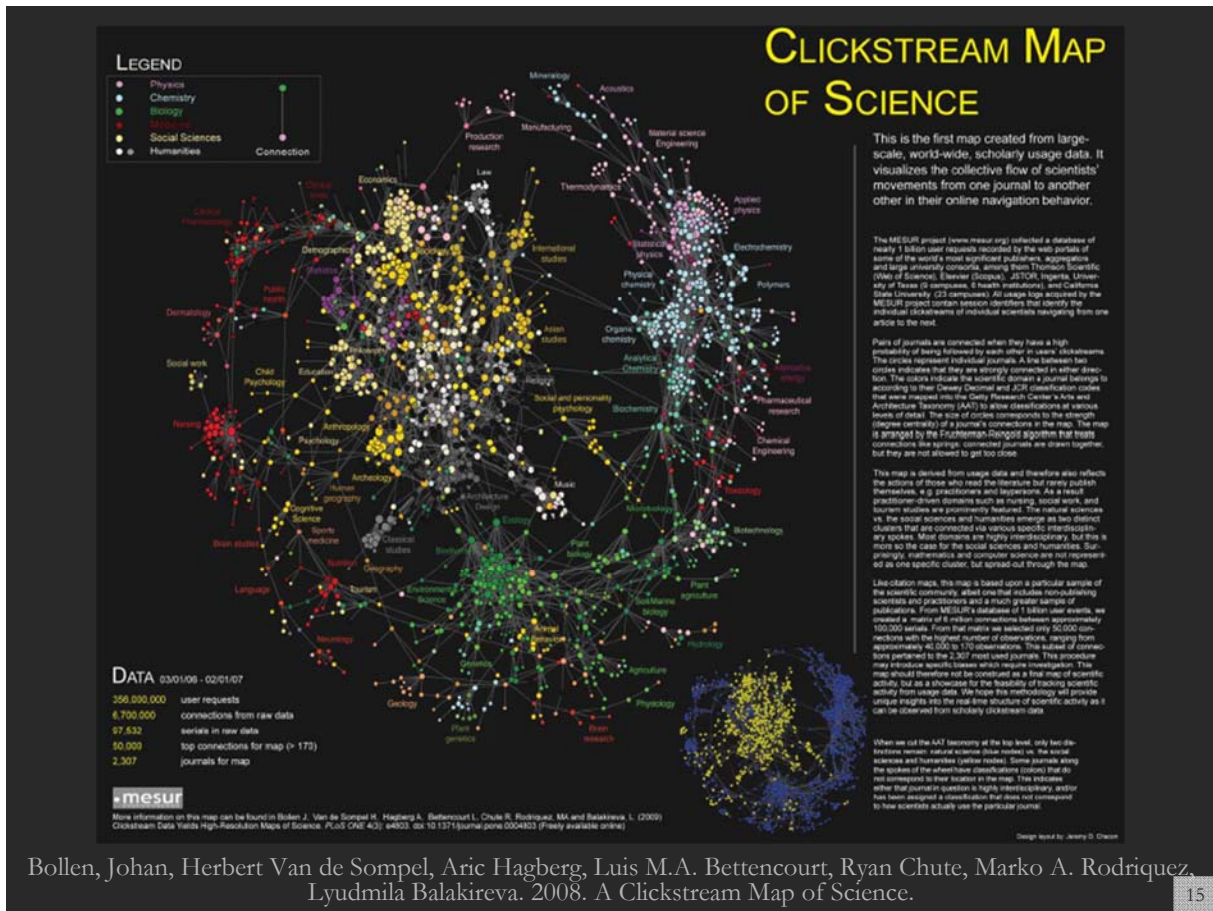
Places & Spaces: Mapping Science Exhibit (<http://scimaps.org>)



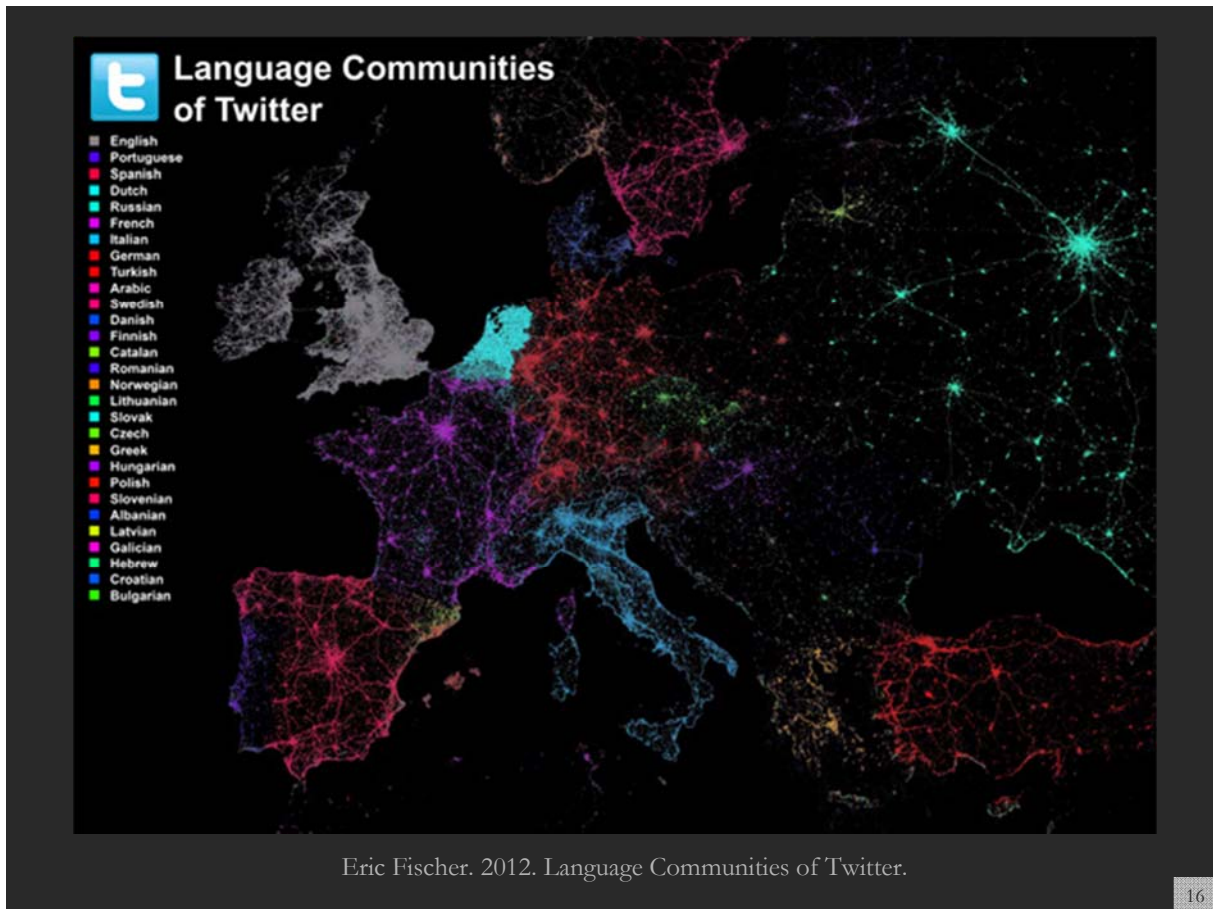
*After eight years, there now exist 80 out of 100 maps.*



Mapping Science Exhibit at MEDIA X, Wallenberg Hall, Stanford University, 2009  
<http://mediax.stanford.edu>, <http://scaleindependentibought.typepad.com/photos/scimaps>



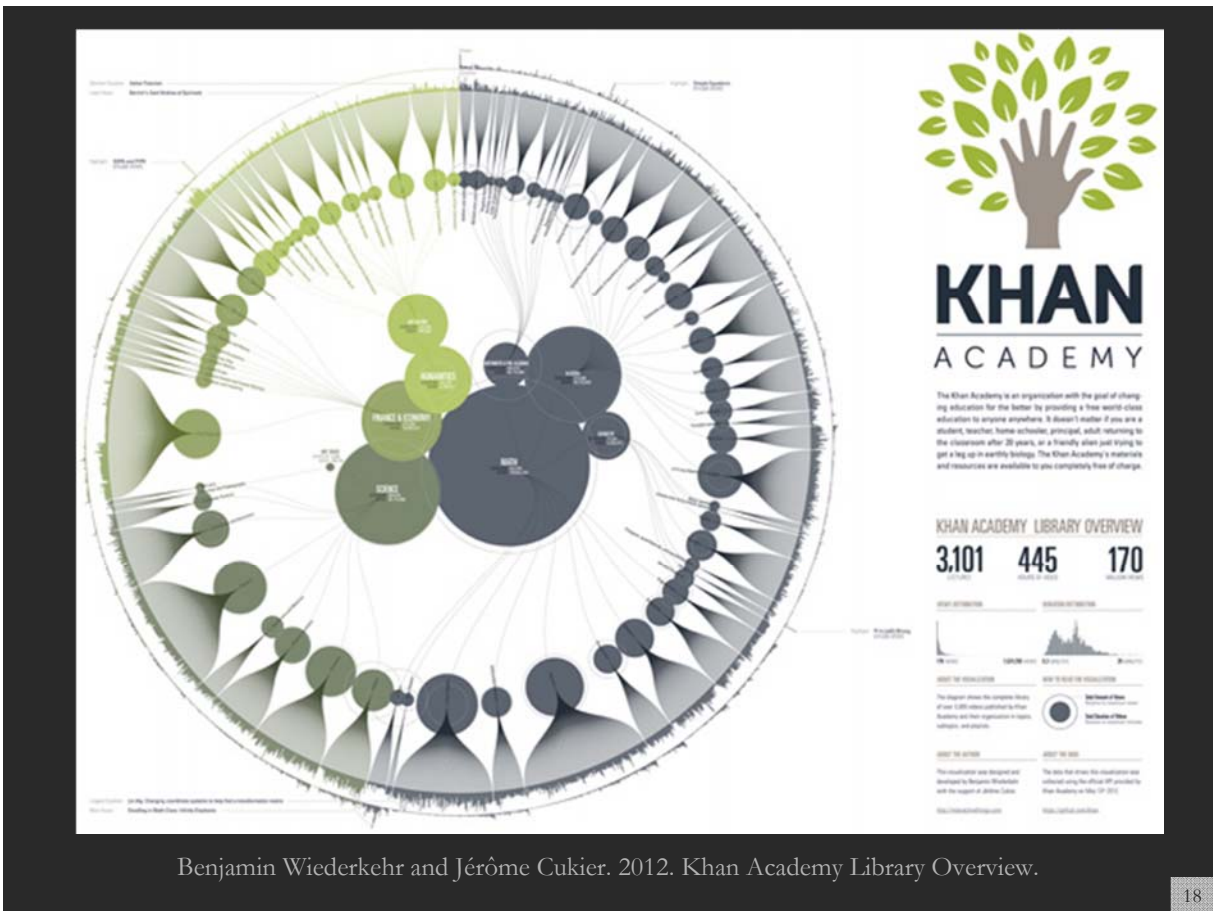
Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriguez, Lyudmila Balakireva. 2008. A Clickstream Map of Science. 15







Facebook Draws a Map of the Connected World. Dec 14, 2010

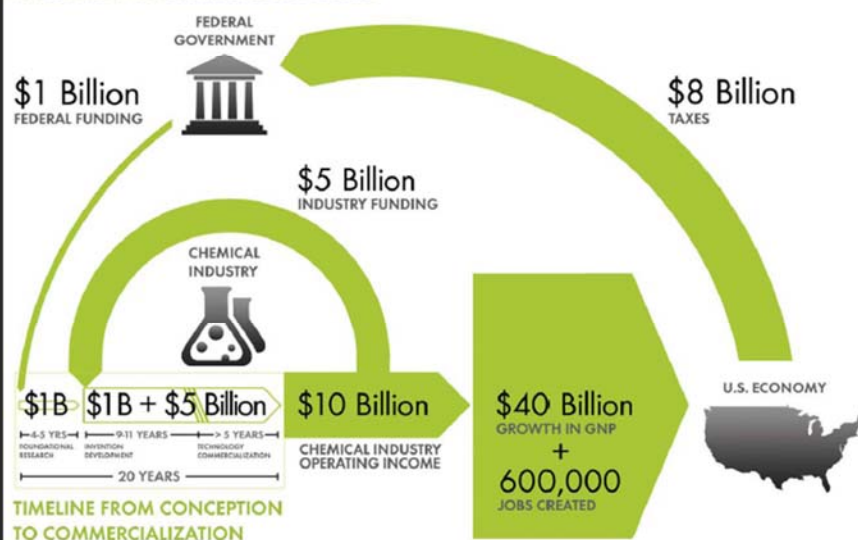


Benjamin Wiederkehr and Jérôme Cukier. 2012. Khan Academy Library Overview.

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

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

## INVESTMENT IN CHEMICAL SCIENCE R&D



### The Council for Chemical Research (CCR)

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

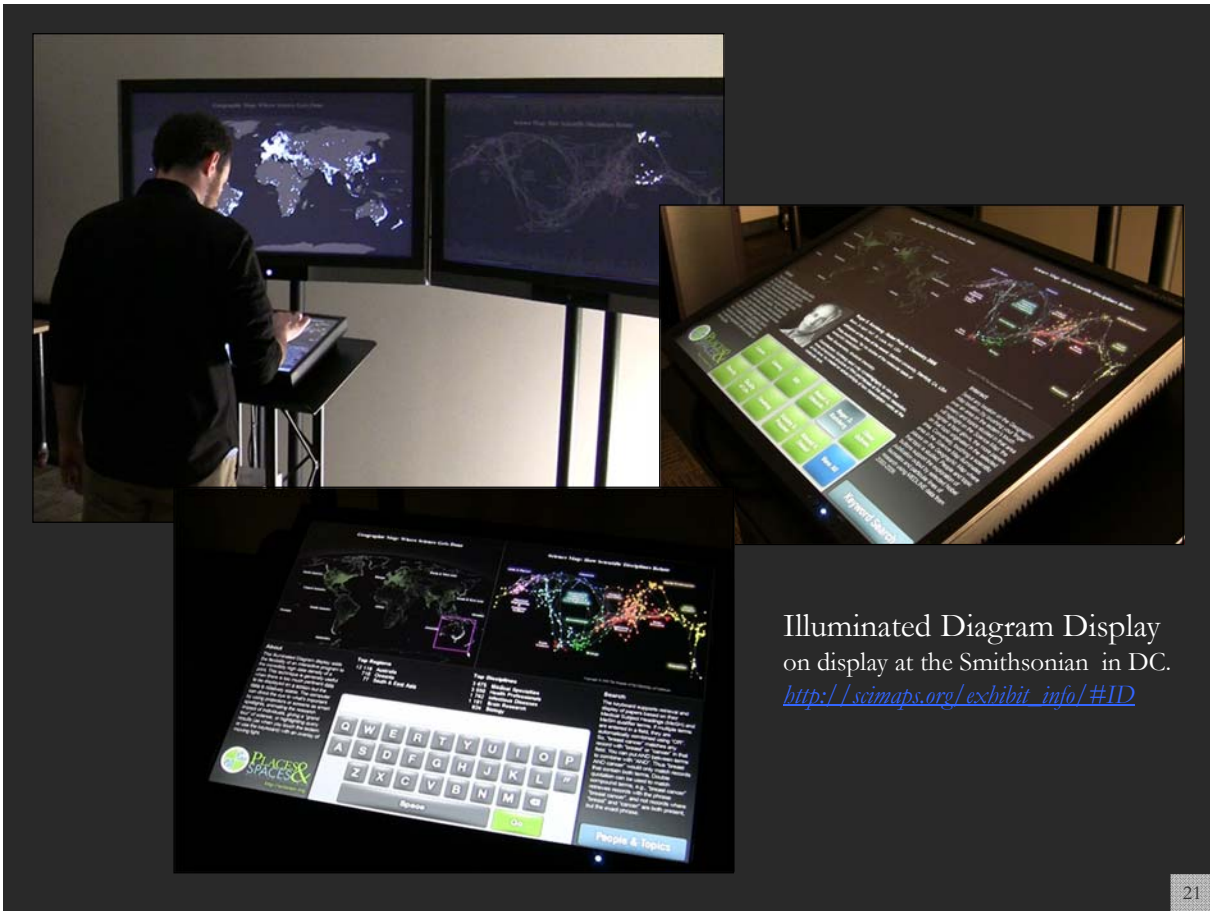


The design shows that an input of \$1B in federal investment, leveraged by \$5B industry investment, brings new technologies to market and results in \$10B of operating income for the chemical industry, \$40B growth in the Gross National Product (GNP) and further impacts the US economy by generating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

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



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



Illuminated Diagram Display  
on display at the Smithsonian in DC.  
[http://scimaps.org/exhibit\\_info/#ID](http://scimaps.org/exhibit_info/#ID)

### Geographic Map: Where Science Gets Done

### Science Map: How Scientific Disciplines Relate

**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.

**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 Professions - 3,589
- Social Sciences - 2,431
- Aeronautical, Chemical, Mechanical & Civil Engineering - 2,208
- Humanities - 1,562

**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.

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

<http://scimaps.org>

**People & Topics**

### Geographic Map: Where Science Gets Done

### Science Map: How Scientific Disciplines Relate

Copyright © 2009 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.

#### 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

### Keyword Search

<http://scmaps.org>

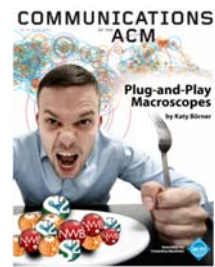
23

**INDIANA UNIVERSITY PLACES AND SPACES EXHIBIT**

24

## Overview

1. **Data + data mining and visualization research** that aims to increase our scientific understanding of the structure and dynamics of science and technology.
2. **Novel approaches and services** that improve information access, researcher networking, and research management.
3. **Data services and plug-and-play macroscope tools** that commoditize data mining and visualization.



Scholarly Database at IU

Supports federated search of 25 million publication, patent, grant records.

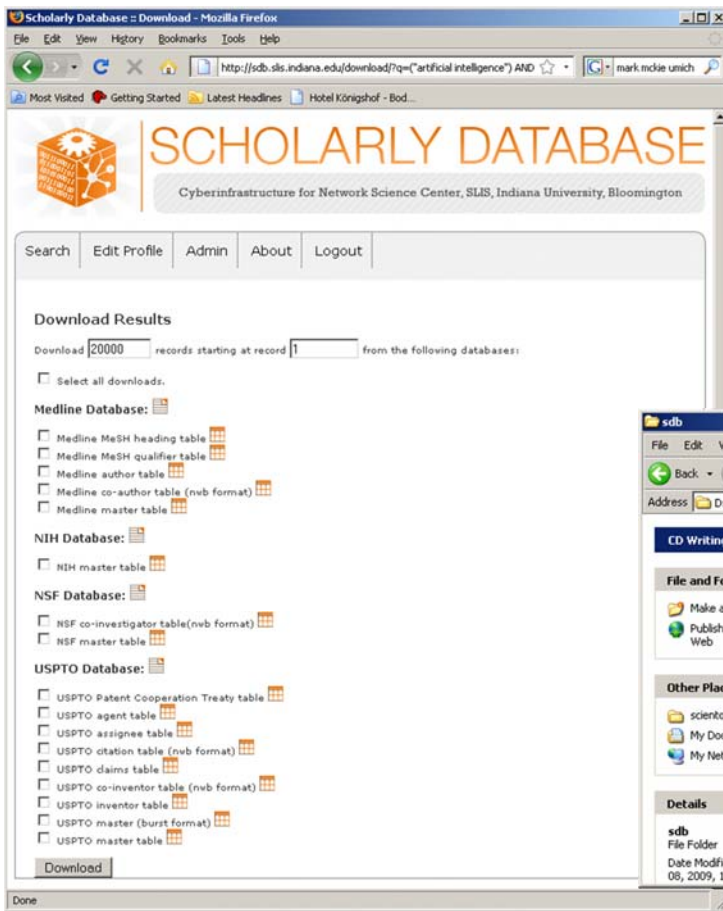
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

The screenshot shows the Scholarly Database homepage. On the left, there are login options for 'IU User' and 'Non-IU User'. The 'IU User' section includes a 'Go to IU Login' button. The 'Non-IU User' section has fields for 'Email' and 'Password' with a 'Login' button. Below the login options, there is a 'Not Registered Yet?' section with a 'Register as a New IU User' link. The main content area features a search interface with a search bar, 'Search' button, and various filters like 'Creators', 'Title', 'Abstract', 'Full Text', 'First Year', and 'Last Year'. There are also checkboxes for database sources: Medline (1999 - 2009), NIH (1961 - 2002), NSF (1985 - 2004), and USPTO (1976 - 2007). A 'Search' button is at the bottom of the search area. On the right side of the search area, there is explanatory text about search syntax, such as using 'OR' for multiple terms and 'AND' to combine terms. At the bottom of the page, there are logos for Indiana University School of Library and Information Science and the James S. McDonnell Foundation.

Register for free access at <http://sdb.cns.iu.edu>

The screenshot shows the Scholarly Database search results page in a Mozilla Firefox browser. The browser's address bar shows the URL: [http://sdb.sls.indiana.edu/search/results?q=\(\"artificial intelligence\"\)](http://sdb.sls.indiana.edu/search/results?q=(\). The page header includes the Scholarly Database logo and the text 'Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Bloomington'. Below the header, there are navigation links: Search, Edit Profile, Admin, About, and Logout. The main content area is titled 'Browse Results' and displays the search results for 'artificial intelligence'. It shows that the search returned 13,231 results in 0.295 seconds. A 'Download' button is visible. Below this, it lists the total results per database: NIH: 2,103, Medline: 10,235, USPTO: 279, NSF: 614. The results are displayed in a table with columns for Source, Authors/Creators, Year, Title, and Score (out of 5.71). The first 20 results are shown, with a 'Next >>' link. The table lists several results from Medline, including 'Artificial intelligence' (1987), 'Artificial intelligence: expert systems.' (1989), 'Artificial intelligence in dentistry' (1990), 'Artificial-intelligence-augmented systems.' (2002), 'Artificial intelligence.' (1980), and 'Artificial intelligence.' (1980).

Source	Authors/Creators	Year	Title	Score (out of 5.71)
Medline	LaCombe	1987	Artificial intelligence.	5.71
Medline		1989	Artificial intelligence: expert systems.	5.71
Medline	Schmitt	1990	Artificial intelligence in dentistry	5.71
Medline	Adlassnig and Adlassnig	2002	Artificial-intelligence-augmented systems.	5.60
Medline	Touretzky	1980	Artificial intelligence.	4.86
Medline	Goldenberg	1980	Artificial intelligence.	4.86

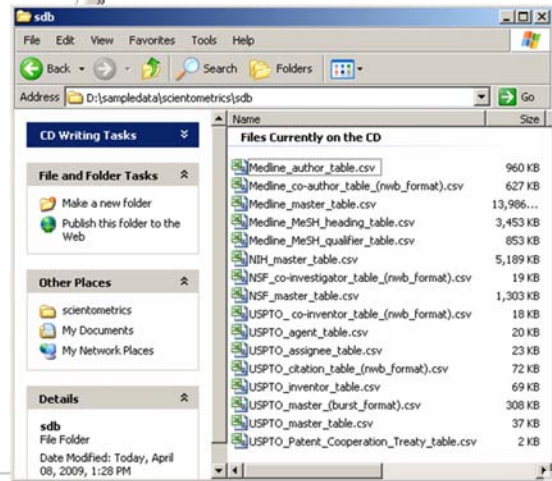


Since March 2009:

Users can download networks:

- Co-author
- Co-investigator
- Co-inventor
- Patent citation

and tables for burst analysis in NWB.

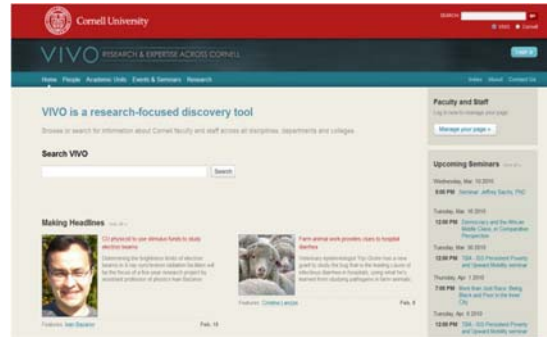


29

VIVO International Researcher  
Network

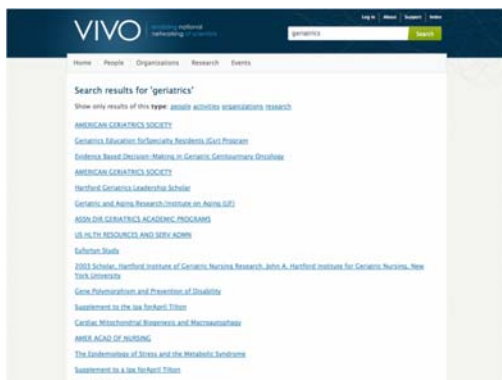
## VIVO: A Semantic Approach to Creating a National Network of Researchers (<http://vivoweb.org>)

- 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 Ferreira, 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 Ping, 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, Michaelen 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.





### University of Florida

How do you want to compare?  
by Grants

Who do you want to compare?  
Search:  X

Records 1 - 10 of 30

Entity Label	Grant Count	Entity Type
<input checked="" type="checkbox"/> Continuing Education	562	UF Department, Agent, Non-Academic Department, Department
<input checked="" type="checkbox"/> Florida Museum of Natural History	203	Museum, Agent
<input checked="" type="checkbox"/> College of Agricultural and Life Sciences	166	Agent, UF College, College
<input checked="" type="checkbox"/> College of Engineering	103	Agent, UF College, College
<input checked="" type="checkbox"/> Evelyn F. and William L. McKnight Brain Institute of the University of Florida	64	UF Center, Agent, Center
<input checked="" type="checkbox"/> International Center	54	UF Department, Agent, Non-Academic Department, Department
<input checked="" type="checkbox"/> Florida Sea Grant	44	UF Center, Agent, Center
<input type="checkbox"/> Whitney Laboratory for Marine Bioscience	42	UF Research Laboratory, Agent, Laboratory, Research Laboratory
<input type="checkbox"/> Water Institute	38	UF Center, Agent, Center
<input type="checkbox"/> College of Dentistry	35	Agent, UF College, College

[Save as CSV](#) [Clear](#)

### Comparing Grants of Organizations in University of Florida

Total Number of Grants

You have selected 7 of a maximum 10 organizations to compare. [Clear](#)

- Florida Sea Grant 44
- International Center 54
- Evelyn F. and William L. McKnight Brain Institute of the University of Florida 64
- College of Engineering 103
- College of Agricultural and Life Sciences 166
- Florida Museum of Natural History 203
- Continuing Education 562

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

enabling national networking of scientists

Index Log in

Search

Home

People

Organizations

Research

Events

### University of Florida

Explore 487 publications activity across 554 scientific sub-disciplines

13 Disciplines | 554 Sub-Disciplines

Search:  X

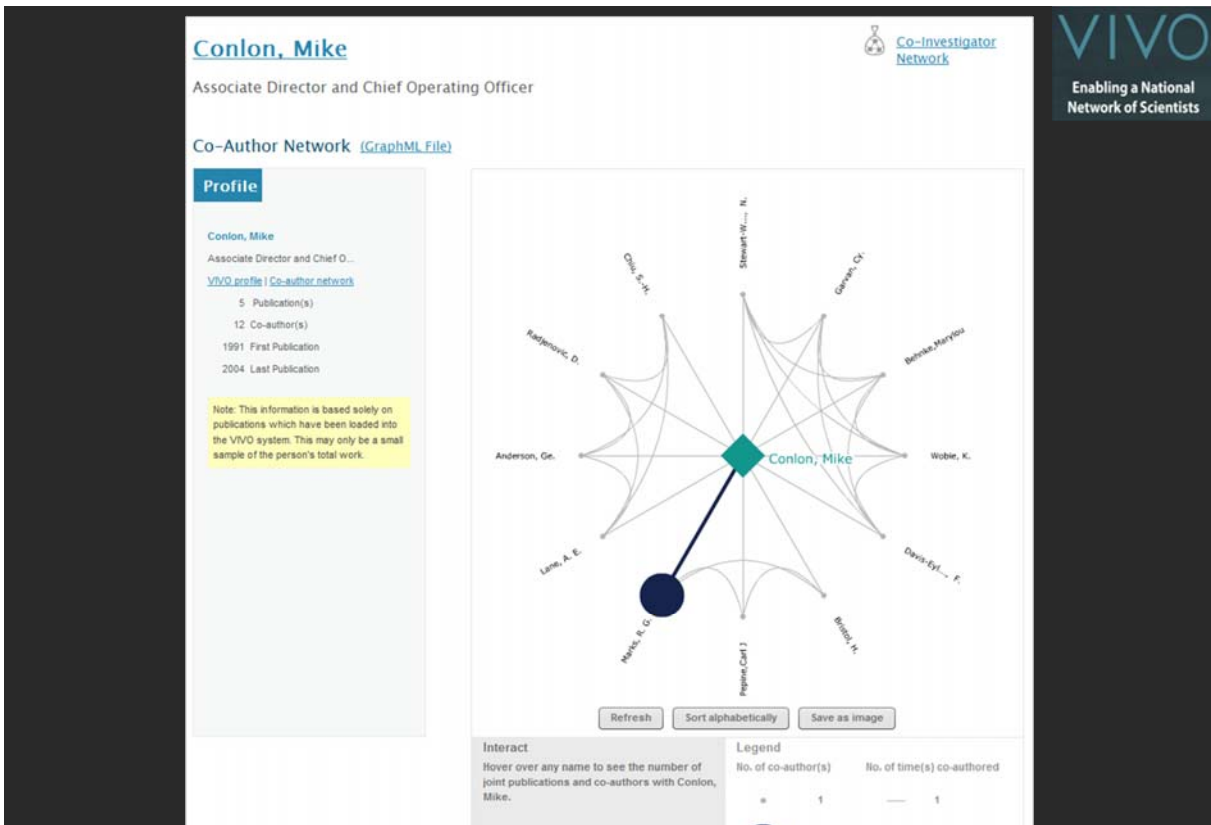
1 - 13 of 554

Sub-Disciplines	# of pubs.	% activity
Pest Management Science	24.2	5.0
Wildlife Research	19.1	3.9
Protein Science	13.1	2.7
Clinical Cancer Research	12.6	2.6
Pain	12.0	2.5
Environmental Contamination	11.2	2.3
Insect Physiology	11.1	2.3
Organic Chemistry	10.9	2.2
Marine Biology	10.3	2.1
Computer Aided Molecular Design	10.2	2.1
BioStatistics	9.0	1.9

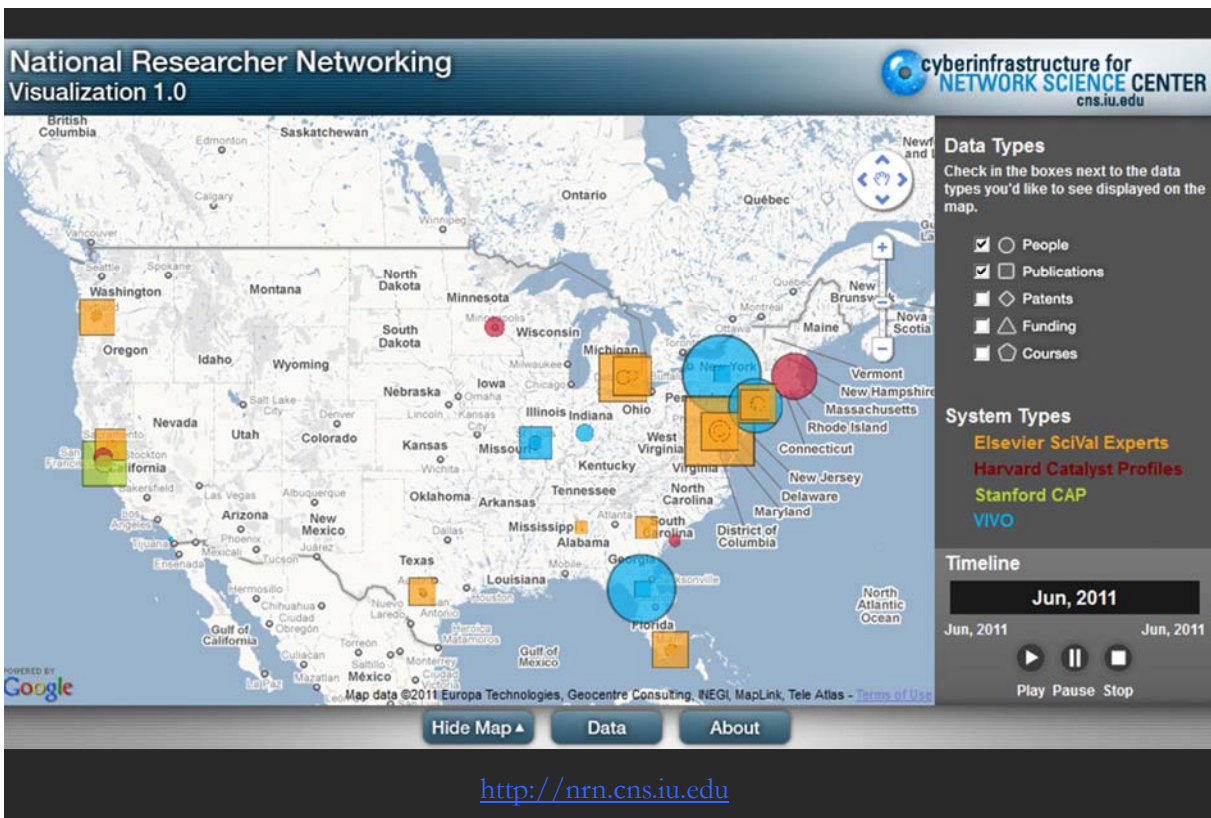
Top 290 disciplines shown

mapped 14.55% of 3,346 publications

**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? 35



**Geospatial Analysis (Where)** Where is what science performed by whom? Science is global and needs to be studied globally. 36

## VIVO On-The-Go

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



## Develop VIVO Visualizations

See also *Visualization in VIVO Workshop on Aug 24, 2011*

<http://wiki.cns.in.edu/display/PRES/VIVO+Presentation>



### VIVO Presentation

4 Added by Chin Hua Kong, last edited by Chintan Tank on Aug 24, 2011 (view change)

#### August, 2011 Workshop

##### Material

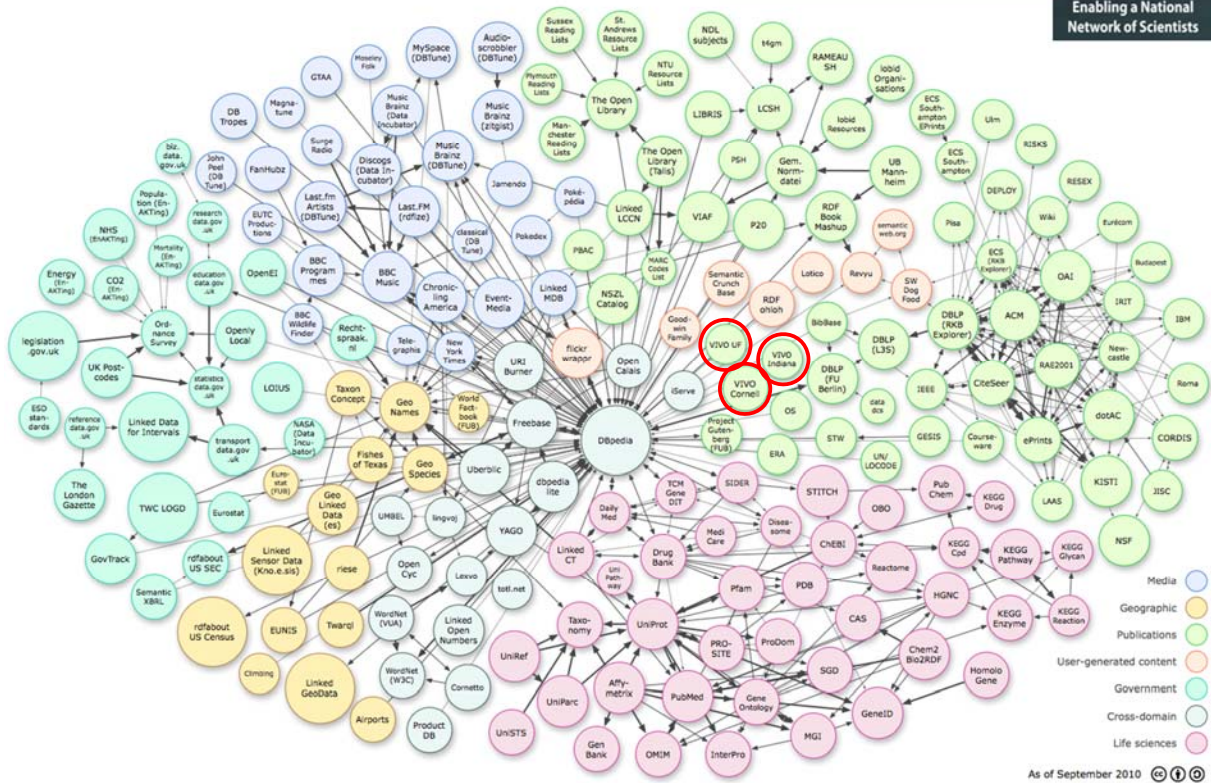
- [Java 1.5 or higher](#) - A programming language and computing platform for developing cross OS softwares.
- [Science of Science tool \(Sci2\)](#) - An desktop application for information analysis and visualization.
- [Gephi](#) - An interactive visualization tool for networks and complex systems, dynamic and hierarchical graphs.
- [VIVO August 2011 workshop data.zip](#) - Hands on workshop data package

##### Slides

- [Tutorial Slides](#) presented at the VIVO Conference 2011
- [Pre-Questionnaire](#) and [Post-Questionnaire](#)

##### Demo Links

- [Map of Science Visualization](#) (dev link)
- [Temporal Graph Visualization](#) (dev link)
- [National Researcher Networking Visualization](#)
- [Word Cloud Visualization](#) dev link



# Mapping Sustainability Research

Geographic Map Science Map



Funding  
 NIH  
 NSF  
 USDA

Publications  
 DOE  
 ISI  
 Medline

Patents  
 USPTO

Citations  Count

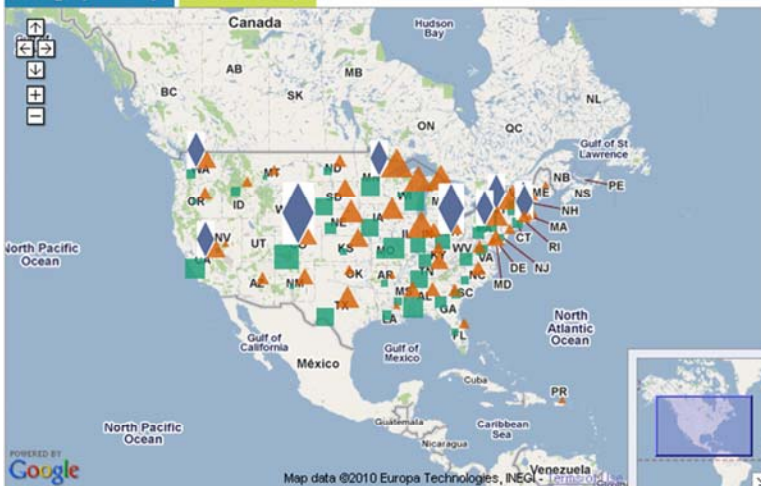
Amount  Count

From year 1901 to year 2009

Search by keyword

<http://mapsustain.cns.iu.edu>

Geographic Map Science Map



Funding  Patents  
 NIH  USPTO  
 NSF  ISI  
 USDA  Medline

Publications  
 DOE  
 ISI  
 Medline

Maps Detail **Data** About

### Datasets

The dataset covers 13,528 records on "biomass" and "biofuel" research and technology from seven different publication, patent, funding datasets for the years 1901 to 2010.

### Funding

National Institutes of Health (NIH) awards retrieved from the Scholarly Database (<http://sdb.slis.indiana.edu>) at Indiana University on 11/20/2010. Search query used was biomass OR biofuel OR "bio mass" OR "bio fuel" in the 'All Text' field.

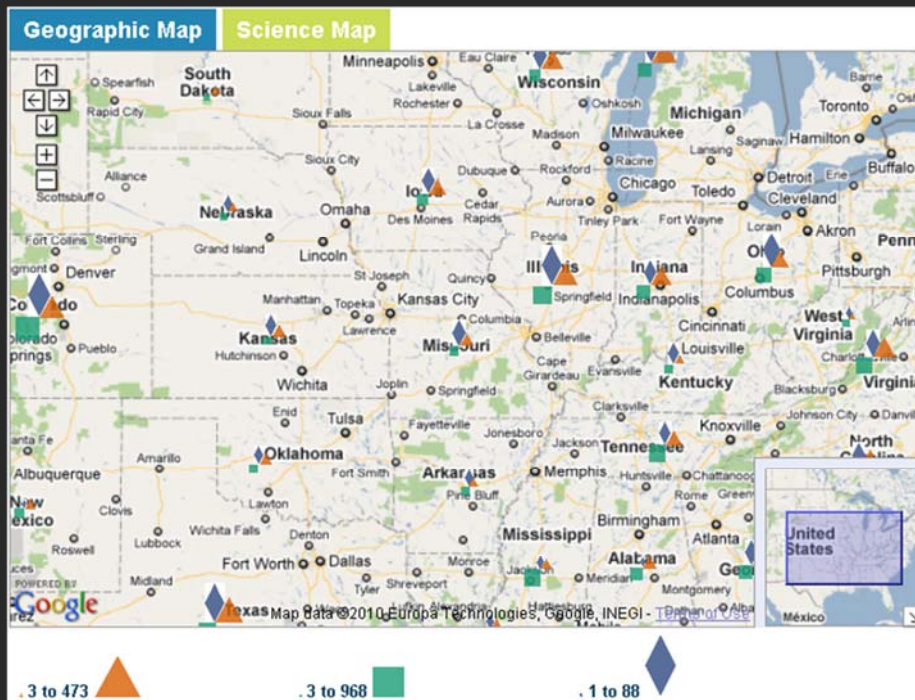
National Science Foundation (NSF) awards retrieved from the Scholarly Database (<http://sdb.slis.indiana.edu>) at Indiana University on 11/20/2010. Search query used was biomass OR biofuel OR "bio mass" OR "bio fuel" in the 'All Text' field.

US Department of Agriculture (USDA) awards made available by a staff member of the Office of Scientific and Technical Information from the US Department of Energy (DOE).

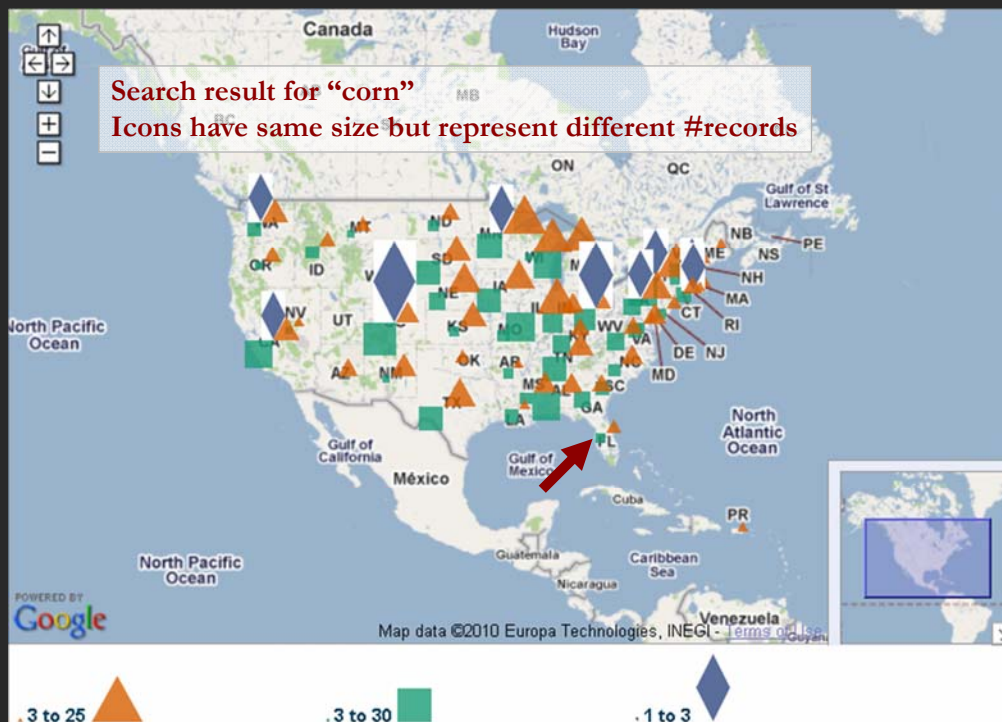
### Publications

MEDLINE papers by the National Library of Medicine retrieved from the Scholarly Database (<http://sdb.slis.indiana.edu>) at Indiana University on 11/20/2010. Search query used was biomass OR biofuel OR "bio mass" OR "bio fuel" in the 'All Text' field.

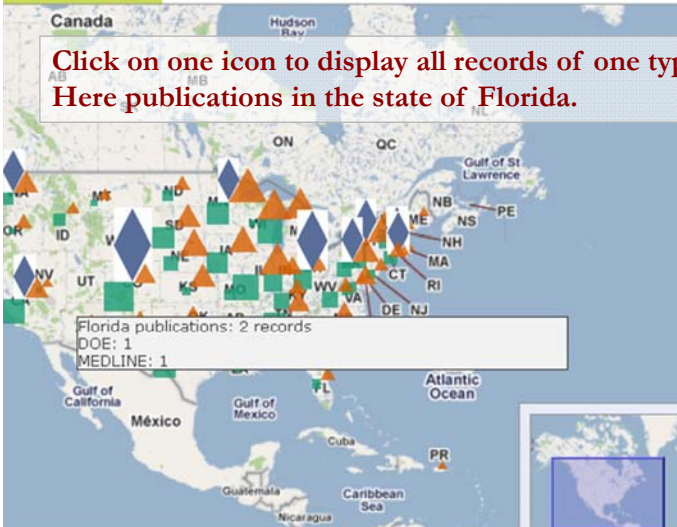
The geographic map at state level.



The geographic map at city level.



## Science Map



## 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](#)

Information Bridge: DOE Scientific and Technical Information - - Document #5789929 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.osti.gov/bridge/product.biblio.jsp?osti\_id=5789929

Most Visited Getting Started Latest Headlines

MapSustain Information Bridge: DOE Scientifi...

DOE Scientific and Technical Information

DOE • OSTI

INFORMATION BRIDGE

Home • Basic Search • Fielded Search • Alerts • Help

FAQ • Widget • Site Map

SHARE

**Bibliographic Citation**

[See/Add Document Discussions](#) [Return to Search Results](#) [Return to Original Search Page](#) [Download as EndNote](#)

**Full Text** Availability information may be found in the Availability, Publisher, Research Organization, Resource Relation and/or Author (affiliation information) fields and/or via the "Full-text Availability" link. For a journal article, please see the Resource Relation field.

**Title** Enzymatic hydrolysis and fermentation of corn for fuel alcohol  
[Word Cloud](#) [More Like This](#)

**Creator/Author** [Mullins, J.T.](#)

**Publication Date** 1985 Jan 01

**OSTI Identifier** OSTI ID: 5789929

**Other Number(s)** Journal ID: CODEN: BIBA

**Resource Type** Journal Article

**Resource Relation** Journal Name: Biotechnol. Bioeng.; (United States); Journal Volume: 27:3

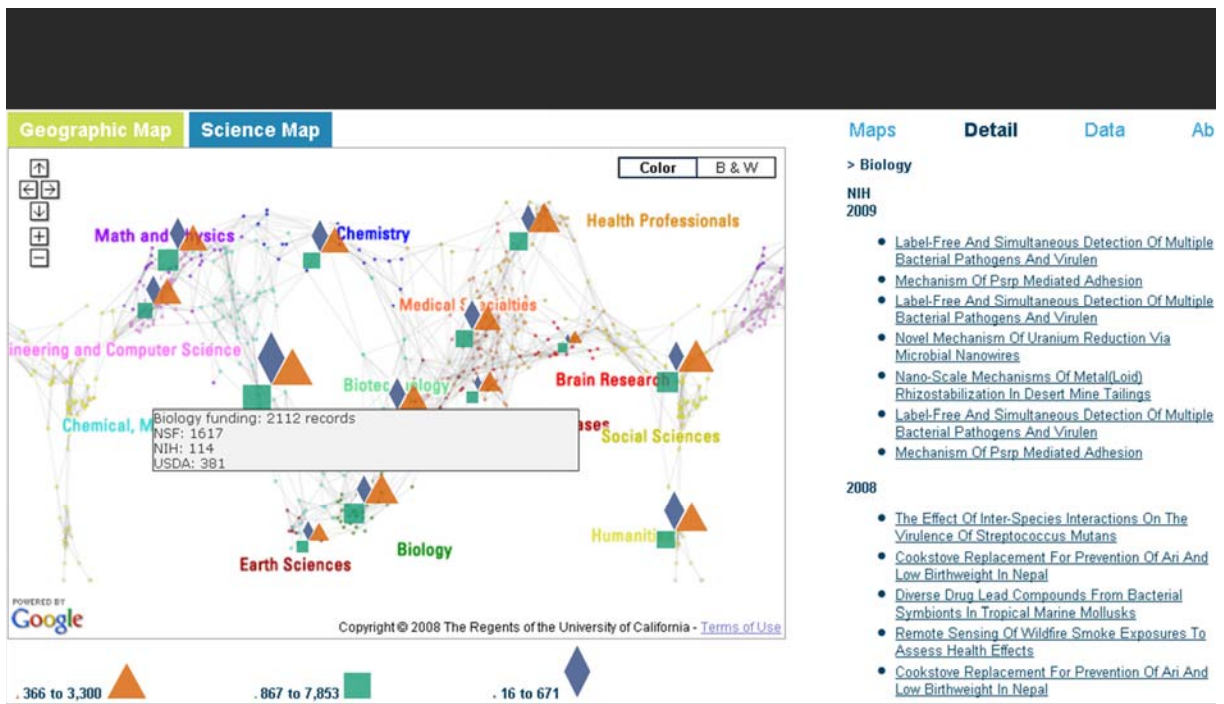
**Research Org** Univ. of Florida, Gainesville

**Subject** 09 BIOMASS FUELS; 32 ENERGY CONSERVATION, CONSUMPTION, AND UTILIZATION; ETHANOL FUELS; BIOSYNTHESIS; MAIZE; ENZYMATIC HYDROLYSIS; FERMENTATION; PRODUCTIVITY; COST; ENERGY EFFICIENCY; EXPERIMENTAL DATA; WASTE PRODUCT UTILIZATION; ALCOHOL FUELS; BIOCONVERSION; CEREALES; CHEMICAL REACTIONS; DATA; DECOMPOSITION; EFFICIENCY; FUEL; GLASS; HYDROLYSIS; INFORMATION; LYSIS; NUMERICAL DATA; PLANTS;

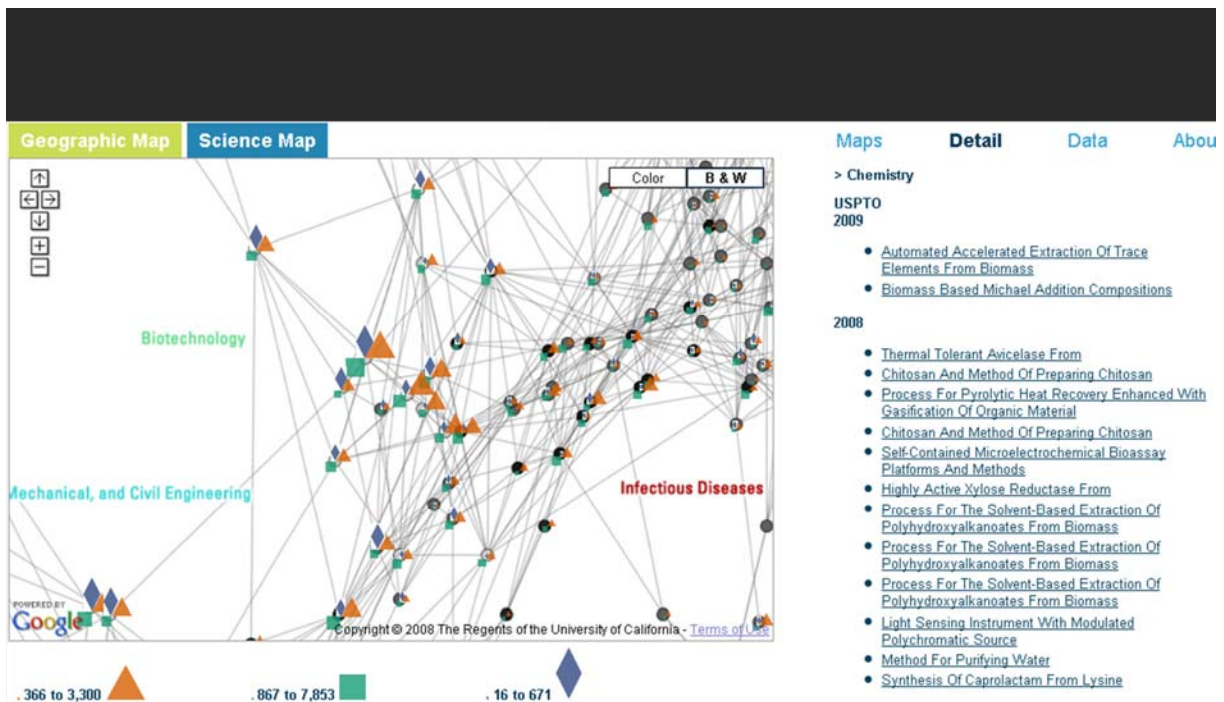
Done

**Detailed information on demand via original source site for exploration and study.**

46

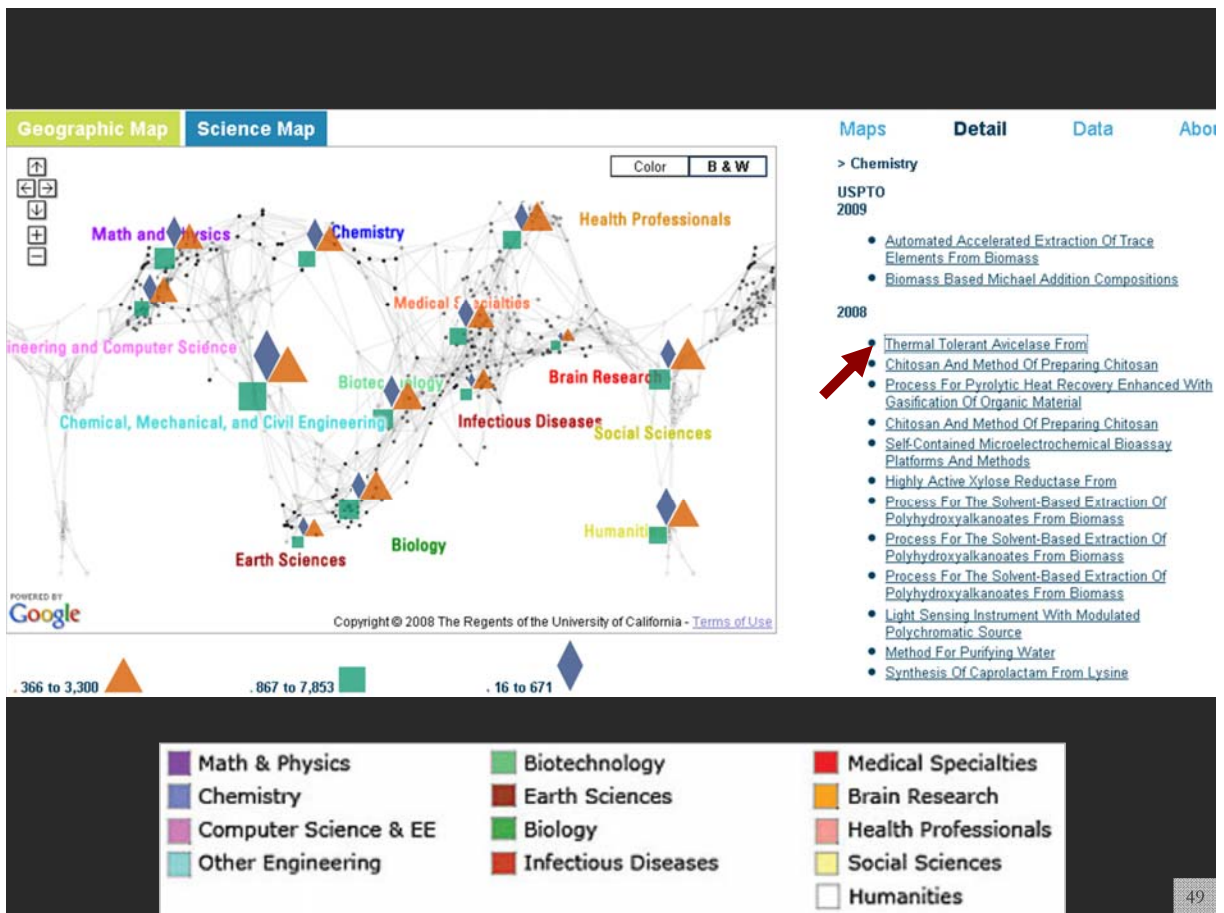


The science map at 13 top-level scientific disciplines level.



The science map at 554 sub-disciplines level.





Maps Detail Data About

> Chemistry

USPTO

2009

- Automated Accelerated Extraction Of Trace Elements From Biomass
- Biomass Based Michael Addition Compositions

2008

- Thermal Tolerant Avicelase From Chitosan And Method Of Preparing Chitosan
- Process For Pyrolytic Heat Recovery Enhanced With Gasification Of Organic Material
- Chitosan And Method Of Preparing Chitosan
- Self-Contained Microelectrochemical Bioassay Platforms And Methods
- Highly Active Xylose Reductase From Process For The Solvent-Based Extraction Of Polyhydroxyalkanoates From Biomass
- Process For The Solvent-Based Extraction Of Polyhydroxyalkanoates From Biomass
- Process For The Solvent-Based Extraction Of Polyhydroxyalkanoates From Biomass
- Light Sensing Instrument With Modulated Polychromatic Source
- Method For Purifying Water
- Synthesis Of Caprolactam From Lysine

United States Patent: 7364890 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=/netacgi/PTO/srch

MapSustain United States Patent: 7364890 United States Patent: 7364890 Information Bridge: DOE Scientific a...

**USPTO PATENT FULL-TEXT AND IMAGE DATABASE**

Home Quick Advanced Pat Num Help

Bottom

View Cart Add to Cart

Images

(1 of 1)

United States Patent 7,364,890  
Ding, et al. April 29, 2008

**Thermal tolerant avicelase from *Acidothermus cellulolyticus***

**Abstract**

The invention provides a thermal tolerant (thermostable) cellulase, AvIII, that is a member of the glycoside hydrolase (GH) family. AvIII was isolated and characterized from *Acidothermus cellulolyticus* and, like many cellulases, the disclosed polypeptide and/or its derivatives may be useful for the conversion of biomass into biofuels and chemicals.

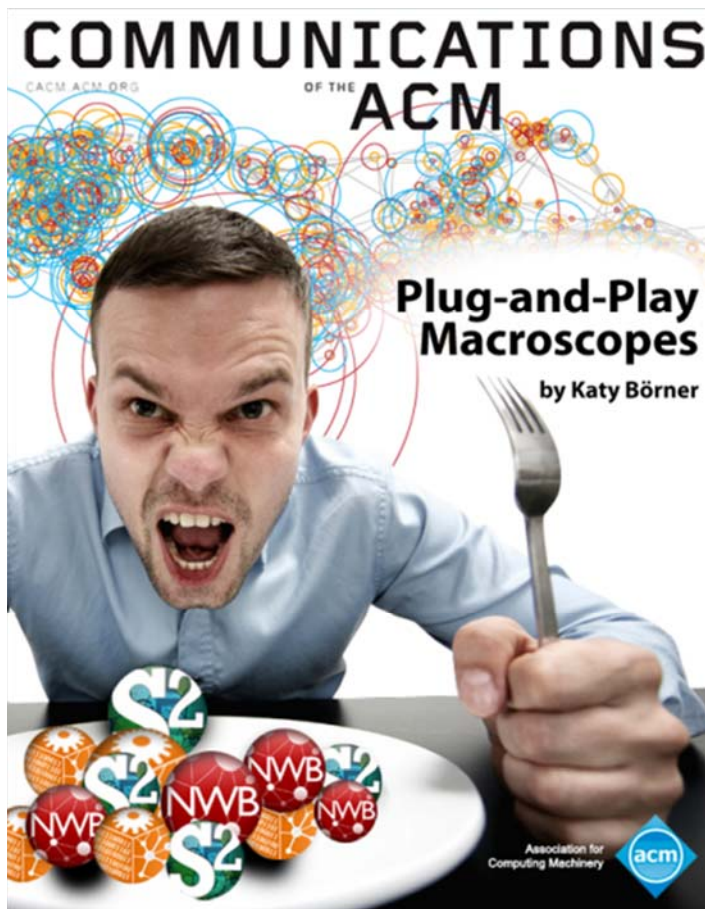
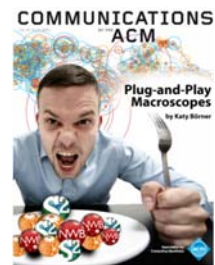
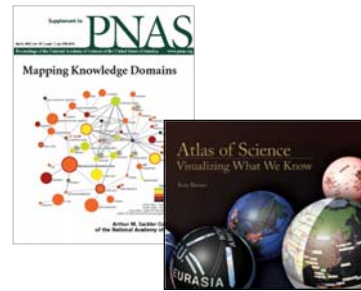
Inventors: Ding; Shi-You (Golden, CO), Adney; William S. (Golden, CO), Vinzant; Todd B. (Golden, CO), Himmel; Michael E. (Littleton, CO)  
Assignee: Midwest Research Institute (Kansas City, MO)  
Appl. No.: 09/017,276

Done

50

## Overview

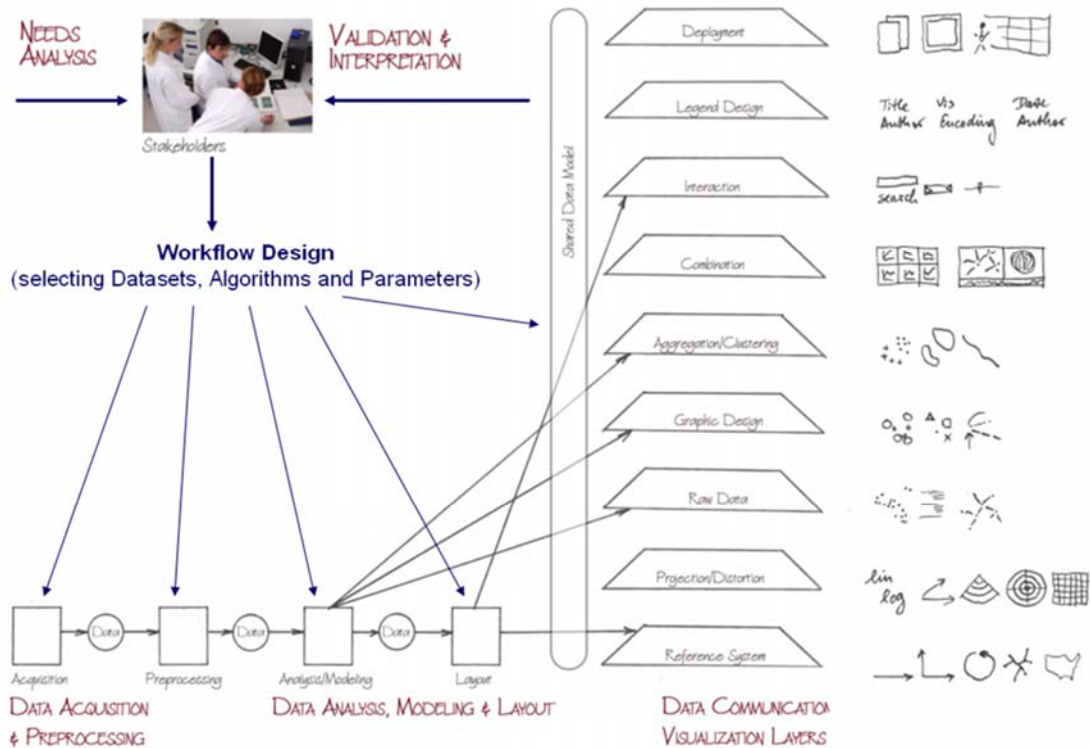
1. **Data + data mining and visualization research** that aims to increase our scientific understanding of the structure and dynamics of science and technology.
2. **Novel approaches and services** that improve information access, researcher networking, and research management.
3. **Data services and plug-and-play macroscope tools** that commoditize data mining and visualization.



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

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

**Needs-Driven Workflow Design** using a modular data acquisition/analysis/ modeling/ visualization pipeline as well as modular visualization layers.

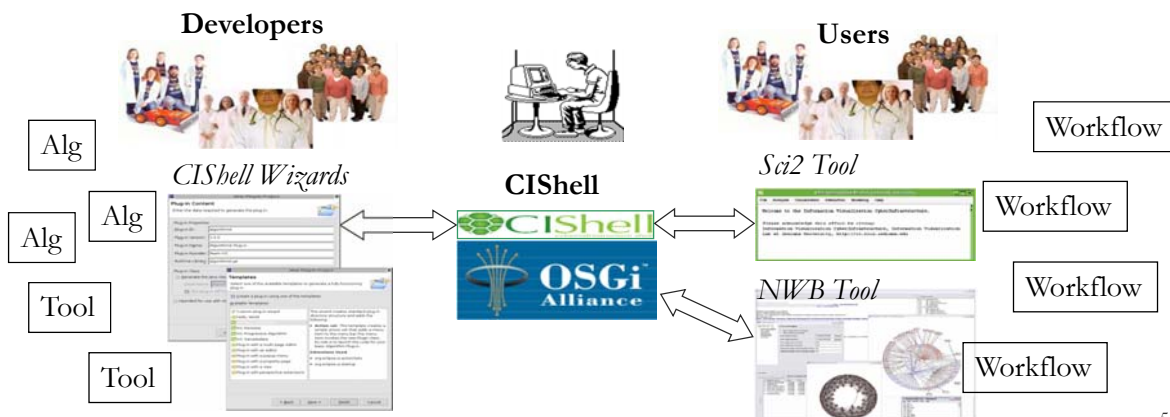


Börner, Katy (2010) *Atlas of Science*. MIT Press. 53



OSGi & CIShell

- CIShell (<http://cishell.org>) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (<http://osgi.org>), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.





## CIShell Developer Guide

(<http://cishell.wiki.cns.iu.edu>)



Edit Add ▾

1 Added by [Micah Linnemeier](#), last edited by [Micah Linnemeier](#) on Mar 16, 2011 ([view change](#))

### About the Cyberinfrastructure Shell

The Cyberinfrastructure Shell (CIShell) is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithm integration support is built in for Java and most other programming languages. Being Java based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License.

CIShell is the basis of [Network Workbench](#), [TexTrend](#), [Sci<sup>2</sup>](#) and the upcoming [EpiC](#) tool.

CIShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

### CIShell Features

**A framework for easy integration of new and existing algorithms written in any programming language**

Using CIShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and

### Learn More...

- [CIShell Papers](#)
- [CIShell Powered Tools](#)
- [Algorithms](#)
- [Plugins \(coming soon\)](#)
- [Misc. Tool Documentation](#)
- CIShell Web Services (coming soon)
- [Screenshots](#)

### Getting Started...

- [Documentation & Developer Resources](#)
- [Download](#)

### Getting Involved...

- [Contact Us](#)

55



## CIShell Portal (<http://cishell.org>)

56

The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

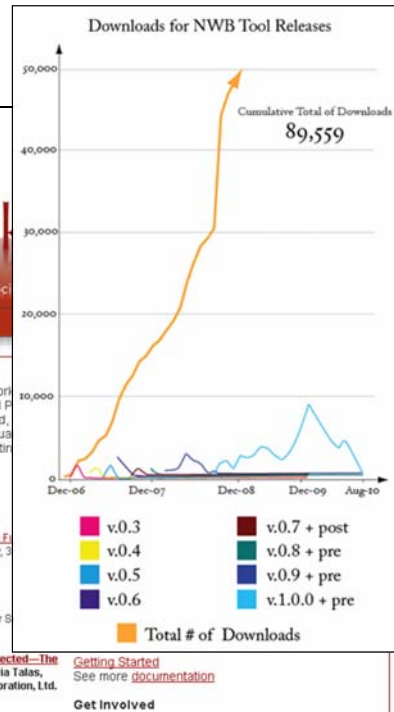
In February 2009, the tool provides more 169 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

**More than 50 of these plugins can be applied or were specifically designed for S&T studies.**

It has been downloaded more than 110,000 times since December 2006.



The screenshot shows the Network Workbench website with a navigation menu (Home, People, Research, Publications) and a 'Summary' section. The summary describes it as a 'Large-Scale Network Toolkit for Biomedical, Social Science and P...' and lists 'News & Updates' with dates and titles.



Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179), *Annals of the New York Academy of Sciences*, Boston, MA.

## Computational Proteomics

What relationships exist between protein targets of all drugs and all disease-gene products in the human protein-protein interaction network?

Yildirim, Muhammed A., Kwan-II Goh, Michael E. Cusick, Albert-László Barabási, and Marc Vidal. (2007). *Drug-target Network*. *Nature Biotechnology* 25 no. 10: 1119-1126.



**Figure 2** Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.

## Computational Economics

Does the type of product that a country exports matter for subsequent economic performance?

C. A. Hidalgo, B. Klinger,  
A.-L. Barabási, R. Hausmann  
(2007) *The Product Space  
Conditions the Development  
of Nations. Science* 317,  
482 (2007).

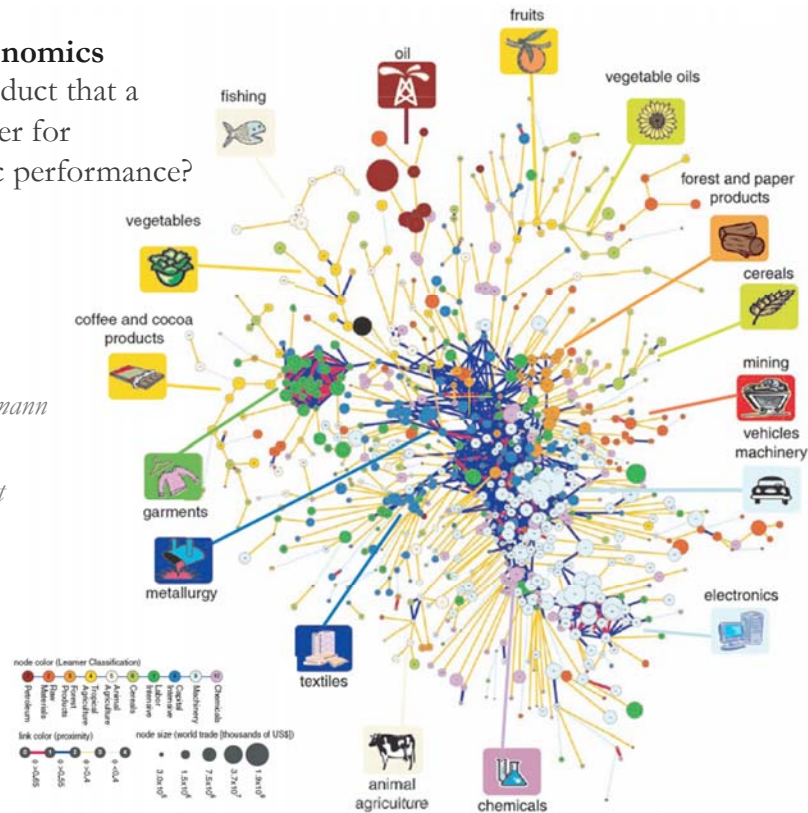


Fig. 1. The product space. (A) Hierarchically clustered proximity matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world trade, and their colors are chosen according to the classification introduced by Leamer.

59

## Computational Social Science

Studying large scale social networks such as Wikipedia

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

<http://www.newscientist.com/newscientist/4277>



## Second sight

Image: Bruce W. Heer and Todd M. Hollaway

### Power struggle

How do you keep track of the bubbling mass of information that is Wikipedia? This chaotic-looking mosaic is one attempt to show which topics are contained in the online encyclopedia.

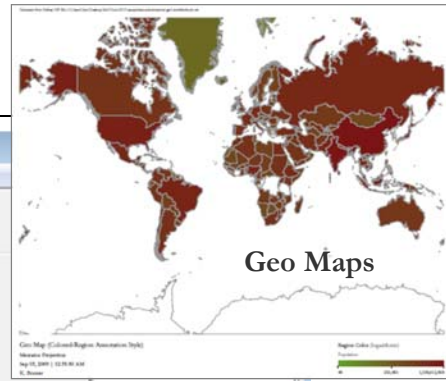


...ed since the most likely topics pages at the time of writing include entries on Sheffield Wednesday football club, Mikhail Gorbachev and pigs). The mosaic has been commended in a competition for images that visualise network dynamics, coinciding with this week's International Workshop and Conference on Network Science in Bloomington.





## Sci² Tool Vis cont.

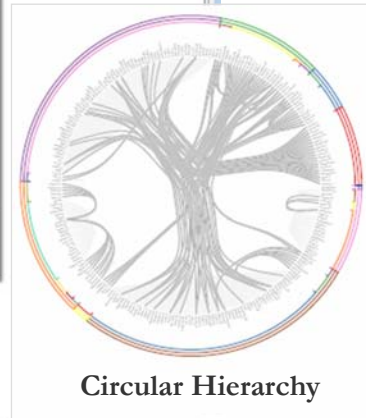


Geo Maps

The screenshot shows the Sci² Tool interface with the 'Visualization' menu open. The menu items are: GUESS, GnuPlot, Radial Tree/Graph (prefuse alpha), Radial Tree/Graph with Annotation (prefuse beta), Tree View (prefuse beta), Tree Map (prefuse beta), Force Directed with Annotation (prefuse beta), Fruchterman-Reingold with Annotation (prefuse beta), DrL (VxOrd), Specified (prefuse beta), Horizontal Line Graph, Circular Hierarchy, Geo Map (circle annotations), Geo Map (region coloring annotations), Image Viewer, and RefMapper.

Below the menu is a table with the following data:

!	Algorithm Name	Date	Time	% Con
<input checked="" type="checkbox"/>	Extract Co-Author Netw...	09/03/2009	00:15:20 AM	<div style="width: 100%;"></div>
<input checked="" type="checkbox"/>	Load and Clean ISI File	09/03/2009	00:15:05 AM	<div style="width: 100%;"></div>



Circular Hierarchy



## Science of Science (Sci²) Tool – Usage

The Sci² Tool is used by the

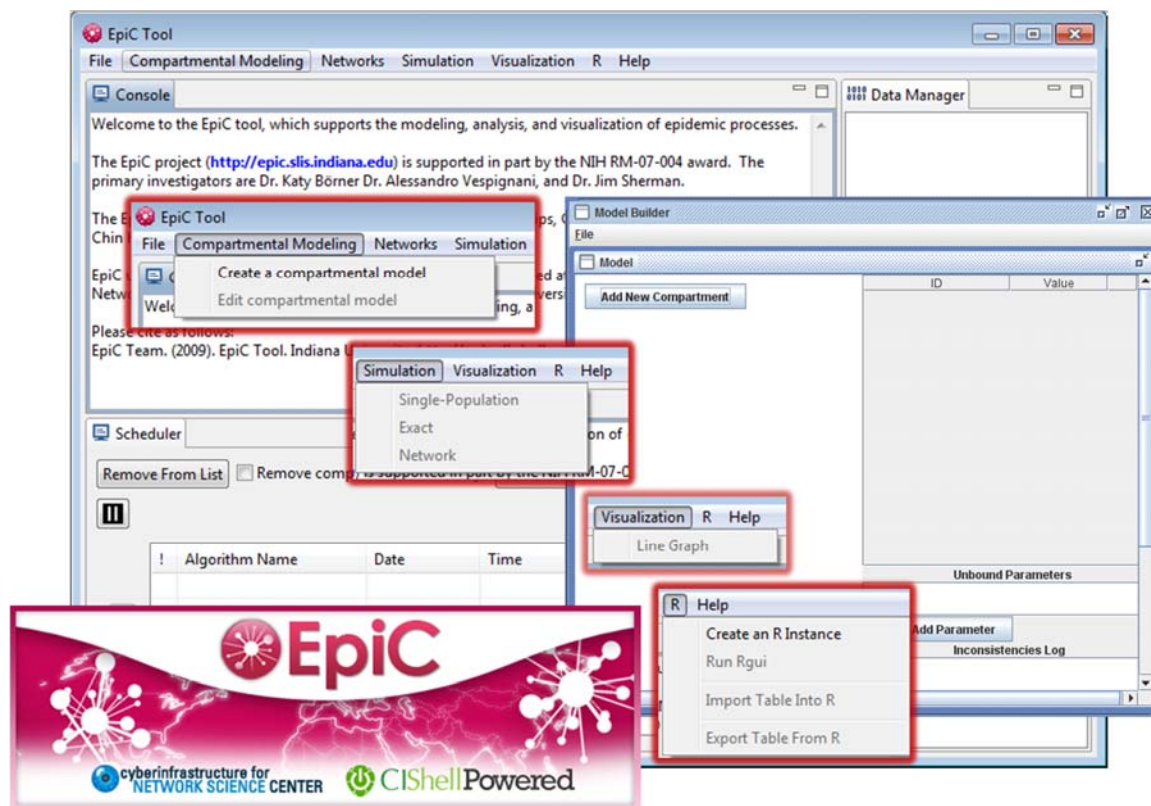
- National Science Foundation,
- National Institutes of Health,
- US Department of Agriculture, and
- National Oceanic and Atmospheric Administration



Tool registrations come from 73 countries and professions such as







65



## OSGi/CIShell Adoption

A number of other projects recently adopted OSGi and/or CIShell:

USA

- *Cytoscape* (<http://cytoscape.org>) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).
- *MAEviz* (<https://wiki.ncsa.uiuc.edu/display/MAE/Home>) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

Europe

- *Taverna Workbench* (<http://taverna.org.uk>) Developed by the myGrid team (<http://mygrid.org.uk>) led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.
- *TEXTrend* (<http://textrend.org>) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.
- *DynaNets* (<http://www.dynanets.org>) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
- *SISOB* (<http://sisob.lcc.uma.es>) An Observatory for Science in Society Based in Social Models.

As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.

66



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>