

# Tutorial: Open Source Tools for S&T Data Analysis and Visualization

**Dr. Katy Börner**

Indiana University, Bloomington, Indiana, USA

<http://cns.iu.edu>



With special thanks to Kevin W. Boyack, Chin Hua Kong, Patrick Phillips, Chintan Tank, Joseph R. Biberstine, Adam Simpson, Steven Corenflos, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Daniel Halsey, Russell J. Duhon, Scott Weingart, Hanning Guo, Mark A. Price, Angela M. Zoss, Ted Polley, and Michael Ginda.

**Please**

- download the Sci2 Tool from <http://sci2.cns.iu.edu> and
- complete the Pre-Tutorial Questionnaire

*OECD, Paris, France*

*June 25, 2014 • 9:30am-13:30pm*

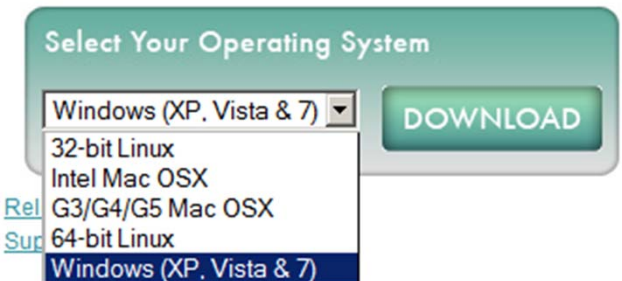


## Software, Datasets, Plugins, and Documentation

- These slides  
<http://cns.iu.edu/docs/presentations/2014-borner-opentoolstutorial-oecd.pdf>
- Sci2 Tool Manual v0.5.1 Alpha, updated to match v1.0 Alpha tool release  
<http://sci2.wiki.cns.iu.edu>
- Sci2 Tool v 1.1 beta (Dec 9, 2013)  
<http://sci2.cns.iu.edu>
- Additional Datasets  
<http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets>
- Additional Plugins  
<http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

### Download

Sci<sup>2</sup> v 1.1 beta  
December 9th, 2013



Make sure you have Java 1.6 (32-bit suffices) or higher installed or download from <http://www.java.com/en/download>. To check your Java version, open a terminal and run 'java -version'.

Some visualizations are saved as Postscript files. A free Postscript to PDF viewer is at <http://ps2pdf.com> and a free PDF Viewer at <http://www.adobe.com/products/reader.html>.

## Tutorial Overview

9:30 Welcome and Overview of Tutorial and Attendees

9:45 Open Data and Tools

- Standardize and federate micro-level datasets of S&T activity
- Open code tools and online services that are interoperable
- Sharing and teaching open datasets and tools

**11-11:30 Networking Break**

11:30 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- Country Collaboration Network, 2011 – **Scoreboard 2013**
- Country Expertise Profiles, 2010 – **New Workflow**
- Country Mobility Network, 1996-2011 – **Scoreboard 2013**
- Acceleration in the Development of Patented Technologies, 2000-11 – **SB 2013**
- Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – **Scoreboard 2013**
- Evolving Country Patent Profiles, 1995-2010 – **New Workflow**

13:00 Outlook and Q&A

**13:30 Adjourn**

## Tutorial Overview

### **9:30 Welcome and Overview of Tutorial and Attendees**

9:45 Open Data and Tools

- Standardize and federate micro-level datasets of S&T activity
- Open code tools and online services that are interoperable
- Sharing and teaching open datasets and tools

11-11:30 Networking Break

11:30 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- Country Collaboration Network, 2011 – Scoreboard 2013
- Country Expertise Profiles, 2010 – New Workflow
- Country Mobility Network, 1996-2011 – Scoreboard 2013
- Acceleration in the Development of Patented Technologies, 2000-11 – SB 2013
- Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – Scoreboard 2013
- Evolving Country Patent Profiles, 1995-2010 – New Workflow

13:00 Outlook and Q&A

13:30 Adjourn

## Tutorial Overview

9:30 Welcome and Overview of Tutorial and Attendees

### **9:45 Open Data and Tools**

- **Standardize and federate micro-level datasets of S&T activity**
- **Open code tools and online services that are interoperable**
- **Sharing and teaching open datasets and tools**

11-11:30 Networking Break

11:30 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- Country Collaboration Network, 2011 – Scoreboard 2013
- Country Expertise Profiles, 2010 – New Workflow
- Country Mobility Network, 1996-2011 – Scoreboard 2013
- Acceleration in the Development of Patented Technologies, 2000-11 – SB 2013
- Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – Scoreboard 2013
- Evolving Country Patent Profiles, 1995-2010 – New Workflow

13:00 Outlook and Q&A

13:30 Adjourn

## Open Data and Tools

**Standardize** and **federate** micro-level datasets of S&T activity

- Data formats
- Access protocols
- DOIs
- Licenses

Open code tools and online services that are **interoperable**

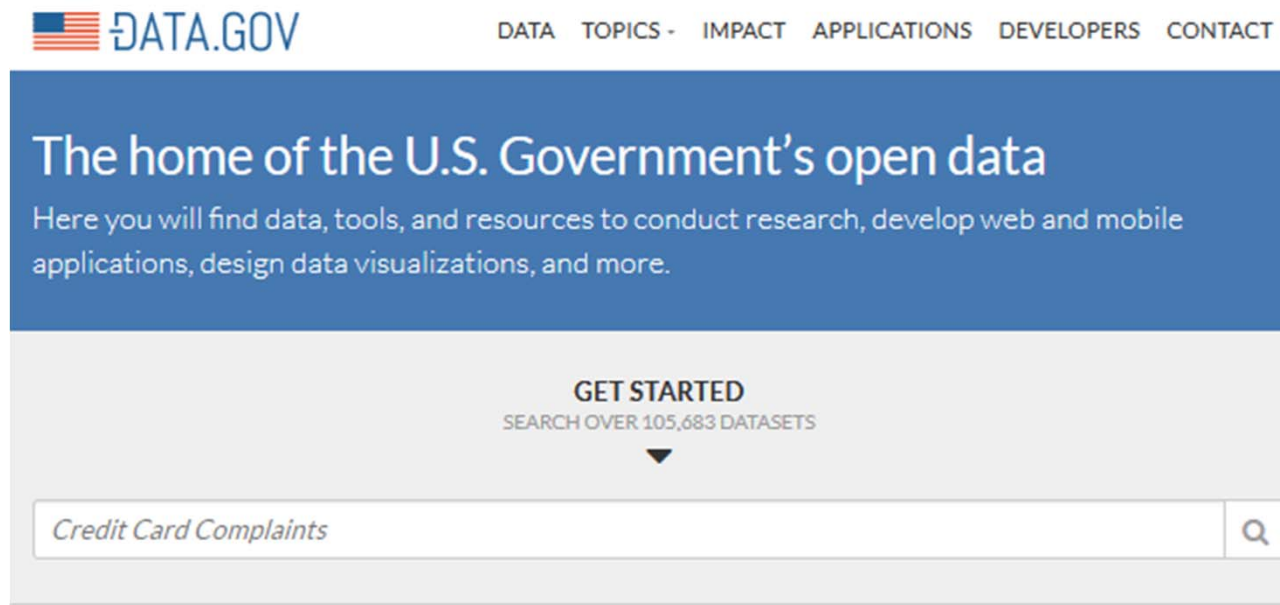
- Libraries and code repositories
- Workflow protocols
- Licenses

**Teaching** open datasets and tools

- (Online) documentation
- MOOCs

# Open Data

## Open Data: Government Data



The screenshot shows the homepage of data.gov. At the top left is the logo with an American flag and the text "DATA.GOV". To the right is a navigation menu with links for "DATA", "TOPICS", "IMPACT", "APPLICATIONS", "DEVELOPERS", and "CONTACT". Below the navigation is a blue banner with the text "The home of the U.S. Government's open data" and a subtext "Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more." Below the banner is a grey section with the text "GET STARTED" and "SEARCH OVER 105,683 DATASETS". Underneath is a search bar containing the text "Credit Card Complaints" and a search icon.

### BROWSE TOPICS

-  Agriculture
-  Climate
-  Education
-  Energy
-  Finance
-  Geospatial
-  Global Development
-  Health
-  Jobs & Skills
-  Public Safety
-  Science & Research
-  Weather

<http://data.gov>



# Open Data: U.S. NIH Funding



**NIH** Research Portfolio Online Reporting Tools  
(RePORT)

Search

HOME | ABOUT RePORT | FAQs | GLOSSARY | CONTACT US

QUICK LINKS | RESEARCH | ORGANIZATIONS | WORKFORCE | FUNDING | REPORTS | LINKS & DATA

**NIH Data Book**

NDB provides basic summary statistics on extramural grants and contract awards, grant applications, the organizations that NIH supports, the trainees and fellows supported through NIH programs, and the national biomedical workforce.

RePORTER | AWARDS BY LOCATION | **NIH DATA BOOK** | FUNDING FACTS | CATEGORICAL SPENDING | REPORT CATALOG | SPECIAL REPORTS | ABOUT REPORT

## Research Portfolio Online Reporting Tools (RePORT)

In addition to carrying out its scientific mission, the NIH exemplifies and promotes the highest level of public accountability. To that end, the Research Portfolio Online Reporting Tools provides access to reports, data, and analyses of NIH research activities, including information on NIH expenditures and the results of NIH supported research.

<http://report.nih.gov>



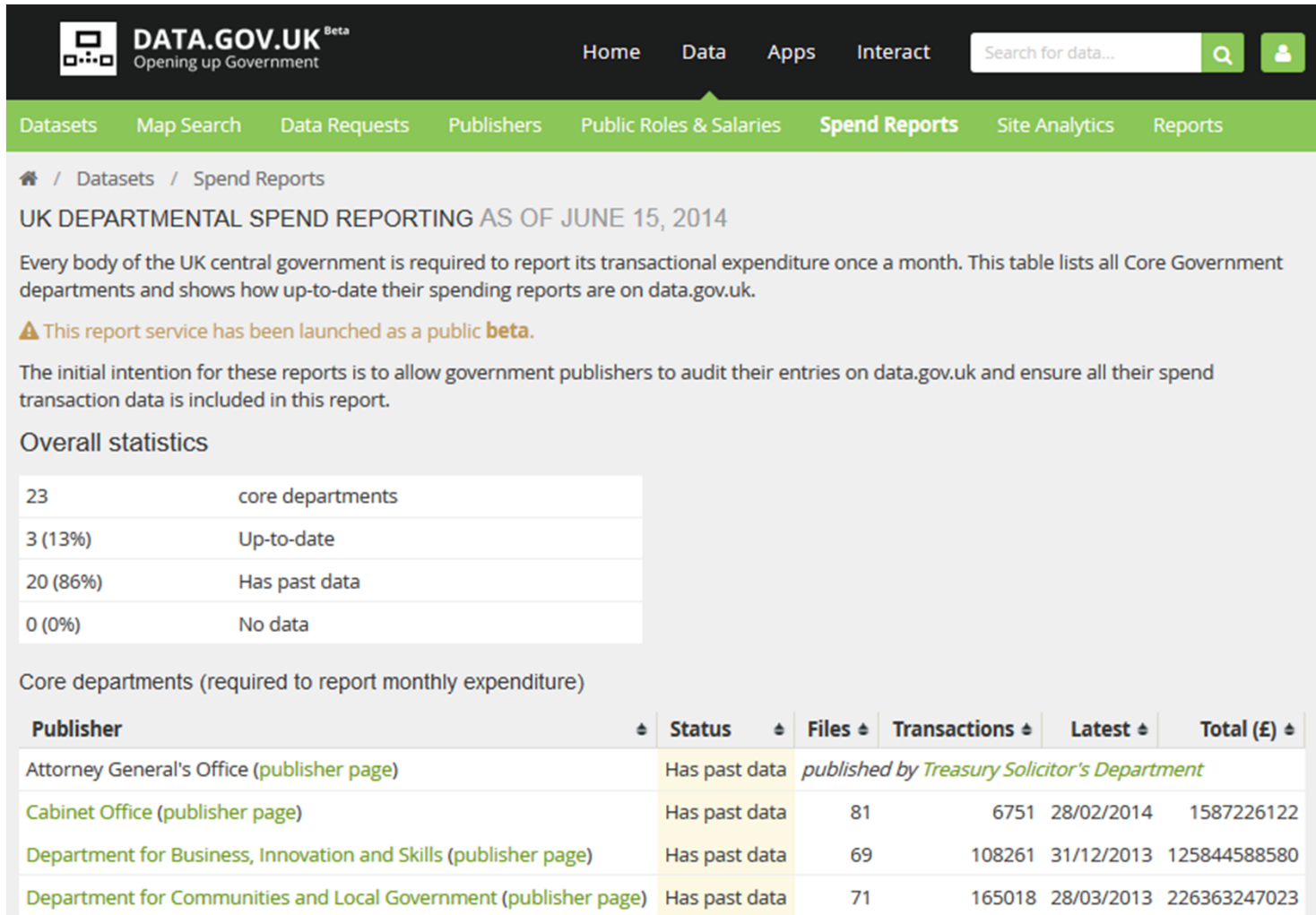
**NIH RePORTER** [ADVANCED SEARCH](#)

Institute/Center: **ALL** | Fiscal Year: **ACTIVE** | Principal Investigator: (Last Name, First Name)



NCI | 2014 |

NEI | 2013 |

# Open Data: UK Funding



**DATA.GOV.UK** <sup>Beta</sup>  
Opening up Government

Home Data Apps Interact Search for data...  

Datasets Map Search Data Requests Publishers Public Roles & Salaries **Spend Reports** Site Analytics Reports

🏠 / Datasets / Spend Reports

## UK DEPARTMENTAL SPEND REPORTING AS OF JUNE 15, 2014

Every body of the UK central government is required to report its transactional expenditure once a month. This table lists all Core Government departments and shows how up-to-date their spending reports are on data.gov.uk.

⚠️ This report service has been launched as a public beta.

The initial intention for these reports is to allow government publishers to audit their entries on data.gov.uk and ensure all their spend transaction data is included in this report.

### Overall statistics

23	core departments
3 (13%)	Up-to-date
20 (86%)	Has past data
0 (0%)	No data

Core departments (required to report monthly expenditure)

Publisher	Status	Files	Transactions	Latest	Total (£)
Attorney General's Office ( <a href="#">publisher page</a> )	Has past data	<i>published by Treasury Solicitor's Department</i>			
Cabinet Office ( <a href="#">publisher page</a> )	Has past data	81	6751	28/02/2014	1587226122
Department for Business, Innovation and Skills ( <a href="#">publisher page</a> )	Has past data	69	108261	31/12/2013	125844588580
Department for Communities and Local Government ( <a href="#">publisher page</a> )	Has past data	71	165018	28/03/2013	226363247023

<http://data.gov.uk/data/openspending-report/index>

# Open Data: Eurostat



The screenshot shows the Eurostat website homepage. At the top, there is a blue header with the European Commission logo and the Eurostat logo, with the tagline "Your key to European statistics". Below the header is a navigation menu with buttons for "Home", "Statistics", "Publications", "About Eurostat", and "User support".

On the left side, there are several sections: "Statistics Database" with a database icon, "Release Calendars" with a calendar icon, "Most popular database tables" listing items like "GDP per capita in PPS", "Real GDP growth rate", "Total population", "Unemployment rate", "Employment rate, by sex", "Inflation (monthly)", and "Inflation rate (annual)", and "Selected Statistics" listing "Macroeconomic Imbalance Procedure (MIP)" and "Europe 2020 indicators".

The main content area is titled "Latest news releases" and includes a search bar, "Press centre | RSS" link, and "Log in | Register | Log off" options. It features a list of news releases with dates and titles, such as "The overall tax-to-GDP ratio in the EU28 up to 39.4% of GDP in 2012" and "Euro area annual inflation down to 0.5%".

On the right side, there is a "Search" bar, a "General government gross debt (% of GDP and million EUR)" section with a bar chart showing data for 2010, 2011, 2012, and 2013, and a "User survey" section with the text "Give us your opinion on".

<http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

# Open Data: People Profiles from NRN Systems

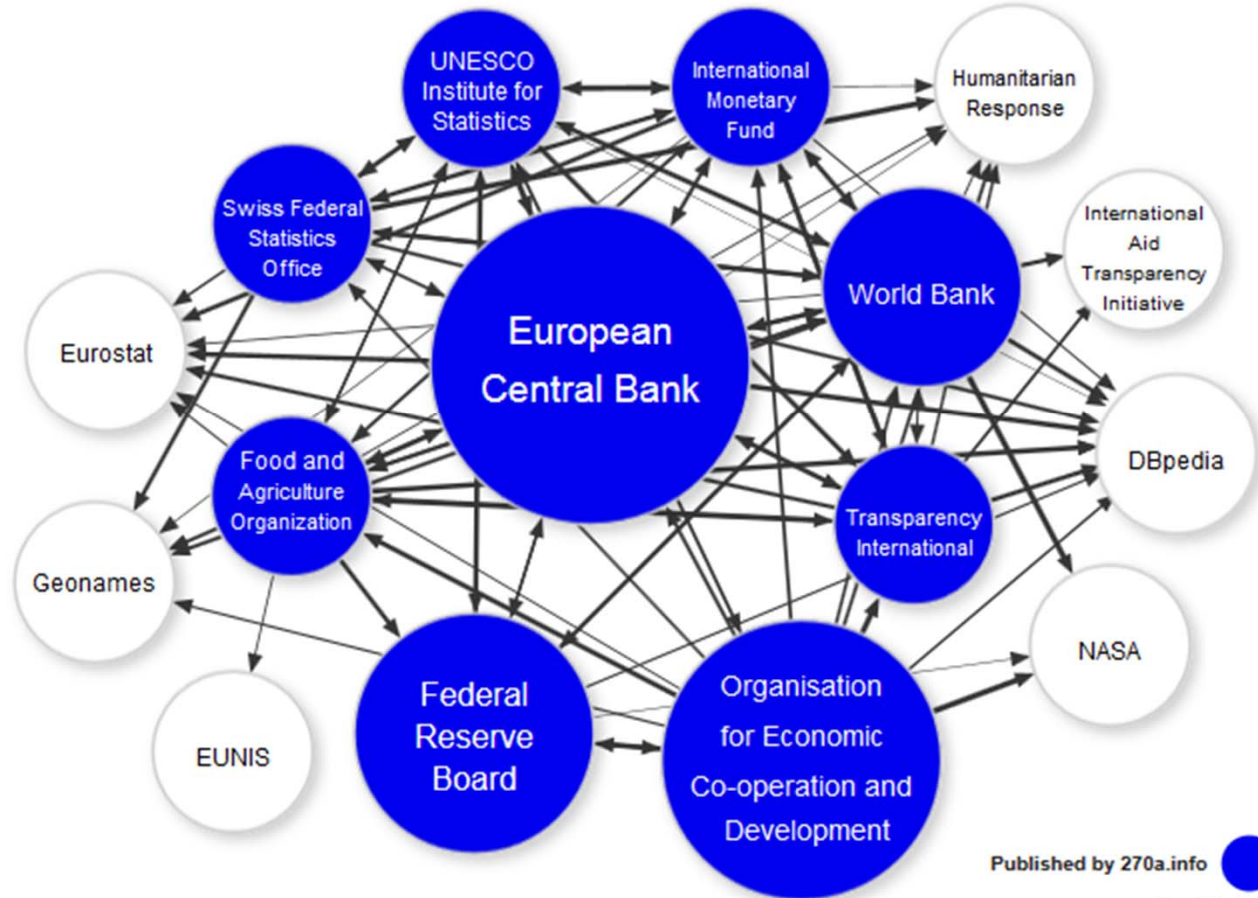
## International Researcher Network Visualization 1.0



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

# Open Data: Linked Data

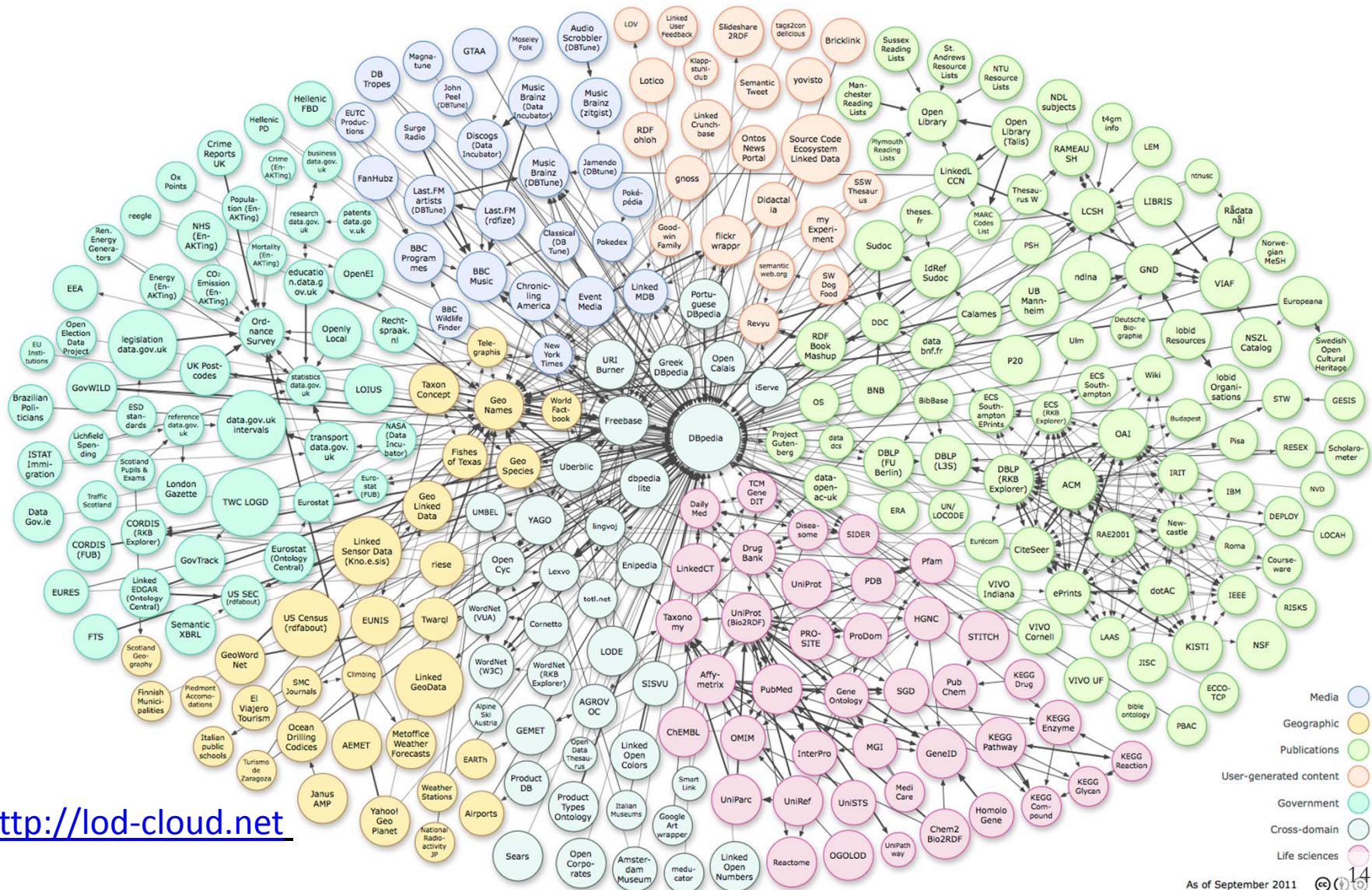
[Statistical Linked Dataspaces](#), where some are based on [Linked SDMX Data](#):



<http://www.270a.info>

2014-05 







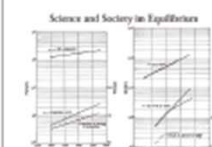
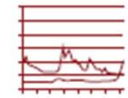




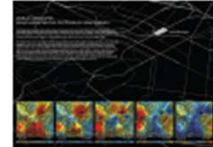
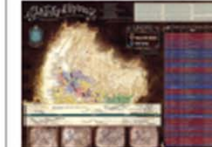









# Open Data: Linked Data, 2011



<http://lod-cloud.net>

# Open Tools

# Types of Analysis and Visualization

		LEVELS		
		MICRO: Individual Level about 1-1000 records	MESO: Local Level about 1001-100,000 records	MACRO: Global Level more than 100,000 records
				
TYPES				
<b>Statistical Analysis</b>		 Knowledge Cartography p. 135	 Productivity of Russian life sciences research teams p. 105	 Science and Society in Equifiters Number of scientists vs. population and R&D costs vs. GNP in U.S. p. 103
<b>WHEN: Temporal Analysis</b>		 Visualizing decision making processes p. 95	 Key events in the development of the video tape recorder p. 85	 Increased travel and communication speeds p. 83
<b>WHERE: Geospatial Analysis</b>		 Cell phone usage in Milan, Italy p. 109	 Victorian poetry in Europe p. 137	 Ecological footprint of countries p. 99
<b>WHAT: Topical Analysis</b>		 Evolving patent holdings of Apple Computers and Jerome Lemelson p. 89	 Evolving journal networks in nanotechnology p. 139	 Product space showing co-export patterns of countries p. 93
<b>WITH WHOM: Network Analysis</b>		 World Finance Corporation network p. 87	 Electronic and new media art networks p. 133	 World-wide scholarly collaboration networks p. 157



# Open Tools: Statistics

Flash-based; does not use DOM convention for representing and interacting with objects in HTML, XHTML and XML documents.

**GAPMINDER** for a fact-based world view

- HOME
- GAPMINDER WORLD**
- DATA
- VIDEOS
- DOWNLOADS
- FOR TEACHERS
- LABS
- IGNORANCE

Reset Open graph menu

GLOBAL TRENDS  
**Wealth & Health of Nations\***

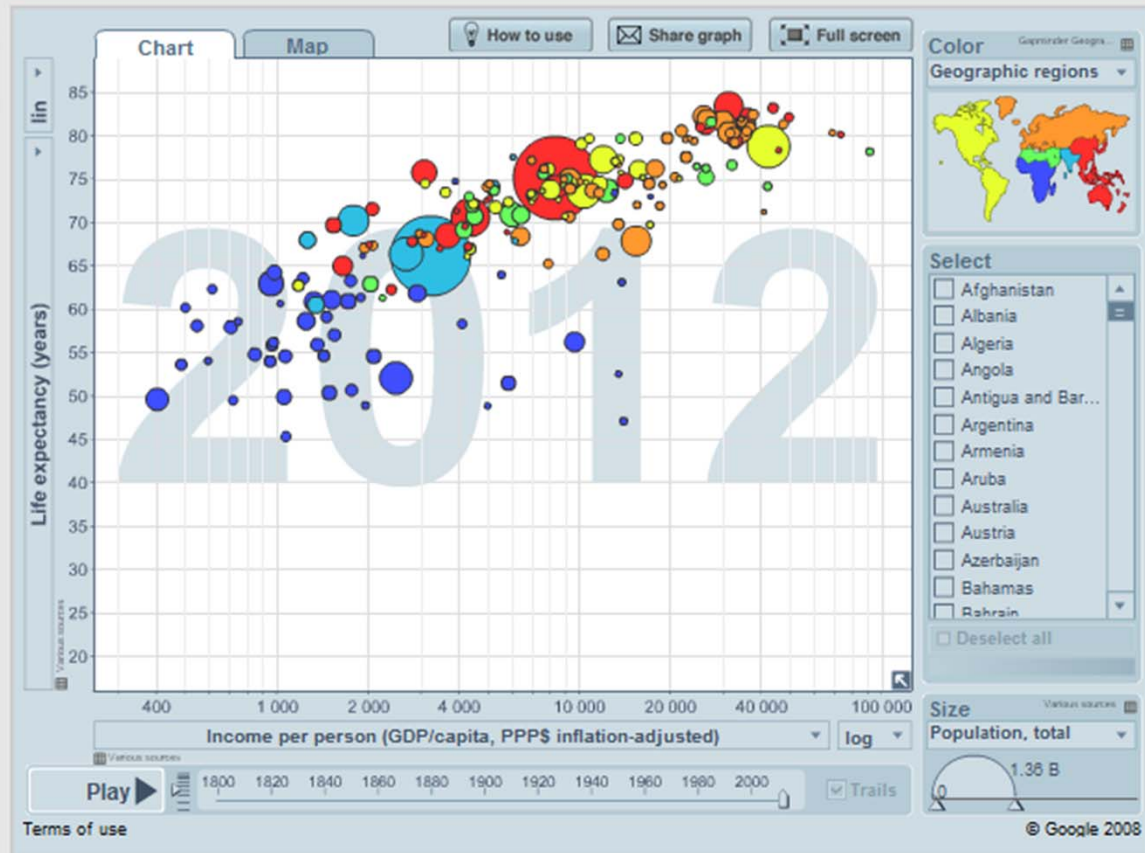
\*This graph was modified from:  
**Wealth & Health of Nations**

Explore Gapminder World on your own:

- by browsing our example graphs
- by selecting indicators and bubbles

See also:

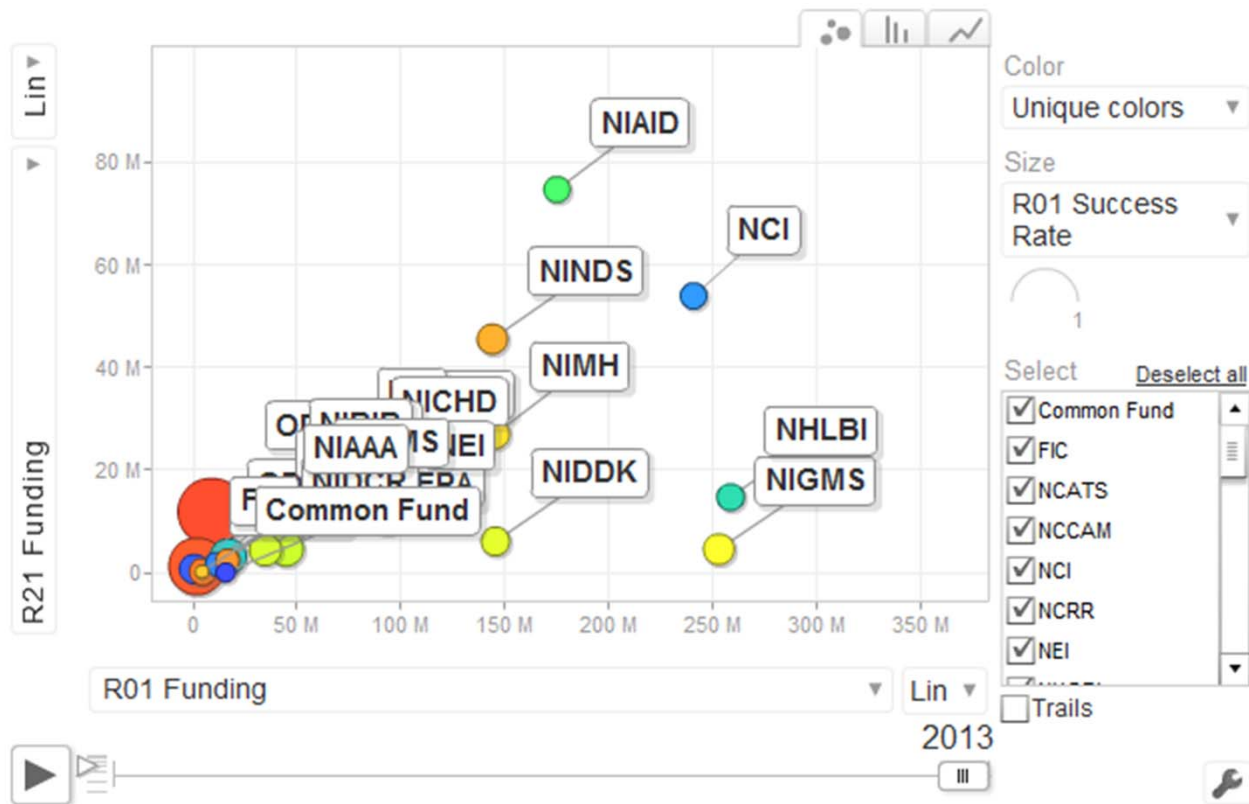
- 200 years that changed the world
- Stop call them "developing countries"



Ask a question

<http://www.gapminder.org>

[http://www.gapminder.org/world/#\\$majorMode=chart\\$is:shi=t:ly=2003;lb=f;il=t;fs=11;al=30;stl=t;st=t;ns1=t;se=t\\$wst;tts=C\\$ts:sp=5.59290322580644;ti=2012\\$zpv:v=0\\$inc\\_x:mmid=XCOORDS:iid=phAwcNAVuyj1jiMAkmq1iMg:by=ind\\$inc\\_y:mmid=YCOORDS:iid=phAwcNAVuyj2tPLxKvvnNPA:by=ind\\$inc\\_s:uniValue=8.21;iid=phAwcNAVuyj0XOoBL\\_n5tAQ:by=ind\\$inc\\_c:niValue=255;gid=CAT1D0:by=grp\\$map\\_x:scale=log:dataMin=283:dataMax=110808\\$map\\_y:scale=lin:dataMin=18:dataMax=87\\$map\\_s:sma=49;smi=2.65\\$cd:bd=0\\$inds=:modified=75](http://www.gapminder.org/world/#$majorMode=chart$is:shi=t:ly=2003;lb=f;il=t;fs=11;al=30;stl=t;st=t;ns1=t;se=t$wst;tts=C$ts:sp=5.59290322580644;ti=2012$zpv:v=0$inc_x:mmid=XCOORDS:iid=phAwcNAVuyj1jiMAkmq1iMg:by=ind$inc_y:mmid=YCOORDS:iid=phAwcNAVuyj2tPLxKvvnNPA:by=ind$inc_s:uniValue=8.21;iid=phAwcNAVuyj0XOoBL_n5tAQ:by=ind$inc_c:niValue=255;gid=CAT1D0:by=grp$map_x:scale=log:dataMin=283:dataMax=110808$map_y:scale=lin:dataMin=18:dataMax=87$map_s:sma=49;smi=2.65$cd:bd=0$inds=:modified=75)



Find more such visualizations in our [ReTools](#) section.  
[Accessibility Statement](#)

Animation Courtesy of Google

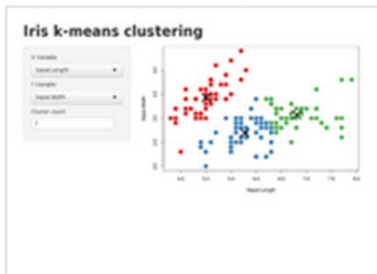
[http://report.nih.gov/retools/competing\\_research\\_project.aspx](http://report.nih.gov/retools/competing_research_project.aspx)



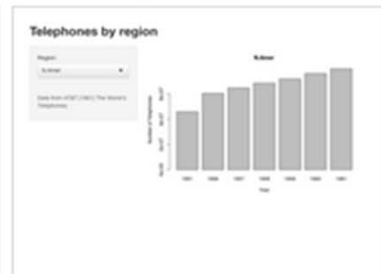
# Gallery

## Start simple

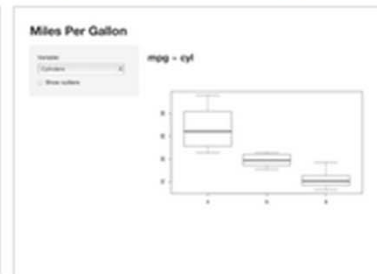
If you're new to Shiny, these simple but complete applications are designed for you to study.



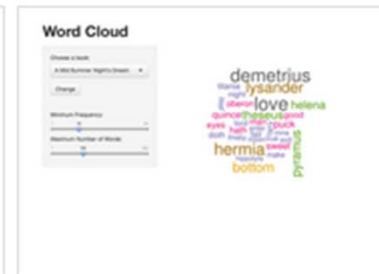
Kmeans example



Telephones by region



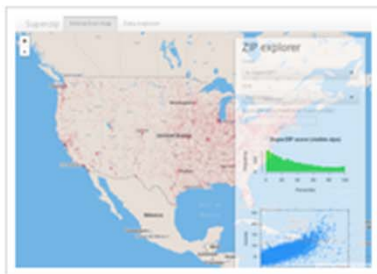
Miles per gallon



Word cloud

## Interactive visualizations

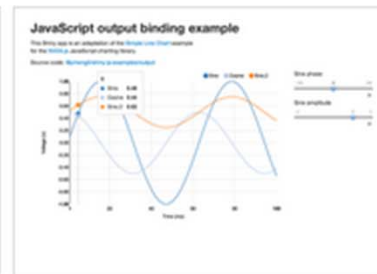
Shiny is designed for fully interactive visualization, using JavaScript libraries like [d3](#), [Leaflet](#), and [Google Charts](#).



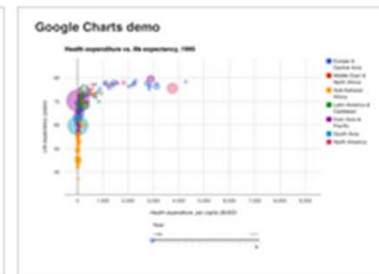
SuperZip example



Movie explorer



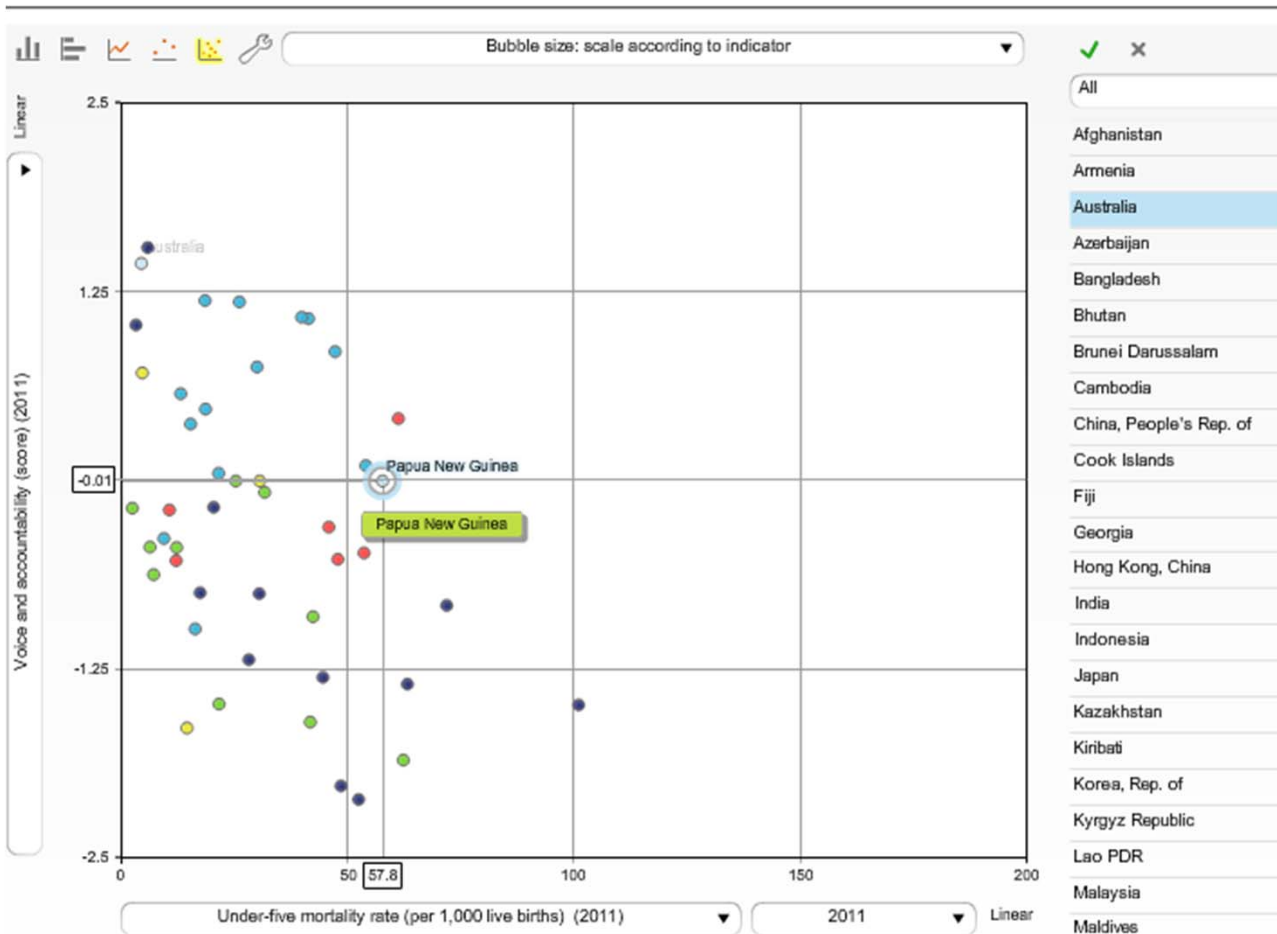
NVD3 line chart output



Google Charts

<http://shiny.rstudio.com/gallery>

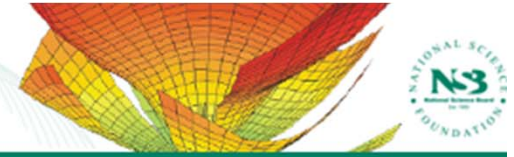
**ADB** Asian Development Bank  
**Key Indicators for Asia and the Pacific**  
**Framework of Inclusive Growth Indicators 2013**



<http://www.adb.org/figi/web/StatTrends.html>

## Open Tools:

# Temporal Analysis and Visualization (WHEN)



## Contents

Preface and Introduction

Global R&D: Measuring Commitment to Innovation

▶ **U.S. R&D: Funding and Performance**

U.S. R&D: Federal Portfolio

U.S. and Global STEM Education

U.S. S&E Workforce: Trends and Composition

Research Outputs: Publications and Patents

Public Research Universities

Glossary and Key to Acronyms

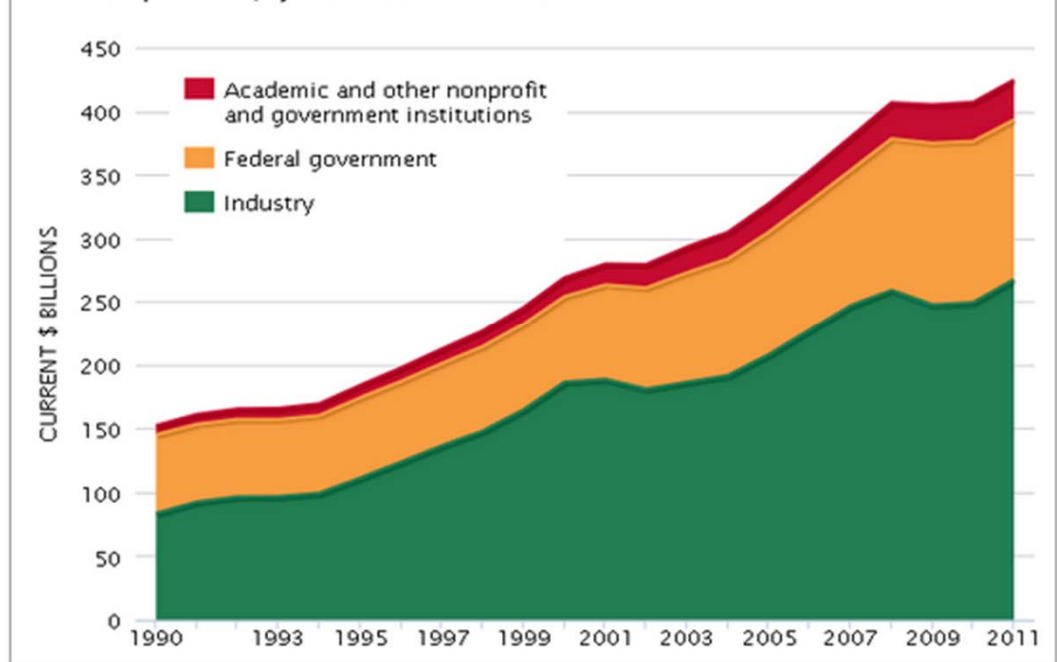
Explore Further

SEI 2014 Online Resources

Acknowledgments

National Science Board Members

U.S. R&D expenditures, by source of funds: 1990-2011

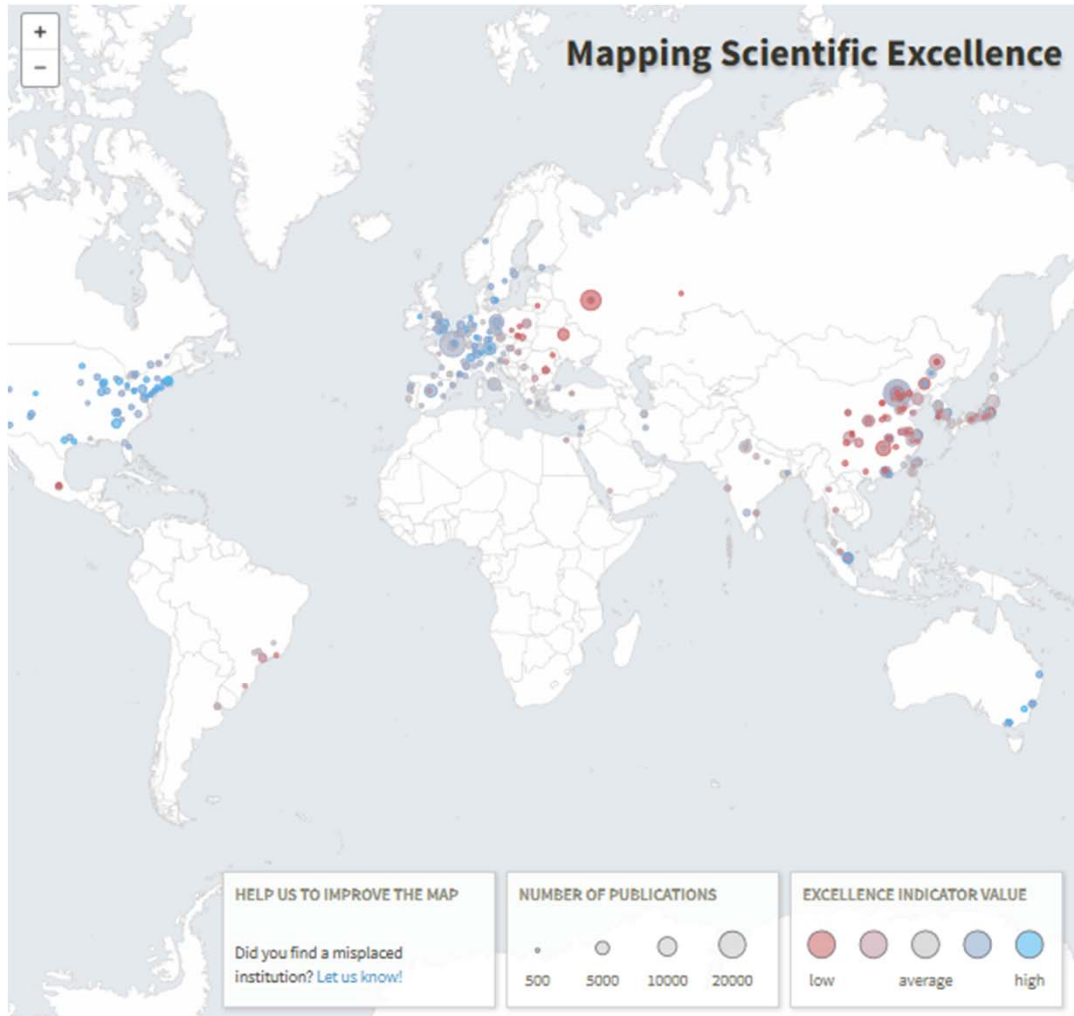


<http://nsf.gov/statistics/seind14/index.cfm/digest>

Open Tools:

Geospatial Analysis and Visualization (WHERE)





This web application visualizes scientific excellence worldwide in 17 subject areas. For each institution (university or research-focused institution), the estimated probabilities of (i) publishing highly cited papers (Best Paper Rate) or (ii) publishing in the most influential journals (Best Journal Rate) are shown. Both probabilities, which can be adjusted by covariates, range from blue (high probability) through grey (average) to red (low probability) at a circle. The circle size corresponds to the institutional number of papers.

[More information](#)

2005 - 2009    2006 - 2010    **2007 - 2011**

SUBJECT AREA ⓘ

Materials Science

COVARIATE ⓘ

- none -

EXCELLENCE INDICATORS ⓘ

**Best Paper Rate**

Best Journal Rate

SIGNIFICANCE ⓘ

Show statistically significant results only

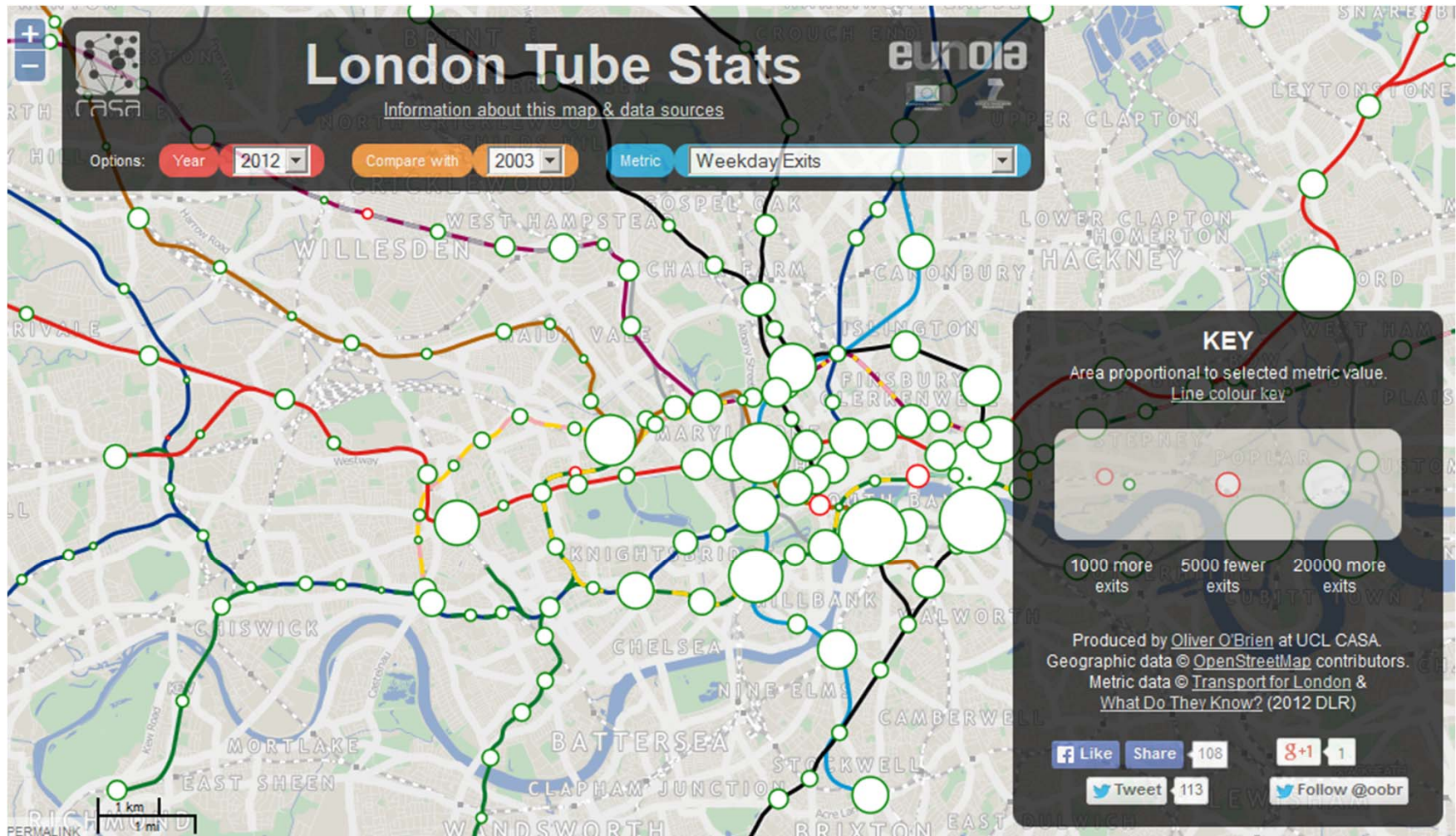
INSTITUTIONAL SCORES

SEARCH:

Institution	Country	Papers	Indicator value
Rice University	USA	843	40.8%
Northwestern University, Evanston	USA	1503	36.4%
Stanford University	USA	1169	35.6%
University of Pennsylvania	USA	722	34.6%
Harvard University	USA	1152	34.2%
University of Washington	USA	1020	33.4%

<http://www.excellencemapping.net>

[http://www.excellencemapping.net/#/view/edition/2014/measure/top10/calculation/a\\_ohne\\_kovariabile/field/materials-science/significant/false](http://www.excellencemapping.net/#/view/edition/2014/measure/top10/calculation/a_ohne_kovariabile/field/materials-science/significant/false)

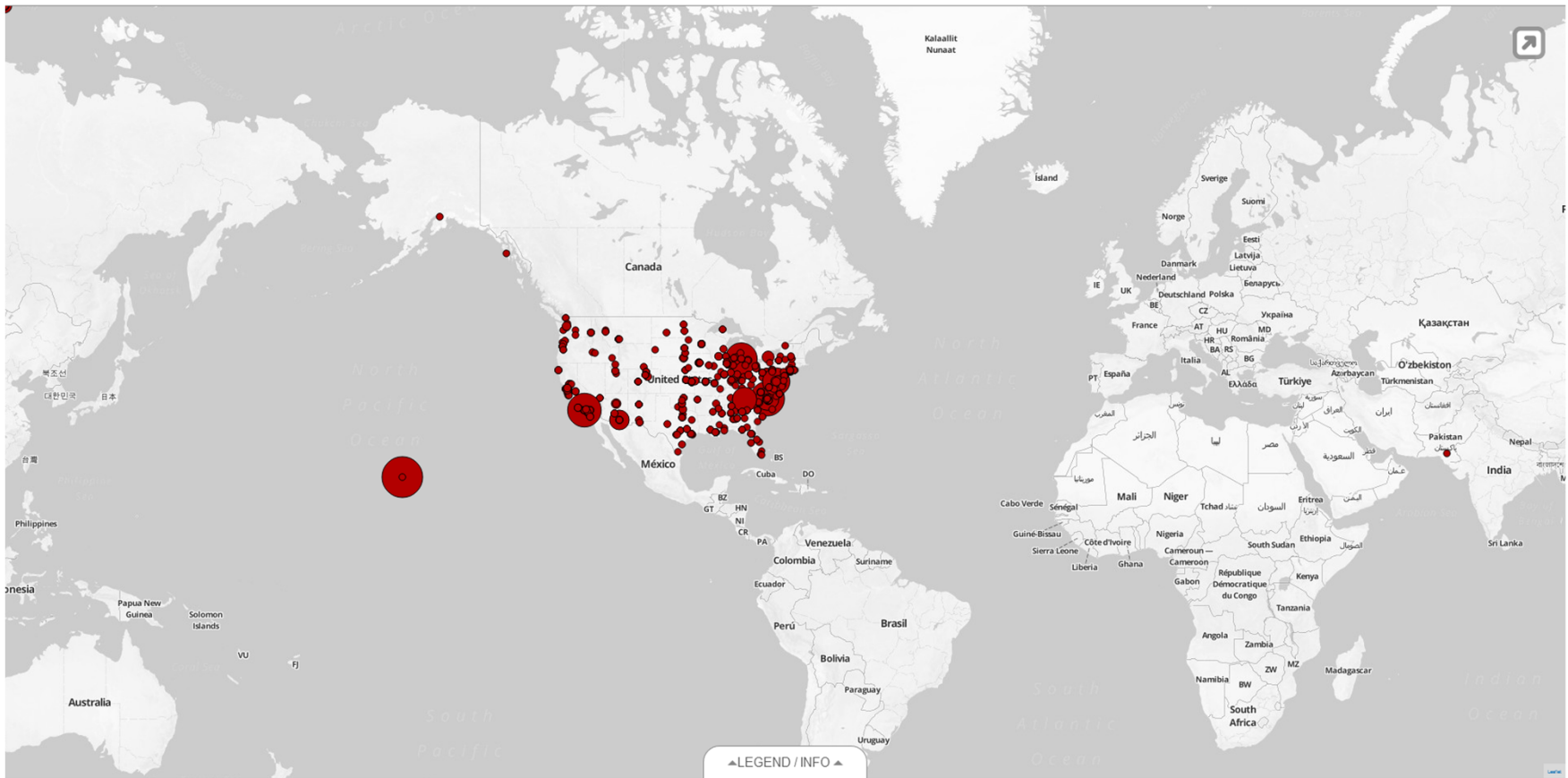


<http://casa.oobrien.com/tube>



Proportional Symbol Map

Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST

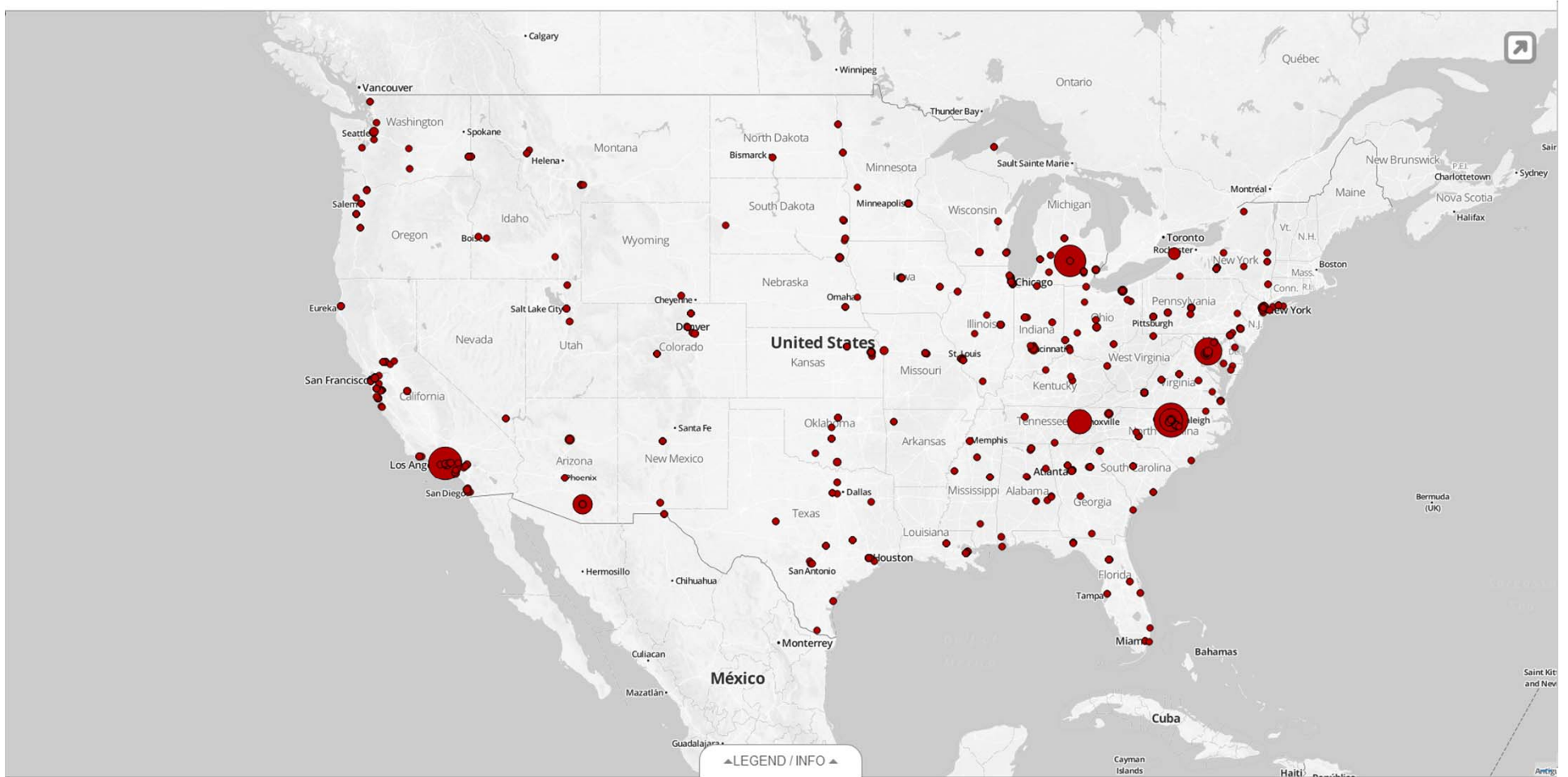


Developed for NIH by CNS and NETE. Responsive design.



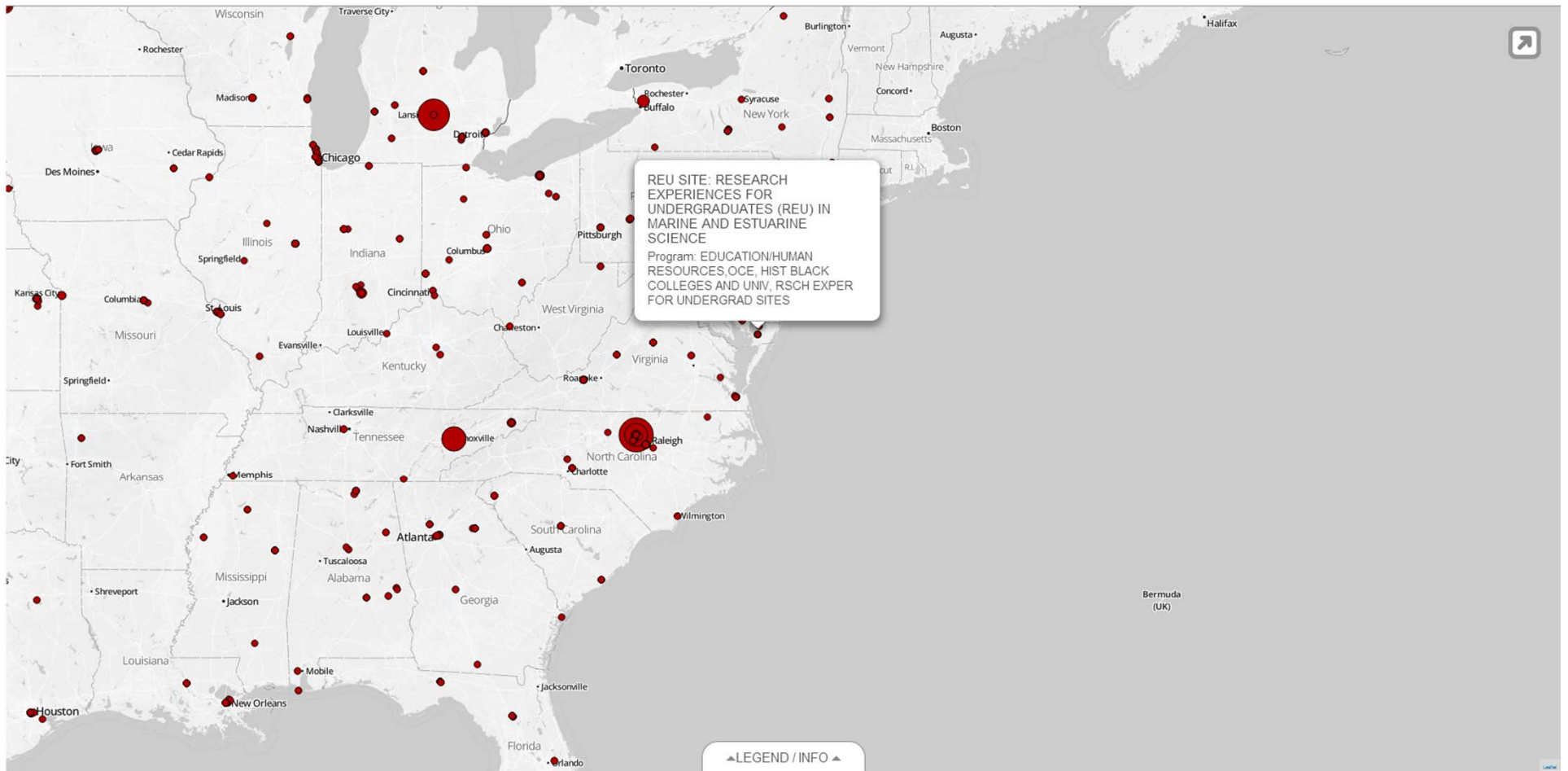
### Proportional Symbol Map

Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST



### Proportional Symbol Map

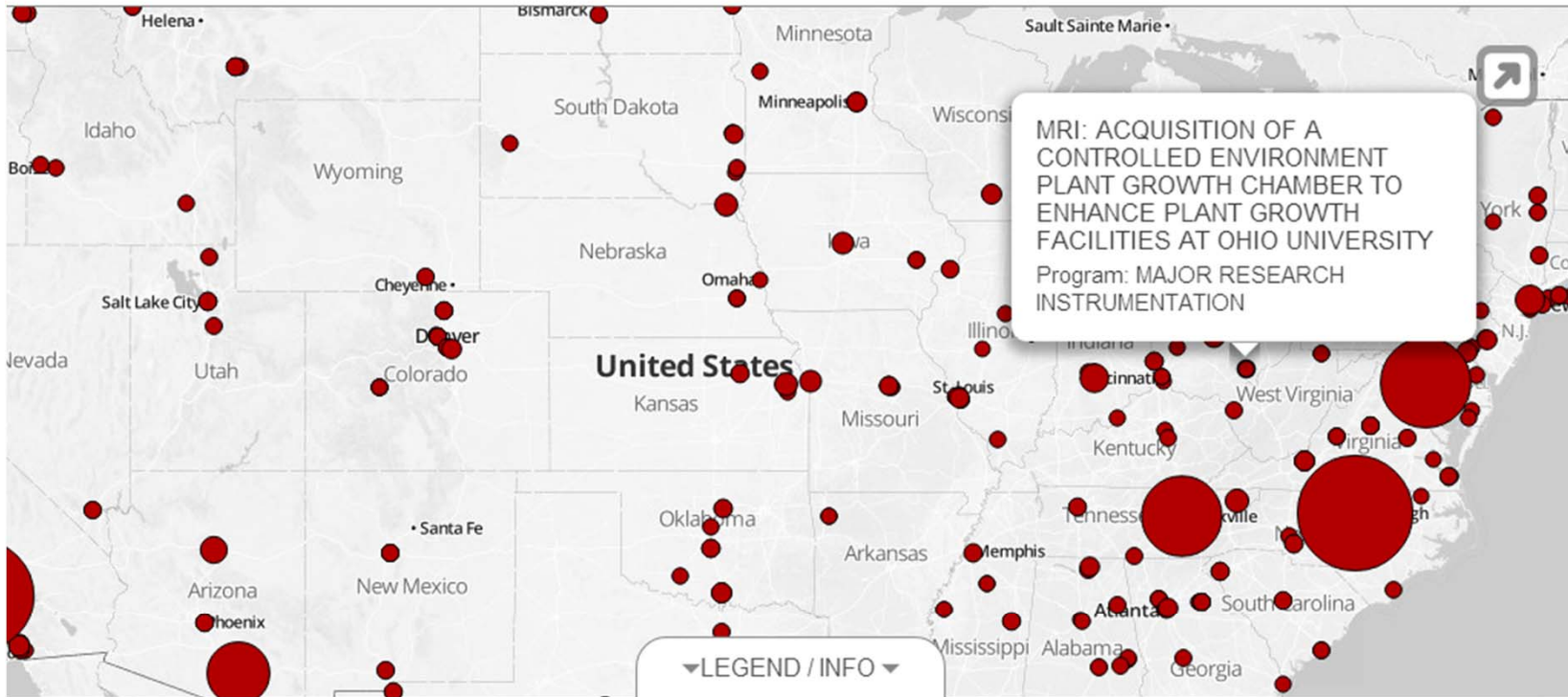
Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST





### Proportional Symbol Map

Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST



#### 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.



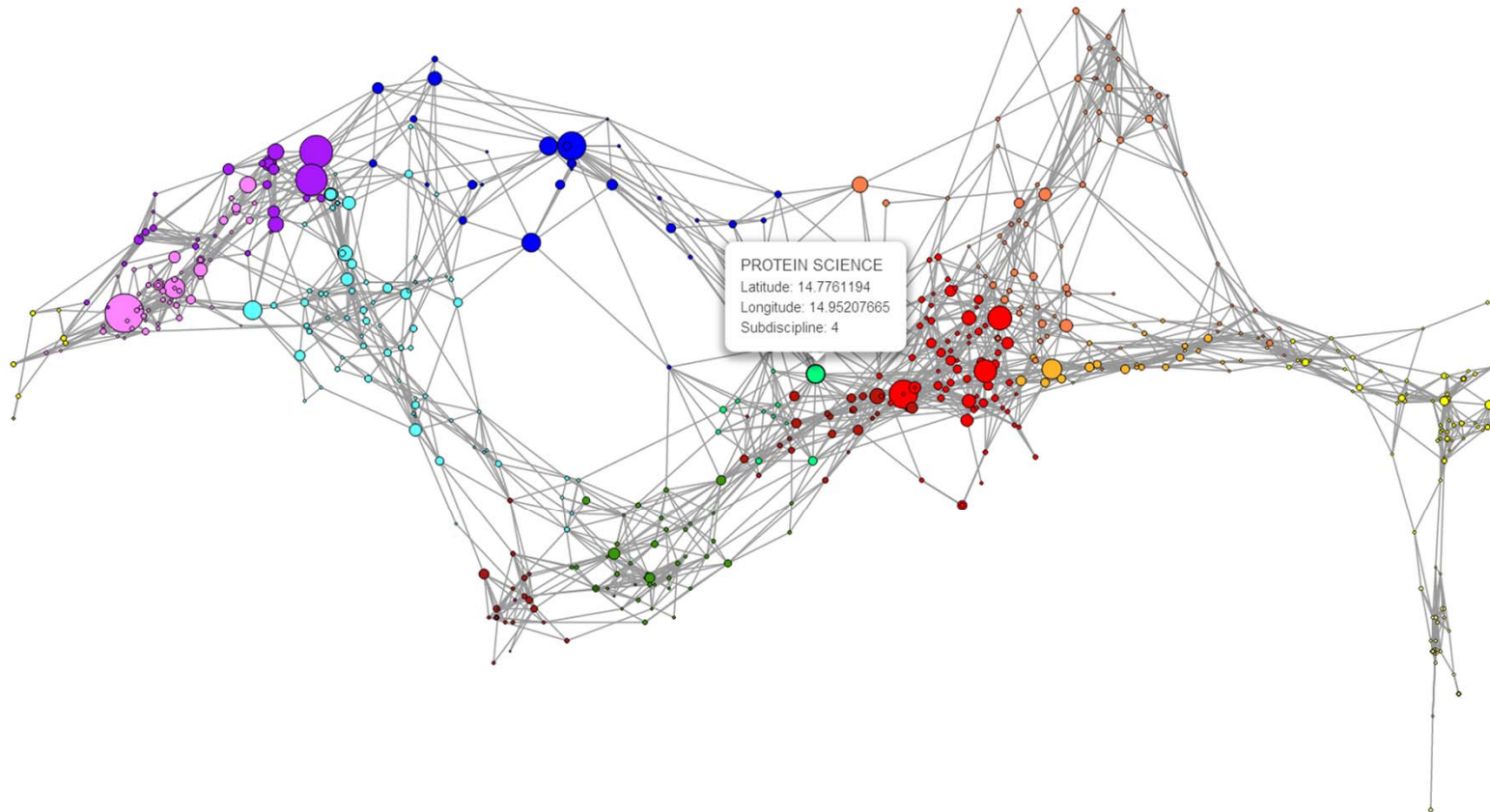
**CNS**

Cyberinfrastructure for  
Network Science Center



### Topic Analysis - Map of Science

Generated from Publications for top 20 projects - Jeffrey R. Alberts, Larry E. Humes, Bernice A. Pescosolido and 9 others; Generated by NETE.

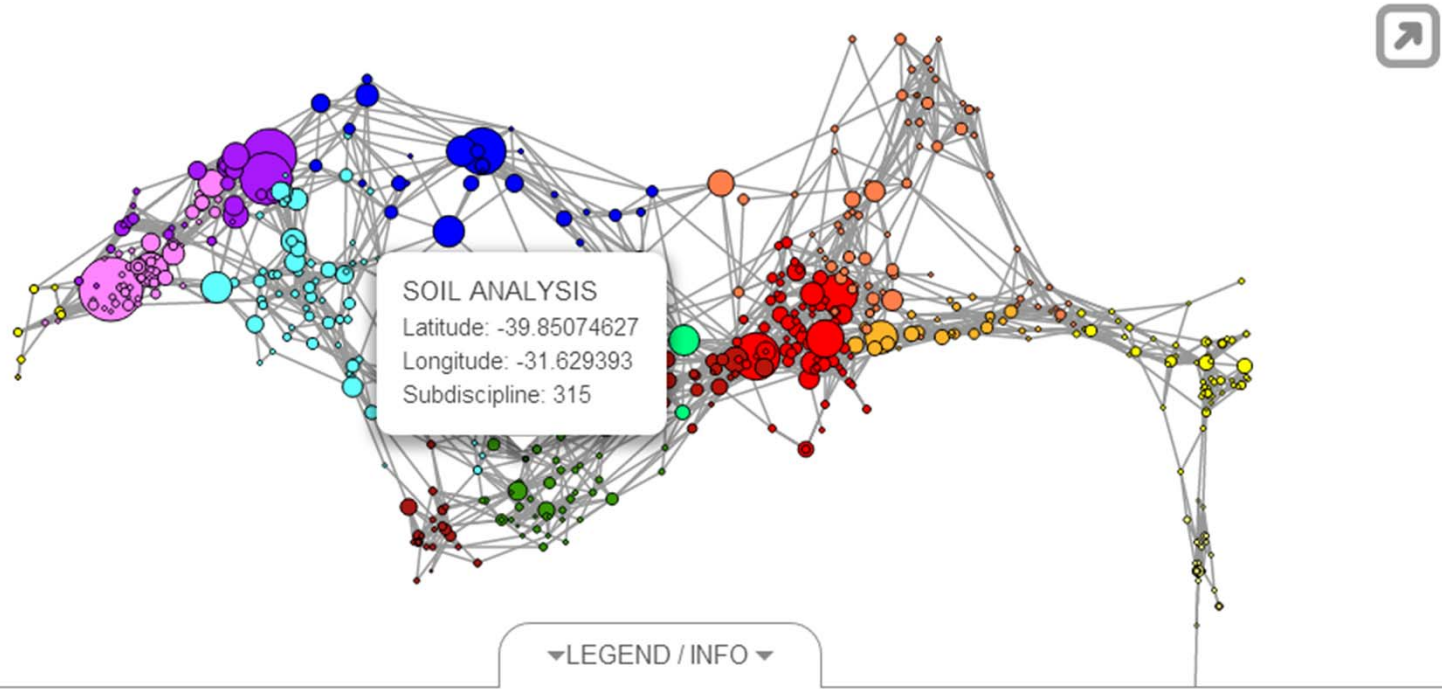


▲LEGEND / INFO ▲



### Topic Analysis - Map of Science

Generated from Publications for top 20 projects - Jeffrey R. Alberts, Larry E. Humes, Bernice A. Pescosolido and 9 others; Generated by NETE.



#### Total Awards



#### How To Read This Map

This map is a visual representation of 554 sub-disciplines within 13 disciplines of science and their relationships to one another, shown as points and lines connecting those points respectively. Over top this visualization is drawn the result of mapping a dataset's journals to the underlying sub-discipline(s) those journals contain. Mapped sub-disciplines are shown with size relative to the number of matching journals and color from the discipline.





**CNS**

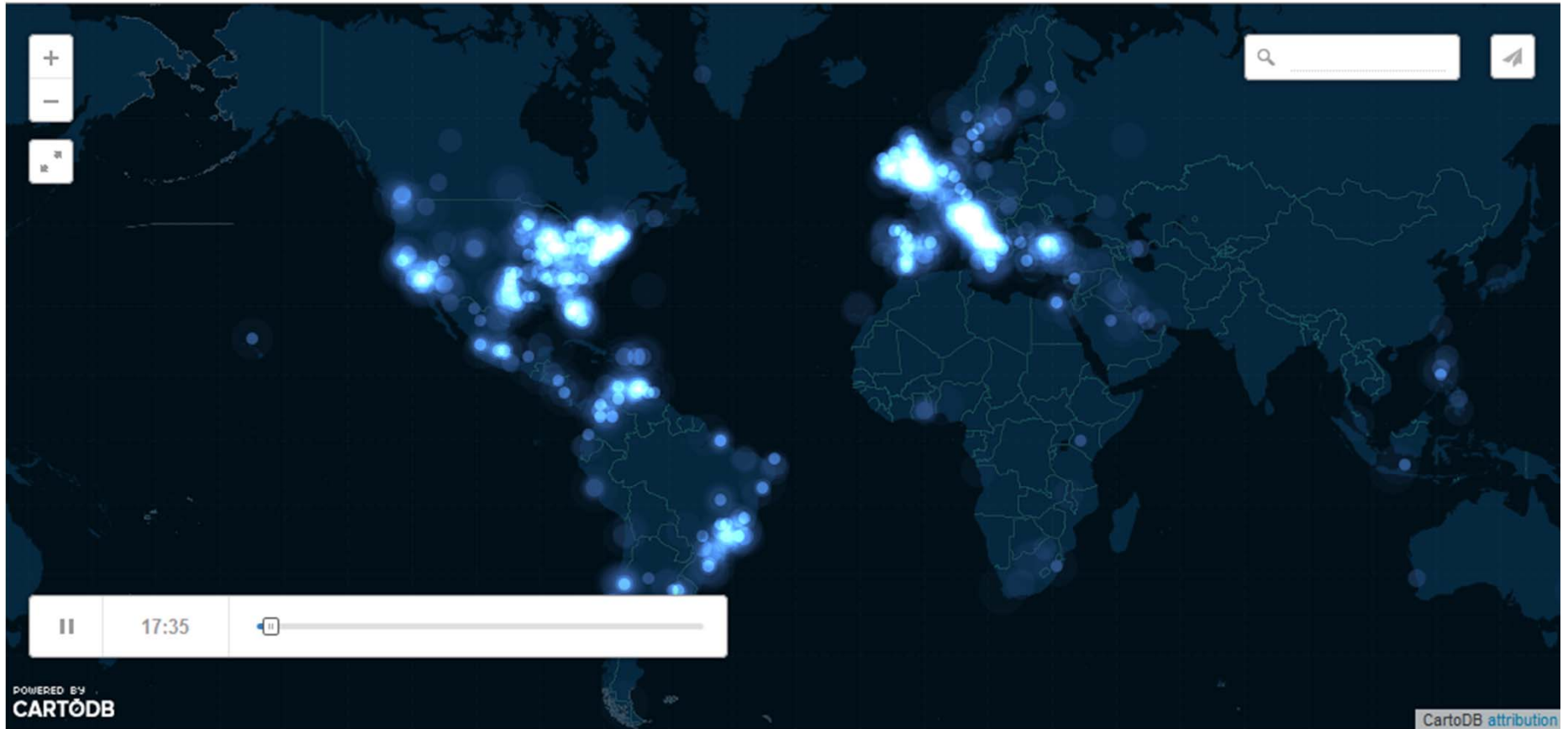
Cyberinfrastructure for  
Network Science Center

Open Tools:

Topical Analysis and Visualization (WHAT)

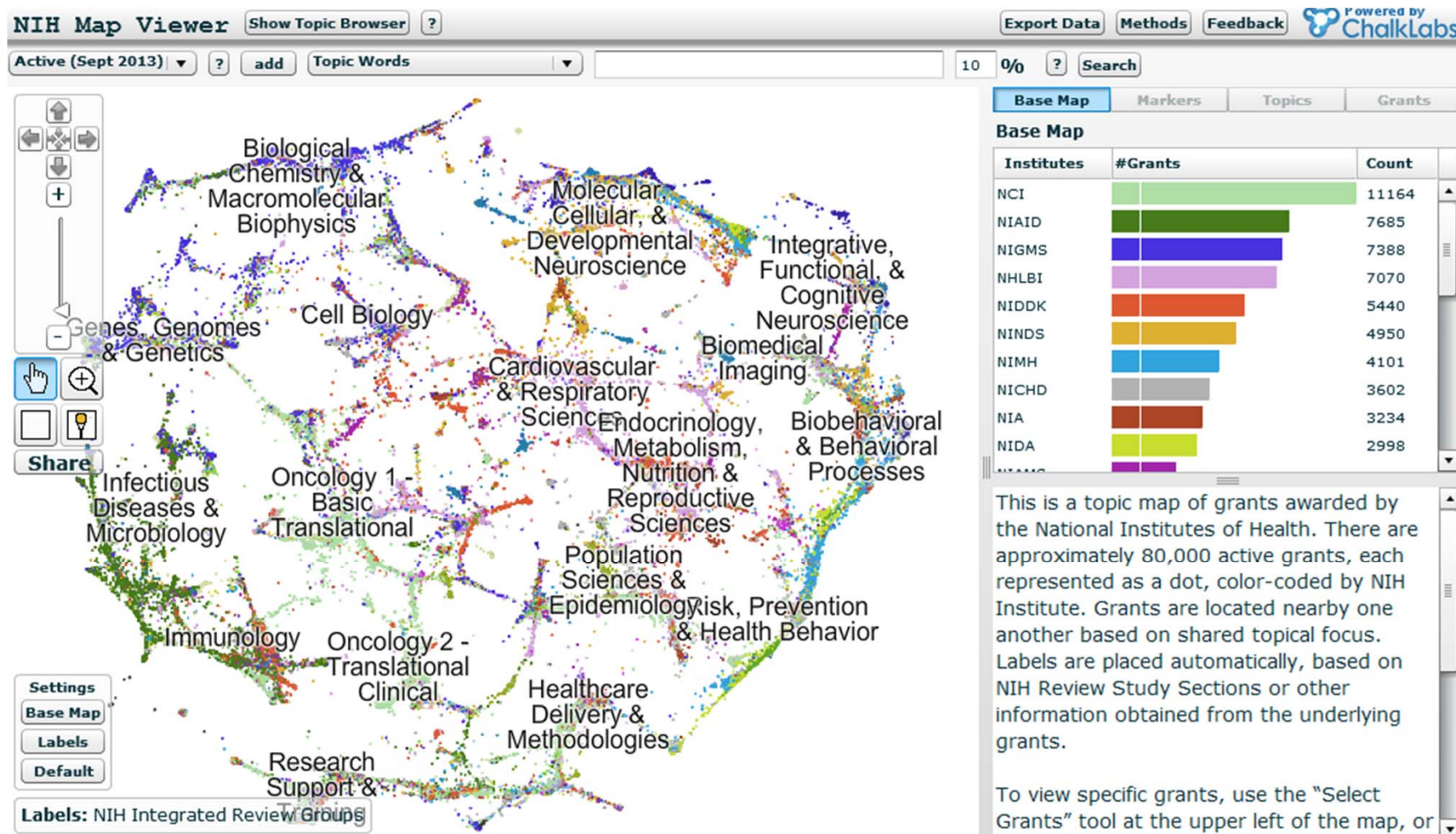
### The #Oscars2014: the awards in 50 seconds

Geo-tagged Tweets about the Oscars, starting March 2, 5pm ET



Create your own custom maps with [CartoDB](#)

[http://srogers.cartodb.com\\_viz/e90c0ef0-a31e-11e3-878e-0e10bcd91c2b/embed\\_map?title=true&description=true&search=true&shareable=true&cartodb\\_logo=true&layer\\_selector=false&legends=false&scrollwheel=true&fullscreen=true&sublayer\\_options=1&sql=&zoom=2&center\\_lat=2.6357885741666065&center\\_lon=-33.57421874999998](http://srogers.cartodb.com_viz/e90c0ef0-a31e-11e3-878e-0e10bcd91c2b/embed_map?title=true&description=true&search=true&shareable=true&cartodb_logo=true&layer_selector=false&legends=false&scrollwheel=true&fullscreen=true&sublayer_options=1&sql=&zoom=2&center_lat=2.6357885741666065&center_lon=-33.57421874999998)



<https://app.nihmaps.org>



Client-side vis framework uses jQuery, Angular, D3, and Leaflet (tiling map library that supplants Google Maps); HTML5, Javascript and SVG.

# MAPSustain

Mapping Biomass Sustainability Research

Geographic Map
Science Map

Funding

NIH

NSF

USDA

Amount  Count

From year  to year

Search by keyword

Publications

DOE

ISI

Medline

Citations  Count

Patents

USPTO

Citations  Count

3 to 467

1 to 765

1 to 88

Maps Detail Data About

### Geographic Visualization

Where is what research performed? This map shows papers, patents, and grants related to biomass and biofuel research and technology. Only work in the US is shown.

Records are aggregated at the state (high) and city (low) level. These levels are accessed by using the zoom feature of the Google Map Interface. Zooming in will break the top level markers apart into the lower level clusters. The maps also allow users to drag the basemap to expose a different area, pan the map along the cardinal directions using arrow controls, and interact directly with the data markers that have been placed on the maps.

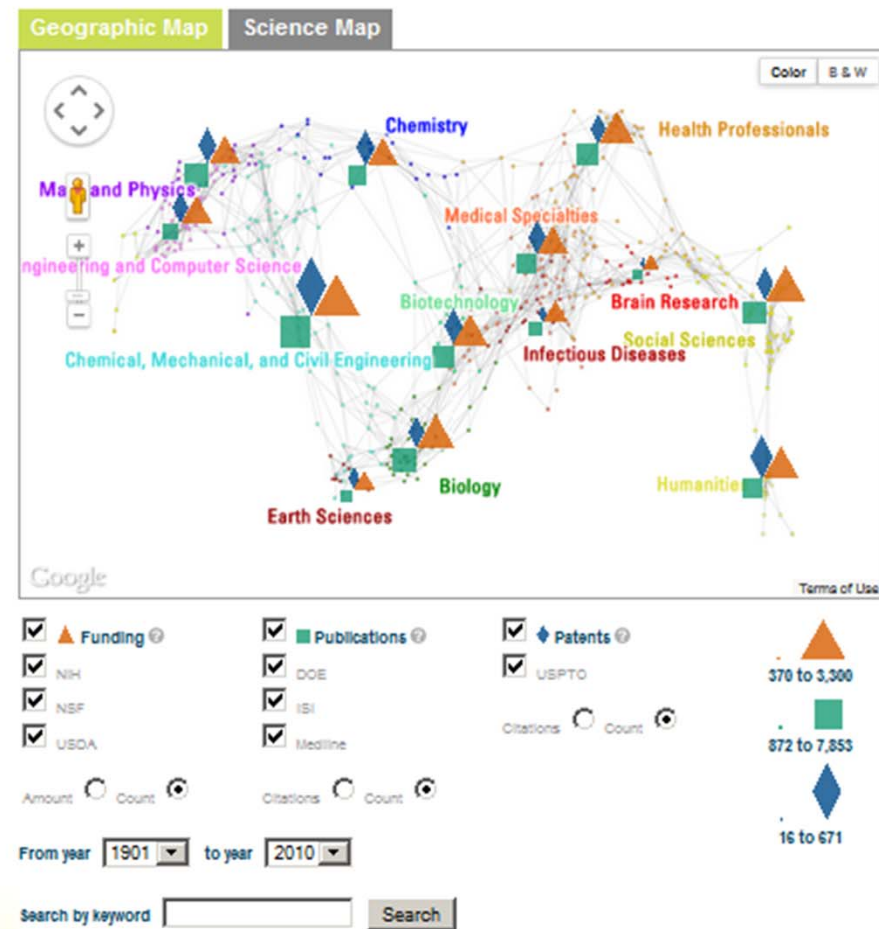
When the mouse hovers over a marker, the data records that are represented by that marker are summarized in terms of marker title, number of records, and distribution of those records across the various data sources. Clicking on the marker updates the detail pane to show a list of the titles of the records, grouped by data source and date. Each title is a link that points to the URL for that record on the data source's website.

The maps also include controls to select subsets of the entire sustainability data set. Checkboxes allow users to limit data sources by Institution or by source type (i.e., funding, publications, and patents). Radio buttons toggle the markers between representing a count of the records or another relevant numerical representation (dollar amount for funding, citations for publications and patents). Records can also be limited by date by selecting start and end years for the query. A keyword search can also restrict the query to only those records that contain a particular word or phrase.

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

# MAPSustain

Mapping Biomass Sustainability Research



Maps Detail Data Abo

## Map of Science

In what areas of science is research performed? This map shows papers, patents, and grants related to biomass and biofuel research and technology overlaid on the UCSD base map of science. Only work in the US is shown.

Records (matched by journal name or keyword) are aggregated both at the upper level of 13 broad scientific disciplines and at the lower level of 554 subdiscipline clusters. Labels for the 13 scientific disciplines appear on the science map and provide a legend for the color of the subdiscipline clusters appearing on the background tiles. The default zoom level on each map will display the high level aggregation of the data. Zooming in will break the top level markers apart into the lower level clusters.

The maps also allow users to drag the basemap to expose a different area, pan the map along the cardinal directions using arrow controls, and interact directly with the data markers that have been placed on the maps.

When the mouse hovers over a marker, the data records that are represented by that marker are summarized in terms of marker title, number of records, and distribution of those records across the various data sources. Clicking on the marker updates the detail pane to show a list of the titles of the records, grouped by data source and date. Each title is a link that points to the URL for that record on the data source's website.

The maps also include controls to select subsets of the entire sustainability data set. Checkboxes allow users to limit data sources by institution or by source type (i.e., funding, publications, and patents). Radio buttons toggle the markers between representing a count of the records or another relevant numerical representation (dollar amount for funding).

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



**CNS**

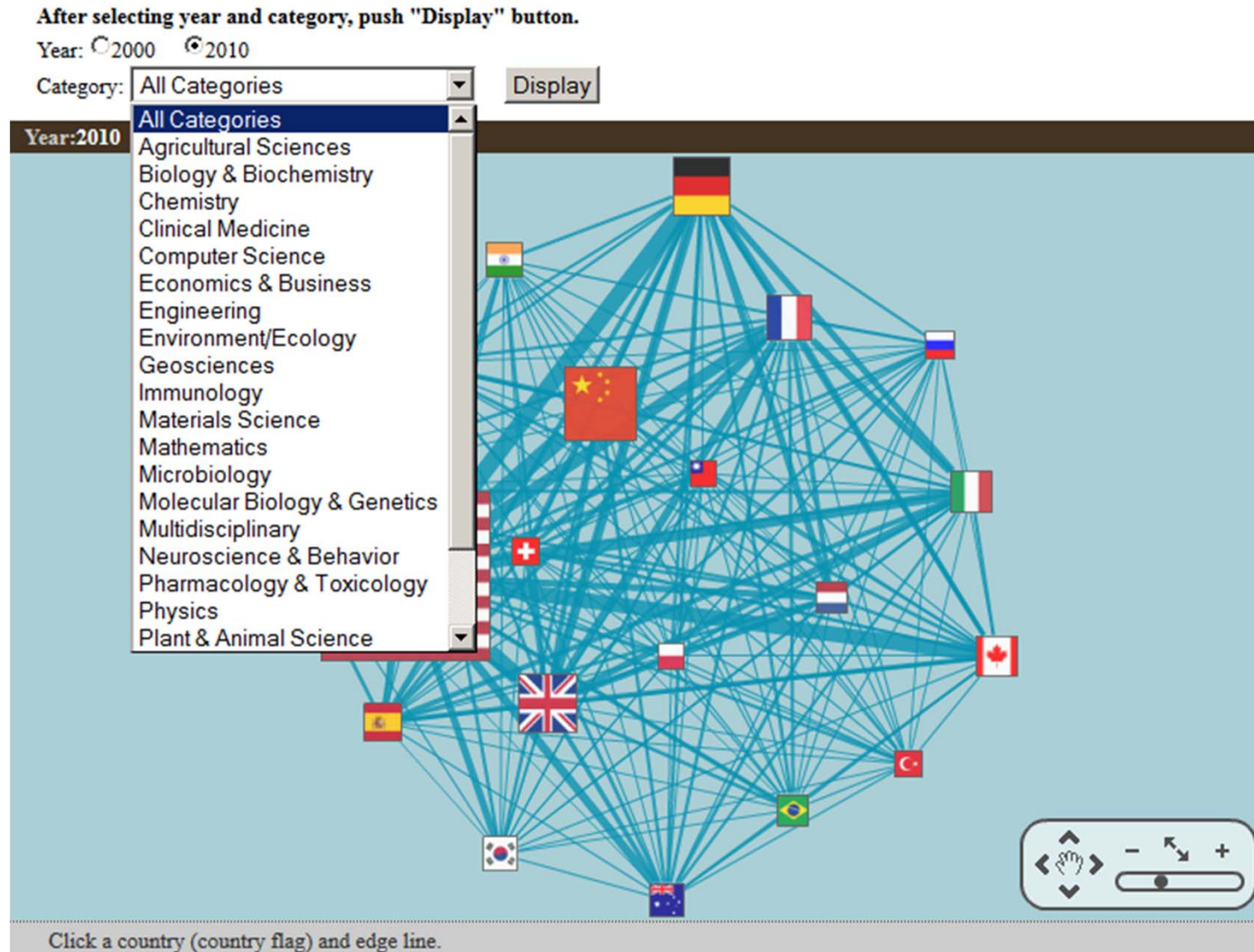
Cyberinfrastructure for  
Network Science Center

Open Tools:

Network Analysis and Visualization  
(WITH WHOM)



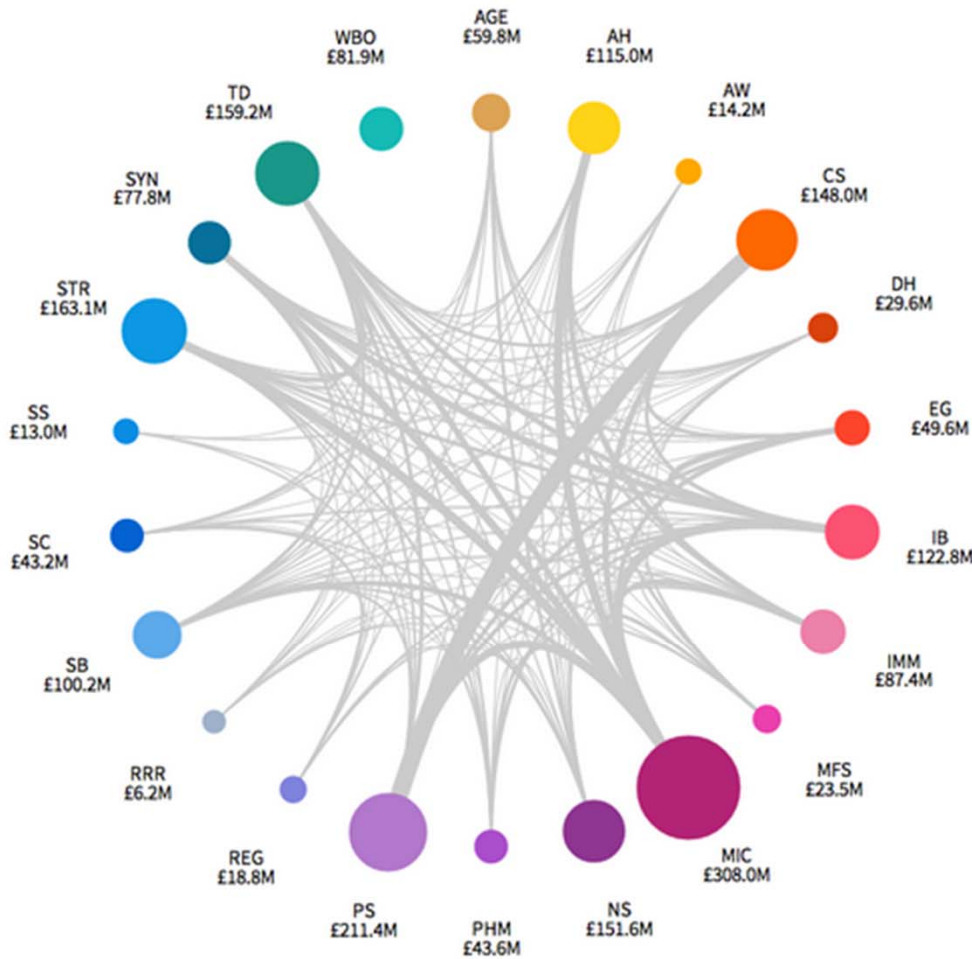
Flash-based. Filtering is handled outside vis client and requires the user to resubmit the page



<http://foresight.jst.go.jp/en/dataranking/collaboration/relationship>



Client-side visualization framework uses jQuery and D3; HTML5, Javascript and SVG.



**BBSRC**  
bioscience for the future

### Advanced Grants Search

**ANALYSIS TYPE:**  
 All  Current  Completed

**FUND TYPE(S):**  
▾ (all)

**INSTITUTION(S):**  
▾ The James Hutton Institute

**RESEARCH TOPIC(S):**  
▾ (all)

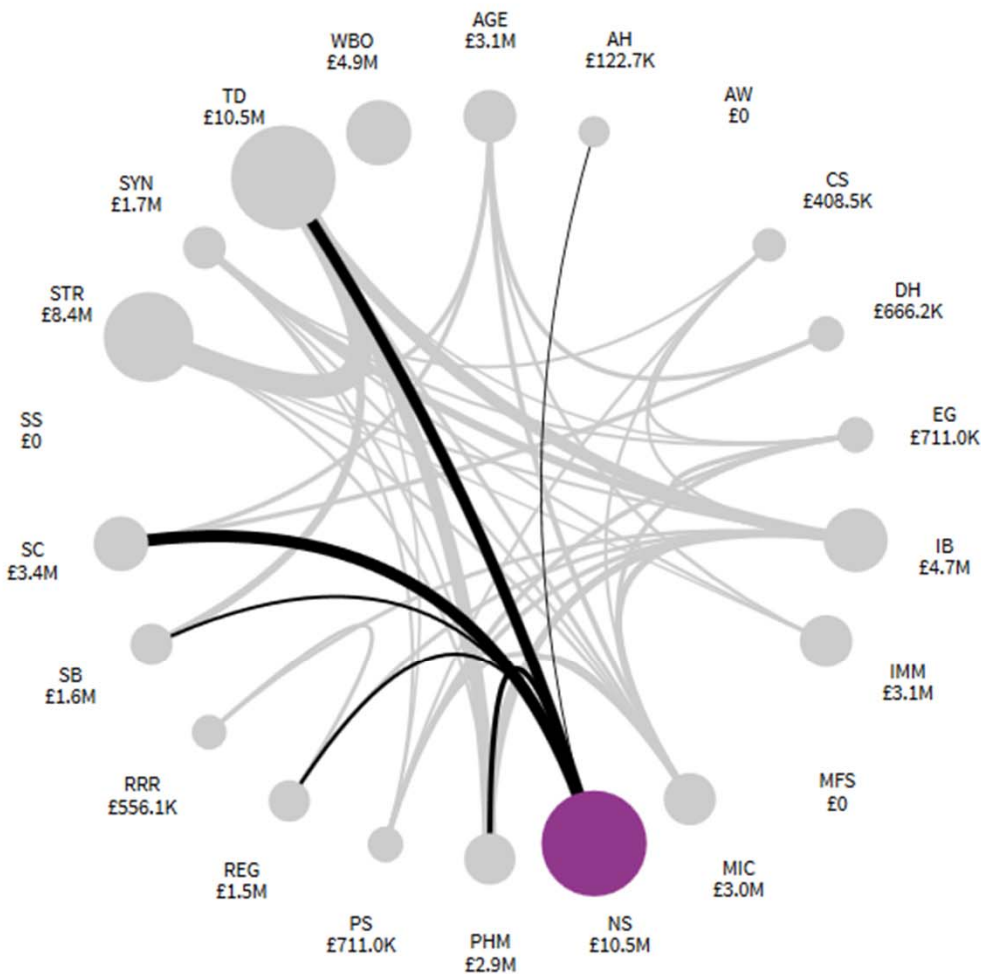
**CATEGORY(S):**  
▾ (all)

**TIME FRAME:**  
 Session Year(s): ▾ (all)  
 Spend Year(s): ▾ (all)

▶

Online interface allows BBSRC research managers to visualize their portfolio across specific topics, universities, and other organizations' funding mechanisms.





### Parameters

#### Analysis Type(s):

Current

#### Fund Type(s):

(all)

#### Institution(s):

University College London

#### Research Topic(s):


(all)

#### Category(ies):

(all)

#### Amount Type:

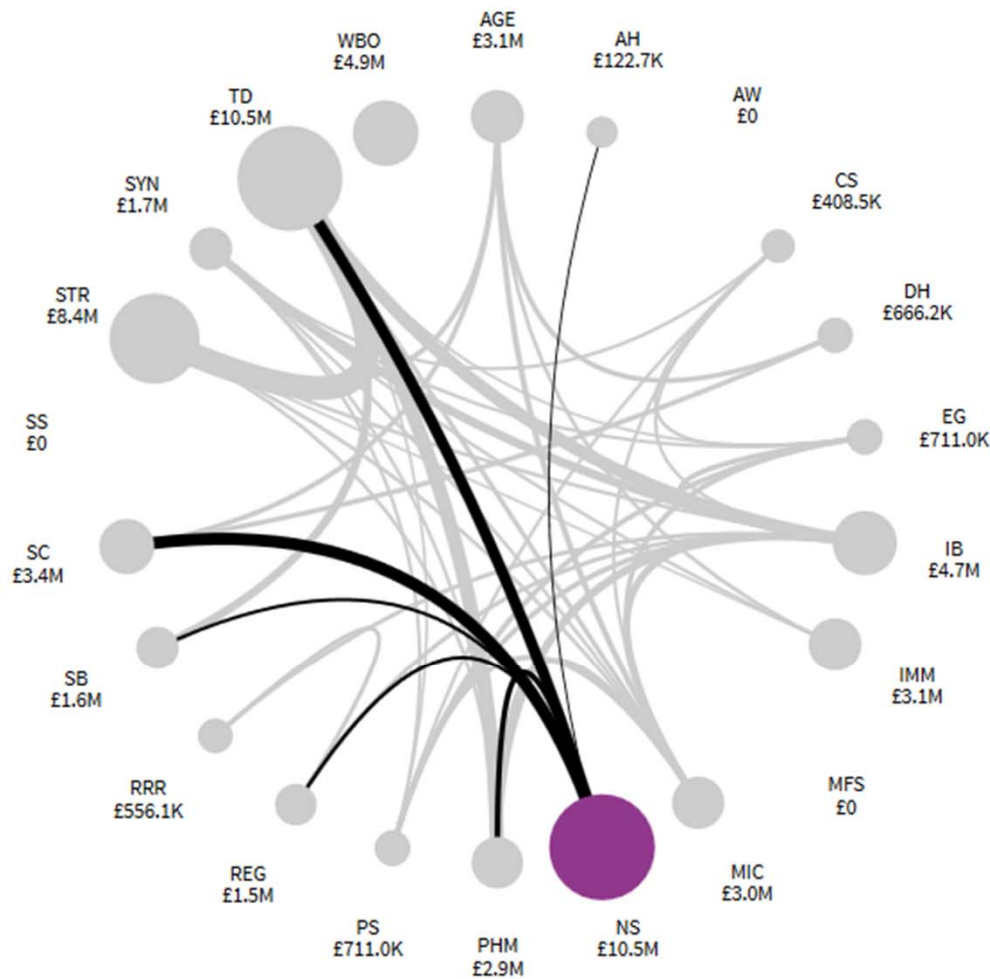
Total Value

 **NS - Neuroscience and Behaviour**  
(including human psychology)  
Total Amount: £10.5M  
Grant Count: 21

AMOUNT



Online interface allows BBSRC research managers to visualize their portfolio across specific topics, universities, and other organizations' funding mechanisms.



**RESEARCH TOPIC:**

Neuroscience and Behaviour (including human psychology) (NS)

**GRANT(S):**

**Title:** Top-down and bottom-up selective mechanisms in attention: subcortical convergence in visual thalamus?

Ref. No: [BB/G022305/1](#)

PI: Professor AM Sillito

Total Value: £1.3M

Institution: University College London

**Title:** Generation of an interactive online atlas of developmental neuroanatomy of the zebrafish brain

Ref. No: [BB/H012516/1](#)

PI: Professor S Wilson

Total Value: £917.4K

Institution: University College London

**Title:** Identifying the signal in the noise: a systems approach for examining invariance in auditory cortex

Ref. No: [BB/H012516/1](#)



Online interface allows BBSRC research managers to visualize their portfolio across specific topics, universities, and other organizations' funding mechanisms.



**CNS**

Cyberinfrastructure for  
Network Science Center

Open Tools:

Multiple Views—Coupled Windows

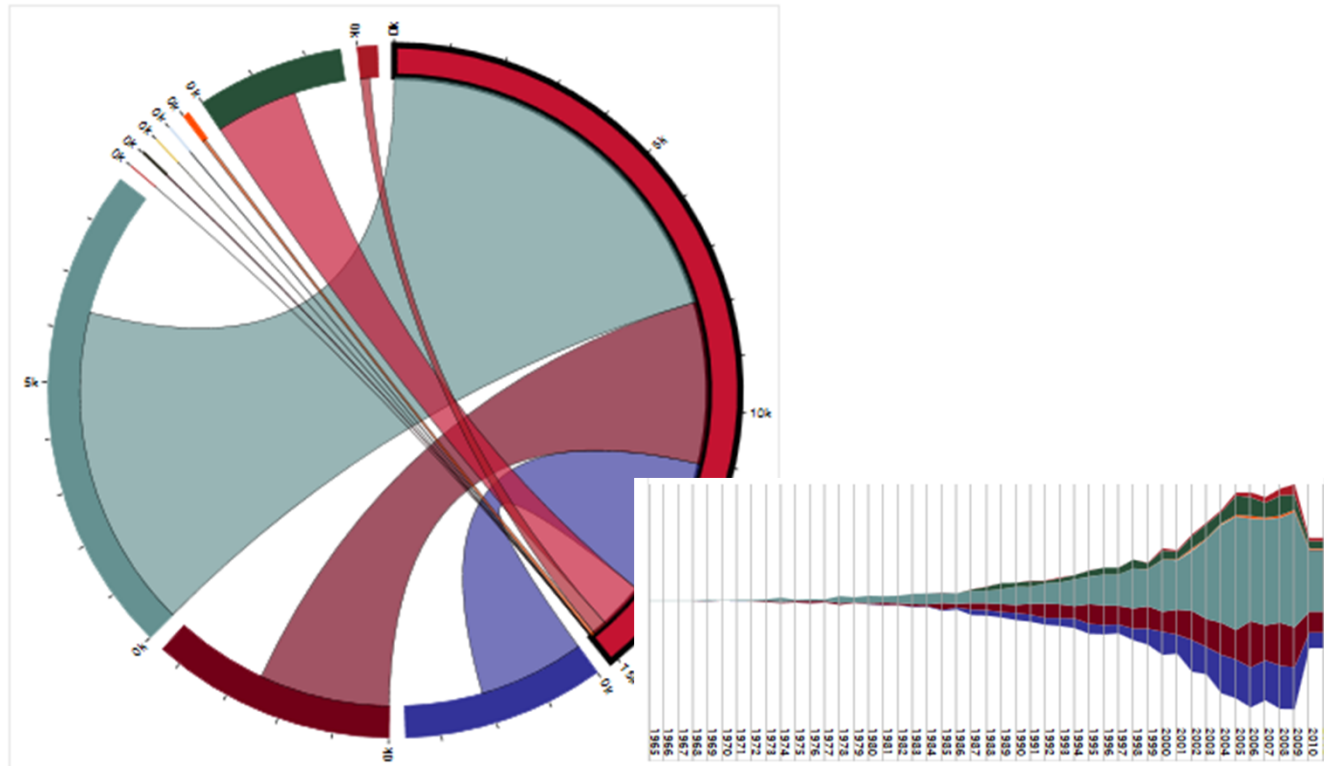
# Inter-Institutional Collaboration Explorer

[View Introduction Video](#) | [Explore](#)

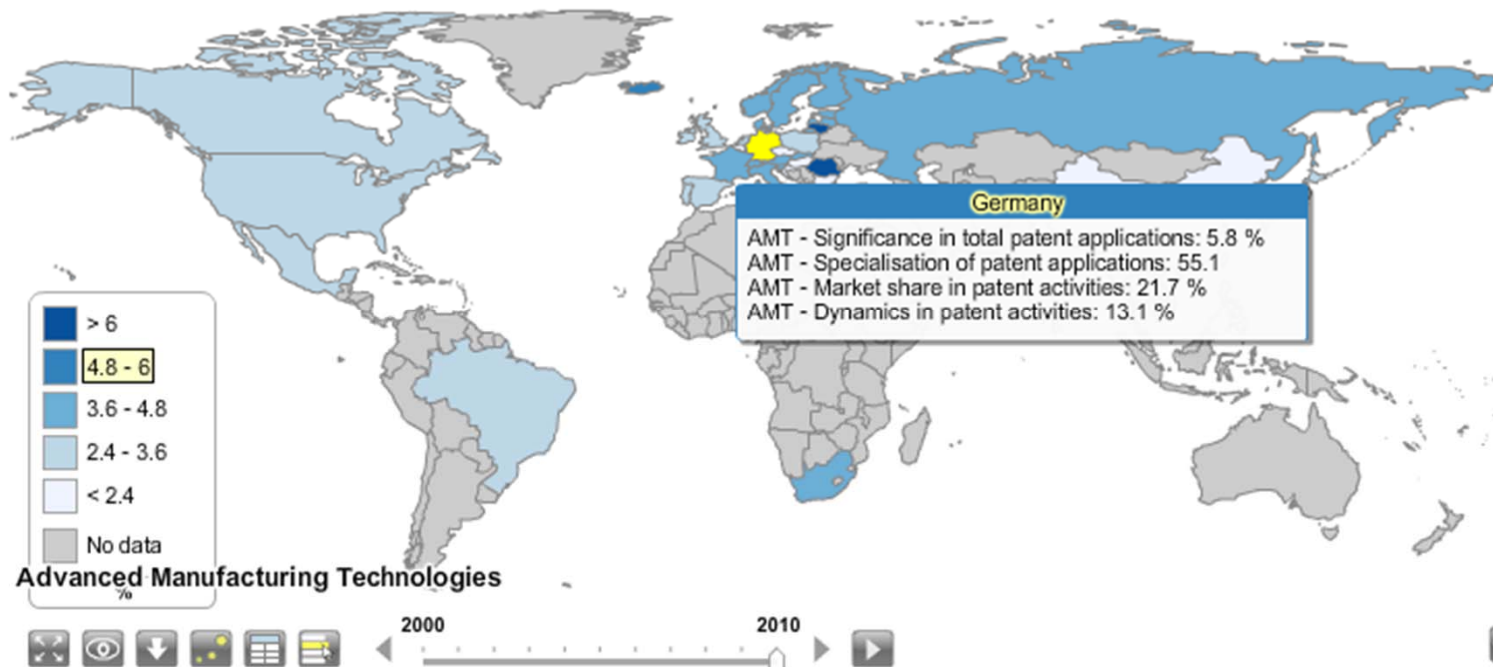
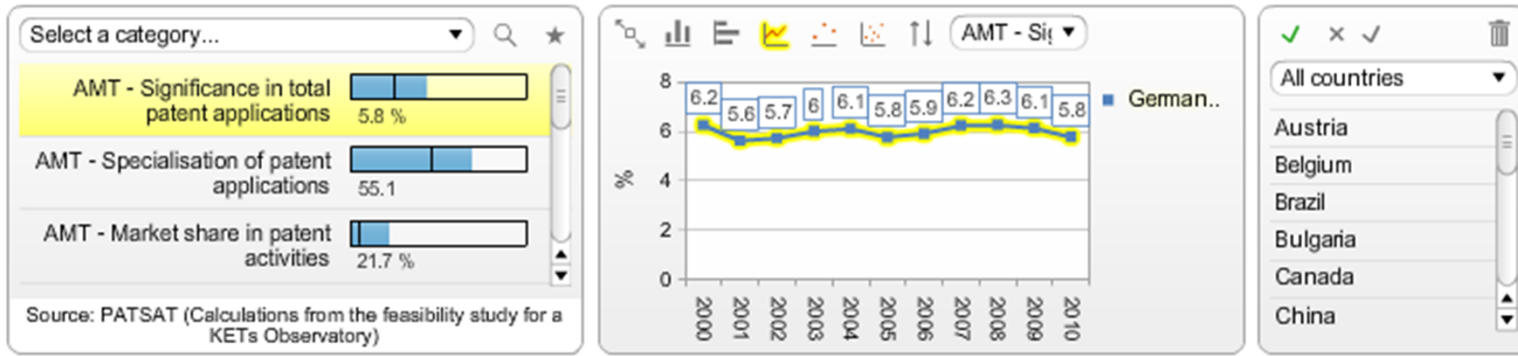
## Collaboration Network

This visualization shows information about collaborative publications – publications that were found at 2 or more Researcher Networking websites. The outer solid colored arcs represent the various institutions. The size of the arc is proportional to the number of collaborative publications found on the site. The inner colored bands represent the number of collaborative publications found between the two institutions that each band connects. Clicking an institution's arc will hide any bands not connected to that institution and will display a timeline of when that institution's collaborative publications were written. Click the institution's arc again to go back to the main visualization. Use the checkboxes on the left to select which institutions are included in the visualization.

Institutions	
<input checked="" type="checkbox"/>	<a href="#">Harvard Med School</a> (15124)
<input checked="" type="checkbox"/>	<a href="#">Northwestern Med School</a> (1813)
<input checked="" type="checkbox"/>	<a href="#">U. of Minnesota</a> (2126)
<input checked="" type="checkbox"/>	<a href="#">U. California at San Fran</a> (5091)
<input checked="" type="checkbox"/>	<a href="#">Cornell</a> (1)
<input checked="" type="checkbox"/>	<a href="#">Cornell Medical</a> (7)
<input checked="" type="checkbox"/>	<a href="#">Ponce School of Med</a> (3)
<input checked="" type="checkbox"/>	<a href="#">Scripps Research Institute</a> (9)
<input checked="" type="checkbox"/>	<a href="#">Univ. of Florida</a> (56)
<input checked="" type="checkbox"/>	<a href="#">Washington U at St. Louis</a> (824)
<input checked="" type="checkbox"/>	<a href="#">Mendeley</a> (174)



<http://xcite.hackerceo.org/VIVOViz/visualization.html>



<https://webgate.ec.europa.eu/ketsobservatory/kets-deployment/technology/timeseries/amt>



# Open Education



# CNS

Cyberinfrastructure for  
Network Science Center

260 courses available

By start date ▾

By subject ▲

- + Computer Science ●
- + Health & Medicine ●
- + Mathematics and Stats ●
- + Business & Management ●
- + Humanities ●
- + Engineering ●
- + Science ●
- Education & Teaching ●
- + Social Sciences ●

Add	Course Name	Start Date	Rating
+	<b>Data Analysis</b> <small>Johns Hopkins University via Coursera</small>	28th Oct, 2013	★★★★★
+	<b>Organizational Analysis</b> <small>Stanford University via Coursera</small>	31st Mar, 2014	★★★★★
+	<b>Exploratory Data Analysis</b> <small>Facebook via Udacity</small>	Self paced	★★★★★
+	<b>Financial Analysis</b> <small>First Finance Institute via First Business MOOC</small>	24th Feb, 2014	★★★★★
+	<b>Exploratory Data Analysis</b> <small>Johns Hopkins University via Coursera</small>	7th Jul, 2014	★★★★★
+	<b>Tourism Industry Analysis</b> <small>University of Central Florida via Canvas.net</small>	23rd Sep, 2013	★★★★★
+	<b>Computational Methods for Data Analysis</b> <small>University of Washington via Coursera</small>	9th Dec, 2014	★★★★★
+	<b>Computing for Data Analysis</b> <small>Johns Hopkins University via Coursera</small>	6th Jan, 2014	★★★★★

<https://www.class-central.com/search?q=analysis>



20 courses available

By start date

By subject

- Computer Science
- + Health & Medicine
- + Mathematics and Stats
- + Humanities
- + Engineering
- + Science
- + Education & Teaching
- + Social Sciences
- Art & Design

Add	Course Name	Start Date	Rating
+	<b>Introduction to Statistics</b> Stanford University via <b>Udacity</b>	Self paced	★★★★★
+	<b>Passion Driven Statistics</b> Wesleyan University via <b>Coursera</b>	25th Mar, 2013	★★★★★
+	<b>Introduction to Infographics and Data Visualization</b> Knight Center for Journalism in the Americas via <b>Independent</b>	6th Oct, 2013	★★★★★
+	<b>Big Data in Education</b> Columbia University via <b>Coursera</b>	24th Oct, 2013	★★★★★
+	<b>Information Visualization</b> Indiana University via <b>Independent</b>	28th Jan, 2014	★★★★★
+	<b>Intro to Data Science</b> via <b>Udacity</b>	Self paced	★★★★★
+	<b>15.071x: The Analytics Edge</b> Massachusetts Institute of Technology via <b>EdX</b>	4th Mar, 2014	★★★★★
+	<b>High Performance Scientific Computing</b> University of Washington via <b>Coursera</b>	31st Mar, 2014	★★★★★

<https://www.class-central.com/search?q=visualization>

## 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.

This year, the course can be taken for three Indiana University credits as part of the [Online Data Science Program](#) just announced by the School of Informatics and Computing. Students interested in applying to the program can [find more information here](#).

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.

Just like last year, students will have the opportunity to collaborate on real-world projects for a variety of clients. [Click here](#) to see this year's list of clients and projects.

Everyone who registers gains free access to the [Scholarly Database](#) (26 million paper, patent, and grant records) and the [Sci2 Tool](#) (100+ algorithms and tools).

Please watch the introduction video to learn more.



[Register for Course](#)

**IVMOOC 2014 course materials will be available until end of November 2014. The IVMOOC 2015 will open in January 2015 with new materials and a cloud computing setup.**

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

## Tutorial Overview

9:30 Welcome and Overview of Tutorial and Attendees

9:45 Open Data and Tools

- Standardize and federate micro-level datasets of S&T activity
- Open code tools and online services that are interoperable
- Sharing and teaching open datasets and tools

### **11-11:30 Networking Break**

11:30 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- Country Collaboration Network, 2011 – Scoreboard 2013
- Country Expertise Profiles, 2010 – New Workflow
- Country Mobility Network, 1996-2011 – Scoreboard 2013
- Acceleration in the Development of Patented Technologies, 2000-11 – SB 2013
- Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – Scoreboard 2013
- Evolving Country Patent Profiles, 1995-2010 – New Workflow

13:00 Outlook and Q&A

13:30 Adjourn

## Tutorial Overview

9:30 Welcome and Overview of Tutorial and Attendees

9:45 Open Data and Tools

- Standardize and federate micro-level datasets of S&T activity
- Open code tools and online services that are interoperable
- Sharing and teaching open datasets and tools

11-11:30 Networking Break

### **11:30 Sci2 Tool Hands-on**

- **Download and run the Sci2 Tool**
- Country Collaboration Network, 2011 – Scoreboard 2013
- Country Expertise Profiles, 2010 – New Workflow
- Country Mobility Network, 1996-2011 – Scoreboard 2013
- Acceleration in the Development of Patented Technologies, 2000-11 – SB 2013
- Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – Scoreboard 2013
- Evolving Country Patent Profiles, 1995-2010 – New Workflow

13:00 Outlook and Q&A

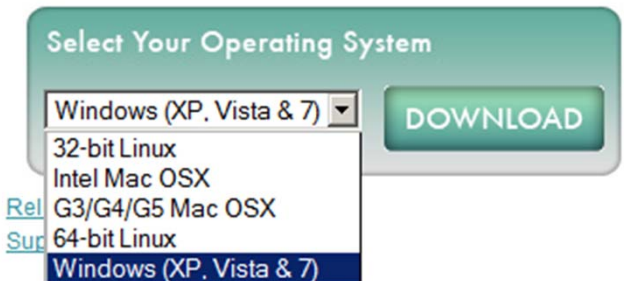
13:30 Adjourn

## Software, Datasets, Plugins, and Documentation

- These slides  
<http://cns.iu.edu/docs/presentations/2014-borner-opentooltutorial-oecd.pdf>
- Sci2 Tool Manual v0.5.1 Alpha, updated to match v1.0 Alpha tool release  
<http://sci2.wiki.cns.iu.edu>
- Sci2 Tool v 1.1 beta (Dec 9, 2013)  
<http://sci2.cns.iu.edu>
- Additional Datasets  
<http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets>
- Additional Plugins  
<http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

### Download

Sci<sup>2</sup> v 1.1 beta  
December 9th, 2013



Make sure you have Java 1.6 (32-bit suffices) or higher installed or download from <http://www.java.com/en/download>. To check your Java version, open a terminal and run 'java -version'.

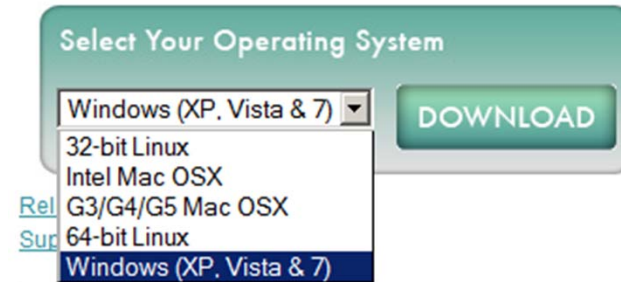
Some visualizations are saved as Postscript files. A free Postscript to PDF viewer is at <http://ps2pdf.com> and a free PDF Viewer at <http://www.adobe.com/products/reader.html>.

## Install and Run Sci2

Sci2 Tool runs on Windows, Mac, and Linux.

### Download

**Sci<sup>2</sup> v 1.1 beta**  
December 9th, 2013



Unzip.



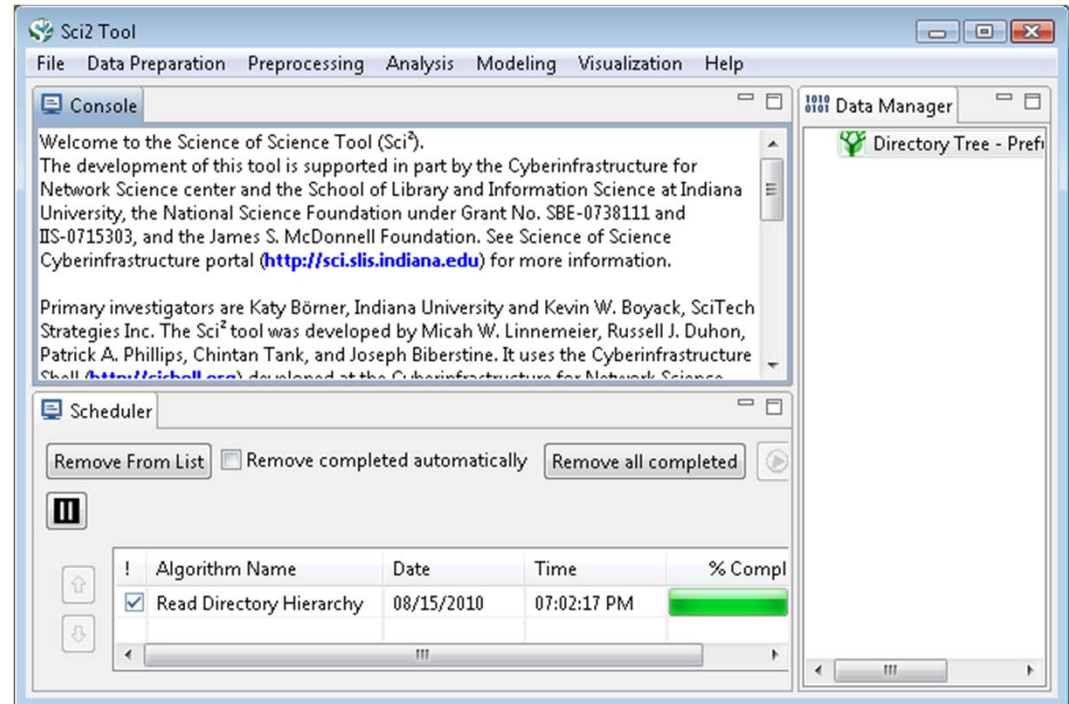
Run /sci2/sci2.exe

## Sci2 Tool Interface Components

See also <http://sci2.wiki.cns.iu.edu/2.2+User+Interface>

### Use

- **Menu** to read data, run algorithms.
- **Console** to see work log, references to seminal works.
- **Data Manager** to select, view, save loaded, simulated, or derived datasets.
- **Scheduler** to see status of algorithm execution.



All workflows are recorded into a log file (see /sci2/logs/...), and can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

All algorithms are documented online; workflows are given in Sci2 Manual at

<http://sci2.wiki.cns.iu.edu>

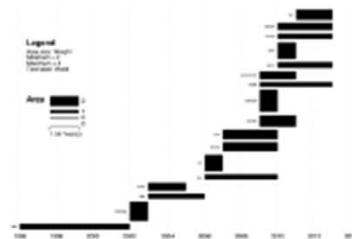
## Load **One** File and Run **Many** Analyses and Visualizations

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, N 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, N 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

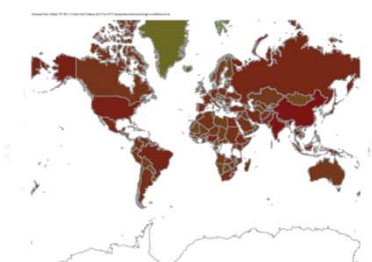
Temporal Burst Analysis—p. 48



Geospatial Analysis—p. 52



Geospatial Analysis—p. 52

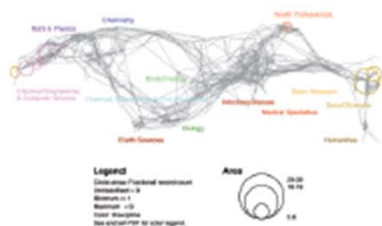




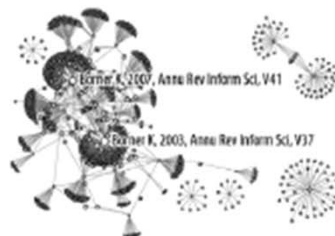
# Load **One** File and Run **Many** Analyses and Visualizations

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, N 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, N Falk-Krzesinski, HJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B

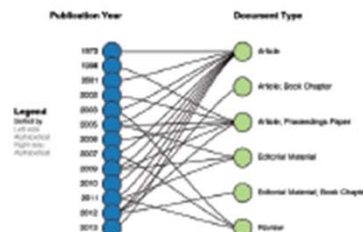
Topical Analysis—p. 56



Paper Citation Network—p. 60



Bi-Modal Network—p. 60



Co-author and many other bi-modal networks.

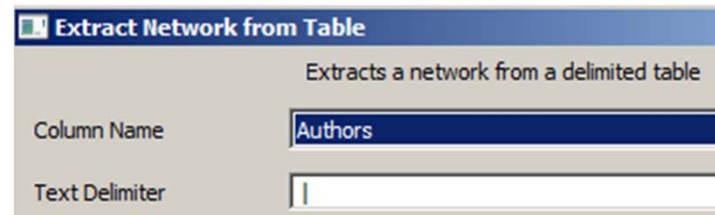
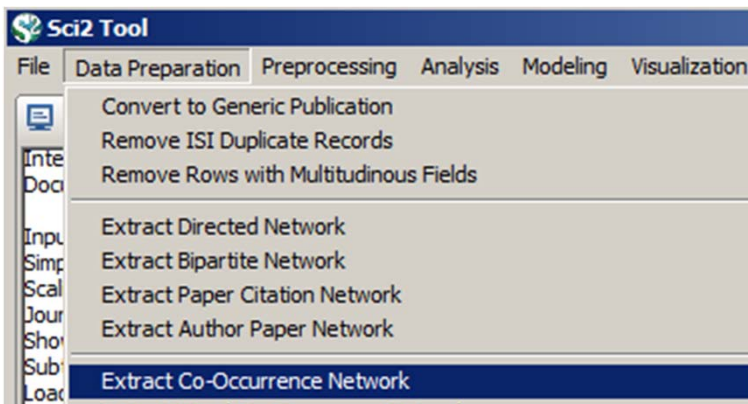
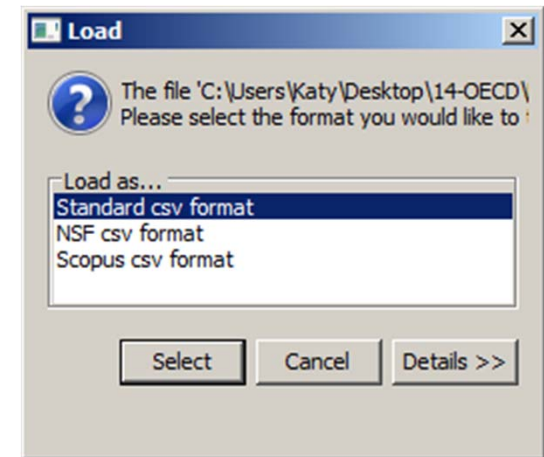
## Load **One** File and Run **Many** Analyses and Visualizations

Download 20publications.csv from

<http://wiki.cns.iu.edu/download/attachments/1245848/20publications.csv?version=1&modificationDate=1403450235951>

In Sci2, use 'File > Load' and load file as 'Standard csv format'.

Run 'Data Preparation > Extract Co-Occurrence Network' with parameters:



Co-author network will appear in **Data Manager**.

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

Run 'Analysis > Network Analysis Toolkit (NAT)' to get basic properties:

Nodes: 65

Isolated nodes: 0

Edges: 404

No self loops were discovered.

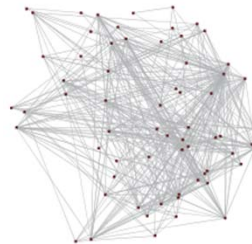
Average degree: 12.4308

The largest connected component consists of 65 nodes.

Density (disregarding weights): 0.1942

Select 'Extracted Network on Column Authors' network in Data Manager and run 'Visualization > GUESS' to open GUESS with file loaded.

Initial layout is random:



In GUESS, apply 'Layout > GEM':



## Tutorial Overview

9:30 Welcome and Overview of Tutorial and Attendees

9:45 Open Data and Tools

- Standardize and federate micro-level datasets of S&T activity
- Open code tools and online services that are interoperable
- Sharing and teaching open datasets and tools

11-11:30 Networking Break

11:30 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- **Country Collaboration Network, 2011 – Scoreboard 2013**
- **Country Expertise Profiles, 2010 – New Workflow**
- **Country Mobility Network, 1996-2011 – Scoreboard 2013**
- **Acceleration in the Development of Patented Technologies, 2000-11 – SB 2013**
- **Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – Scoreboard 2013**
- **Evolving Country Patent Profiles, 1995-2010 – New Workflow**

13:00 Outlook and Q&A

13:30 Adjourn

# Country Collaboration Network, 2011

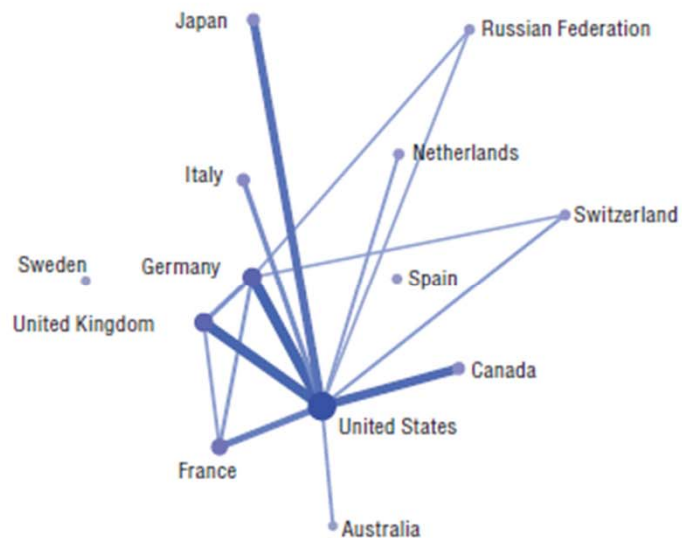
see OECD Scoreboard 2013, p. 58 and 59

## Collaboration in scientific research

In the global landscape of scientific research, scientific output has grown rapidly and collaboration between institutions in different countries has intensified. The emergence of new players has changed the structure of global collaboration networks.

**54a. International collaboration networks in science, 1998**

*Whole counts of internationally co-authored documents*



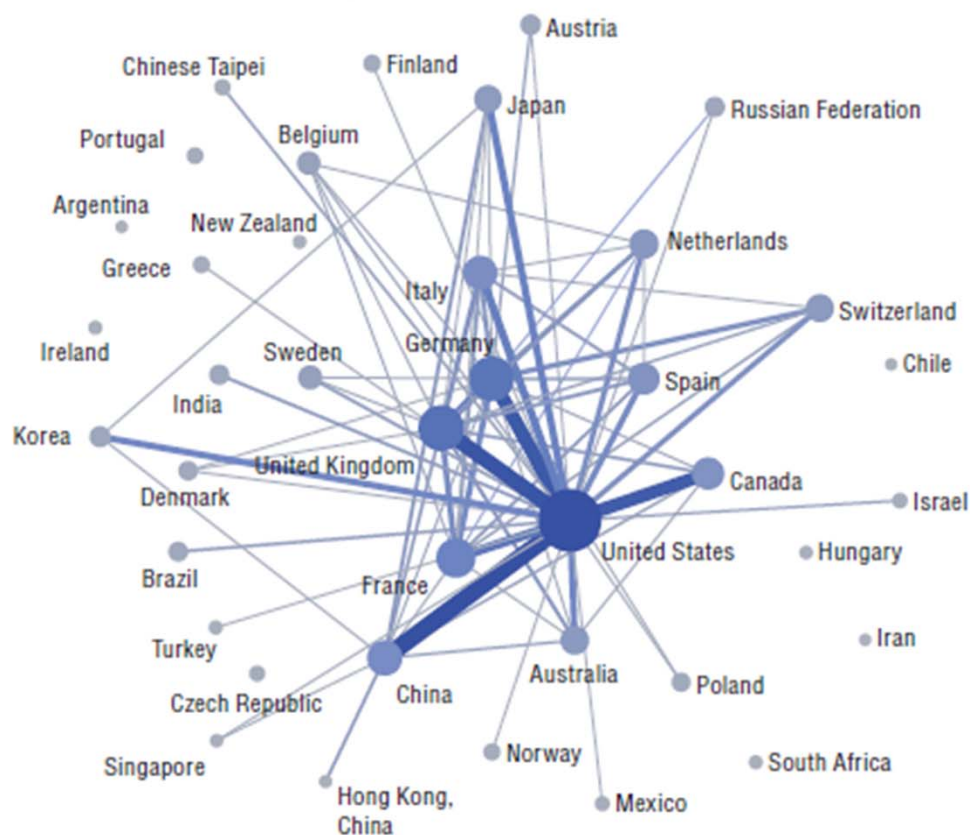
Source: OECD calculations based on Scopus Custom Data, Elsevier, version 5.2012, June 2013. See chapter notes.

## Collaboration in scientific research

China and several other economies have become increasingly integrated in the global science system. China accounted for more than 74 000 collaborations in 2011 compared with only 9 000 in 1998. Over the period, its number of co-authored documents with US-based institutions increased from nearly 2 000 to more than 22 000. The United States continues to be at the centre of the international research network, accounting in 2011 for nearly 15% of all scientific collaborations documented in peer-reviewed scientific publications.

### 54b. International collaboration networks in science, 2011

*Whole counts of internationally co-authored documents*



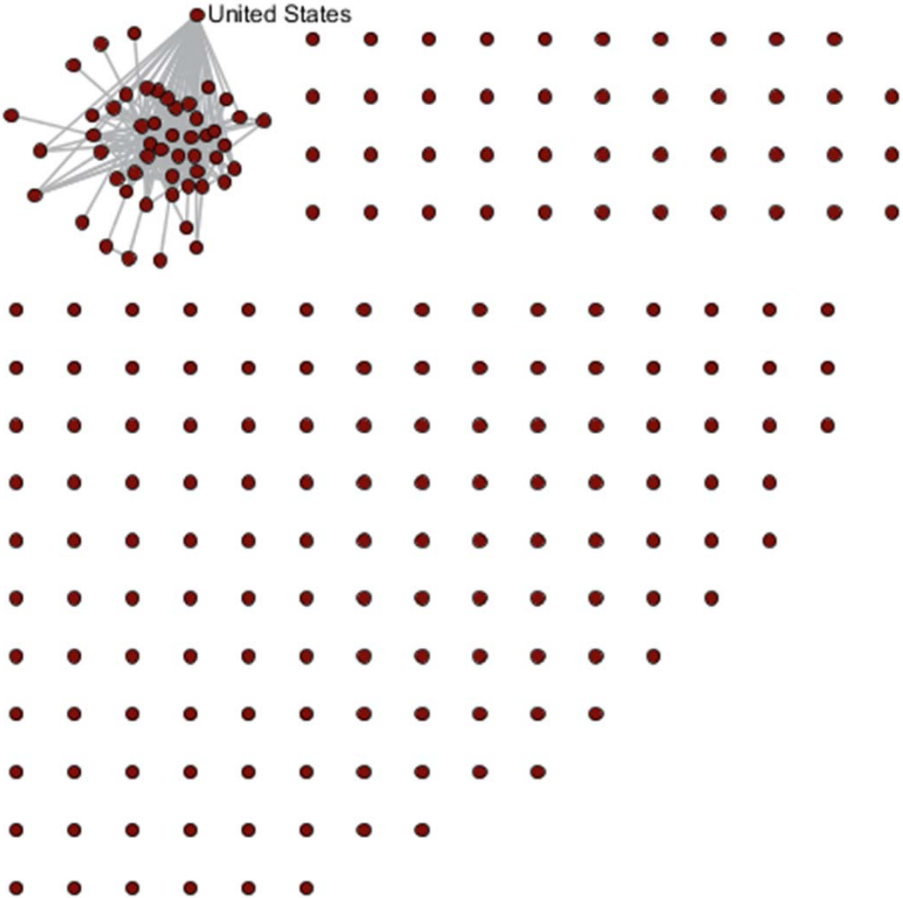
Source: OECD calculations based on Scopus Custom Data, Elsevier, version 5.2012, June 2013. See chapter notes.

# Compile data

	A	B	C
1	A	B	count
2	United States	China	22179
3	China	United States	22179
4	United States	United Kingdom	19427
5	United Kingdom	United States	19427
6	United States	Germany	17791
7	Germany	United States	17791
8	United States	Canada	16697
9	Canada	United States	16697
10	United States	France	11689
11	France	United States	11689
12	United Kingdom	Germany	10181
13	Germany	United Kingdom	10181
14	United States	Italy	10027
15	Italy	United States	10027
16	United States	Japan	9245
17	Japan	United States	9245
18	United States	Australia	8446
19	Australia	United States	8446
20	United States	South Korea	8429
21	South Korea	United States	8429

2011 nodes Unique matrix final nodes

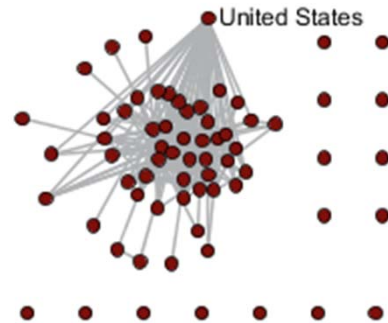
# Extract network, undirected but weighted





## Identify essential nodes

Keep nodes with more than 10,000 and edges with more than 2,000.

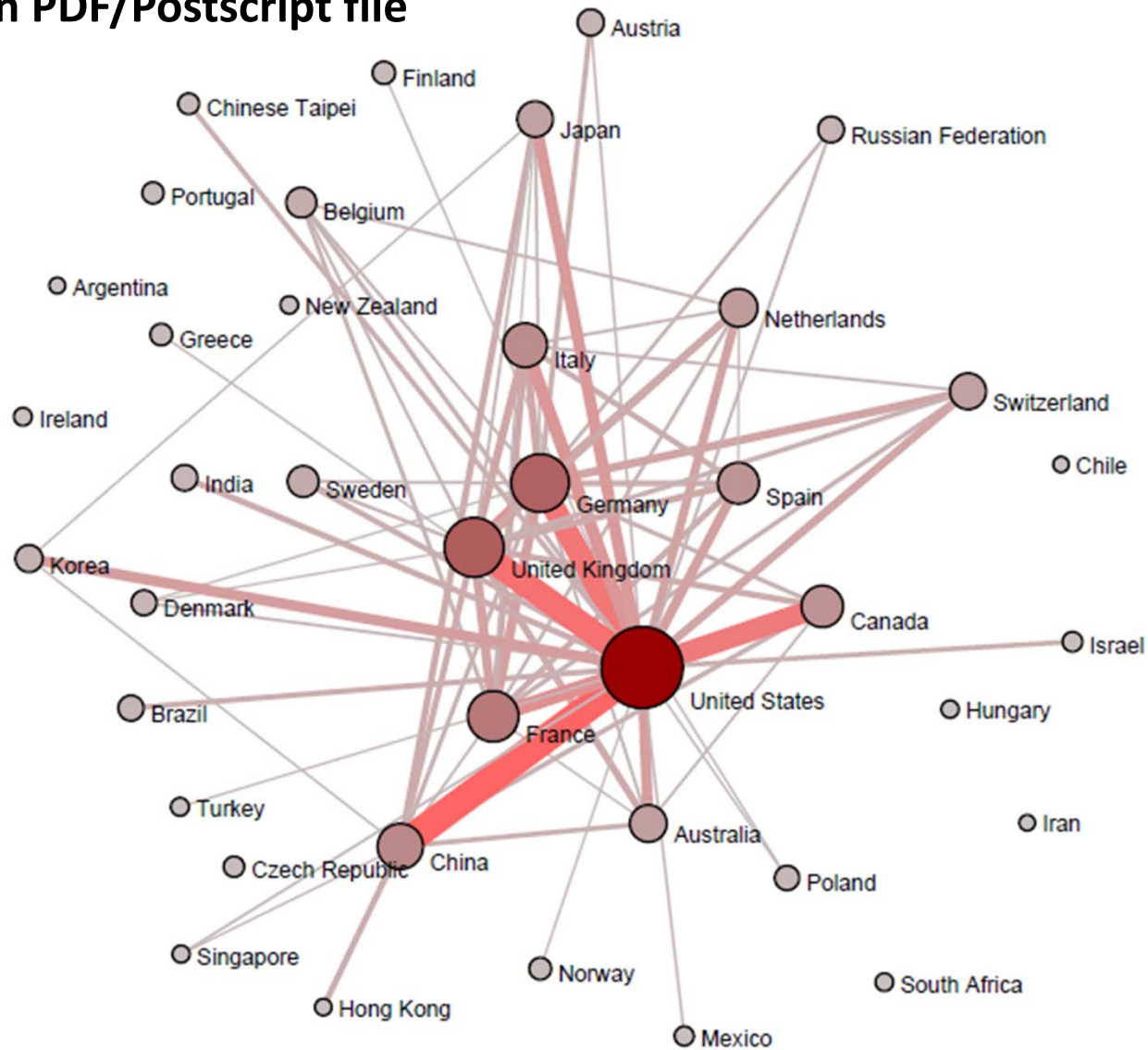


## Identify essential edges

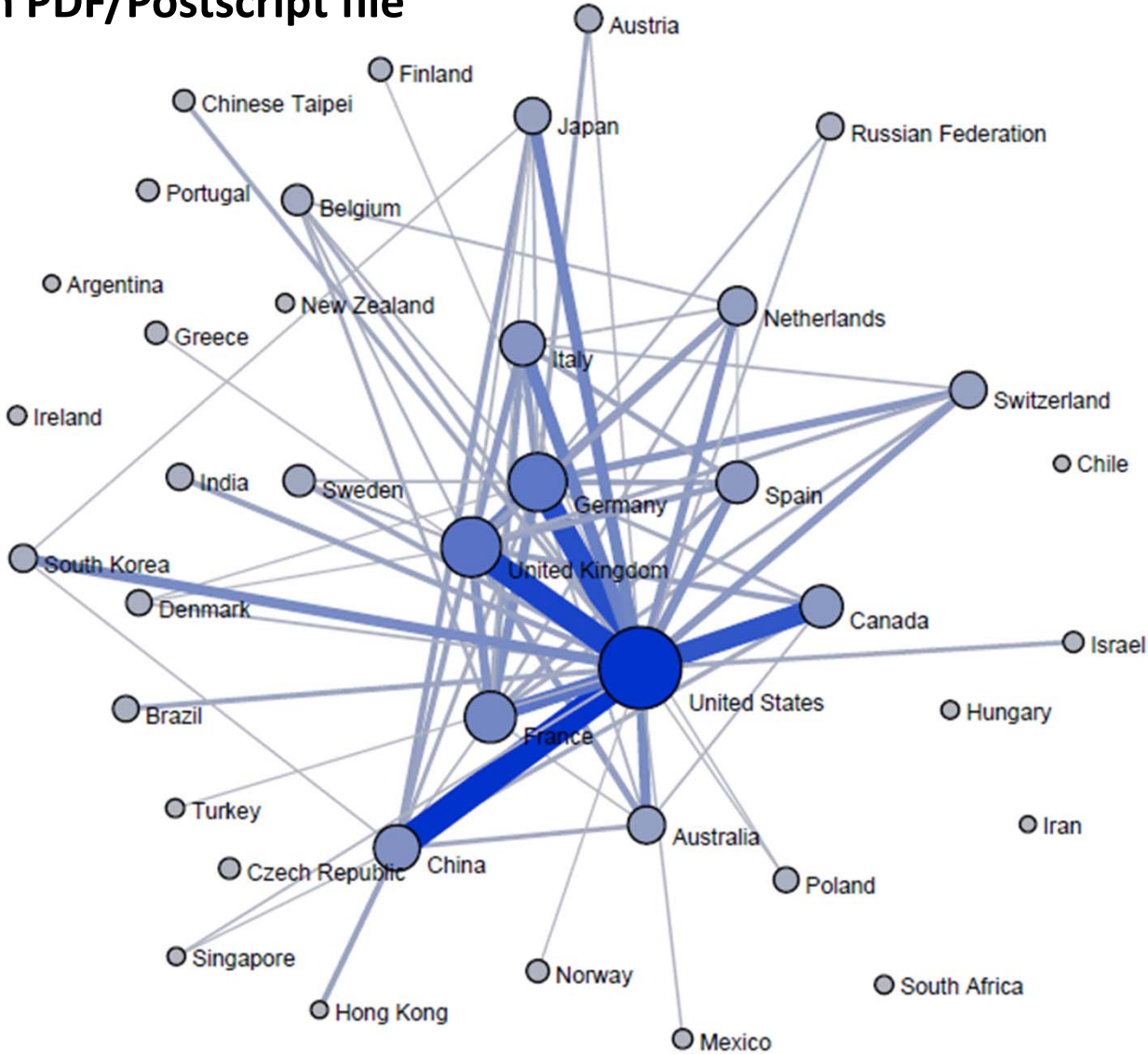
- Apply thresholds to edge weights
- Keep top-n edges per node
- Use Pathfinder Network Scaling

## Layout network, apply color and size coding

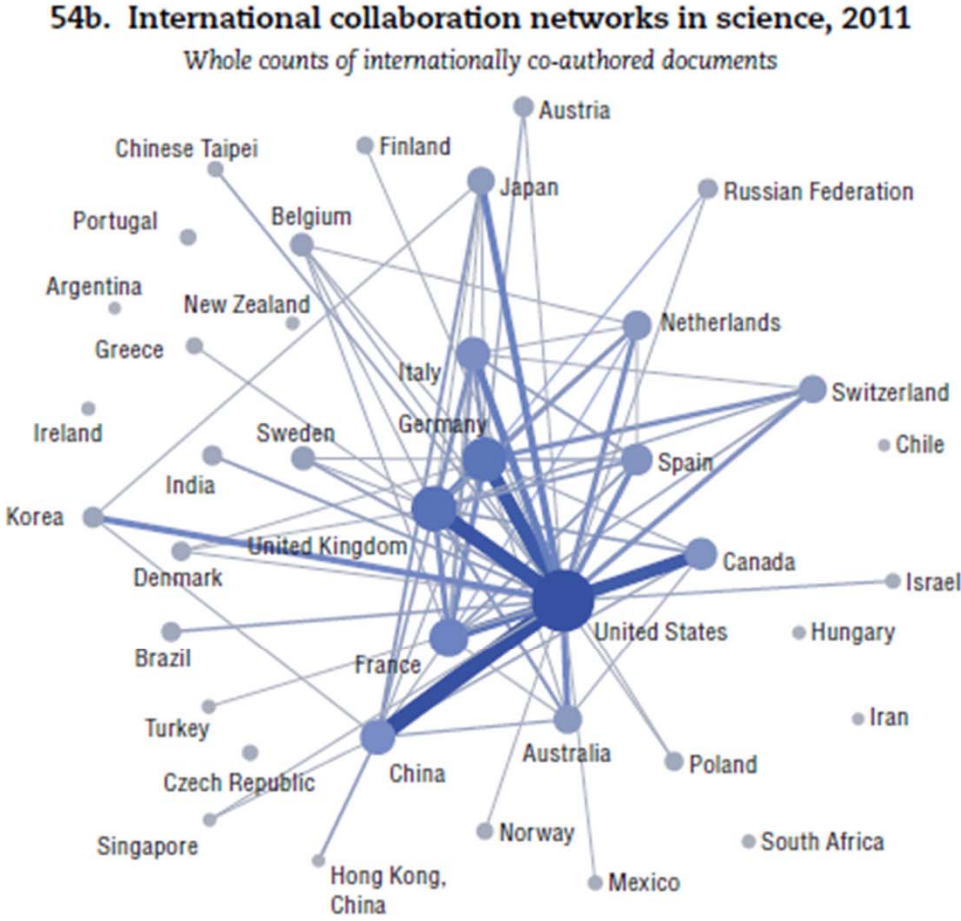
Save result in PDF/Postscript file



**Apply color and size coding**  
**Save result in PDF/Postscript file**



**Adjust colors and type font, add title, legend, data source info  
in your favorite graphic design and layout program**



Source: OECD calculations based on Scopus Custom Data, Elsevier, version 5.2012, June 2013. See chapter notes.

## Country Expertise Profiles: Germany vs. United Kingdom

Objective: Map out and compare country scientific profiles on UCSD map of science.

OECD version of Elsevier's Scopus custom database - scientific publication data for 2010. Analysis carried out within OECD – based on query carried out in Dec 2013 by Brigitte van Beuzekom (OECD), two countries selected for demonstration purposes: 192,583 publications from Germany; 194,221 from United Kingdom. 386,804 total records.

There are exactly 12,690 unique journal names. 2696 of these cannot be located in UCSD map of science and classification system that covers 8 years (2001-2008) of Scopus data. These might be journals created/renamed in 2009 and 2010.

See <http://sci.cns.iu.edu/ucsdmap> for Scopus journal names used in the Sci2 SciMap.

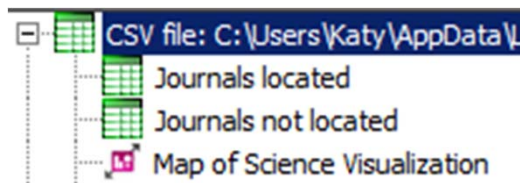
*Börner, Katy, Richard Klavans, Michael Patek, Angela Zoss, Joseph R. Biberstine, Robert Light, Vincent Larivière, and Kevin W. Boyack (2012) Design and Update of a Classification System: The UCSD Map of Science. PLoS ONE 7(7)*

## Prepare data:

Save publication records for each country in one csv file

Load files into Sci2; Render UCSD map overlay using

Three files are generated

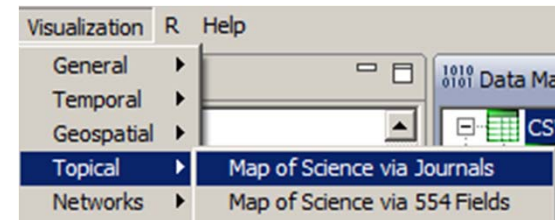


Review "Journals not located"

Correct journal names if possible

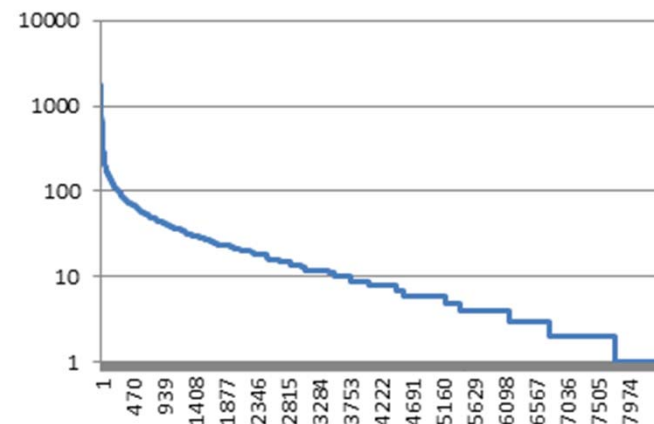
Right click "Map of Science Visualization" in Data Manager and save it as a PostScript file.

Convert to pdf or open in Illustrator or similar to view.



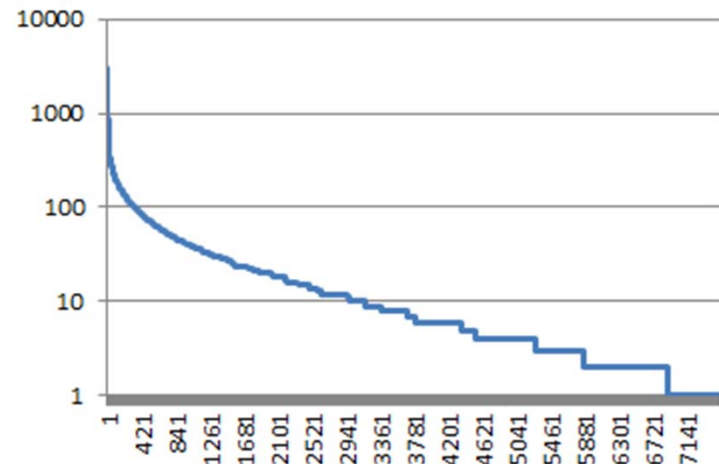
## United Kingdom: Top-20 Most Frequent Journals (8,438 unique)

Journal name	Frequency
lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)	1790
plos one	1365
chemical communications	1360
proceedings of spie - the international society for optical engineering	1210
monthly notices of the royal astronomical society	782
journal of biological chemistry	640
bioorganic and medicinal chemistry letters	624
langmuir	600
journal of the american chemical society	584
blood	456
bioinformatics	420
journal of the royal society interface	415
international journal of std and aids	408
proceedings of the royal society b: biological sciences	405
vaccine	365
journal of theoretical biology	364
british journal of oral and maxillofacial surgery	357
astronomy and astrophysics	354
biochemical journal	350
acm international conference proceeding series	344



## Germany: Top-20 Most Frequent Journals (7,548 unique)

Journal name	Frequency
proceedings of spie - the international society for optical engineering	3080
lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)	3038
plos one	1041
chemical communications	946
astronomy and astrophysics	838
journal of the american chemical society	788
journal of biological chemistry	767
langmuir	766
physical review b - condensed matter and materials physics	676
<a href="#">unfallchirurg</a>	656
<a href="#">nervenarzt</a>	611
acm international conference proceeding series	580
materials science forum	568
journal of physical chemistry c	548
macromolecules	540
<a href="#">chemie-ingenieur-technik</a>	489
<a href="#">konstruktion</a>	436
blood	408
progress in biomedical optics and imaging - proceedings of spie	404



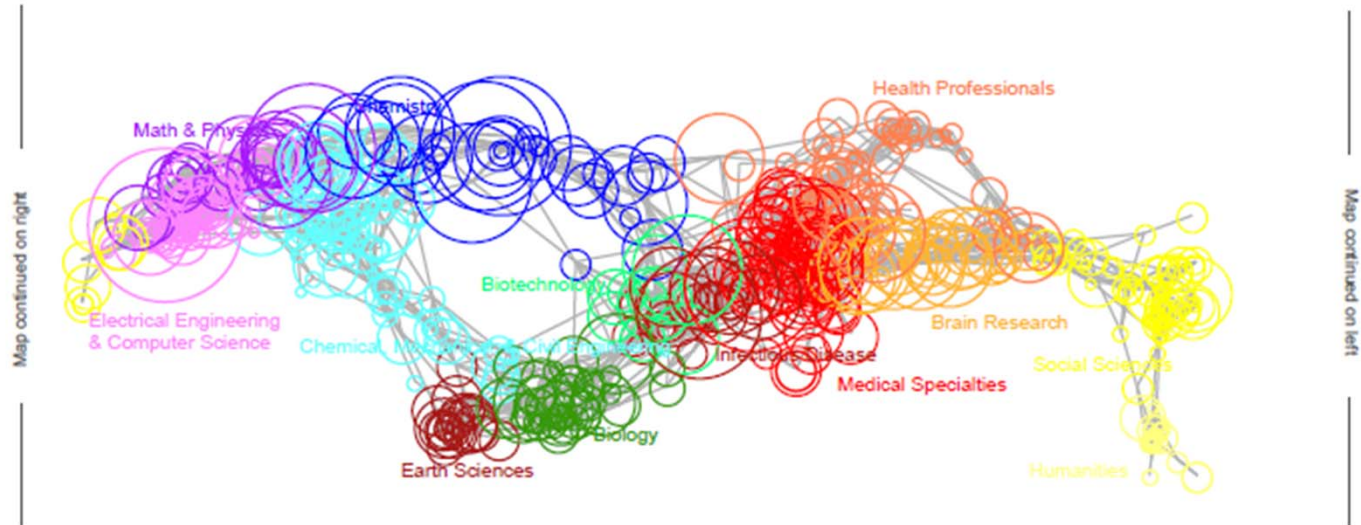


# United Kingdom vs. Germany: Science Map Overlay, 13 Colored Disciplines

Generated from CSV Germany.csv, Scaling Factor 0.01  
7,548 out of 9,351 records were mapped to 544 subdisciplines and 13 disciplines.  
June 21, 2014 | 07:26 PM CEST

**DE:**

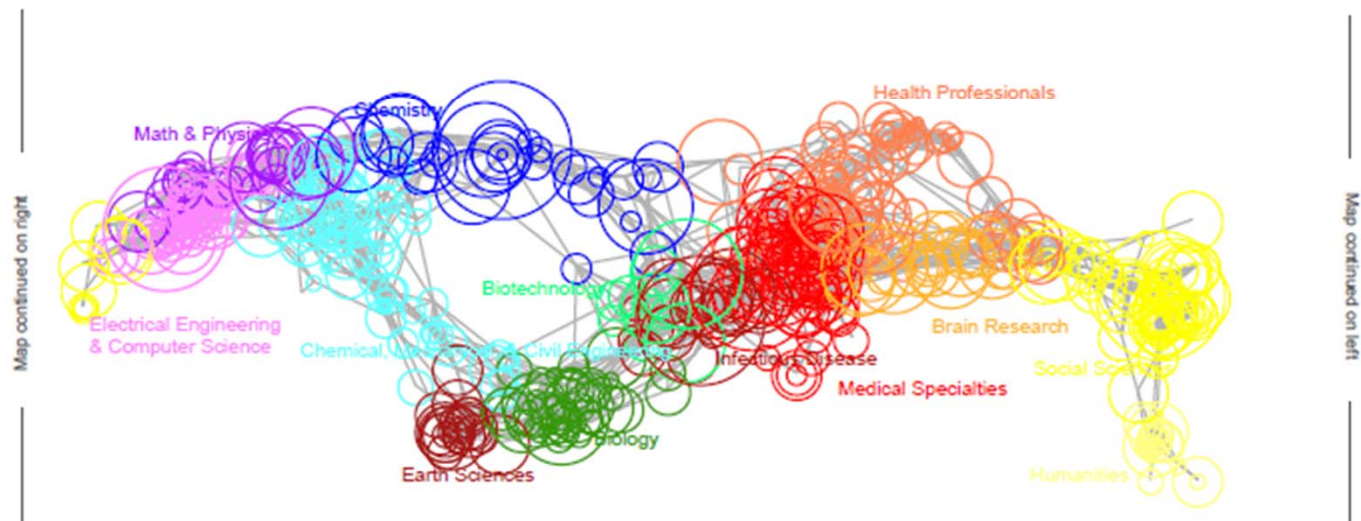
SocSci: 9,003



Generated from CSV United-Kingdom.csv, Scaling Factor 0.01  
8,438 out of 10,538 records were mapped to 548 subdisciplines and 13 disciplines.  
June 21, 2014 | 07:28 PM CEST

**UK:**

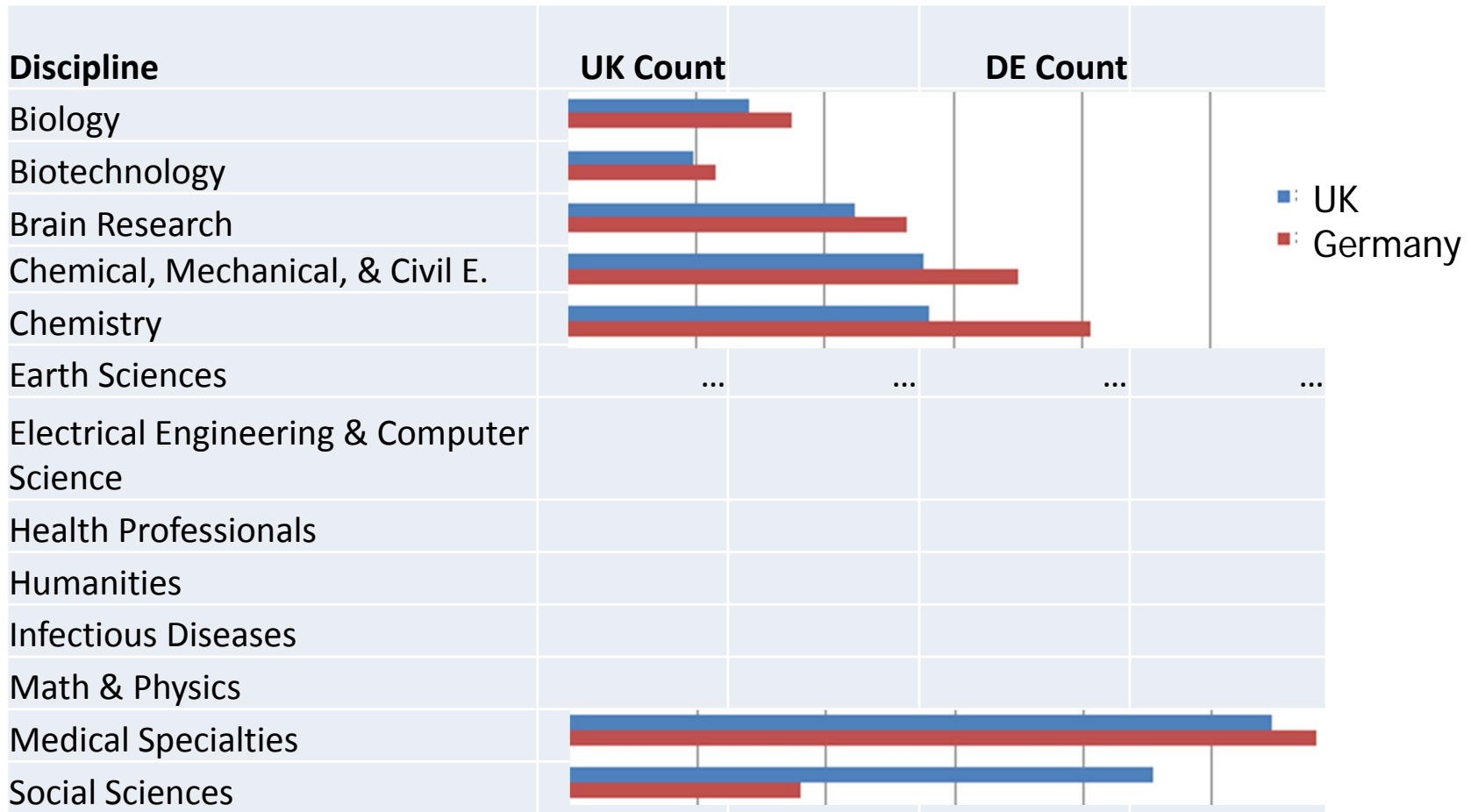
SocSci: 22,731



## United Kingdom vs. Germany: Number of Papers per Discipline

<b>Discipline</b>	<b>UK Count</b>	<b>UK Unique</b>	<b>DE Count</b>	<b>DE Unique</b>
Biology	7108	437	8717	599
Biotechnology	4882	167	5773	171
Brain Research	11187	541	13160	519
Chemical, Mechanical, & Civil E.	13824	707	17531	678
Chemistry	14051	406	20314	417
Earth Sciences	...	...	...	...
Electrical Engineering & Computer Science				
Health Professionals				
Humanities				
Infectious Diseases				
Math & Physics				
Medical Specialties	27294	1181	29055	1107
Social Sciences	22731	1668	9003	1111

## United Kingdom vs. Germany: Number of Papers per Discipline



Study should be re-run using citation counts.

Country Mobility Network, 1996-2011  
see OECD Scoreboard 2013, p.62

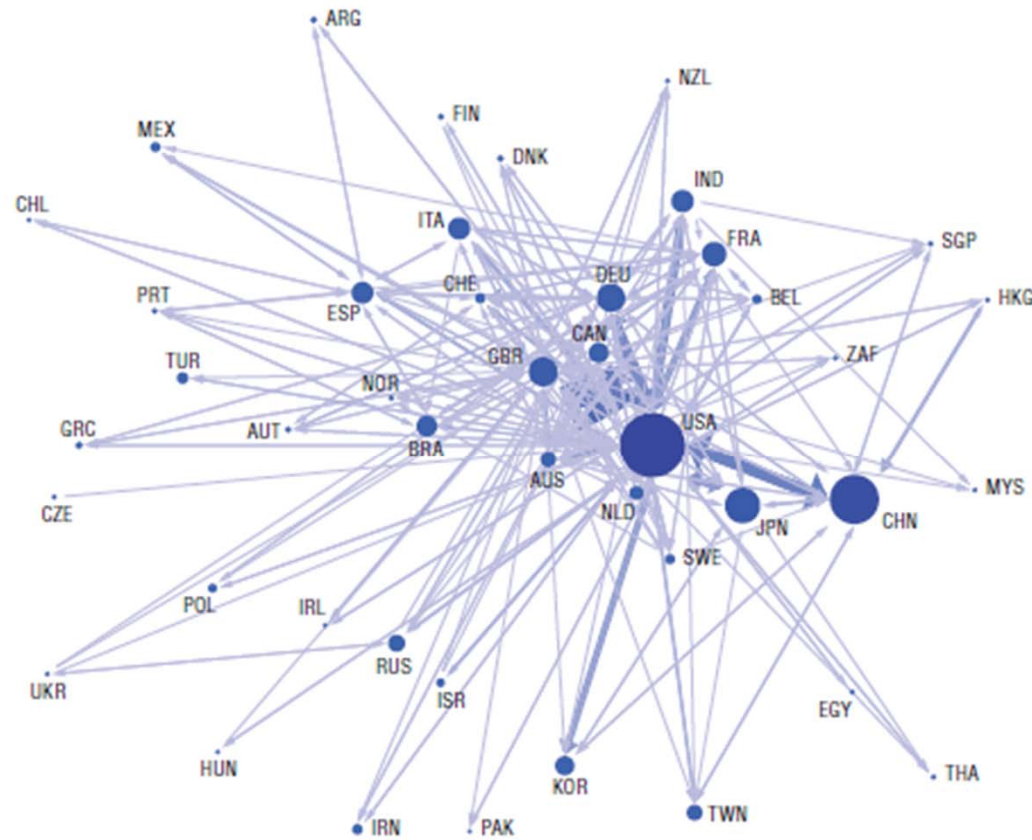
Uses Sci2 and GUESS

## Knowledge networks

The mobility of researchers contributes to the diffusion of scientific and technological knowledge across institutions, at a national and international level. The trail of affiliation changes left by scientific authors in their scholarly publication records provides a partial means of identifying the international network of researcher flows. As expected, leading research countries tend to attract more scientific authors from abroad than they have authors who leave. Flows within each pair of countries tend to be of a similar order of magnitude in both directions, suggesting the existence of complex patterns of knowledge circulation representing the mobility of individuals at different stages of their careers, from students to established professors. The international mobility network also displays a number of interesting patterns that reveal affinities between different economies based on linguistic, historical as well as political and cultural linkages, such as the link between Spain and Latin America countries.

### 57. International mobility network, 1996-2011

*Counts of bilateral flows, by first and last affiliation*



Source: OECD calculations based on Scopus Custom Data, Elsevier, version 5.2012, June 2013.

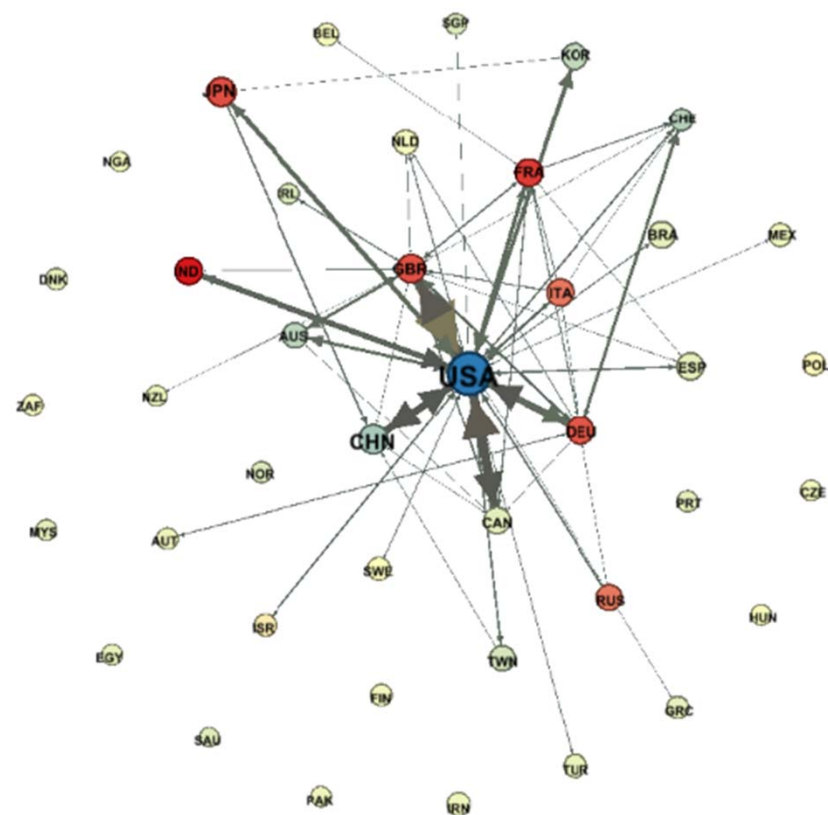
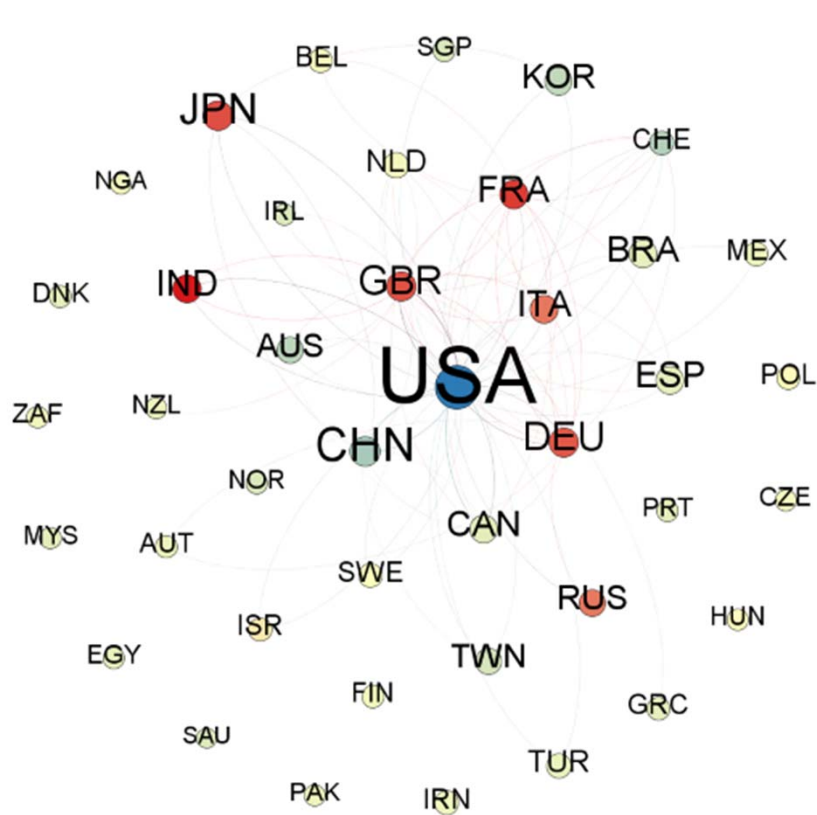
## Compile data

	A	B	C	D	E	F	G
1	code	Source	Target	FLOWS	sum_sou	sum_end	scaled_d
2	2788	GBR	USA	12739	42989	88450	0.2065893
3	1141	CAN	USA	10932	22116	88450	0.2471709
4	7147	USA	GBR	10323	88450	42989	0.1674088
5	7113	USA	CAN	8769	88450	22116	0.1982658
6	7116	USA	CHN	8537	88450	16393	0.2241956
7	1885	DEU	USA	8042	31212	88450	0.1530574
8	1395	CHN	USA	7978	16393	88450	0.2095154
9	3409	IND	USA	6550	13867	88450	0.1870256
10	7130	USA	DEU	6210	88450	31212	0.1181903
11	4011	JPN	USA	5668	14935	88450	0.1559475
12	2584	FRA	USA	4913	24207	88450	0.1061762
13	7182	USA	KOR	4769	88450	4827	0.2308021
14	7176	USA	JPN	4039	88450	14935	0.1111277
15	7092	USA	AUS	3934	88450	11981	0.1208478
16	7476	YUX	SRB	3825	4415	128	5.088164
17	1759	DEU	CHE	3813	31212	9619	0.22006
18	2631	GBR	AUS	3634	42989	11981	0.1601252
19	314	AUS	USA	3596	11981	88450	0.1104649
20	7167	USA	IND	3365	88450	13867	0.0960826
21	3818	ITA	USA	3331	13098	88450	0.0978641

edges Source Target Totals nodes



Identify essential nodes and edges , apply color and size coding  
Layout network but keep in mind that layout is non-deterministic  
Different tools render the same network differently: Gephi Layouts

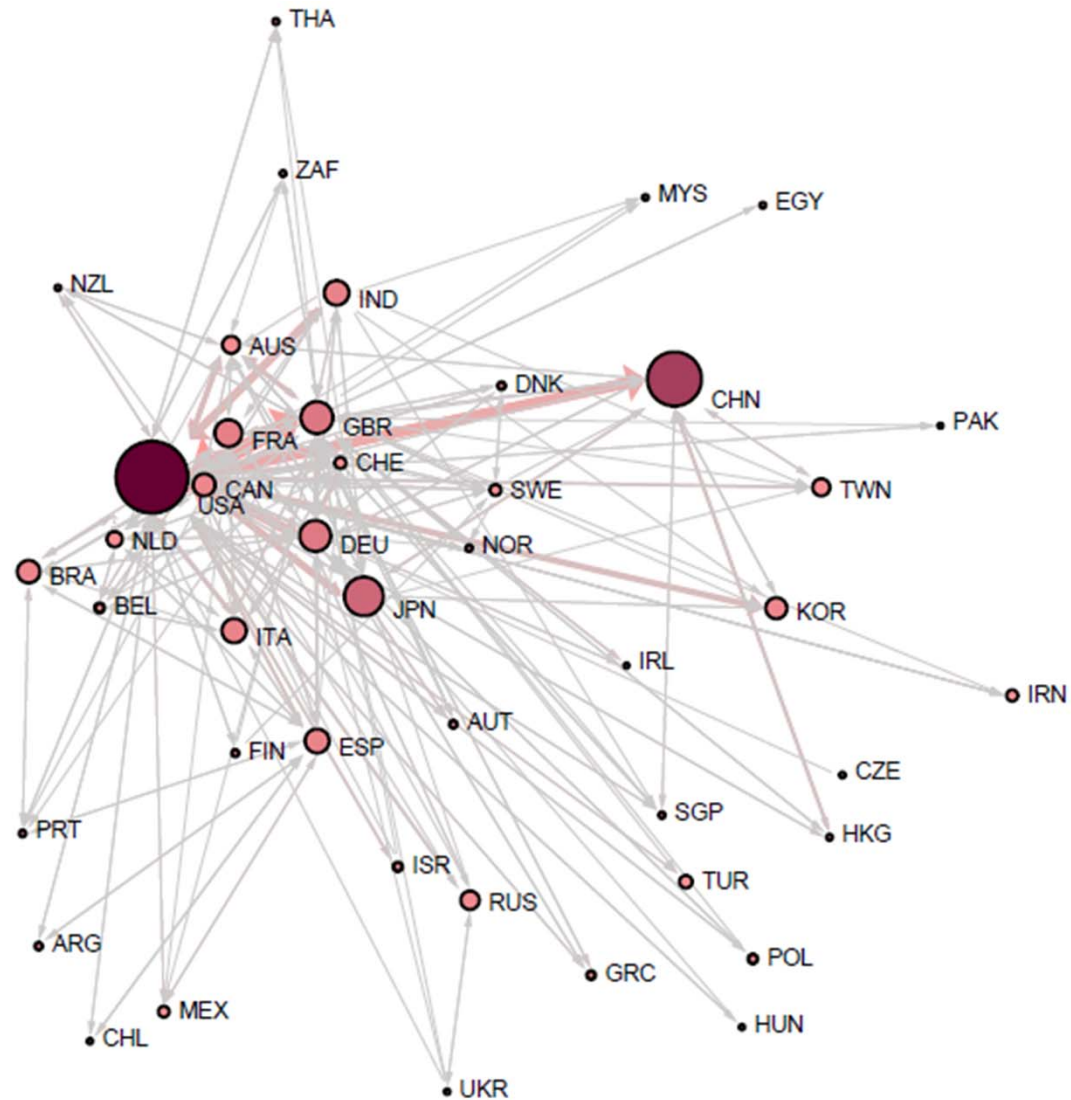




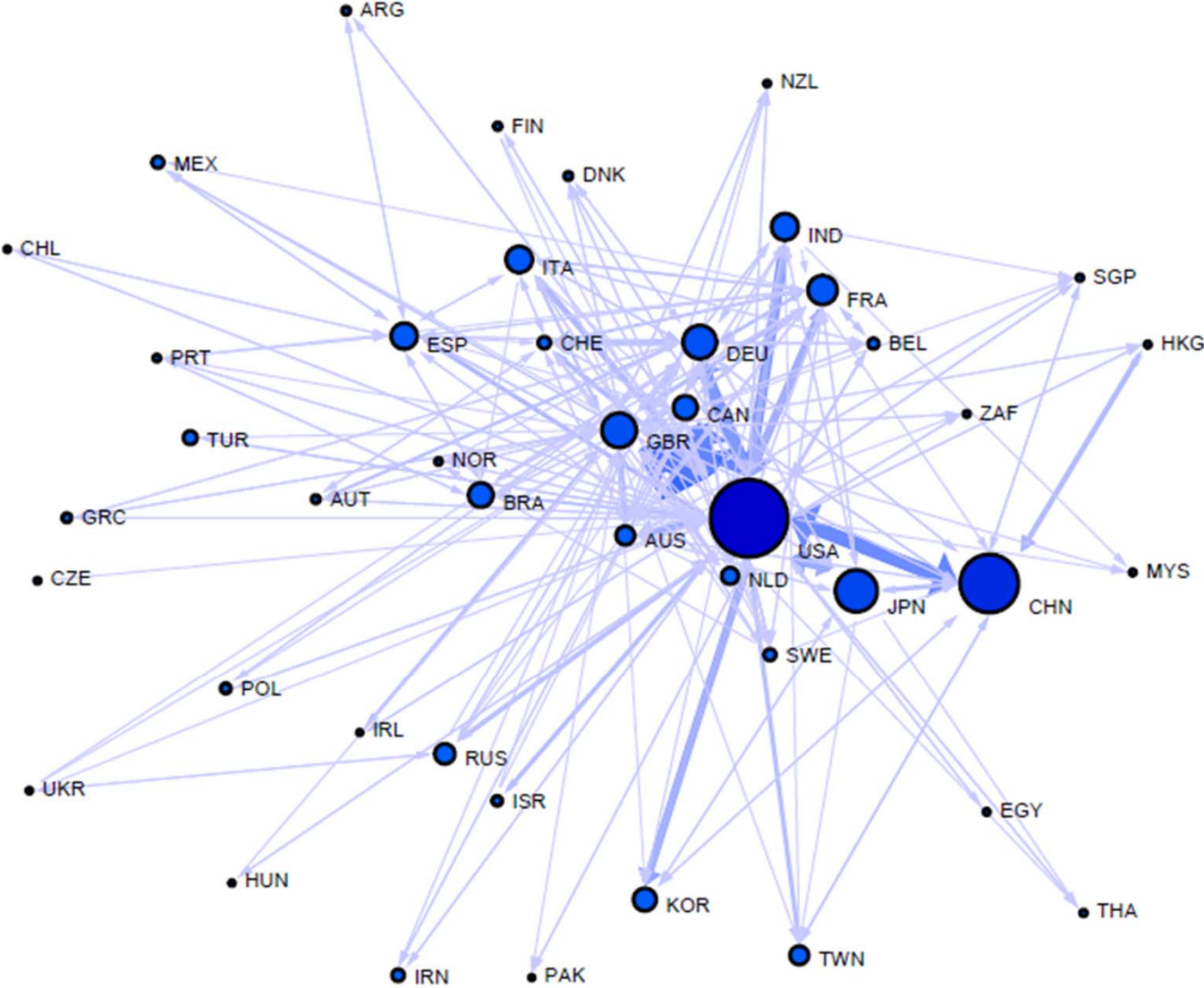
Identify essential nodes and edges, apply color and size coding

Layout network but keep in mind that layout is non-deterministic

Different tools render the same network differently: Guess Layout

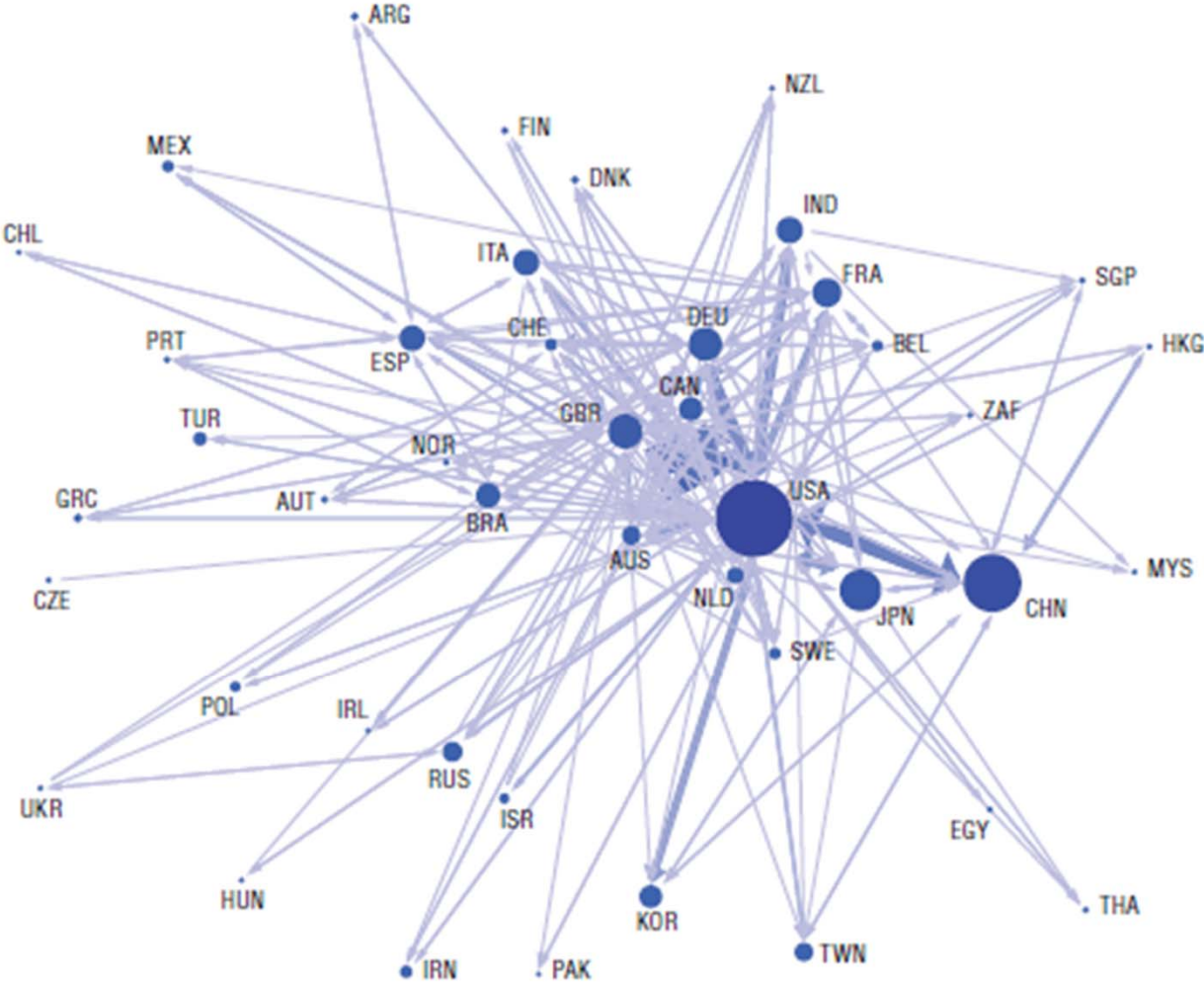


Apply final color and size coding  
Save result in PDF/Postscript file



**Adjust colors and type font, add title, legend, data source info  
in your favorite graphic design and layout program**

**57. International mobility network, 1996-2011**  
*Counts of bilateral flows, by first and last affiliation*



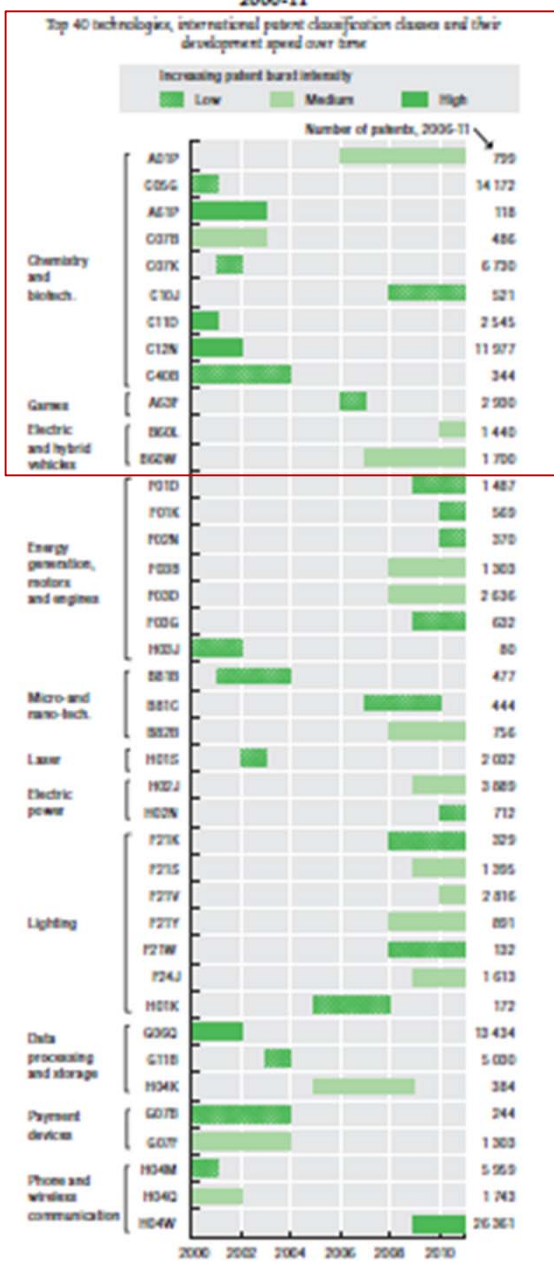
Source: OECD calculations based on Scopus Custom Data, Elsevier, version 5.2012, June 2013.

Acceleration in The Development of Patented Technologies,  
2000-11, see OECD Scoreboard 2013, p.170

Top 40 technologies, international patent classification classes and their  
development speed over time

Uses Kleinberg's burst detection

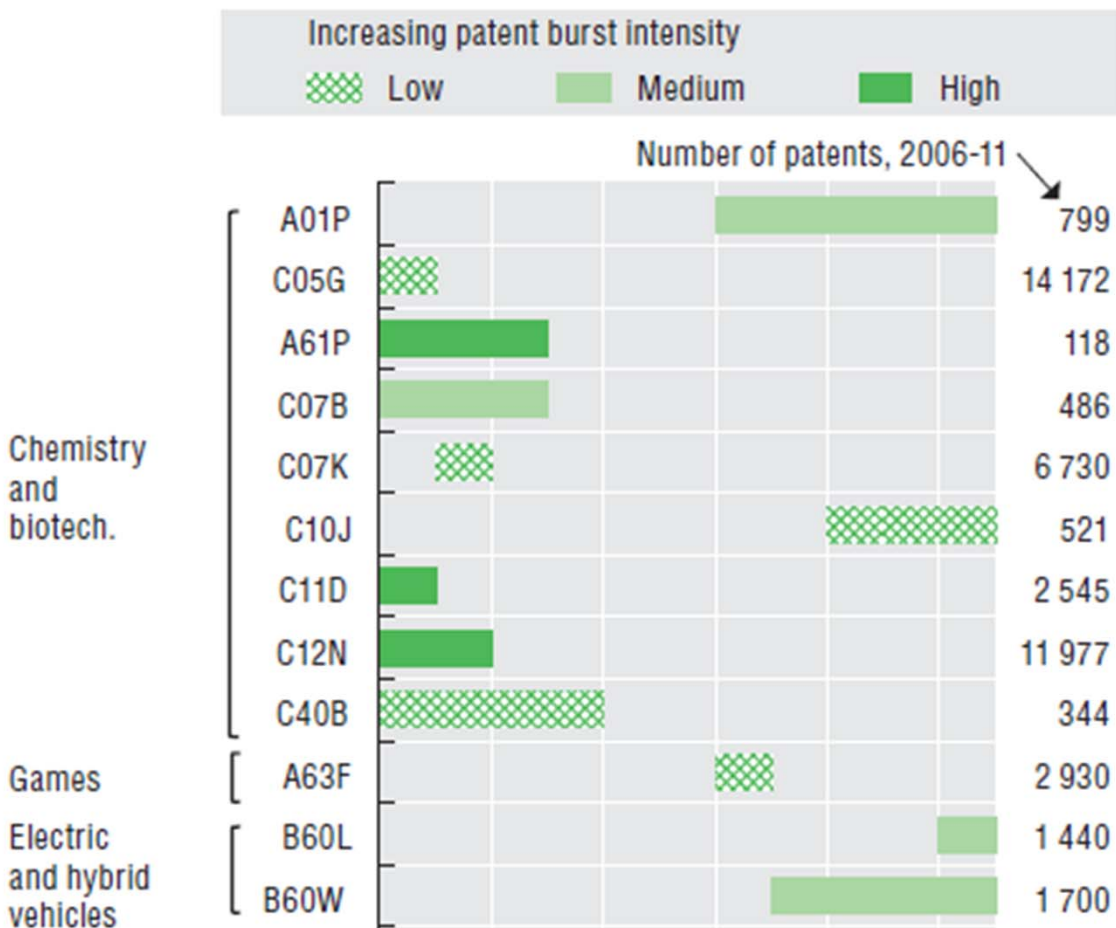
Acceleration in the development of patented technologies,  
2000-11



Source: OECD calculations based on the Worldwide Patent Statistical Database, CPO, April 2013. See chapter notes.

DOI: <http://dx.doi.org/10.1787/888932892366>

Top 40 technologies, international patent classification classes and their  
development speed over time



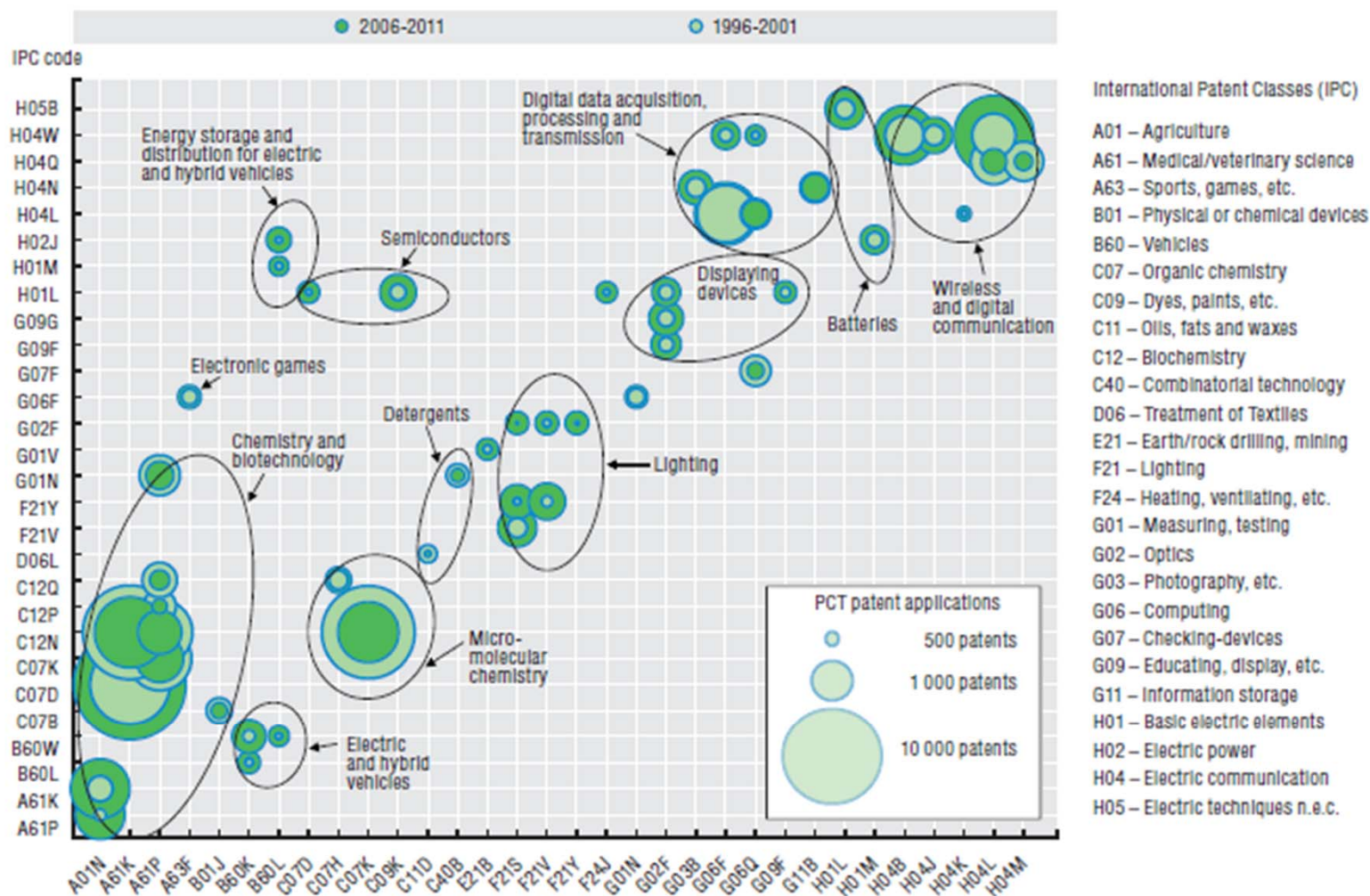
Acceleration in The Co-Development of Patented Technologies, 1996-2001 and 2006-11, see OECD Scoreboard 2013, p.171

Top 50 co-developments of IPC classes by development speed observed in the 2000s

Uses Kleinberg's burst detection algorithm

## Acceleration in the co-development of patented technologies, 1996-2001 and 2006-11

Top 50 co-developments of IPC classes by development speed observed in the 2000s



Note: The technologies experiencing an acceleration in co-development can be identified at the intersection of the x and y axes (e.g. electronic games arise from the combination of A63F, sports, games etc., and G06F, computing). Co-developments that have increased in importance over time are characterised by dark bubbles that are bigger than light bubbles (e.g. biotechnologies arising from the combination of A61P, medical/veterinary science, and A01N, agriculture). Co-development for which the intensity has been fading is characterised by light bubbles that are bigger than dark bubbles (e.g. micro-molecular chemistry, at the intersection of C07K, organic chemistry, and C12N, biochemistry).

Source: OECD calculations based on the Worldwide Patent Statistical Database, EPO, April 2013. See chapter notes.

StatLink <http://dx.doi.org/10.1787/888932892385>

## Evolving Country Patent Profiles, 1995-2010

B07C,B41J,B41K,G02F,G03G,G05F,G06,G07,G09G,G10L,G11C,H03K,H03L

Patents in the ICT sector, Computers, office machinery

Application dates 1995-2010

Germany, Japan, and United States

Data extracted from <http://stats.oecd.org/index.aspx?queryid=22010> on 17  
Jun 2014 14:34 UTC (GMT).

Requires alignment of IPC and UCSD Map of Science classes.



# Original data

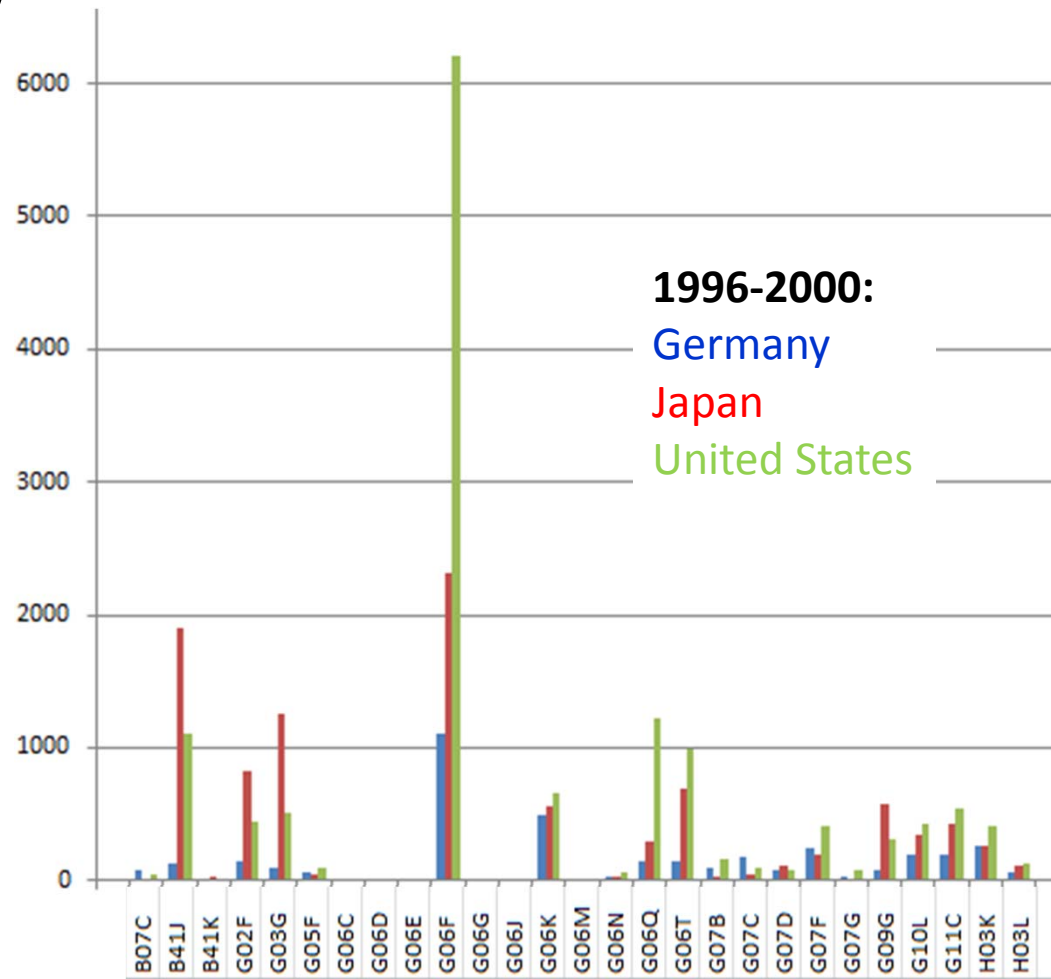
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
2	<b>Patents by</b>														
3				<b>Reference Date</b>	Application date										
4				<b>Patents Office &amp; Triadic Patents Families</b>	Patent applications to the EPO										
5				<b>Reference country</b>	Inventor(s)'s country(ies) of residence										
6				<b>Time</b>	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	
7	<b>Country</b>	<b>Technology domains &amp; IPC</b>													
8	Germany	Total Patents			12622.8175	13240.111	15779.17	18101.586	19841.17	21722.356	22473.45	22284.287	22220.19	22667.702	
9		Total Patents	<u>Total Patents</u>		..	..	..	..	..	..	..	..	..	..	
10			B07C		7.6463	17.5516	20.1179	17.4956	9.4948	10.5912	11.3071	17.9187	15.7548	22.7262	
11			B41J		22.7899	25.8097	18.6103	31.9822	24.4942	29.8127	27.6222	21.2729	18.2979	19.2072	
12			B41K		0.2778	0	0.5	0	2	0.25	2.4444	2.3294	0.6667	0.65	
13			G02F		14.8301	18.2806	20.6442	28.5525	43.725	39.6363	39.1086	35.5051	26.9451	17.7201	
14			G03G		20.6317	18.01	30.821	14.6151	17.7553	13.0656	37.7532	20.5526	26.0204	14.8134	
15			G05F		11.6048	11.5254	6.6429	15.219	8.2617	12.669	16.2897	14.9153	11.7583	9.2867	
16			G06C		0	0	0	0	0	0	0	0	0	0.2	
17			G06D		0	0	0	0	0	0	0	0	0	0	
18			G06E		0	0	0	1	0.2	0	1	0	0	0.25	
19			G06F		91.2608	96.4931	165.9575	208.1392	284.5826	358.2143	423.8494	544.4798	521.0575	525.5357	
20			G06G		0.2	0.6222	1.6429	4.1755	0.8333	1.25	2.6117	2.3409	0	2.0595	
21			G06J		0	0.3333	0.5	0.2	0.2	0	0.1429	0	0	0	
22			G06K		55.2017	67.652	97.7645	91.4249	109.6034	127.8255	89.2896	101.0009	114.6108	95.1039	
23			G06M		1.9524	2.019	1.4583	0	2	4.1	1.6429	1.4	0.5	1.1667	
24			G06N		9.8917	2.9625	8.1563	7.869	5.773	8.5329	9.1802	11.1333	4.9056	6.7333	
25			G06Q		6.5409	8.6289	18.7844	12.8064	34.828	76.9067	120.5416	109.6937	128.4806	127.5137	
26			G06T		14.0468	13.5735	25.8964	36.1194	36.4983	38.6907	59.8127	56.6558	70.7853	90.9361	
27			G07B		14.8179	13.3544	13.8357	16.5667	20.2083	25.7833	22.2952	12.9202	12.1845	19.7361	
28			G07C		16.3054	29.1837	29.7984	36.5401	35.3748	43.117	38.9489	33.8147	42.0437	40.8933	

## Prepare data

Three 5-year bins: 1996-2000, 2001-2005, 1996-2010

Three files: One for each country

	A	B	C	D	E
1	IPC	1996-2000	2001-2005	2006-2010	Country
2	B07C	75.2511	96.8664	98.0071	Germany
3	B41J	130.7091	114.1695	141.4017	Germany
4	B41K	2.75	7.4238	2.4667	Germany
5	G02F	150.8386	145.6149	92.9569	Germany
6	G03G	94.267	120.1174	51.3974	Germany
7	G05F	54.318	57.65	38.3015	Germany
8	G06C	0	0.2	0.3333	Germany
9	G06D	0	0	0.3333	Germany
10	G06E	1.2	1.25	0.25	Germany
11	G06F	1113.3867	2533.2586	2116.259	Germany
12	G06G	8.5239	8.0121	1.2778	Germany
13	G06J	1.2333	0.1429	0	Germany
14	G06K	494.2703	511.6949	600.8935	Germany
15	G06M	9.5773	9.6763	6.9557	Germany
16	G06N	33.2937	34.2024	39.4688	Germany
17	G06Q	151.9544	636.5412	440.5302	Germany



# Align 16 IPC Classes with 554 UCSD Map of Science Classes

Retrieving IPC text from WIPO, <http://web2.wipo.int/ipcpub>

IPC	IPC Name
B07C	POSTAL SORTING; SORTING INDIVIDUAL ARTICLES, OR BULK MATERIAL FIT TO BE SORTED PIECE-MEAL, e.g. BY PICKING
B41J	TYPEWRITERS; SELECTIVE PRINTING MECHANISMS, i.e. MECHANISMS PRINTING OTHERWISE THAN FROM A FORME; CORRECTION OF TYPOGRAPHICAL ERRORS
B41K	STAMPS; STAMPING OR NUMBERING APPARATUS OR DEVICES
G02F	DEVICES OR ARRANGEMENTS, THE OPTICAL OPERATION OF WHICH IS MODIFIED BY CHANGING THE OPTICAL PROPERTIES OF THE MEDIUM OF THE LIGHT RAY
G03G	ELECTROGRAPHY; ELECTROPHOTOGRAPHY; MAGNETOGRAPHY
G05F	SYSTEMS FOR REGULATING ELECTRIC OR MAGNETIC VARIABLES
G06C	DIGITAL COMPUTERS IN WHICH ALL THE COMPUTATION IS EFFECTED MECHANICALLY
G06D	DIGITAL FLUID-PRESSURE COMPUTING DEVICES
G06E	OPTICAL COMPUTING DEVICES
G06F	ELECTRIC DIGITAL DATA PROCESSING
G06G	ANALOGUE COMPUTERS
G06J	HYBRID COMPUTING ARRANGEMENTS
G06K	RECOGNITION OF DATA; PRESENTATION OF DATA; RECORD CARRIERS; HANDLING RECORD CARRIERS
G06M	COUNTING MECHANISMS; COUNTING OF OBJECTS NOT OTHERWISE PROVIDED FOR
G06N	COMPUTER SYSTEMS BASED ON SPECIFIC COMPUTATIONAL MODELS
G06Q	DATA PROCESSING SYSTEMS OR METHODS, SPECIALLY ADAPTED FOR ADMINISTRATIVE, COMMERCIAL, FINANCIAL, MANAGERIAL, SUPERVISORY OR FORECASTING PURPOSES
G06T	IMAGE DATA PROCESSING OR GENERATION, IN GENERAL

The screenshot shows the WIPO IP Services website interface. At the top, the WIPO logo is on the left, and 'IP SERVICES International Patent Classification (IPC) Official Publication' is in the center. On the right, there is a search bar and a 'Contact us | Accessibility' link. Below the header is a navigation bar with 'WORLD INTELLECTUAL PROPERTY ORGANIZATION' and several menu items: 'Scheme', 'RCL', 'Compilation', 'Catchwords', and 'Guide to the IPC'. The 'Scheme' menu is active, showing a list of IPC classes. The class 'B41J' is selected and highlighted in orange. To the left of the class list, there are controls for 'Version' (set to 2014.01), 'Current symbol' (set to B41J), and 'Language' (set to English). The main content area displays the detailed description for class B41J: 'TYPEWRITERS; SELECTIVE PRINTING MECHANISMS, i.e. MECHANISMS PRINTING OTHERWISE THAN FROM A FORME; CORRECTION OF TYPOGRAPHICAL ERRORS'. The description includes examples of subclasses and their applications, such as B41B for composing, B41F for printing on special surfaces, B41K for laundry marking, B43L 19/00 for fluid media, C09D 10/00 for coating, G01 for measuring, G06K for recognition, G07B for apparatus, H01H 13/70 and H03K 17/94 for keyboards, H03M 11/00 for coding, H04L for transmission, and H04N 1/00 for facsimile transmission.

## Align 16 IPC Classes with 554 UCSD Map of Science Classes

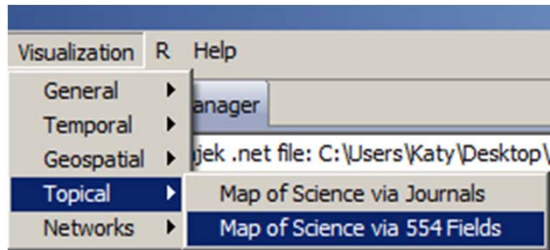
Retrieving UCSD Map data from <http://sci.cns.iu.edu/ucsdmap>

IPC	IPC Name	UCSDMapS	UCSDMapSubdiscipline	1996-2000	2001-2005	2006-2010
B07C	POSTAL SORTING; SORTING INDIVIDUAL ARTICLES, OR BULK MA	236	General Practice	75.2511	96.8664	98.0071
B41J	TYPEWRITERS; SELECTIVE PRINTING MECHANISMS, i.e. MECHAN	488	Social History	130.7091	114.1695	141.4017
B41K	STAMPS; STAMPING OR NUMBERING APPARATUS OR DEVICES	396	Materials Processing	2.75	7.4238	2.4667
G02F	DEVICES OR ARRANGEMENTS, THE OPTICAL OPERATION OF WH	46	Applied Optics	150.8386	145.6149	92.9569
G03G	ELECTROGRAPHY; ELECTROPHOTOGRAPHY; MAGNETOGRAPHY	258	Electrocardiography	94.267	120.1174	51.3974
G05F	SYSTEMS FOR REGULATING ELECTRIC OR MAGNETIC VARIABLES	378	Radiology	54.318	57.65	38.3015
G06C	DIGITAL COMPUTERS IN WHICH ALL THE COMPUTATION IS EFFE	5	Signal Processing	0	0.2	0.3333
G06D	DIGITAL FLUID-PRESSURE COMPUTING DEVICES	5	Signal Processing	0	0	0.3333
G06E	OPTICAL COMPUTING DEVICES	46	Applied Optics	1.2	1.25	0.25
G06F	ELECTRIC DIGITAL DATA PROCESSING	3	Data Mining	1113.3867	2533.2586	2116.259
G06G	ANALOGUE COMPUTERS	306	Computer Systems Desig	8.5239	8.0121	1.2778
G06J	HYBRID COMPUTING ARRANGEMENTS	5	Signal Processing	1.2333	0.1429	0
G06K	RECOGNITION OF DATA; PRESENTATION OF DATA; RECORD CAR	164	Library Science; Infomat	494.2703	511.6949	600.8935
G06M	COUNTING MECHANISMS; COUNTING OF OBJECTS NOT OTHERV	5	Signal Processing	9.5773	9.6763	6.9557
G06N	COMPUTER SYSTEMS BASED ON SPECIFIC COMPUTATIONAL MOI	5	Signal Processing	33.2937	34.2024	39.4688
G06Q	DATA PROCESSING SYSTEMS OR METHODS, SPECIALLY ADAPTED	164	Library Science; Infomat	151.9544	636.5412	440.5302
G06T	IMAGE DATA PROCESSING OR GENERATION, IN GENERAL	136	Image Processing	150.7783	366.8153	382.5599
G07B	TICKET-ISSUING APPARATUS; TAXIMETERS; ARRANGEMENTS OR	306	Computer Systems Desig	89.7484	98.511	87.0863
G07C	TIME OR ATTENDANCE REGISTERS; REGISTERING OR INDICATIN	96	Automatic Control	174.014	200.8899	227.7802
G07D	HANDLING OF COINS OR OF PAPER CURRENCY OR SIMILAR VALU	96	Automatic Control	76.3624	168.8894	169.2234
G07F	COIN-FREED OR LIKE APPARATUS	96	Automatic Control	236.4634	239.0997	191.0469
G07G	REGISTERING THE RECEIPT OF CASH, VALUABLES, OR TOKENS	359	Database Design & Mar	19.7012	18.2833	19.8826
G09G	ARRANGEMENTS OR CIRCUITS FOR CONTROL OF INDICATING D	359	Database Design & Mar	72.8215	146.2022	58.9442
G10L	SPEECH ANALYSIS OR SYNTHESIS; SPEECH RECOGNITION; SPEEC	3	Data Mining	197.5959	322.0039	336.7404
G11C	STATIC STORES (information storage)	359	Database Design & Mar	192.3607	188.7806	77.0363
H03K	PULSE TECHNIQUE (sinusoidal oscillations; transmission of	118	Broadband Communicat	263.128	283.1114	298.6634
H03L	AUTOMATIC CONTROL, STARTING, SYNCHRONISATION, OR STAB	153	Power Distribution	66.9469	60.4425	33.7656

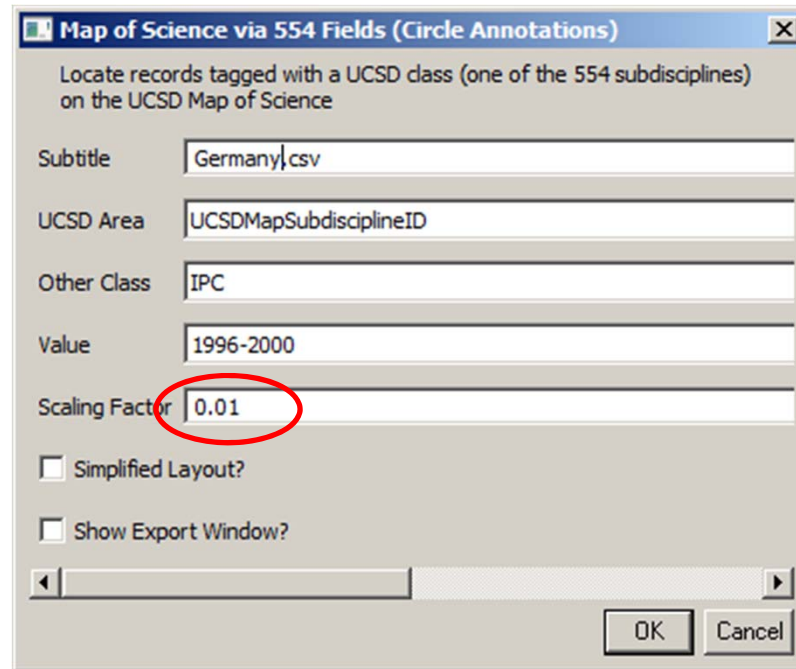
## Align 16 IPC Classes with 554 UCSD Map of Science Classes

Retrieving UCSD Map data from <http://sci.cns.iu.edu/ucsdmap>

Run *Visualization > Topical > Science Map via 554 Fields*



using parameters

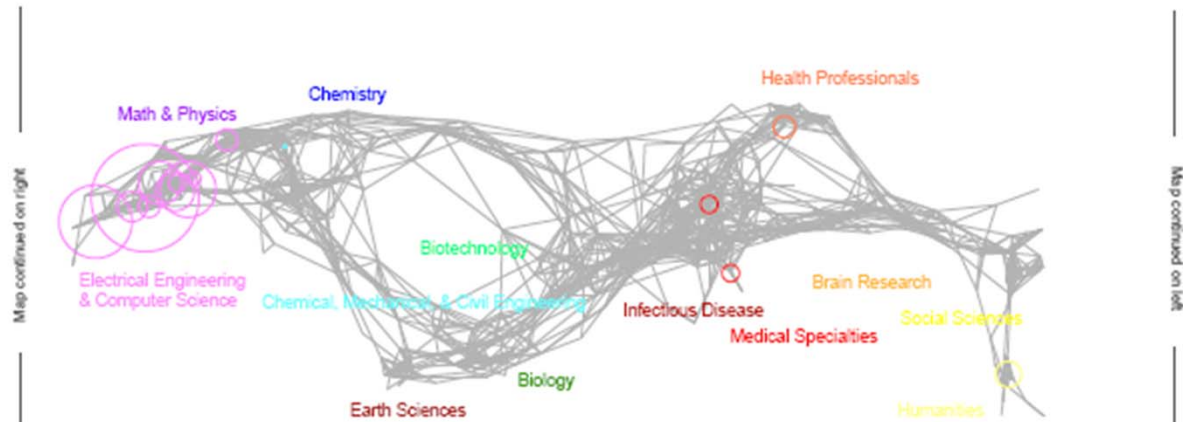


Postscript file will appear in Data Manager.

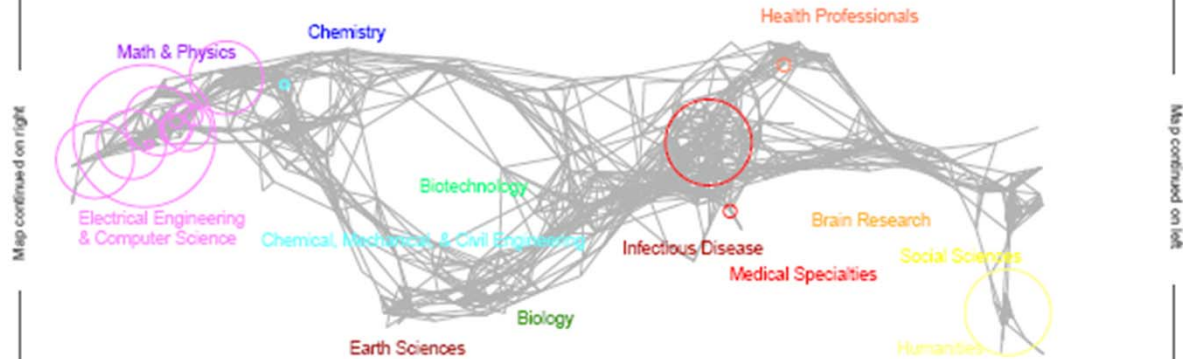
Save and open with a Postscript Viewer.

1996-2010

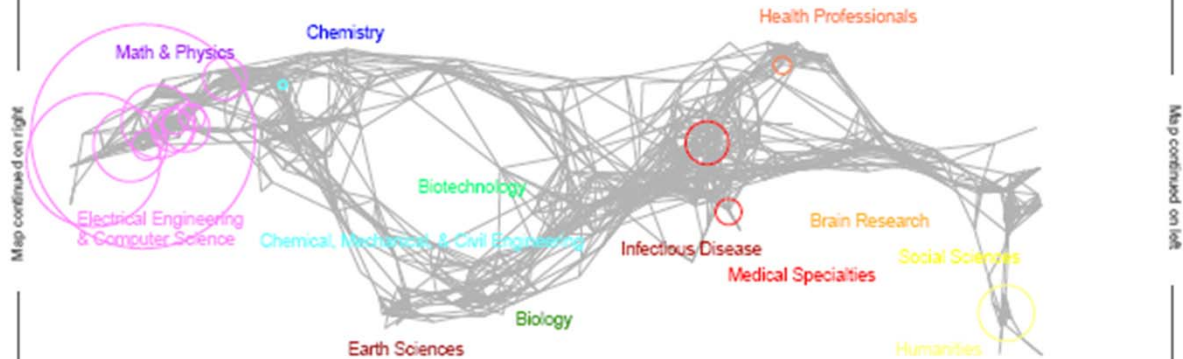
Germany



Japan

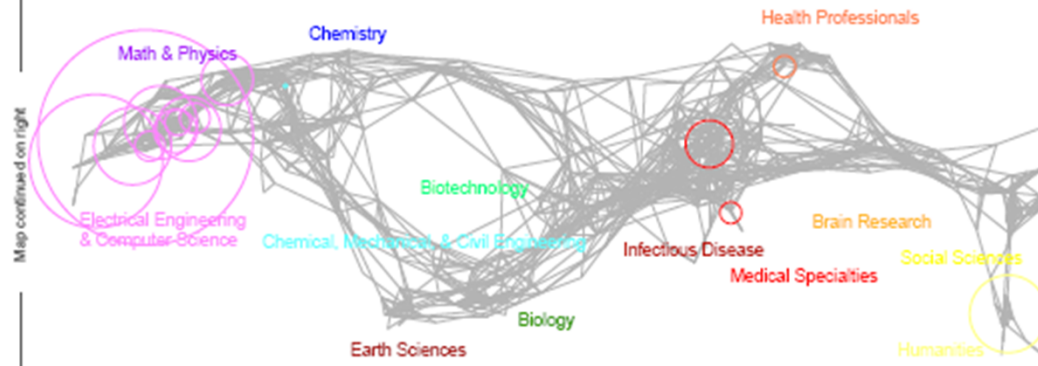


United States

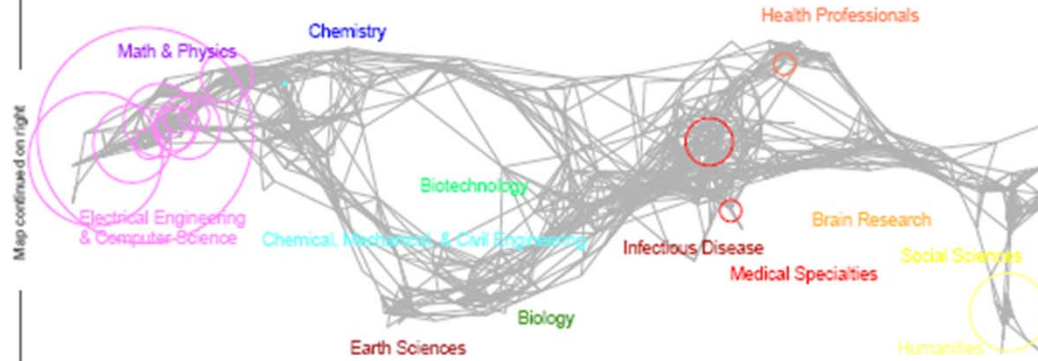


# United States

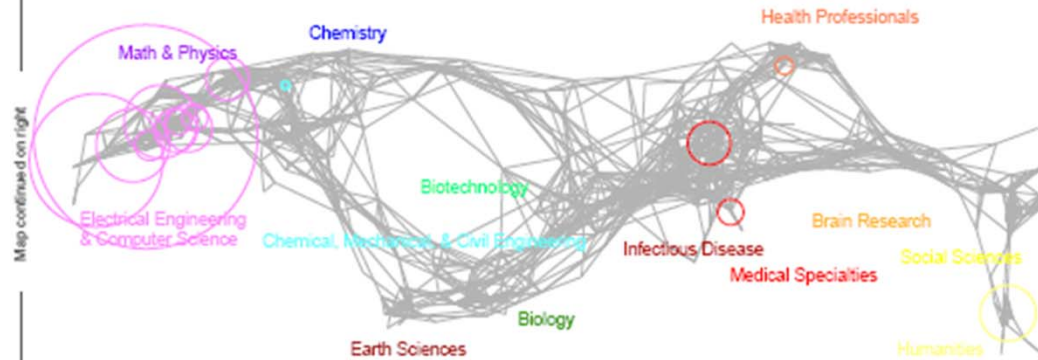
1996-2000



2001-2005



1996-2010

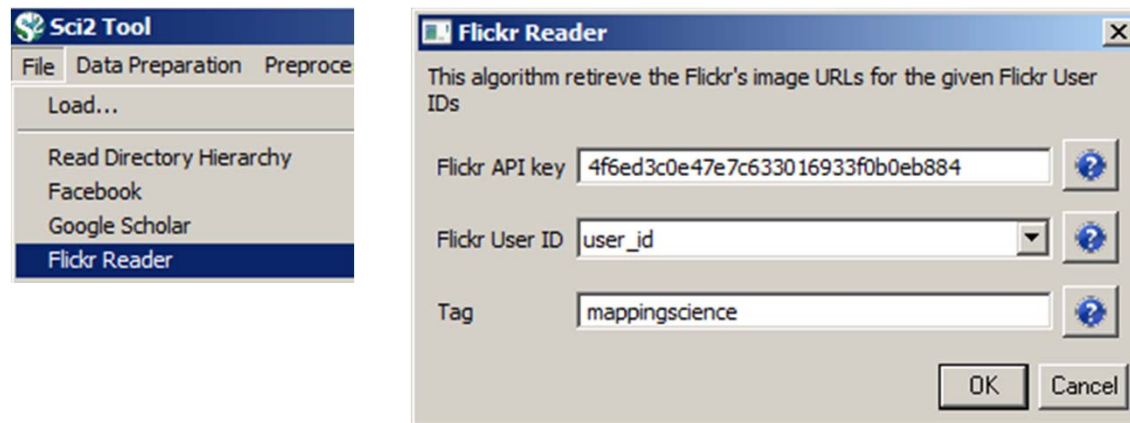


## Read and Analyze Flickr Data Based on User Names

Provide file with user names:

user_id
89182765@N03
91138678@N08

Load file into Sci2, select it in Data Manager, run 'File > Flickr Reader'



To generate a file with information what images these users uploaded when:

Flickr UserID	User Name	Title	Taken At	Image URL
91138678@N08	mappingscience	IX.6 Visualizing Trends and Dynamics	2/18/2014 6:46	http://www.flickr.com/photos/91138678@N08/12613878944/
91138678@N08	mappingscience	IX.2 Hurricanes & Tropical Storms--Lc	2/18/2014 6:45	http://www.flickr.com/photos/91138678@N08/12613852844/
91138678@N08	mappingscience	IX.4 Pulse of the Nation (2010), by Al	2/18/2014 6:46	http://www.flickr.com/photos/91138678@N08/12613472175/
91138678@N08	mappingscience	IX.7 The Hewlett Foundation Grant V	2/18/2014 6:46	http://www.flickr.com/photos/91138678@N08/12613881314/
91138678@N08	mappingscience	IX.5 Map of Complexity (2013), by Bri	2/18/2014 6:46	http://www.flickr.com/photos/91138678@N08/12613576633/
91138678@N08	mappingscience	IX.8 Who Really Matters in the Worlc	2/18/2014 6:46	http://www.flickr.com/photos/91138678@N08/12613582433/
91138678@N08	mappingscience	IX.9 Identifying Emerging Topics in S	2/18/2014 6:46	http://www.flickr.com/photos/91138678@N08/12613482325/

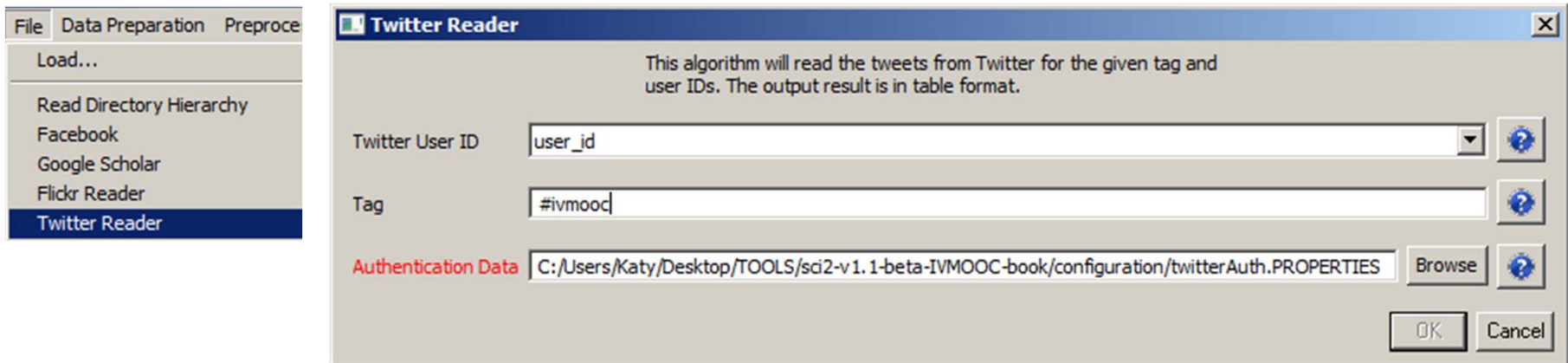


## Read and Analyze Flickr Data

Provide file with user names:

user\_id  
katycns  
CNSTed

Load file into Sci2, select it in Data Manager, run 'File > Twitter Reader'



Generate authentication data using process explained at

<http://wiki.cns.iu.edu/display/CISHELL/Authenticating+with+Twitter%27s+API>

Press OK to generate a file with tweets by those users and with the listed #tags.

## Read and Analyze Twitter Data Based on Keywords

Collect data on a specific keyword using

<http://mashe.hawksey.info/2013/02/twitter-archive-tagsv5/>

or similar.

### MASHe

The musing of Martin Hawksey (EdTech Explorer)

[HOME](#)

[GOOGLE APPS SCRIPT EXAMPLES](#)

[WORDPRESS PLUGINS](#) ▾

[TWITTER ARCHIVE](#)

## Twitter Archiving Google Spreadsheet TAGS v5

Share this post on: [g+](#) | [f](#) | [t](#) | [in](#)

Posted in Featured, Google Apps Script, Google Spreadsheet, Twitter on February 15, 2013 by Martin Hawksey. • 311 Comments

For a couple of years now to support my research in Twitter community analysis/visualisation I've been developing my Twitter Archiving Google Spreadsheet (TAGS). To allow other to explore the possibilities of data generated by Twitter I've released copies of this template to the community.

In September 2012 [Twitter announced the release of a new version of their API](#) (the

## Read and Analyze Twitter Data

A
B
C

### TAGS v5.1

Created by mhawksey. Read more about this at:  
<http://mashe.hawksey.info/twitter-archive-tagsv5/>

**With this spreadsheet you can:**

- automatically pull results from a Twitter Search into a Google Spreadsheet

**Instructions:**

1. Click this button twice (once to authenticate, second to add custom menu) --> Enable custom menu
2. Open TAGS > Twitter API Authentication and follow the instructions
3. Enter search term #science AND  
#education <- you can use search operators like AND OR as well as from: and from:BarackObama' (without quotes)
4. Make a one off collection with TAGS > Run now! or set a trigger to collect every hour TAGS > Update archive frequency open Tools -> Script Editor then Triggers -> Current script's triggers... and adjust

**Advanced Settings:**

Period	default	<- if NOT using default switch to 'paged' below
Continuous/paged	continuous	<- continuous will add results to a single sheet, paged sheet
Followler count filter	1	<- if search term is being spammed you can set the max number of people per person must have to be included in archive

Top Tweeters	No.	@'s	% RT	Twitter Activity
ScienceWatchout	49	2	110%	
UKSchoolTweet	26		104%	
SciProject4Kids	15	18		
Shyam17	15	11		
homeschoolcurr	15	14		
rikafan_p	14	2		
Randirobics	9	14		
SAHSScience	8	13		
TizonAgustin	7		100%	
middlereaders	7		114%	
cucumberjuice	6	3	67%	
mkahramankoleji	6	9		
neuroblogs	6	5		
InnovationAfric	5	7		

e/Settings Archive Summary Dashboard C

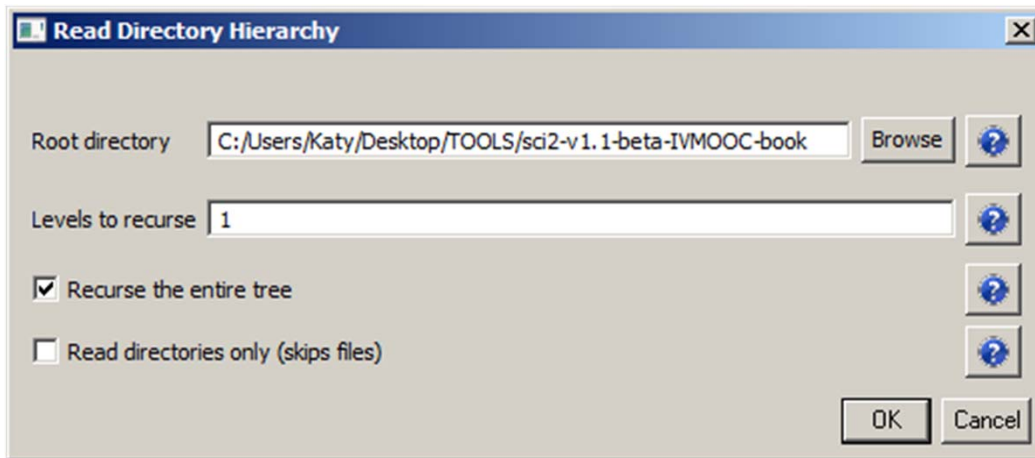
## Read and Analyze Twitter Data

Download data and extract follower counts, map geolocations, analyze reply networks, run burst over text or tags or user activity, e.g., to study public perception and discussion of science.

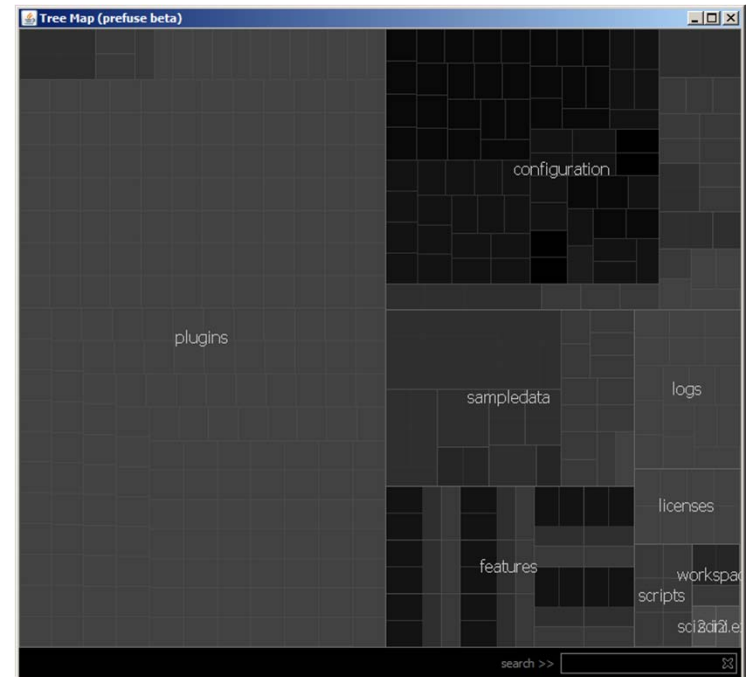
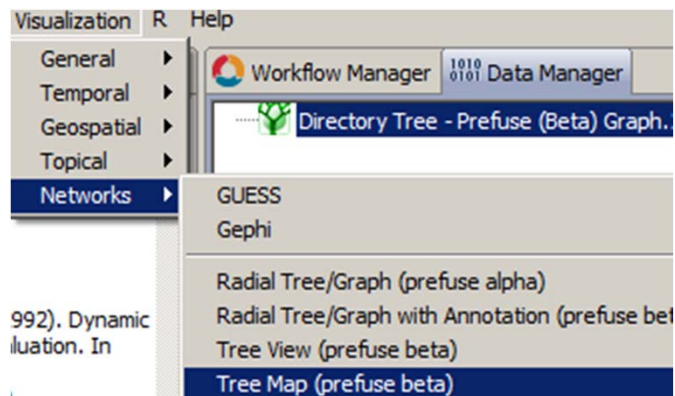
id_str	from_user	text	created_at	time/geo_coord	user_in_rein_re	from_user_in_rep	source	profile_in	user_followers_c	user_friends_c	
4.81544E+17	Yusuf_Trn_	RT @Sanc	Tue Jun 24 21:07:5	24/06/2014 22:0	tr			9.58E+08	<a href=http://pbs	1042	841
4.81544E+17	TeijoLeppamaki	RT @Sanc	Tue Jun 24 21:06:2	24/06/2014 22:0	fi			2.35E+09	<a href=http://pbs	1030	1174
4.81544E+17	ScienceWatchout	RT @Sanc	Tue Jun 24 21:04:5	24/06/2014 22:0	en			2.43E+09	<a href=http://pbs	4133	4190
4.81543E+17	Sancaktepebilim	#sancakte	Tue Jun 24 21:04:3	24/06/2014 22:0	tr			2.48E+09	<a href=http://pbs	23	172
4.81543E+17	Sancaktepebilim	#sancakte	Tue Jun 24 21:04:0	24/06/2014 22:0	tr			2.48E+09	<a href=http://pbs	23	172
4.81543E+17	Sancaktepebilim	#sancakte	Tue Jun 24 21:03:2	24/06/2014 22:0	tr			2.48E+09	<a href=http://pbs	23	172
4.81543E+17	Sancaktepebilim	#sancakte	Tue Jun 24 21:02:4	24/06/2014 22:0	tr			2.48E+09	<a href=http://pbs	23	172
4.81542E+17	Bulls_DL	My future	Tue Jun 24 20:58:1	24/06/2014 21:5	en			26852128	<a href=http://pbs	174	384
4.81541E+17	Petchary	RT @SAHS	Tue Jun 24 20:56:4	24/06/2014 21:5	en			1.75E+08	<a href=http://pbs	3245	3346
4.81541E+17	cucumberjuice	RT @SAHS	Tue Jun 24 20:56:2	24/06/2014 21:5	en			92181782	<a href=http://pbs	1357	593
4.81541E+17	sciencekidfun	Our future	Tue Jun 24 20:55:2	24/06/2014 21:5	en			2.39E+09	<a href=http://pbs	45	233
4.8154E+17	Ultra_Rush	RT @SAHS	Tue Jun 24 20:52:0	24/06/2014 21:5	en			43895582	<a href=http://pbs	752	729
4.8154E+17	SAHSScience	We have t	Tue Jun 24 20:51:1	24/06/2014 21:5	en			2.48E+09	<a href=http://pbs	33	92
4.81536E+17	ndamsgaard	RT @Pauli	Tue Jun 24 20:35:5	24/06/2014 21:5	sv			5E+08	<a href=http://pbs	210	187
4.81532E+17	sunnybeamy	RT @cucu	Tue Jun 24 20:19:5	24/06/2014 21:1	en			2.45E+09	<a href=http://pbs	141	305
4.81532E+17	ScienceWatchout	RT @cucu	Tue Jun 24 20:19:4	24/06/2014 21:1	en			2.43E+09	<a href=http://pbs	4133	4190
4.81532E+17	Jherane_	RT @cucu	Tue Jun 24 20:19:2	24/06/2014 21:1	en			14805551	<a href=http://pbs	826	167
4.81531E+17	cucumberjuice	Just a few	Tue Jun 24 20:17:0	24/06/2014 21:1	en			92181782	<a href=http://pbs	1357	593
4.8153E+17	AlienNeighbor	RT @Rand	Tue Jun 24 20:12:1	24/06/2014 21:1	en			4.14E+08	<a href=http://pbs	1218	39
4.8153E+17	AlienNeighbor	RT @Rand	Tue Jun 24 20:12:1	24/06/2014 21:1	en			4.14E+08	<a href=http://pbs	1218	39

## Re-Run Workflows

Run 'File > Read Directory Hierarchy' using parameters:



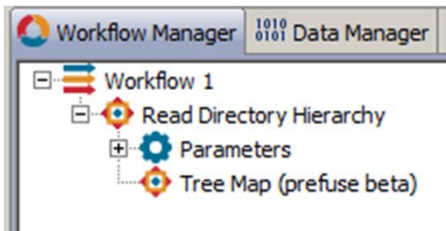
'Visualize > Networks > Tree Map':



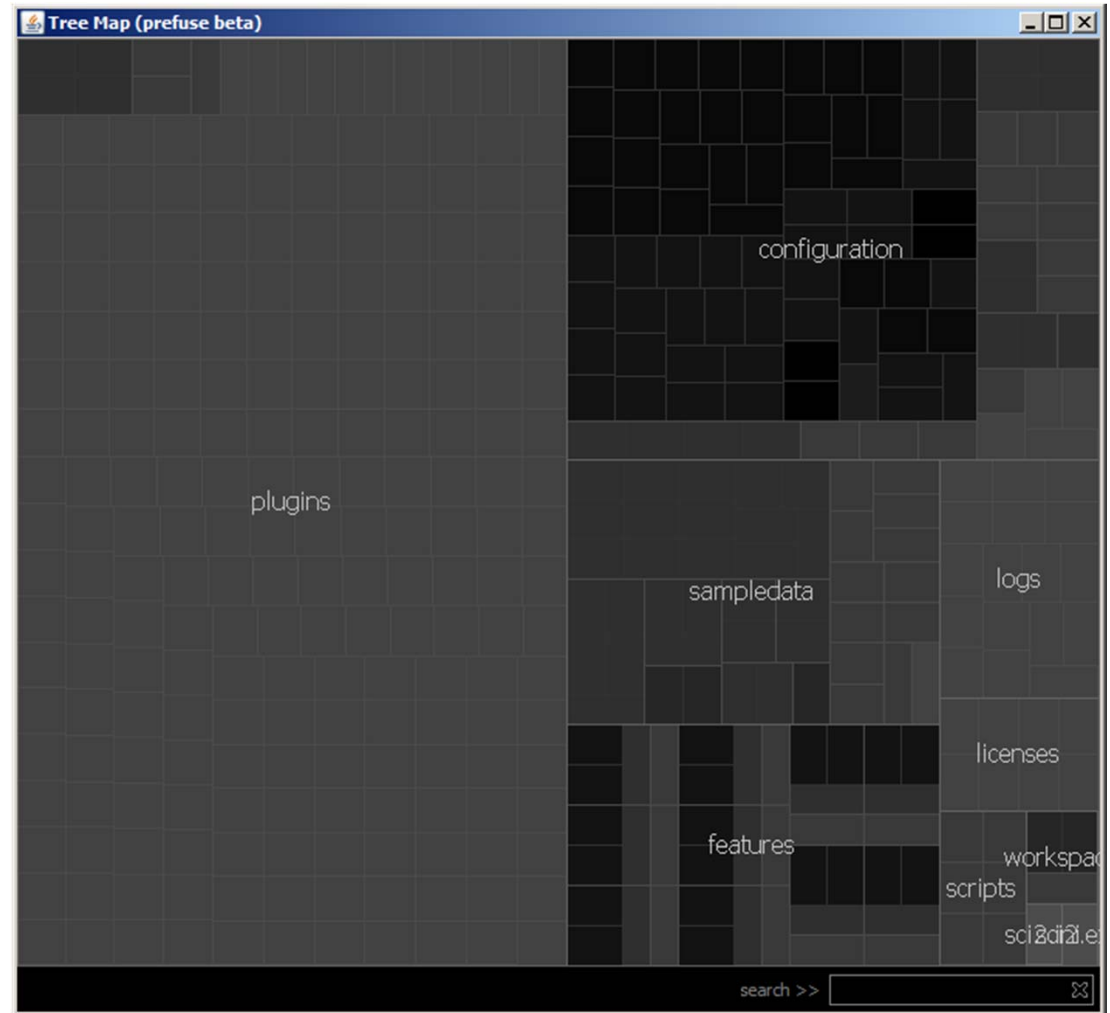
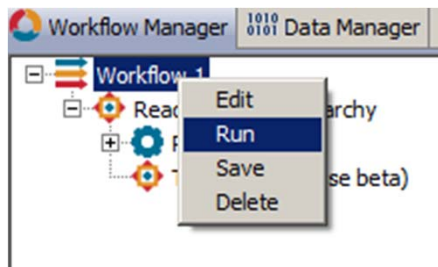
## Re-Run Workflows

Delete file in **Data Manager**

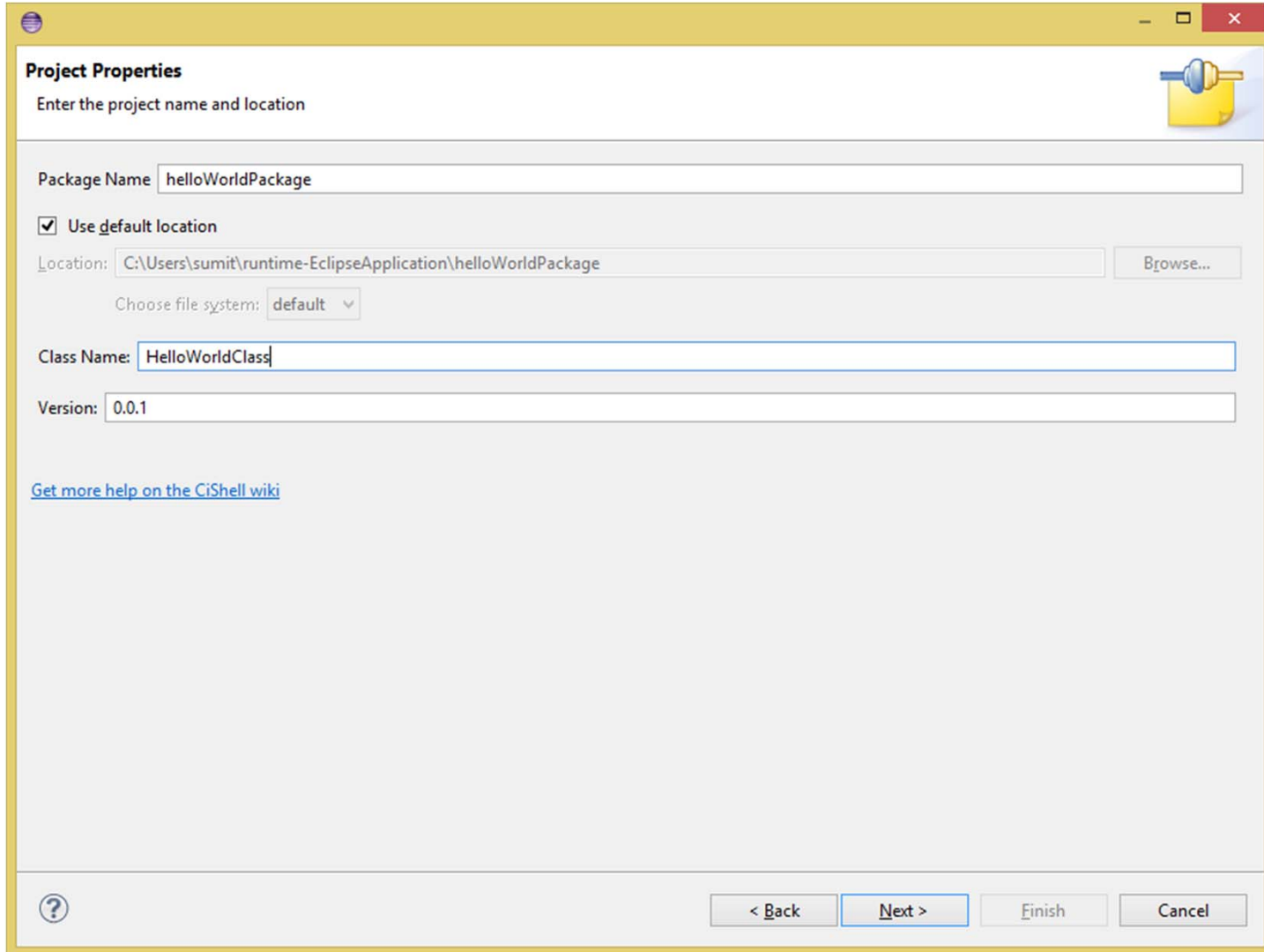
In **Workflow Manager**



Right click Workflow and 'Run':



Adding a new algorithm to Sci2 is easy. Simply use the Wizard driven process:

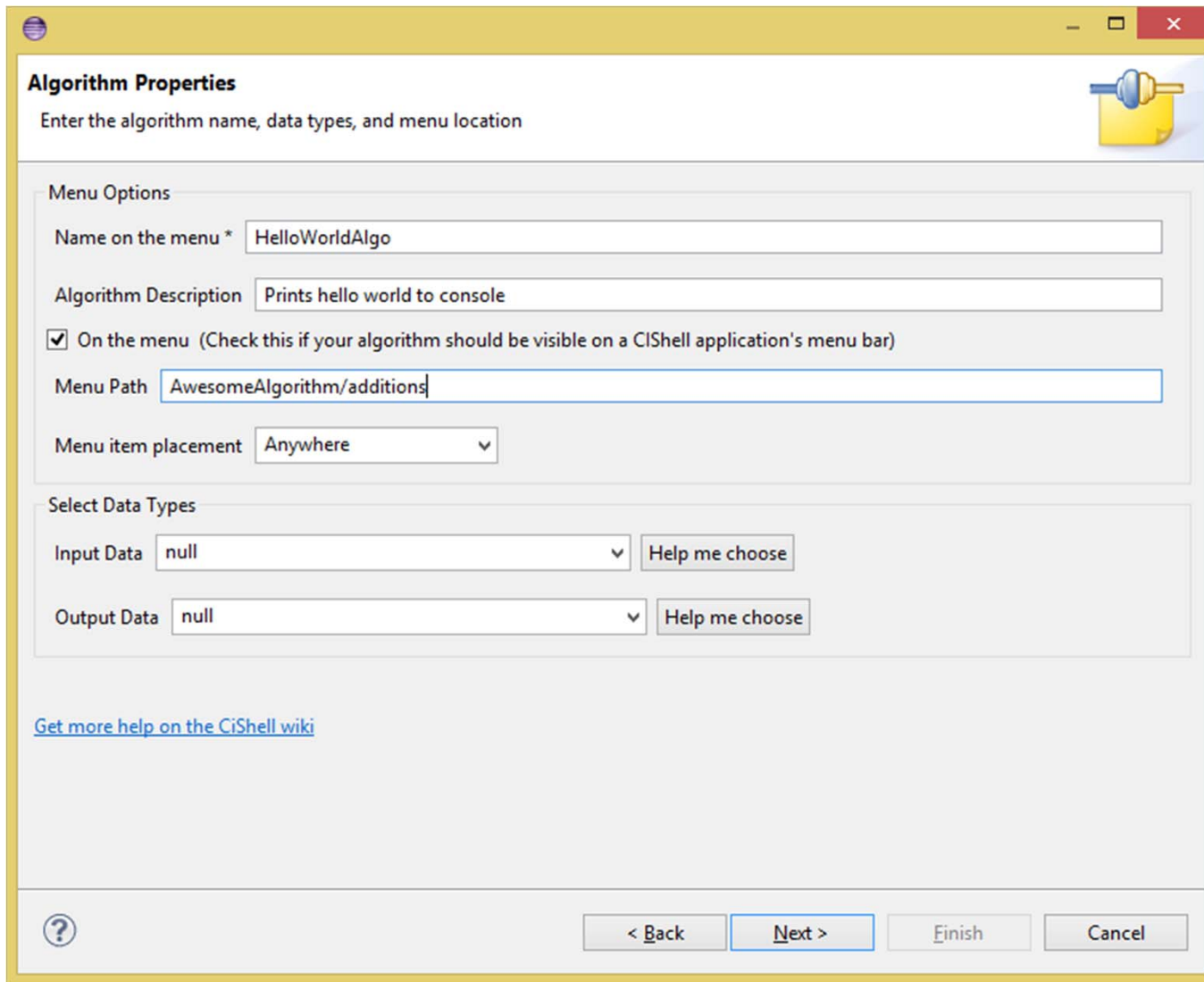


The screenshot shows the "Project Properties" dialog box in Eclipse, titled "Enter the project name and location". The dialog is set to create a new project with the following configuration:

- Package Name:** helloWorldPackage
- Use default location**
- Location:** C:\Users\sumit\runtime-EclipseApplication\helloWorldPackage (with a "Browse..." button)
- Choose file system:** default
- Class Name:** HelloWorldClass
- Version:** 0.0.1

At the bottom left, there is a link: [Get more help on the CiShell wiki](#). At the bottom right, there are navigation buttons: "< Back", "Next >" (highlighted), "Finish", and "Cancel".

Adding a new algorithm to Sci2 is easy. Simply use the Wizard driven process:



**Algorithm Properties**  
Enter the algorithm name, data types, and menu location

Menu Options

Name on the menu \*

Algorithm Description

On the menu (Check this if your algorithm should be visible on a CiShell application's menu bar)

Menu Path

Menu item placement

Select Data Types

Input Data

Output Data

[Get more help on the CiShell wiki](#)

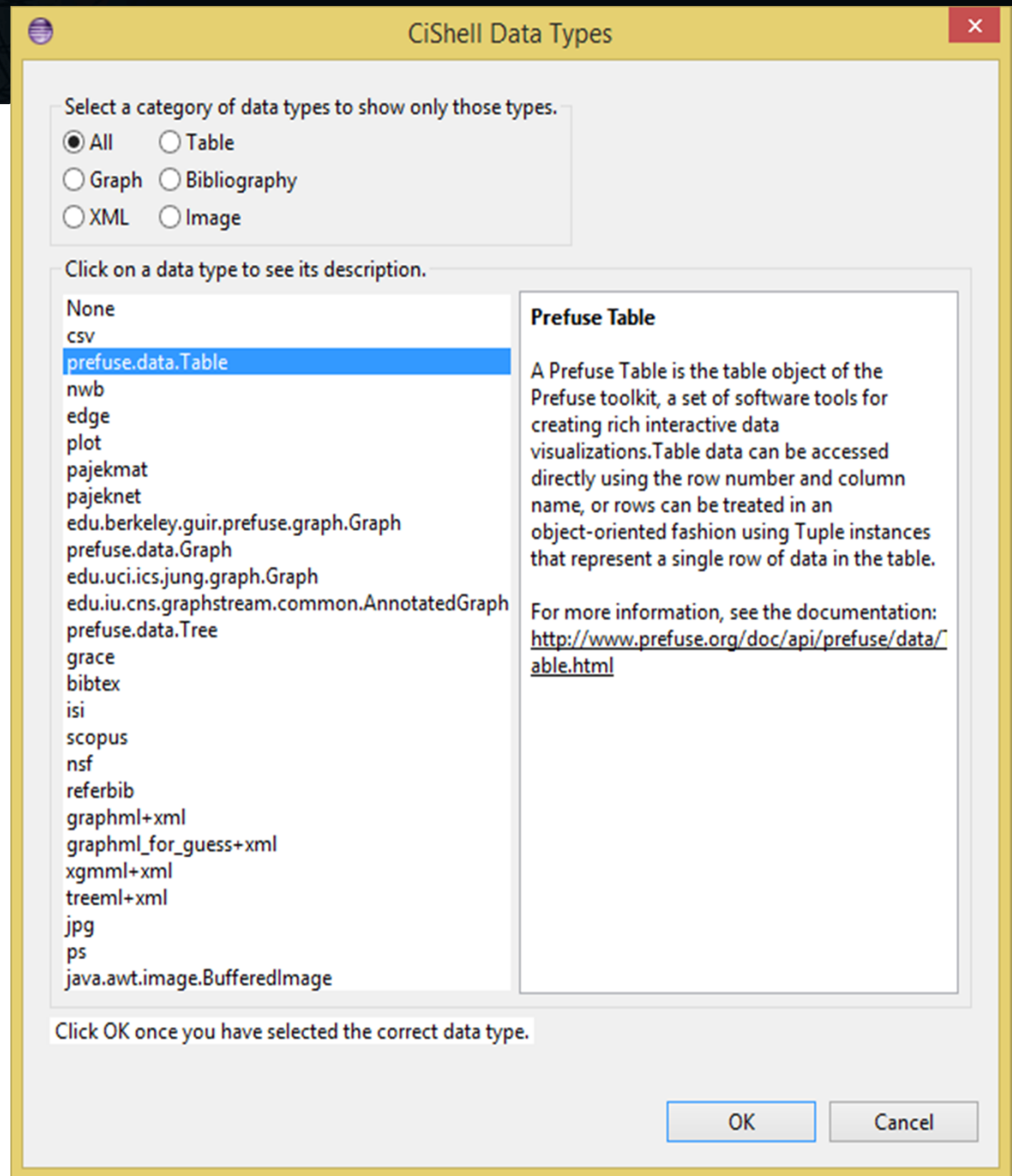


Adding a new algorithm to Sci2 is easy. Simply use the Wizard driven process.

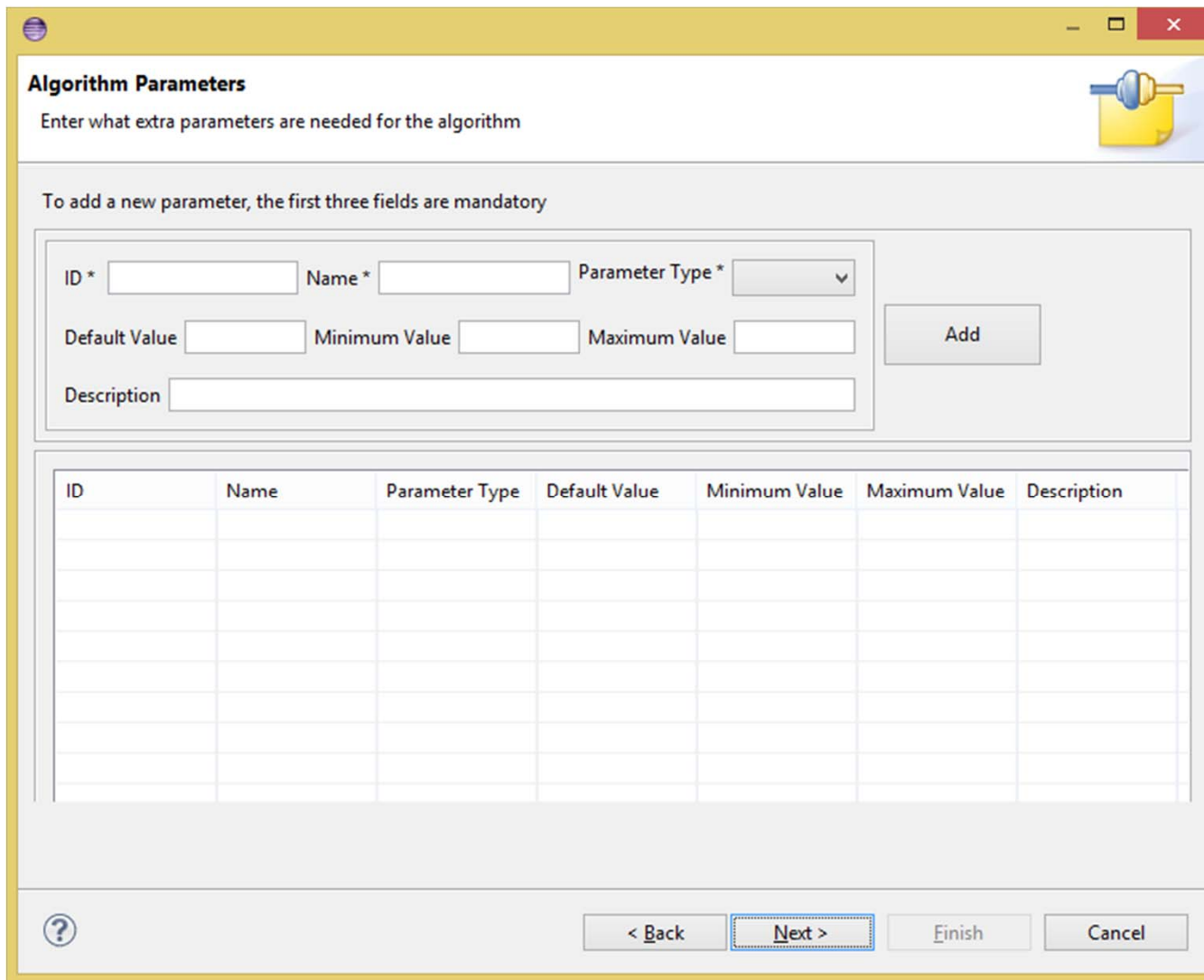
See also

<http://wiki.cns.iu.edu/display/CISHELL/Hello+World+Tutorial>

<http://cishell.wiki.cns.iu.edu/Home>



Adding a new algorithm to Sci2 is easy. Simply use the Wizard driven process:



**Algorithm Parameters**  
Enter what extra parameters are needed for the algorithm

To add a new parameter, the first three fields are mandatory

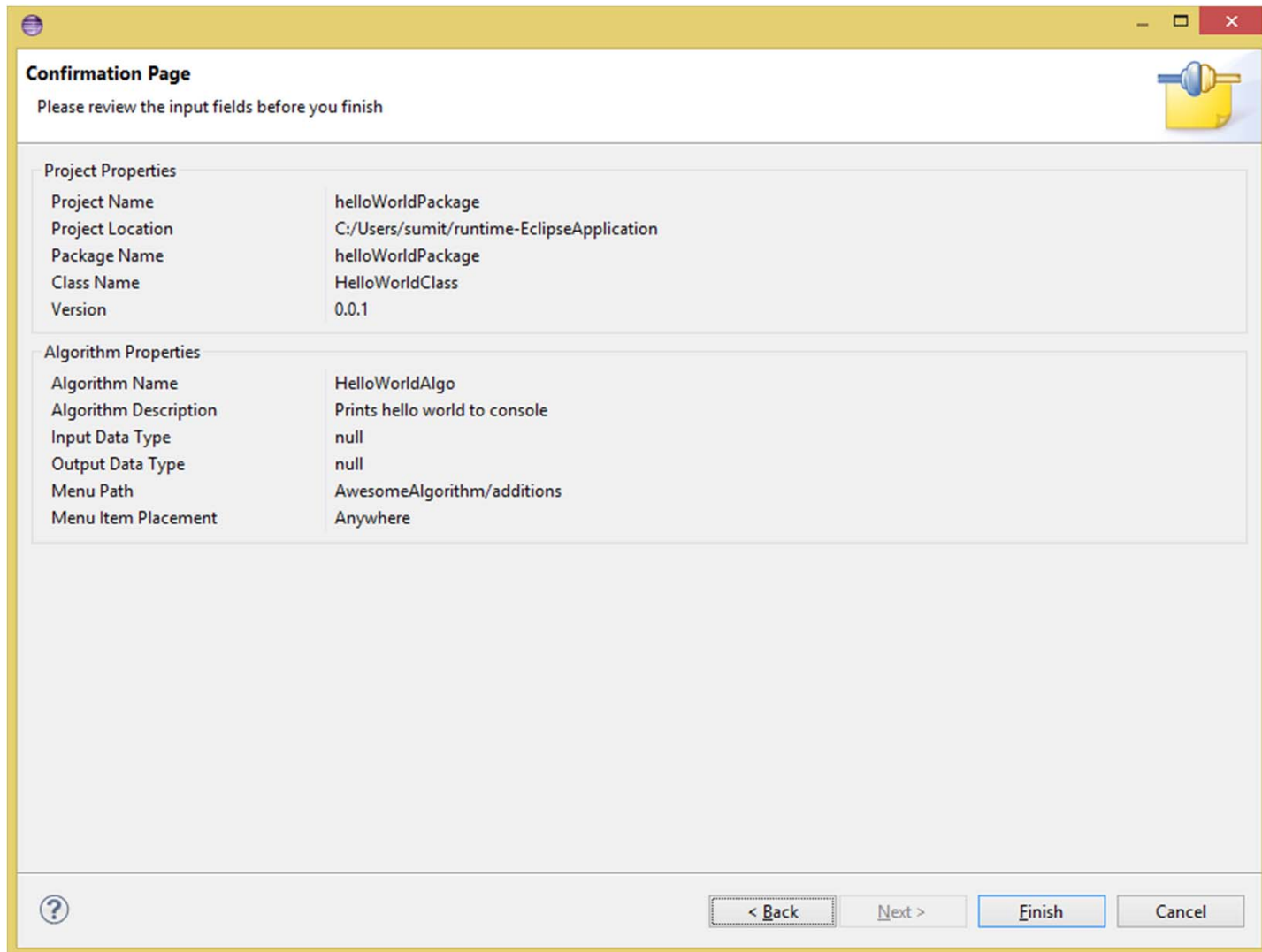
ID \*  Name \*  Parameter Type \*

Default Value  Minimum Value  Maximum Value

Description

ID	Name	Parameter Type	Default Value	Minimum Value	Maximum Value	Description

Adding a new algorithm to Sci2 is easy. Simply use the Wizard driven process:



## 2012 Tutorial Covered Many General Workflows

Covered in <http://cns.iu.edu/docs/presentations/2012-borner-sci2tutorial-oecd.pdf>

**Temporal** analysis and vis: Slides 49-52

**Geospatial** analysis and vis: Slides 81-93

**Topical** analysis and vis: Slides

**Network** analysis and vis:

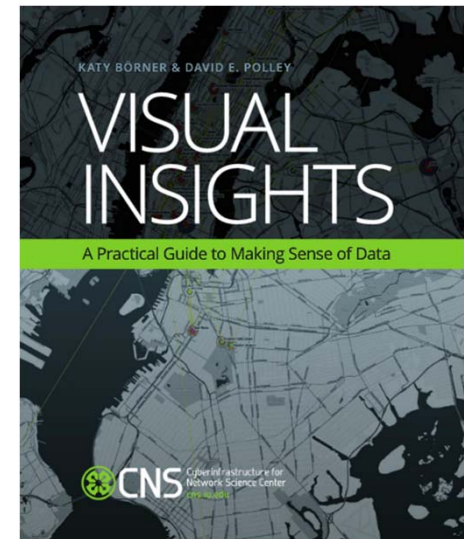
- Network Extraction: Slides 73-78
- Backbone identification: Slide 67
- Hierarchical clustering/community detection: S. 68-70
- Animating network growth: Slides 97-101

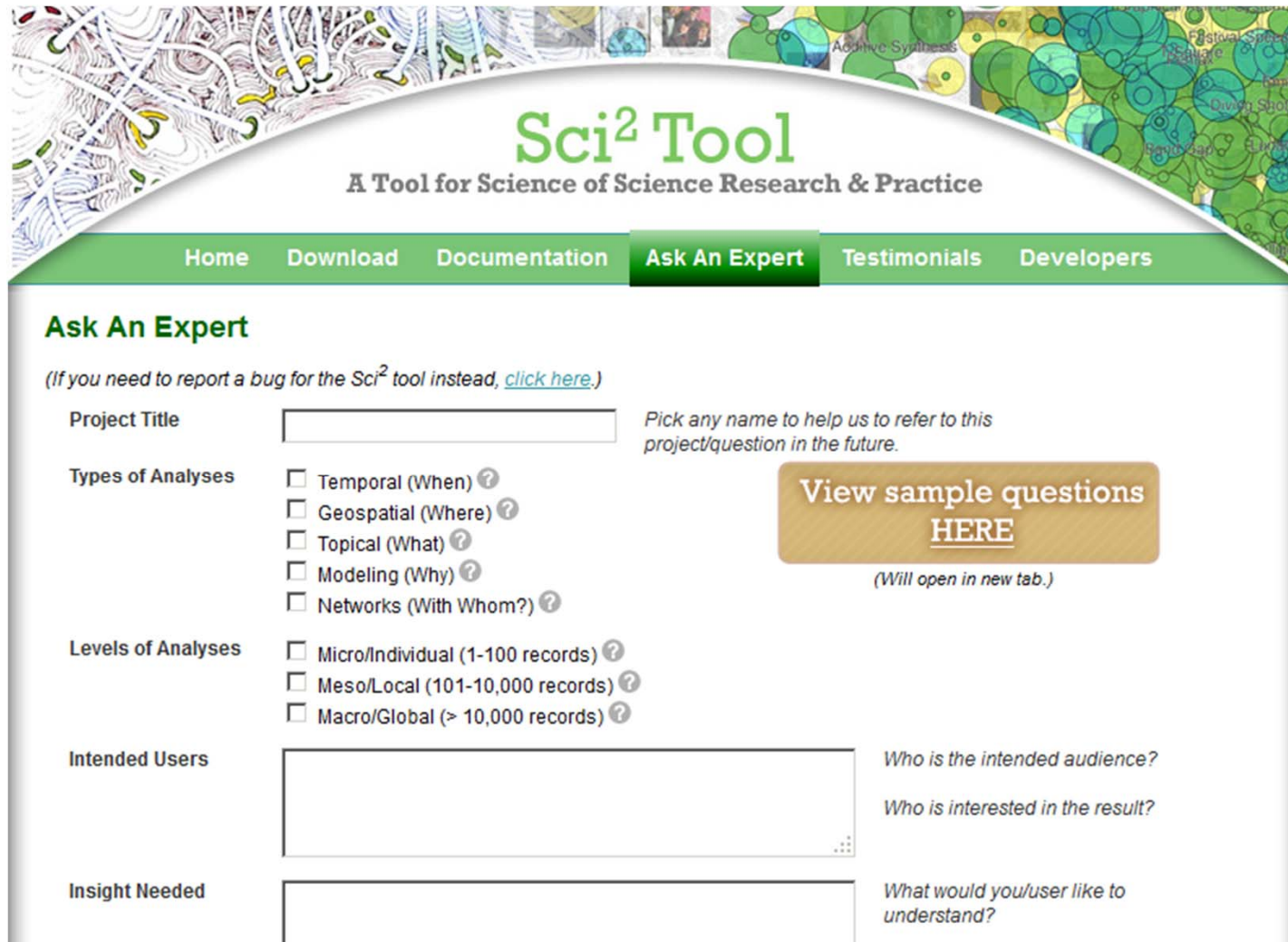
But also see details and novel workflows in

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

plus hands-on online tutorials in the Information Visualization MOOC at

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





The screenshot shows the Sci² Tool website interface. At the top, there is a navigation bar with links for Home, Download, Documentation, Ask An Expert (highlighted), Testimonials, and Developers. Below the navigation bar is the main heading "Sci² Tool" and the subtitle "A Tool for Science of Science Research & Practice". The "Ask An Expert" section contains a form with several fields and checkboxes. The "Project Title" field is a text input box. The "Types of Analyses" section has five checkboxes: Temporal (When), Geospatial (Where), Topical (What), Modeling (Why), and Networks (With Whom?). The "Levels of Analyses" section has three checkboxes: Micro/Individual (1-100 records), Meso/Local (101-10,000 records), and Macro/Global (> 10,000 records). The "Intended Users" field is a text input box. The "Insight Needed" field is a text input box. To the right of the form, there is a button labeled "View sample questions HERE" and a note "(Will open in new tab.)".

**Sci<sup>2</sup> Tool**  
A Tool for Science of Science Research & Practice

Home Download Documentation **Ask An Expert** Testimonials Developers

### Ask An Expert

*(If you need to report a bug for the Sci<sup>2</sup> tool instead, [click here.](#))*

Project Title

Types of Analyses

- Temporal (When) ?
- Geospatial (Where) ?
- Topical (What) ?
- Modeling (Why) ?
- Networks (With Whom?) ?

Levels of Analyses

- Micro/Individual (1-100 records) ?
- Meso/Local (101-10,000 records) ?
- Macro/Global (> 10,000 records) ?

Intended Users

Insight Needed

*Pick any name to help us to refer to this project/question in the future.*

**View sample questions  
HERE**

*(Will open in new tab.)*

*Who is the intended audience?*

*Who is interested in the result?*

*What would you/user like to understand?*

<https://sci2.cns.iu.edu/user/ask.php>

## Tutorial Overview

9:30 Welcome and Overview of Tutorial and Attendees

9:45 Open Data and Tools

- Standardize and federate micro-level datasets of S&T activity
- Open code tools and online services that are interoperable
- Sharing and teaching open datasets and tools

11-11:30 Networking Break

11:30 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- Country Collaboration Network, 2011 – Scoreboard 2013
- Country Expertise Profiles, 2010 – New Workflow
- Country Mobility Network, 1996-2011 – Scoreboard 2013
- Acceleration in the Development of Patented Technologies, 2000-11 – SB 2013
- Acceleration in the Co-Development of Patented Technologies, 1996-2001 and 2006-11 – Scoreboard 2013
- Evolving Country Patent Profiles, 1995-2010 – New Workflow

**13:00 Outlook and Q&A**

13:30 Adjourn

## 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.

This year, the course can be taken for three Indiana University credits as part of the Online Data Science Program just announced by the School of Informatics and Computing. Students interested in applying to the program can find more information here.

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.

Just like last year, students will have the opportunity to collaborate on real-world projects for a variety of clients. Click here to see this year's list of clients and projects.

Everyone who registers gains free access to the Scholarly Database (26 million paper, patent, and grant records) and the Sci2 Tool (100+ algorithms and tools).

Please watch the introduction video to learn more.



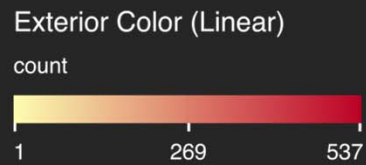
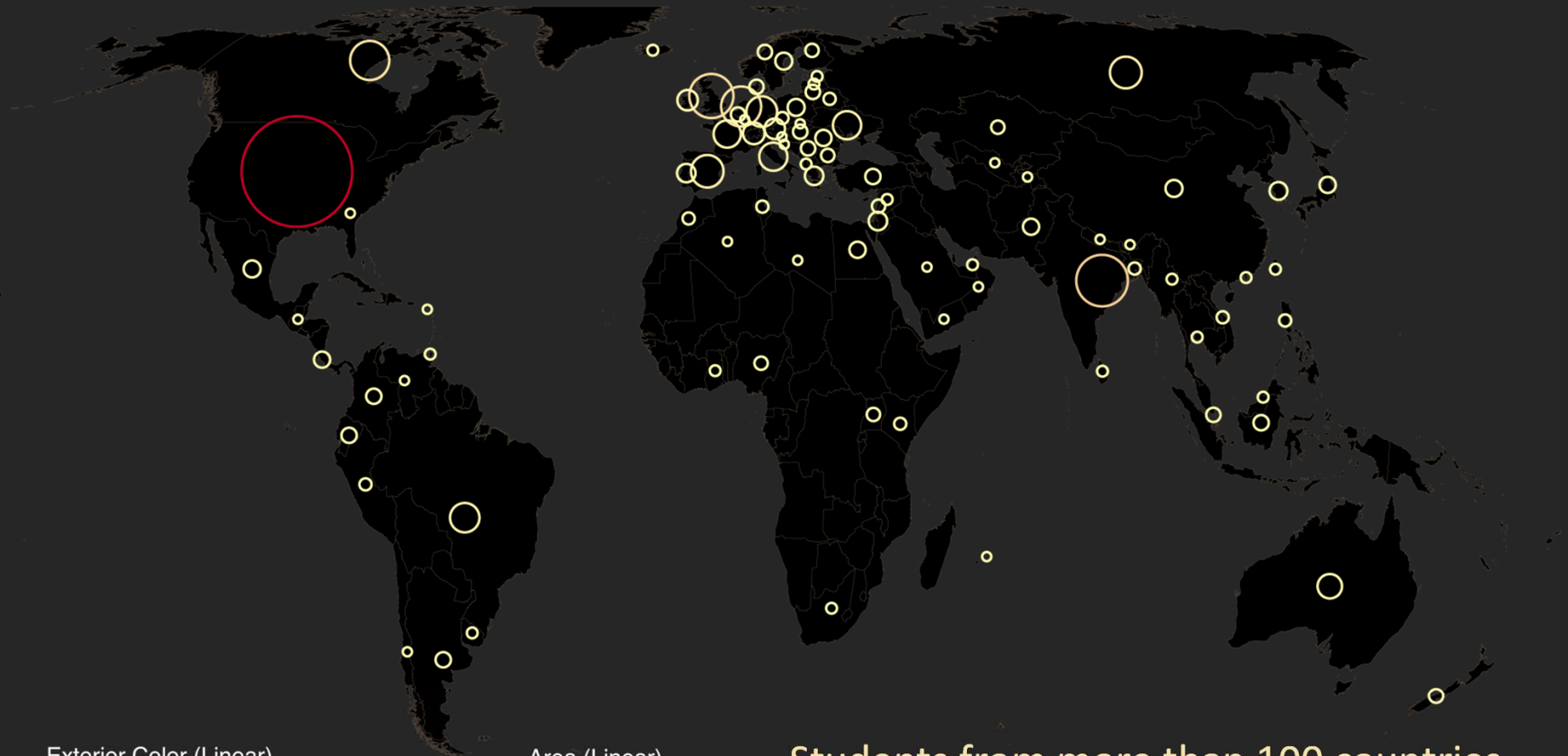
[Register for Course](#)

**IVMOOC 2014 course materials will be available until end of November 2014. The IVMOOC 2015 will open in January 2015 with new materials and a cloud computing setup.**

Register for free at <http://ivmooc.cns.iu.edu>. Class will restart in January 2015.

# The Information Visualization MOOC

[ivmooc.cns.iu.edu](http://ivmooc.cns.iu.edu)



Students from more than 100 countries  
350+ faculty members  
#ivmooc



# Course Schedule

- **Session 1** – Workflow design and visualization framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

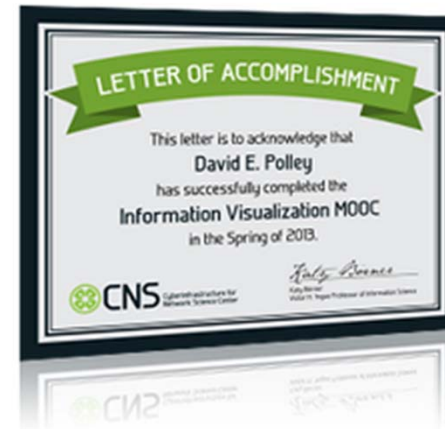
## Mid-Term

### Students work in teams with clients.

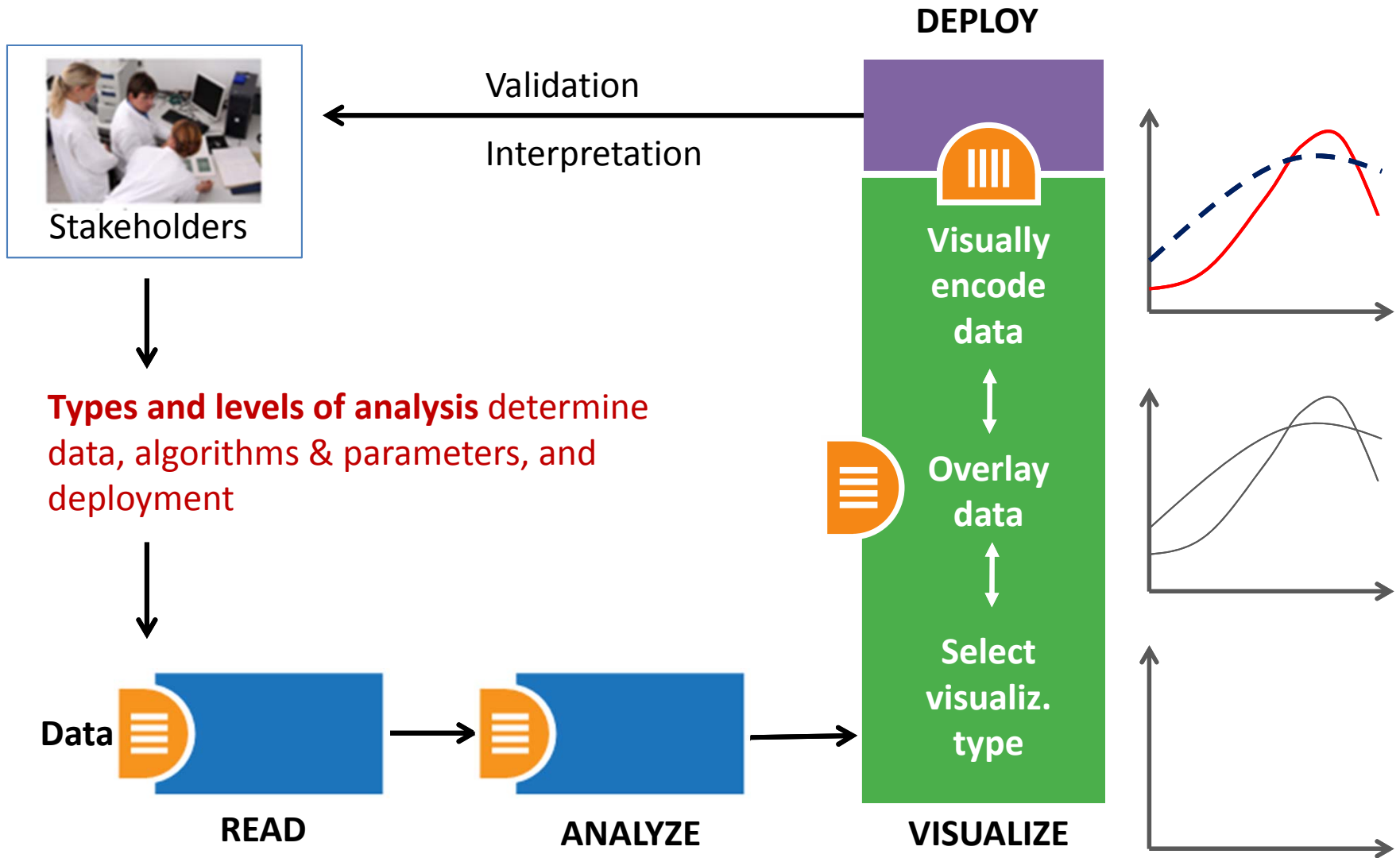
- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

## Final Exam

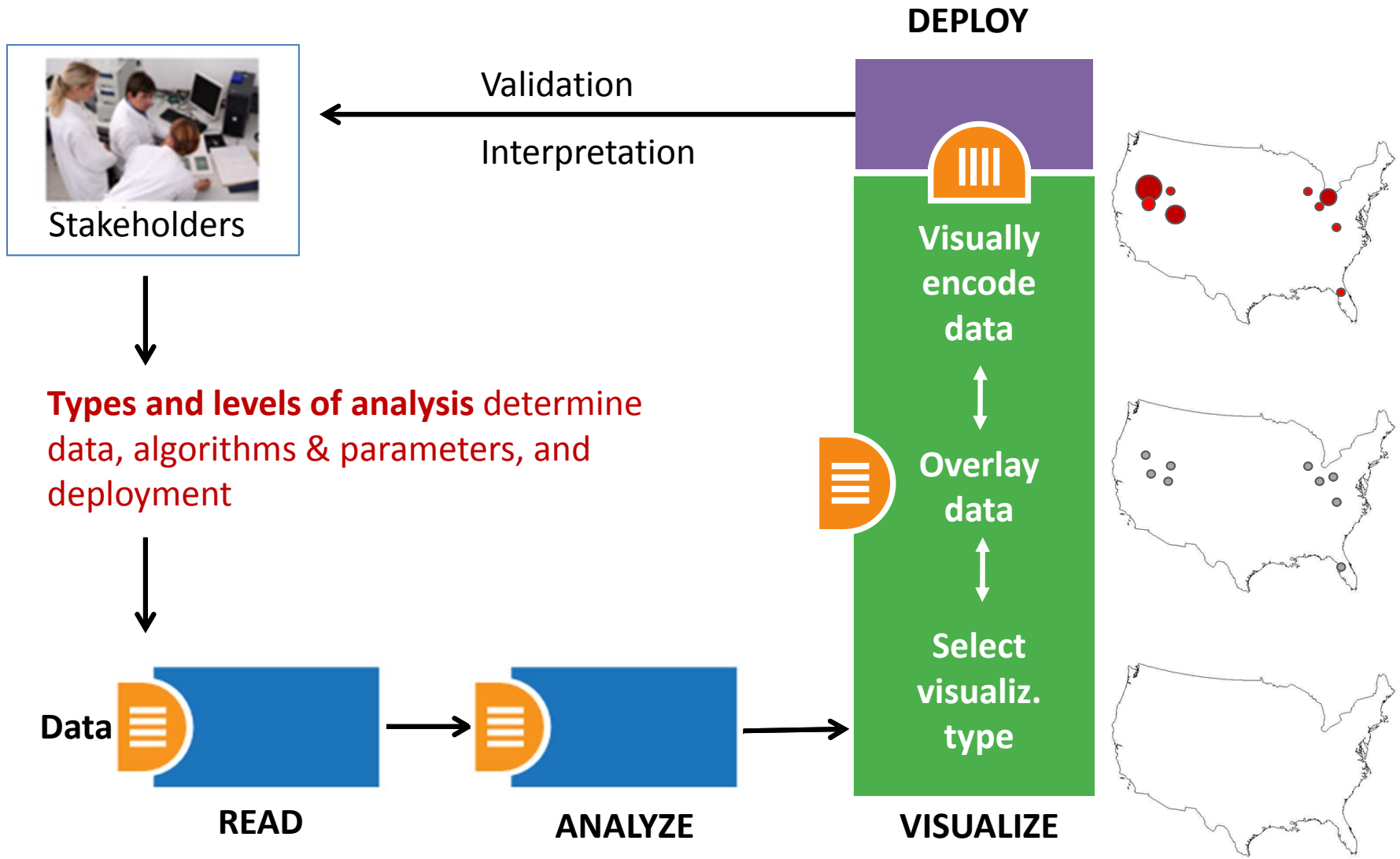
Final grade is based on Midterm (**30%**), Final (**40%**), Client Project (**30%**).



# Needs-Driven Workflow Design



# Needs-Driven Workflow Design



# Clients

## List of Clients



**Project Title:** Isis: 100 Years

**Client Name:** Jay Malone

**Project goal/scientific or practical value:** A visual representation Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic picture of how scholarship in the history of science has shifted over the past century.

**Information on dataset(s) to be used:** Citation information, author locale, and issue number for Isis publications.

**Relevant publications, websites, etc:** <http://www.press.uchicago.edu/ucp/journals/journal/isis.html>

**Conditions under which students can publish results and/or add project results to their resume:** Client would like to approve results.



**Project Title:** e-Xploration

**Client Name:** Luyi

**Project goal/scientific or practical value:** e-Xploration is an agent-based model for the ethnographic observation and the registry, analysis, and interpretation of social practices in virtual communities for intervention in the development of collaboration and cooperation. This project will analyze the interactions between subjects and objects in a platform collaborative community called OYCIB, a project based on e-Xploration ([e-crick.net](http://e-crick.net)).

**Information on dataset(s) to be used:** I can provide a data base in .graphml format for the students. The file .graphml contains the interactions between subjects and objects in a platform collaborative community called OYCIB. In the level of practice, it is not necessary that students know agent-based models for using the database. But, in another level, for example: the collaborate level for the OYCIB development, it is necessary to have basic knowledge in AMS or MAS and another competences like PHP and MySQL.

**Relevant publications, websites, etc:** <http://www.e-crick.net/logs>

**Conditions under which students can publish results and/or add project results to their resume:** If any person or institution use my dataset or another info about eXploration ([e-crick.net](http://e-crick.net), [oycib.net](http://oycib.net)), I need to approve the results and appear as co-author.

<http://ivmooc.cns.iu.edu/clients.html>

If you enjoyed the tutorial, please thank Fernando GALINDO-RUEDA, Mariagrazia SQUICCIARINI, and Alessandra COLECCHIA at OECD and Daniel Halsey, Robert Light, Michael Ginda at CNS.

## Q & A

Please complete the Post-Tutorial Questionnaire so that we can further improve these tutorials.

\* \* \*

Bug reports and all comments are welcome.

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

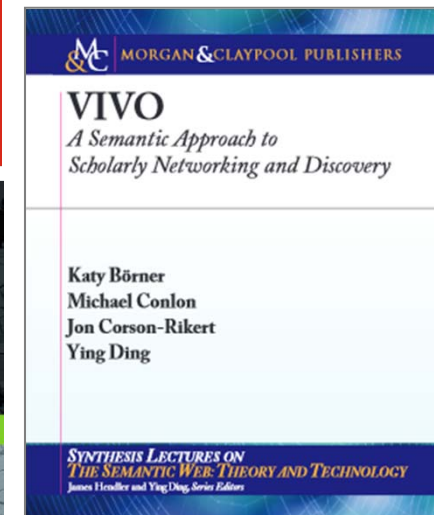
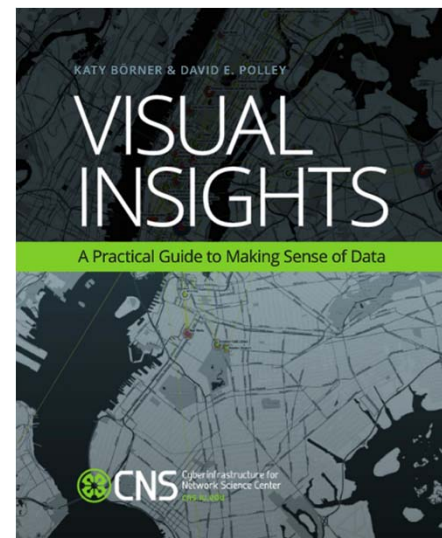
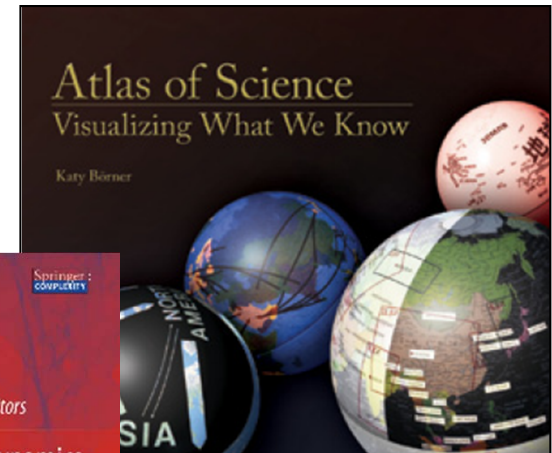
Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., Volume 41, Chapter 12, pp. 537-607. <http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

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.

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**. MIT Press.





**CNS** Cyberinfrastructure for Network Science Center

search  Search

About Us Research Development Teaching Outreach Videos News & Events Connect With Us

We work closely with clients to provide custom-made data, visualization, and software solutions

**Research**  
Open Data and Open Code for Big Science of Science Studies

**Latest News**  
Put your money where your citations are: a proposal for a new funding system (website accessed 9/05/13)

**Upcoming Events**  
OCT 1 Katy Börner attends PIUG 2013 Northeast Conference  
10.13 Katy Börner presents Mapping Science Exhibit at WSSF  
10.15 Ted Polley & Google Team present IVMOOC at EDUCAUSE  
10.22 Katy Börner presents at the SciELO 15 Years Conference

**Development**  
Behind the scenes of the design and development of *AcademyScope*

**Outreach**  
See some of the most fascinating data visualizations in the world.

**Videos**  
Watch Katy Börner's full presentation from TEDxBloomington

**Teaching**  
Successful IVMOOC will be offered again in January of 2014

**Our Products**  
We work closely with clients to provide custom-made data, visualization, and software solutions

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>