

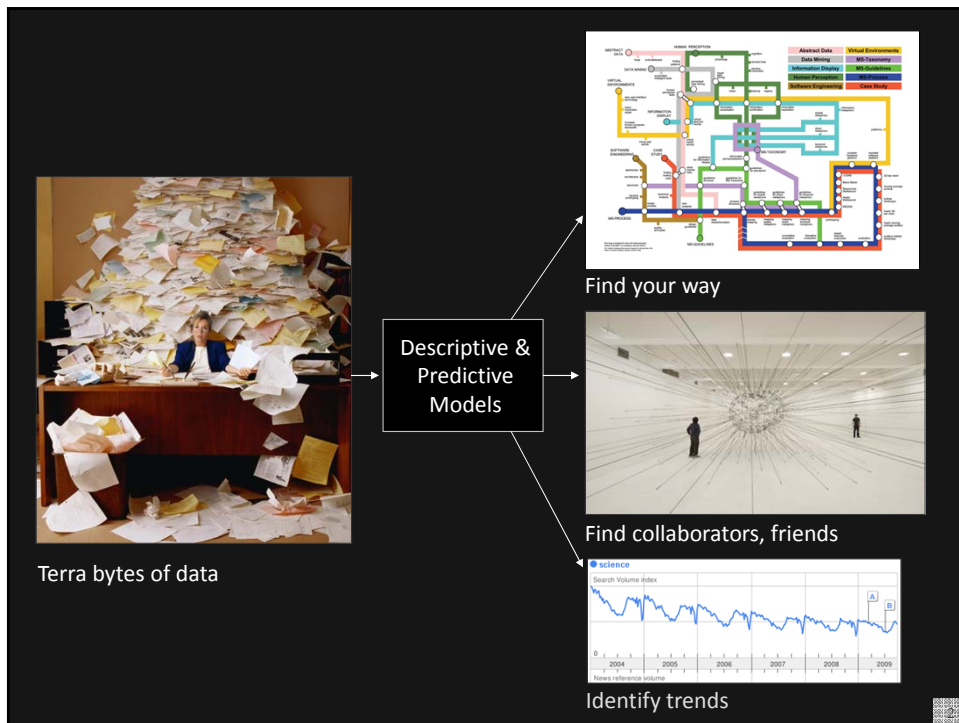
# Scalable Multi-Scale Visual Analytical Tools for NCATS

Katy Börner

Victor H. Yngve Professor of Information Science  
Director, Cyberinfrastructure for Network Science Center  
School of Informatics and Computing, Indiana University, USA

Washington DC  
April 2, 2015

*Language Communities of Twitter - Eric Fischer - 2012*



### **Scalable Multi-Scale Visual Analytical Tools for NCATS— Questionnaire Results**

Number of responses: 5 NIH, 4 NCATS, possible overlap

Gender: 5 female, 4 male

Median age: 38 (average Twitter user is 37.3 years old)

Academic background: Microbiology, bioinformatics, informatics,  
computer science, epidemiology, pharma,  
physics, physician

Native Language: 9 English

3

### **Scalable Multi-Scale Visual Analytical Tools for NCATS— Questionnaire Results**

#### **NIH: In your daily work, what datasets do you use?**

Medline Publications, NIH Awards, USPTO Patents, Clinical Trials,  
Program Management, bioactivity data, LinkedIn, ResearchGate

#### **NCATS: What systems/databases do you use?**

Medline Publications, eRA (IMPACT, QVR, xTrain), Clinical Trials,  
Excel, APR, LinkedIn, Research Gate

4

## Scalable Multi-Scale Visual Analytical Tools for NCATS— Questionnaire Results

NIH or NCATS	WHEN	WHAT AREAS	WITH WHOM	WHERE
NIH	1	1		
NIH	1	1	1	
NIH		1	1	1
NIH	1	1	1	1
NIH	1	1	1	1
NCATS	1	1	1	1
NCATS	1	1	1	1
NCATS	1	1	1	1
NCATS	1	1	1	
<b>TOTALS</b>	<b>8</b>	<b>9</b>	<b>8</b>	<b>6</b>

In your daily work, what general questions do you need to answer?

- Temporal questions, e.g., WHEN something happens or temporal distributions
- Geospatial questions, e.g., WHERE something happens or geospatial distributions
- Topical questions, e.g., WHAT areas of science are emerging, active, or affected
- Network questions, e.g., WITH WHOM people collaborate or how experts migrate

5

## Scalable Multi-Scale Visual Analytical Tools for NCATS— Questionnaire Results

NCATS/NIH	Bursts of	Diffusion of				
	Tweets/Topics	\$	Experts	Translation	Ideas	Experts
NIH	1	1				
NIH						
NIH	1	1	1			
NIH	1	1	1			
NIH	1	1	1			
NCATS						
NCATS	1	1	1	1	1	1
NCATS	1	1	1	1	1	1
NCATS				1	1	1
<b>TOTALS</b>	<b>6</b>	<b>6</b>	<b>5</b>	<b>3</b>	<b>3</b>	<b>3</b>

Tools now exist that detect bursts of activity, i.e., sudden increases in the usage frequency of words, such as names, institutions, keywords, citation references. How might you use such a tool to improve your daily decision making?

- Identify surges of interest in publication or social media (e.g., Twitter data) on certain topics (e.g., specific diseases)
- Identify sudden changes in the number of dollars spent on a topic (across agencies)
- Identify steep increases in the number of experts that work on a topic

Another set of tools can be used to depict branching and diffusion processes using network visualizations. How might you use such a tool to improve your daily decision making?

- Analyze and depict the translation of research results into bedside advances of health
- Visualize the diffusion of ideas between different areas of science
- Study the circulation of experts between institutions, geolocations, or scientific areas

6

## *Descriptive Models*

*Multiple levels: Micro ... Macro*

*Answering: When? Where? What? With Whom?*



## **Different Levels of Abstraction/Analysis**

Macro/Global  
Population Level



Meso/Local  
Group Level



Micro  
Individual Level

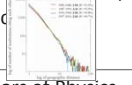
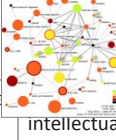





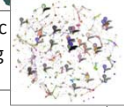

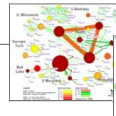
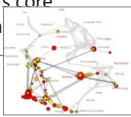


## Type of Analysis vs. Level of Analysis

	<b>Micro/Individual (1-100 records)</b>	<b>Meso/Local (101-100,000 records)</b>	<b>Macro/Global (100,000 &lt; records)</b>
<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

9

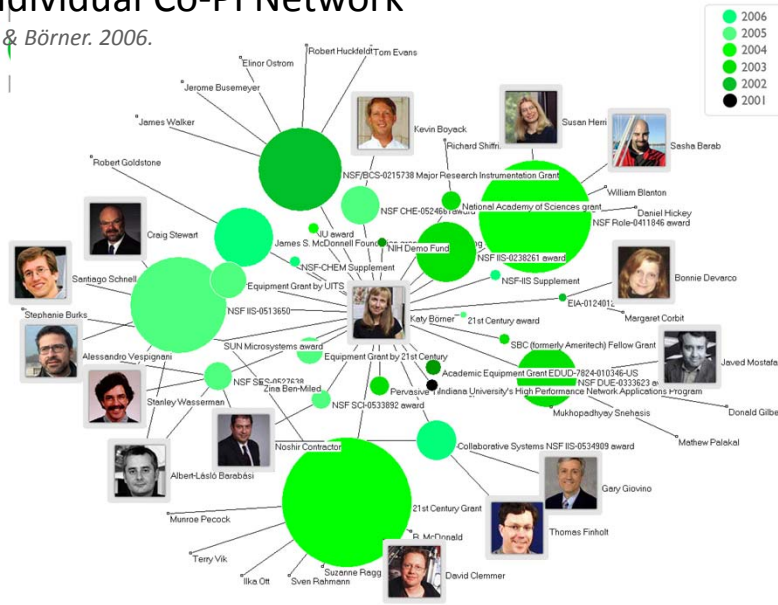
## Type of Analysis vs. Level of Analysis

	<b>Micro/Individual (1-100 records)</b>	<b>Meso/Local (101-100,000 records)</b>	<b>Macro/Global (100,000 &lt; records)</b>
<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 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 

10

# Individual Co-PI Network

Ke & Börner. 2006.

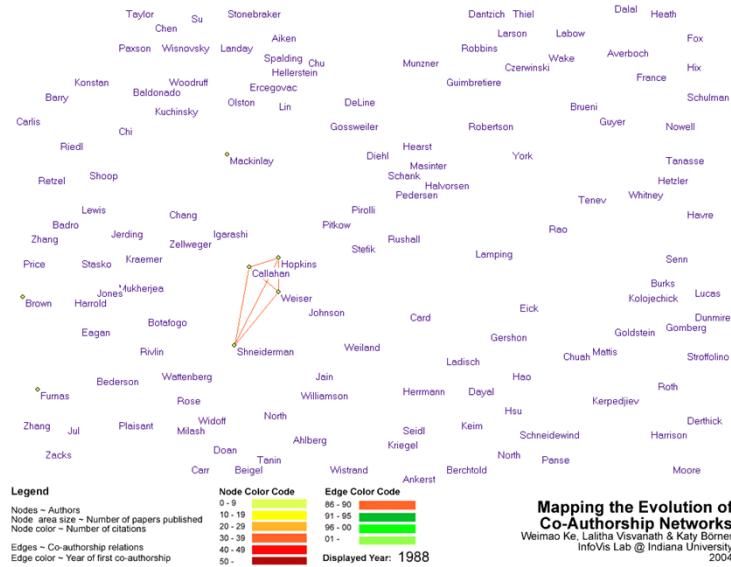


1

11

# Mapping the Evolution of Co-Authorship Networks

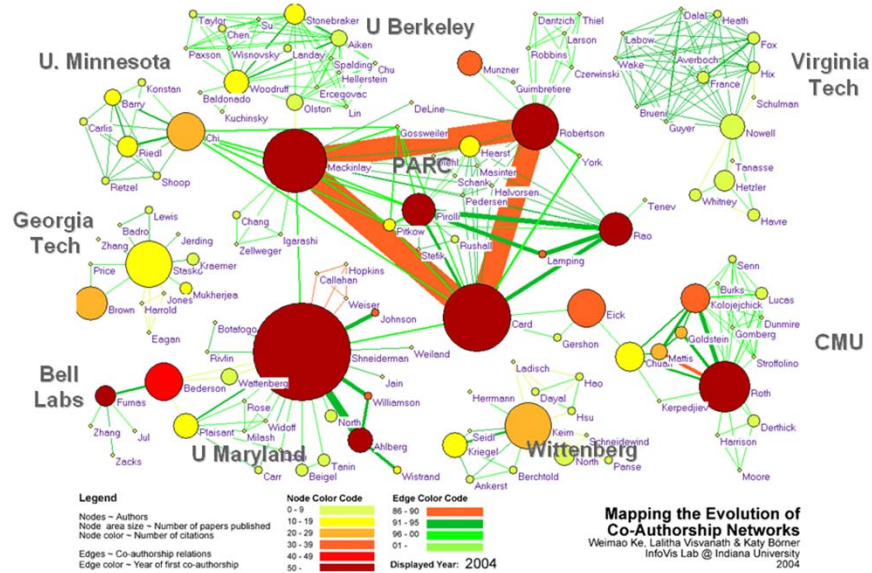
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.



12

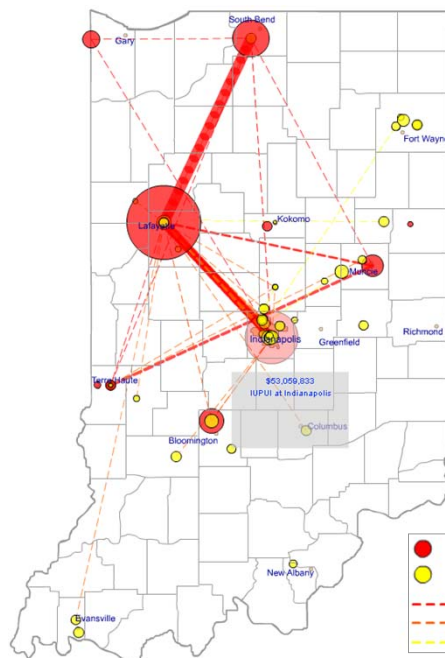
## Mapping the Evolution of Co-Authorship Networks 2

Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.



13

## Mapping Indiana's Intellectual Space



Identify

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

14

## Mapping Transdisciplinary Tobacco Use Research Centers Publications

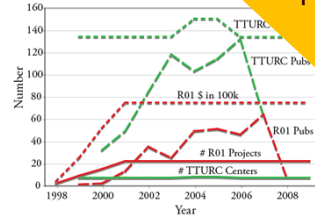
Compare R01 investigator-based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Stipelman, Hall, Zoss, Okamoto, Stokols, Börner, 2014.

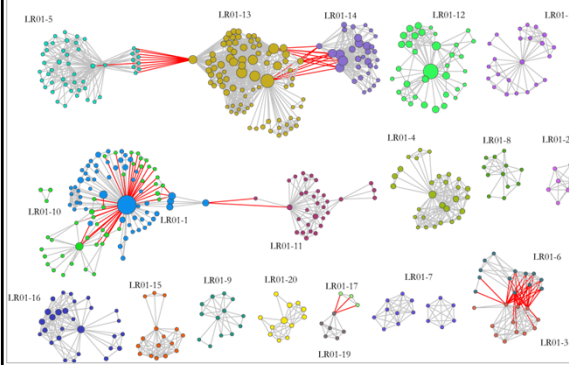
Supported by NIH/NCI Contract HHSN261200800812

R01 & TTURC Project Information

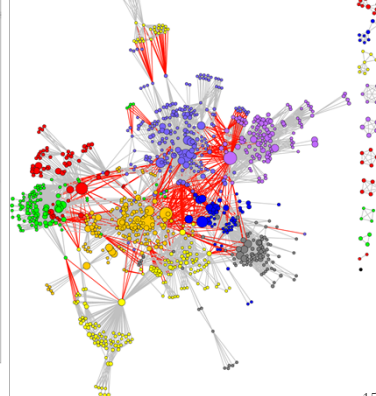
4



Longitudinal R01 Co-Authorship Network



TTURC Co-Authorship Network



15

## Research Collaborations by the Chinese Academy of Sciences

Huang, Duhon, Hardy & Börner

5

北京地区中科院各院所



### 中科院与世界各地的研究合作关系

黄维霞, Russell J. Duhon, Elisha F. Hardy, Katy Börner, Indiana University, USA

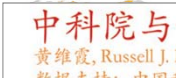
数据支持: 中国科学院国家科学图书馆科学前沿分析中心

金碧辉, 岳焯

甘肃省中科院各院所



上海地区中科院各院所



吉林省中科院各院所

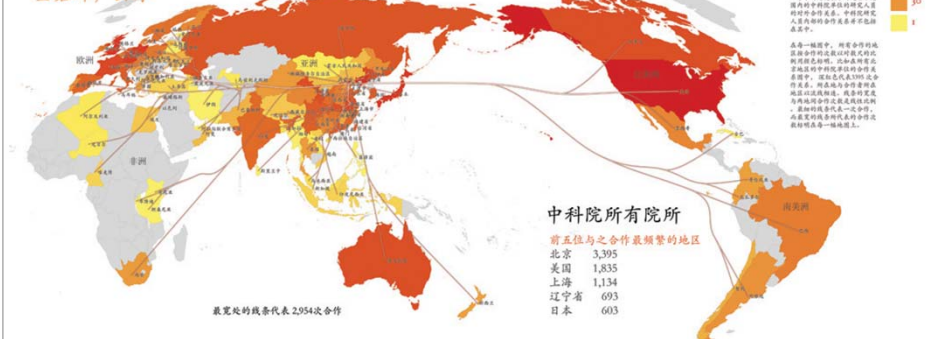


### 中科院与世界各地的研究合作关系

黄维霞, Russell J. Duhon, Elisha F. Hardy, Katy Börner, Indiana University, USA

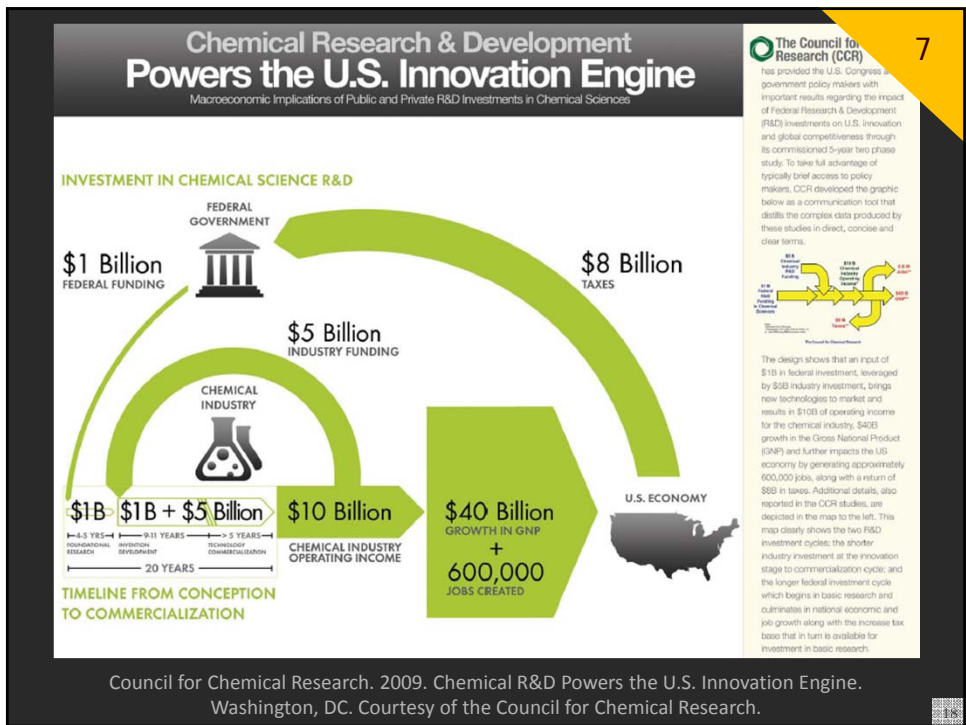
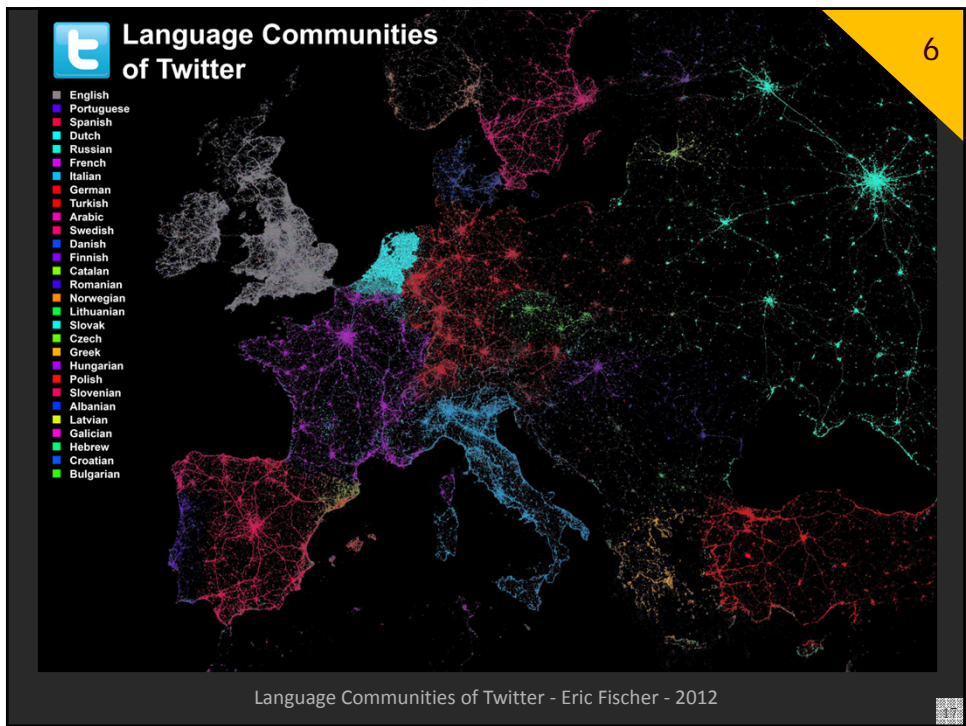
数据支持: 中国科学院国家科学图书馆科学前沿分析中心

金碧辉, 岳焯



16





## *Empowering Anyone to Visualize STI*

Example: The Information Visualization MOOC

Information Visualization MOOC 2015

INDIANA UNIVERSITY CNS

Information Visualization MOOC 2015

### Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

The course can be taken for three Indiana University credits as part of the [Online Data Science Program](#), as part of the Information and Library Science M.S. program, and as part of the online Data Science M.S. Program offered by the School of Informatics and Computing. Students seeking enrollment information should contact Rhonda Spencer at 812-855-2018, [ilsmain@indiana.edu](mailto:ilsmain@indiana.edu) or [datasci@indiana.edu](mailto:datasci@indiana.edu).

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.

Information Visualization MOOC

[Register for Course](#)

Already registered? [Click here to go to the course.](#)  
Forgot your password? [Click here to reset it.](#)

Register for free at <http://ivmooc.cns.iu.edu>. Class restarted in January 13, 2015.

20

## Course Schedule

### Part 1: Theory and Hands-On

- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

### Mid-Term

- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

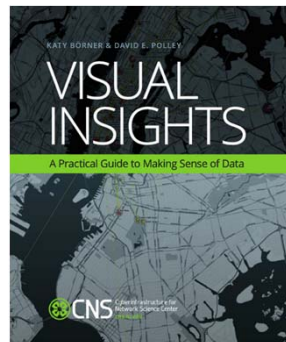
### Final Exam

### Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

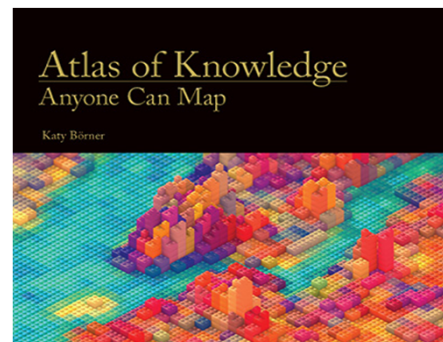


## Books Used in the IVMOOC





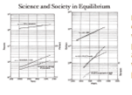


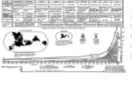
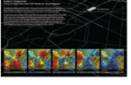
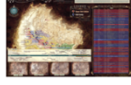







### Teaches timely knowledge:

Advanced algorithms, tools, and hands-on workflows.



### Teaches timeless knowledge:

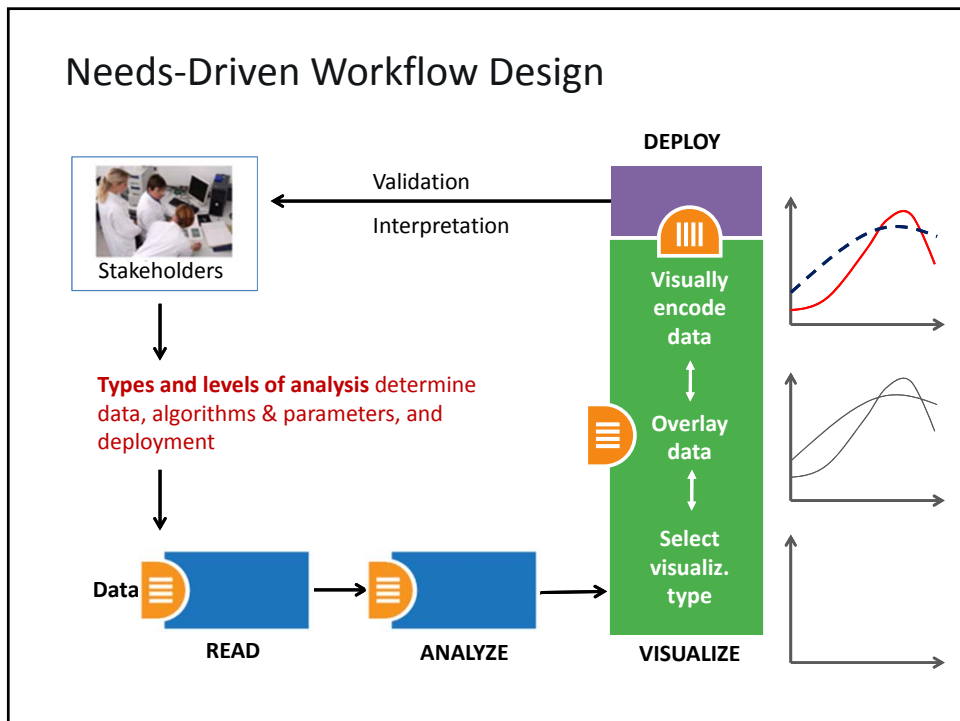
Visualization framework—exemplified using generic visualization examples and pioneering visualizations.

Tasks	LEVELS		
	MICRO: Individual Level about 1–1,000 records page 6	MESO: Local Level about 1,001–100,000 records page 8	MACRO: Global Level more than 100,000 records page 10
<b>TYPES</b>			
<b>Statistical Analysis</b> page 44	 Knowledge cartography page 135	 Productivity of Russian life sciences research teams page 105	 Number of scientists versus population and R&D costs page 105
<b>WHEN: Temporal Analysis</b> page 48	 Visualizing decision-making processes page 95	 Key events in the development of the video tape recorder page 85	 Increased travel and communication speeds page 83
<b>WHERE: Geospatial Analysis</b> page 52	 Cell phone usage in Milan, Italy page 109	 Victorian poetry in Europe page 137	 Ecological footprint of countries page 99
<b>WHAT: Topical Analysis</b> page 56	 Evolving patent holdings of Apple Computer, Inc. and Jerome Lemelson page 89	 Evolving journal networks in nanotechnology page 139	 Product space showing co-export patterns of countries page 95
<b>WITH WHOM: Network Analysis</b> page 60	 World Finance Corporation network page 87	 Electronic and new media art networks page 133	 World-wide scholarly collaboration networks page 157

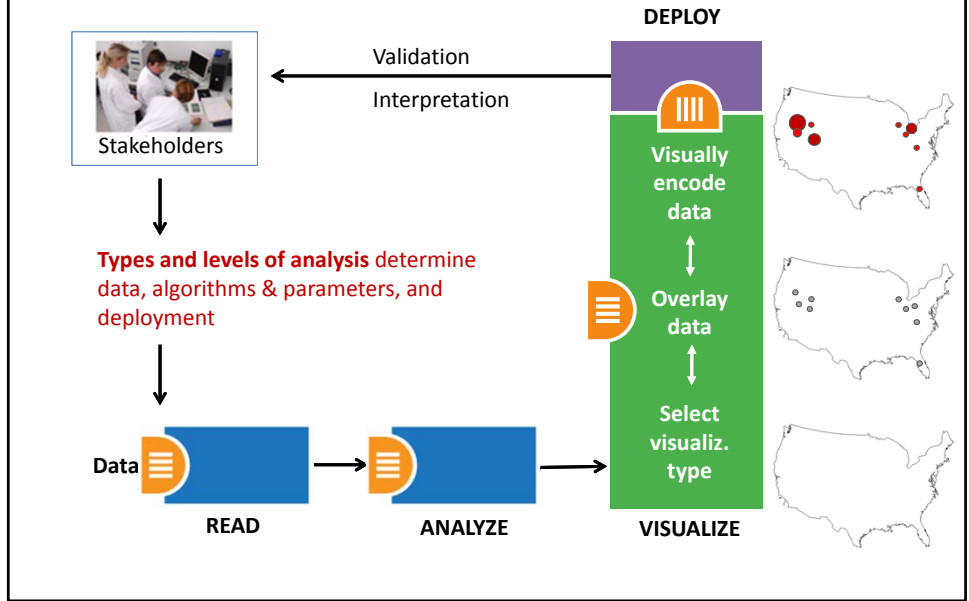
Atlas of Knowledge  
Know-Can-Map

See page 5

23



# Needs-Driven Workflow Design

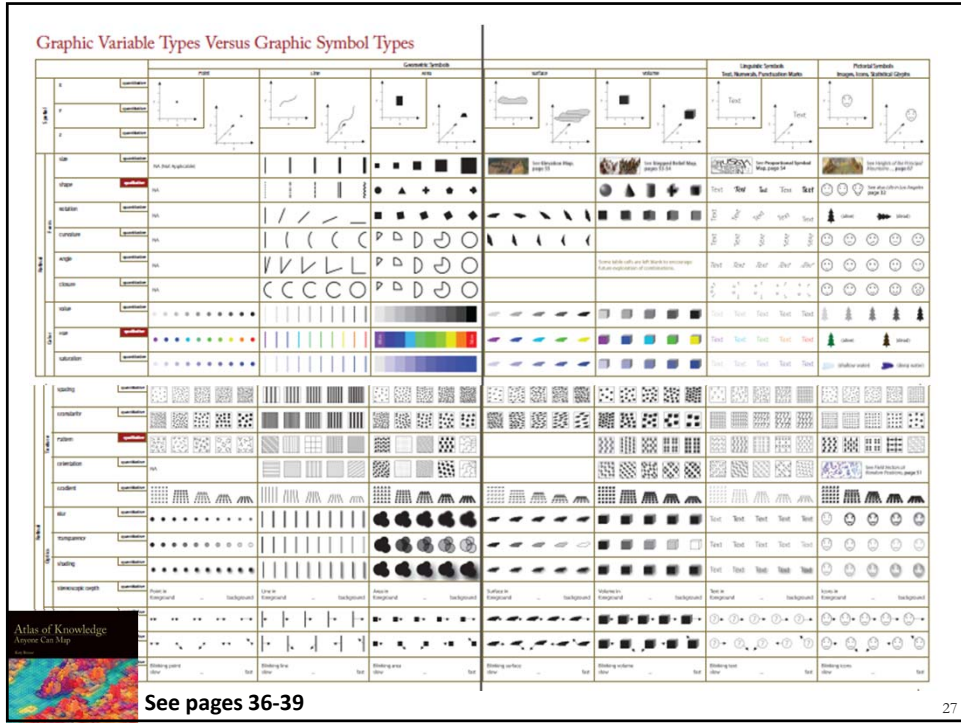


## Graphic Variable Types Versus Graphic Symbol Types

		Geometric Symbols				
		Point	Line	Area		
Spatial	x	quantitative				
	y	quantitative				
	z	quantitative				
Retinal	Form	Size	quantitative	NA (Not Applicable)		
		Shape	qualitative	NA		
		Rotation	quantitative	NA		
		Curvature	quantitative	NA		
		Angle	quantitative	NA		
		Closure	quantitative	NA		
Color	Value	quantitative				
	Hue	qualitative				
	Saturation	quantitative				



See page 36



## Load **One** File and Run **Many** Analyses and Visualizations

Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	Research & Experimental Medicine	Falk-Krzesinski, HJ Borner, K Contractor, NJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B
13	2010	WASHINGTON	USA	SCIENCE TRANSLATIONAL MEDICINE	A Multi-Level Systems Perspective for the Science of Team Science	Cell Biology Research & Experimental Medicine	Borner, K Contractor, NJ Falk-Krzesinski, HJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B

Statistical Analysis--p. 44

Location	Count	# Citations
Netherlands	13	292
United States	9	318
Germany	11	36
United Kingdom	1	2

Temporal Burst Analysis--p. 48

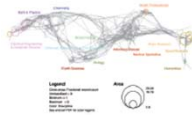
Geospatial Analysis--p. 52

Geospatial Analysis--p. 52

## Load **One** File and Run **Many** Analyses and Visualizations

Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	Research & Experimental Medicine	Falk-Krzesinski, HJ Borner, K Contractor, NJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B
13	2010	WASHINGTON	USA	SCIENCE TRANSLATIONAL MEDICINE	A Multi-Level Systems Perspective for the Science of Team Science	Cell Biology Research & Experimental Medicine	Borner, K Contractor, NJ Falk-Krzesinski, HJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B

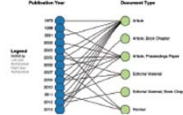
Topical Analysis—p. 56



Paper Citation Network—p. 60



Bi-Modal Network—p. 60



Co-author and many other bi-modal networks.

29

## READINGS

### Papers

- Stipelman, Brooke A., Hall, Kara L., Zoss, Angela, Okamoto, Janet, Stokols, Dan, and Börner, Katy (2014) Mapping the Impact of Transdisciplinary Research: A Visual Comparison of Investigator Initiated and Team Based Tobacco Use Research Publications. *The Journal of Translational Medicine and Epidemiology*.
- Bollen, Johan, David Crandall, Damion Junk, Ying Ding, and Katy Börner. 2014. [From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review](#). *EMBO Reports* 15 (1): 1-121.
- Mazloumian, Amin, Dirk Helbing, Sergi Lozano, Robert Light, and Katy Börner. 2013. [Global Multi-Level Analysis of the 'Scientific Food Web'](#). *Scientific Reports* 3, 1167.

### Books

- Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. The MIT Press.
- Börner, Katy, and David E. Polley. 2014. *Visual Insights: A Practical Guide to Making Sense of Data*. Cambridge, MA: The MIT Press.
- Scharnhorst, Andrea, Katy Börner, and Peter van den Besselaar, eds. 2012. *Models of Science Dynamics*. Springer Verlag.
- Börner, Katy, Mike Conlon, Jon Corson-Rikert, and Ying Ding, eds. 2012. *VIVO: A Semantic Approach to Scholarly Networking and Discovery*. Morgan & Claypool Publishers LLC.
- Börner, Katy. 2010. *Atlas of Science: Visualizing What We Know*. The MIT Press.

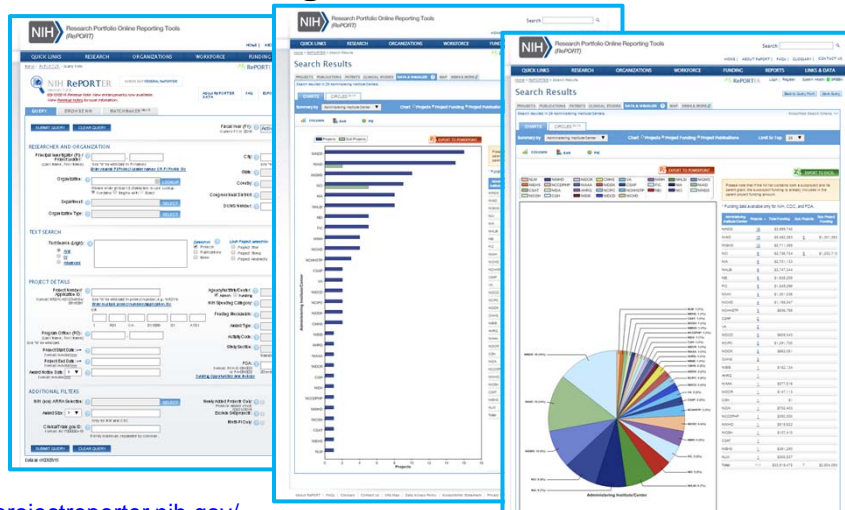
30

## Existing Interfaces for VATS-like data

- 9) NIH RePORTER: Visual Interface to Biomedical Funding Data in U.S.
- 10) CISHell/Sci2 World and Science Visualizations of NIH RePORTER Data
- 11) NIH RePORTER: NIH Map
- 12) BBSRC: Visual Interface to Biomedical Funding Data in UK
- 13) IAI Multidimensional Analysis
- 14) Scraawl: Twitter Analysis
- 15) Illuminated Diagram: Searchable World and Science Maps

## NIH RePORTER: Visual Interface to Biomedical Funding Data in US

9

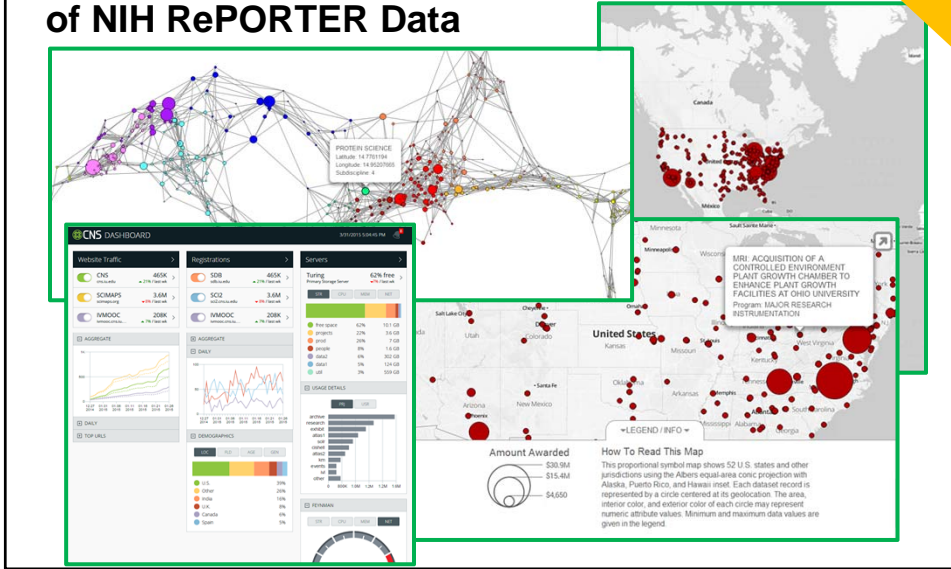


<http://projectreporter.nih.gov/>

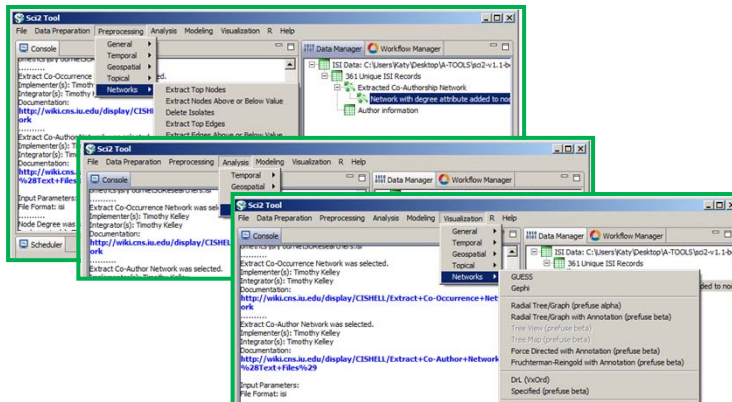


# CIShell/Sci2 World and Science Visualizations of NIH RePORTER Data

10

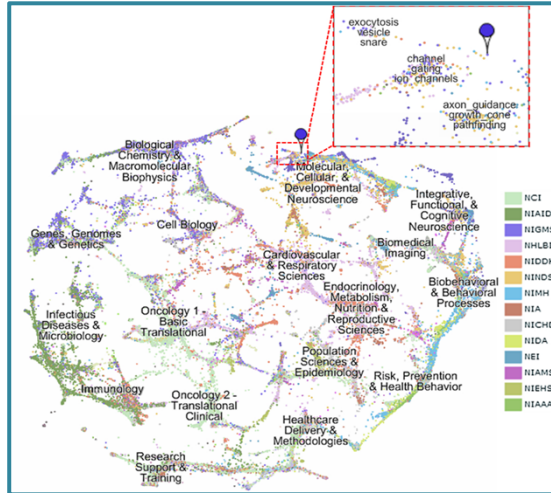


# Sci2 Desktop



**NIH RePORTER: NIH Map**

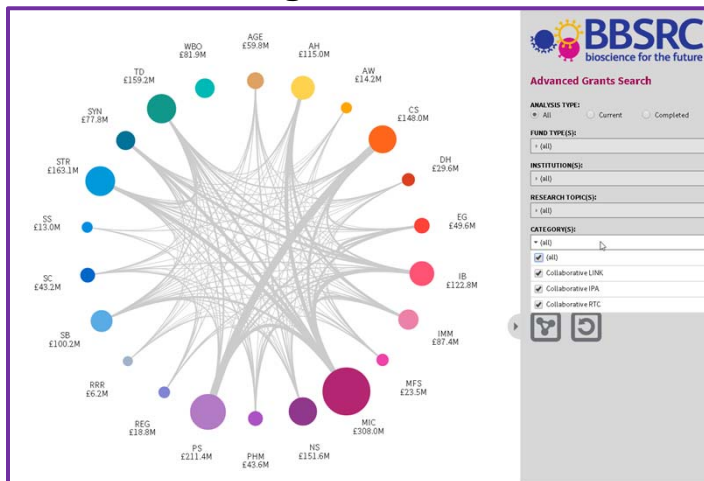
11



<http://nihmaps.org/>

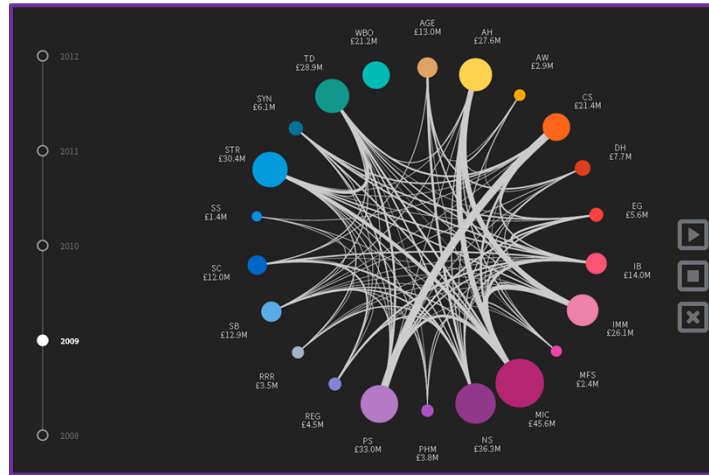
**BBSRC: Visual Interface to Biomedical Funding Data in UK**

12



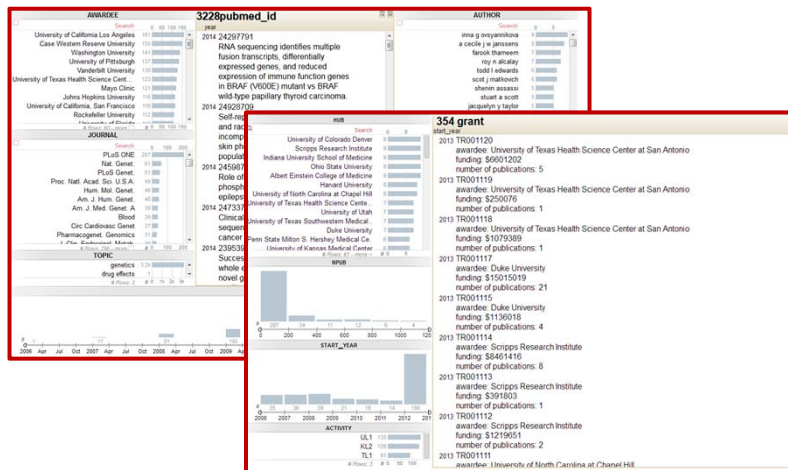
<http://www.bbsrc.ac.uk/>

**BBSRC: Temporal animation**



**IAI Multidimensional Analysis**

13



## Scraawl: Twitter Analysis

**Edit Report**  
Report Name: NIHNCATS Test - Search Terms: @NIH, #ncats

**Search terms**  
@NIH #ncats

**Filter report**  
When the report is completed, automatically create a new report and continue collecting.

**Update Report**

**Name (optional)**  
NIHNCATS Test

**Optional name for the report**

**Data sources**  
 Twitter (Posting soon)
  Instagram
  Tumblr
  LinkedIn
  Flickr

**Top Users**

@frankbaitm	4
@dougblumghd	2
@frankbaitm	2
@genresearch	2
@medsci_news	2
80 total users	

**Top Words**

Cancer	45
Research	37
Medical	25
Mary	25
Lasker	25
197 total words	

**Top Hashtags**

#genetics	24
#nih	22
#cancer	22
#nicabnl	6
#precisionmedicine	4
23 total hashtags	

**Top Mentions**

@nih	100
@conanderennes	6
@nicabnl	5
@frankbaitm	3
@healthmedary	3
29 total mentions	

**Top URLs**

http://genetics.nih.gov/...	14
http://nih.gov/ncats/med...	9
http://www.nih.gov/ncats/...	8
http://www.cancer.gov/ncats/...	7
http://www.nih.gov/health/...	5
26 total URLs	

**Top Retweets**

RT @NIH: Mary Lasker, cent...	20
RT @NIH: Largely thanks to...	6
RT @NIH: President Nixon st...	6
RT @NIH: Cancer is a gene...	6
RT @NIH: #Cancer occurs w...	5
28 retweeted tweets	

**Top Languages**

95% English  
5% French  
2 total languages

**Top Locations**

1 location  
Ligette

**Tweet Timeline**

<https://www.scraawl.com/>

## Scraawl: Advanced analysis

**Community Detection Results**

Number of Communities	19
Average Communities Size	8
Maximum Community Size	50

**Communities**

Rank	Community Label	Community Size
1	@nih #cancerfilm #cancer	50
2	#nih #frankbaitman #precisionmedicine	26
3	@invstong100 @jymajrkkkanen @who	8
4	@conanderennes @tom_l_giordano @sauravguha	7
5	@mesamalania @riqueira_andre @mantaray	7
6	@be_counseling @mehsp4u #worldbipolarday	6
7	@markdilleys @ncats_nags @nascarhometrack	5
8	@thirtyonacran @wtop @dc	4
9	@rosshemingway @moran_ats @sportslegacy	4
10	@anuacharya @paul_somnier @heartdocsharon	4

**Visualization of Communities**

# Illuminated Diagram: Searchable World and Science Maps



[http://cns.iu.edu/interactive\\_displays.html](http://cns.iu.edu/interactive_displays.html)

# Illuminated Diagram: Search detail

**Geographic Maps: Where Science Gets Done**

**Science Maps: 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 is important by using projections or screens as smart spotlights, directing the research instead of individuals, giving a "grand tour" of research, or highlighting queries made (as when you touch the screen) 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."

**Interest**  
Select any location on the Geographic Map (optionally by dragging your finger over an area on the screen'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 categorical lines of research using MEDLINE data from 2000-2009.

Cancer

Cloning

HIV

Robert G. Edwards

Roger D. Kramberg

Elinor Ostrom

Obesity

Quality of Life

Smoking

Stanley B. Prusiner

Ahmed H. Zewail

View All

## 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.sjis.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/](http://www.pnas.org/content/vol101/suppl_1/)

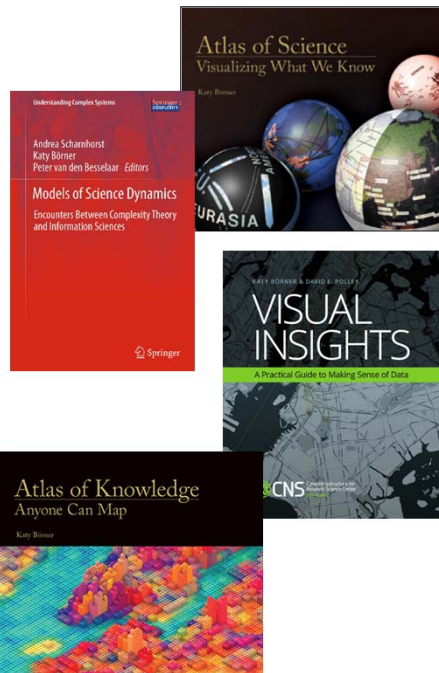
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

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

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. <http://scimaps.org/atlas2>



43

All papers, maps, tools, talks, press are linked from <http://cns.iu.edu>  
 These slides will soon be at <http://cns.iu.edu/docs/presentations>  
 CNS Facebook: <http://www.facebook.com/cnscenter>  
 Mapping Science Exhibit Facebook: <http://www.facebook.com/mappingscience>

44