Atlas of Science: Envisioning Scholarly Data

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

With special thanks to the members at the Cyberinfrastructure for Network Science Center, the Mapping Science exhibit map makers, the exhibit advisory board, and the VIVO Collaboration.

HUBbub 2011: The HUBzero Conference (http://hubzero.org/hubbub2011)
April 6, 2011
Early Maps of the World VERSUS Early Maps of Science

3D
Physically-based
Accuracy is measurable
Trade-offs have more to do with granularity
2-D projections are very accurate at local levels
Centuries of experience
Geo-maps can be a template for other data

n-D
Abstract space
Accuracy is difficult
Trade-offs indirectly affect accuracy
2-D projections neglect a great deal of data
Decades of experience
Science maps can be a template for other data
Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org

The Power of Maps (2005)

Science Maps for Economic Decision Makers (2008)


Science Maps for Science Policy Makers (2009)

The Power of Forecasts (2007)

Science Maps for Scholars (2010)

Science Maps as Visual Interfaces to Digital Libraries (2011)

Science Maps for Kids (2012)

Science Forecasts (2013)

How to Lie with Science Maps (2014)

Exhibit has been shown in 72 venues on four continents. Currently at
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA
- Center of Advanced European Studies and Research, Bonn, Germany
- University of Michigan, Ann Arbor, MI

http://scimaps.org
Today used as a science "base map", see later slides.
Chemical Research & Development
Powers the U.S. Innovation Engine
Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences

INVESTMENT IN CHEMICAL SCIENCE R&D

$1 Billion
FEDERAL GOVERNMENT

$5 Billion
INDUSTRY FUNDING

$1B + $1B + $5 Billion

$10 Billion
CHEMICAL INDUSTRY OPERATING INCOME

$40 Billion
GROWTH IN GNP

600,000
JOBS CREATED

U.S. ECONOMY

The Council for Chemical Research (CCR) provides the U.S. Congress and government policy makers with important statistics regarding the impact of Federal Research & Development (R&D) investment on U.S. innovation and global competitiveness. Through its commissioned 5-year study, "Toward a Full Advantage of Typically Fossil Fuel Policy Makers," CCR developed the graphics below to communicate complex data points by the studies in direct, concise and clear terms.

The design shows that an initial $1B in Federal Investment, leveraged by $1B in Industry Investment, jointly new technologies to market and result in a $10B in operating income for the chemical industry, which growth in the Gross National Product (GNP) and further impacts the U.S. economy by generating approximately $40 billion GNP, along with a return of $600,000 in taxes. Additional benefits, also captured in the CCR studies, are not included in this map to reflect. This map clearly shows the high R&D investment pays for the shorter industry investment at innovative start-up companies to the manufacturing facilities and the larger Federal investment scale which begins in basic research and culminates in national economic and job growth along with the intangible benefits that in turn will be available for investment in future research.
Debut of 5th Iteration of the Mapping Science Exhibit at MEDIA X was in 2009 at Wallenberg Hall, Stanford University, [http://mediax.stanford.edu](http://mediax.stanford.edu), [http://scaleindependentthought.typepad.com/photos/scimaps](http://scaleindependentthought.typepad.com/photos/scimaps)
Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months, 12 coaches, 300 m long.  [http://www.expedition-zukunft.de](http://www.expedition-zukunft.de)

Interactive S&T Maps
Scholarometer is a social tool to facilitate citation analysis and help evaluate the impact of an author's publications.

Install the browser extension and start querying and tagging authors!

http://scholarometer.indiana.edu

http://mapofscience.com and SciVal by Elsevier
Interactive Maps of Science – Philanthropy

http://www.philanthropyinsight.org

Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010)

Map of Science

Scientific domains are highly interconnected. The boundaries between different domains are often fuzzy. One way of thinking about the relationships between domains is to connect them by the flow of information and co-authorship patterns observed in a large network of top researchers. Creating a network of scientific research can be accomplished by looking at scientific journals and their authors. The U.S. Map of Science used here is the product of a large study by researchers of the Network Science Lab at the University of California, Los Angeles, using 7.2 million papers and over 100,000 researchers, proceedings, and single-authored grants. Scientific and technological co-citation over the five-year period from 1995 to 2000. The connections were defined by the flows of citations between papers, with each paper connected to the set of all papers that cite it, plus a small group of highly related journals.

These clusters are represented by 754 individual nodes in the network. The link thickness indicates the strength of the connections between clusters. Each cluster is related to other clusters but is not tightly connected to the journals that make up each cluster. Then the clusters are scaled to the context area shared by the authors in the cluster, and the remaining scientific and technological clusters are connected to the cluster (encircled to one of 13 colors).

Map of science is one can be used to understand the flow of information and co-authorship patterns observed in a large network of researchers. It can be used to identify the key areas of research and the connections between them.
Google Map JavaScript API was used to implement both maps with two aggregation layers for each. The geographic map aggregates to the state level and the city level. The science map has a high level of aggregation of 13 top-level scientific disciplines and a low level of 554 sub-disciplines.
The geographic map at state level.

The geographic map at city level.
Search result for “corn"
Icons have same size but represent different #records

Click on one icon to display all records of one type.
Here publications in the state of Florida.
Detailed information on demand via original source site for exploration and study.

Search result for “Miscanthus,” a special energy biomass crops for second generation biofuel.
The science map at 13 top-level scientific disciplines level.

The science map at 554 sub-disciplines level.
Thermal tolerant avicelase from Acidothermus cellulolyticus

Abstract

The invention provides a thermal tolerant (thermostable) cellulase, AvIII, that is a member of the glycoside hydrolase (GH) family. AvIII was isolated and characterized from Acidothermus cellulolyticus and, like many cellulases, the disclosed polypeptide and/or its derivatives may be useful for the conversion of biomass into biofuels and chemicals.
https://app.nihmaps.org
OBESITY, INSULIN RESISTANCE, IGF’S, AND BREAST CANCER RISK IN AFRICAN AMERICANS

The purpose of this study is to better understand how lifestyle factors and their interaction with genetic factors influence a woman’s risk of developing breast cancer. In order to learn more about the causes of breast cancer, we need to compare the lifestyles of people who have breast cancer with those who do not. 608 women are expected to be enrolled.
S&T Studies Using Semantic Web Data
Linked Open Data

- Interlinking existing data silos and
- Exposing them as structured data
- Adding new high quality data relevant for S&T studies

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

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

Funded by $12 million NIH award.

**Cornell University:** Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistelbauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Mageaux Johnson, Sara Kreinin, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raun, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University:** Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Dunlop, Jon Dunn, Micah Linneimer, Nianli Ma, Robert McDonald, Barbara Ann O’Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michael Trimarchi. **Washington University, St. Louis:** Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cherrif, Oscar Cruz, Dan Dickinson, Chris Huang, Ifty Kla, Peter Michelini, Grace Migliorini, John Ruffing, Jason Speckand, Tru Tran, Jesse Turner, Vinay Varughese.
<table>
<thead>
<tr>
<th>Type of Analysis vs. Level of Analysis</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Analysis/Profiling</td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of science</td>
</tr>
<tr>
<td>Temporal Analysis (When)</td>
<td>Funding portfolio of one individual</td>
<td>113 Years of PNAS Research</td>
<td></td>
</tr>
<tr>
<td>Geospatial Analysis (Where)</td>
<td>Career trajectory of one individual</td>
<td>Mapping an intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td>Topical Analysis (What)</td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td>Network Analysis (With Whom?)</td>
<td>NSF Co-PI network of one individual</td>
<td>NIH’s core competency</td>
<td></td>
</tr>
</tbody>
</table>

**Temporal Analysis (When)** Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level
**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)

**Network Analysis (With Whom?)** Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?
This XML file does not appear to have any style information associated with it. The document tree is shown below.

- <rdf:RDF>
  - <rdf:Description rdf:about="http://vivo.all.e-ea.cheneau.kn.562096">
    <rdf:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Member</rdf:label>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdf:memberRoleOf rdf:resource="http://vivo.all.e-ea.cheneau.kn.562096#MemberRole"/>
    <rdf:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1997</rdf:startDate>
    <rdf:Description rdf:about="http://vivo.all.e-ea.cheneau.kn.787654">
      <rdf:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Member</rdf:label>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
      <rdf:memberRoleOf rdf:resource="http://vivo.all.e-ea.cheneau.kn.787654#MemberRole"/>
      <rdf:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1980</rdf:startDate>
      <rdf:Description rdf:about="http://vivo.all.e-ea.cheneau.kn.255620">
        <rdf:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Member</rdf:label>
        <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
        <rdf:memberRoleOf rdf:resource="http://vivo.all.e-ea.cheneau.kn.255620#MemberRole"/>
        <rdf:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1997</rdf:startDate>
      </rdf:Description>
    </rdf:Description>
    <rdf:Description rdf:about="http://vivo.all.e-ea.cheneau.kn.255620">
      <rdf:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Member</rdf:label>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
      <rdf:memberRoleOf rdf:resource="http://vivo.all.e-ea.cheneau.kn.255620#MemberRole"/>
      <rdf:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1997</rdf:startDate>
    </rdf:Description>
  </rdf:Description>
</rdf:RDF>
Data Download Support

General Statistics
- 36 publication(s) from 2001 to 2010 (.CSV File)
- 80 co-author(s) from 2001 to 2010 (.CSV File)

Co-Author Network
(GraphML File)

Save as Image (.PNG file)

Tables
- Publications per year (.CSV File)
- Co-authors (.CSV File)
36 publication(s) from 2001 to 2010 ([CSV File])

80 co-author(s) from 2001 to 2010 ([CSV File])

<table>
<thead>
<tr>
<th>Year</th>
<th>Count</th>
<th>Co-Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>1</td>
<td>Chen C.</td>
</tr>
<tr>
<td>2002</td>
<td>2</td>
<td>Chen C., Mahnke T., Fang Y.</td>
</tr>
<tr>
<td>2003</td>
<td>2</td>
<td>Chen C., Boyack K.W.</td>
</tr>
<tr>
<td>2004</td>
<td>17</td>
<td>Serguya A., Penyukh S., Tashkur S., Sambamuthu R., Maru J.T., Stallic R.M., Mare K., Moor K.A.</td>
</tr>
</tbody>
</table>

Co-author network ([GraphML File])

Save as Image ([PNG file]), see top file.

Publications per year ([CSV File]), see top file.

Co-authors ([CSV File])

<table>
<thead>
<tr>
<th>Co-Author</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andriopoulos G.</td>
<td>7</td>
</tr>
<tr>
<td>Andriopoulos N.</td>
<td>1</td>
</tr>
<tr>
<td>Ben-Meled Z.</td>
<td>1</td>
</tr>
<tr>
<td>Blackwell A.</td>
<td>1</td>
</tr>
<tr>
<td>Boyack K.W.</td>
<td>4</td>
</tr>
<tr>
<td>Boden M.</td>
<td>1</td>
</tr>
<tr>
<td>Breslow J.</td>
<td>1</td>
</tr>
<tr>
<td>Burkhard R.J.</td>
<td>6</td>
</tr>
<tr>
<td>Chen C.</td>
<td>6</td>
</tr>
</tbody>
</table>

Run Science of Science (Sci2) Tool and load Co-Author Network ([GraphML File])

Network Analysis Toolkit
Nodes: 81
Edges: 390

Visualize the file using Radial Graph layout.

Click on node to focus on it.
Hover over a node to highlight its co-authors.

Code and tutorials are linked from [http://sci.slis.indiana.edu/sci2](http://sci.slis.indiana.edu/sci2)
Scholarly Database: 25 million scholarly records
http://sdb.slis.indiana.edu

VIVO Research Networking
http://vivoweb.org

Network Workbench Tool & Community Wiki
http://nwb.cns.iu.edu

Science of Science (Sci²) Tool
http://sci2.cns.iu.edu

Epidemics Cyberinfrastructure
http://epic.cns.iu.edu

VIVO National Level Visualizations

Video and paper at
http://www.scivee.tv/node/27704
Science is global. World view of VIVO activity.
Web site visits are aggregated at the country level.

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

Shown are the
- Number of people profiles in the 7 different VIVO installation sites plus CAS and U Melbourne.
- Email contacts by data and service providers as well as institutions interested to adopt VIVO.
- The number of visitors on [http://vivoweb.org](http://vivoweb.org)

Circles are area size coded using a logarithmic scale.
VIVO 1.0 source code was publicly released on April 14, 2010. 87 downloads by June 11, 2010.
The more institutions adopt VIVO, the more high quality data will be available to understand, navigate, manage, utilize, and communicate progress in science and technology.
Second Annual VIVO Conference
August 24-26, 2011
Gaylord National, Washington D.C.

http://vivoweb.org/conference

VIVO is supported by NIH Award U24 RR029822

References


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

CNS Facebook: http://www.facebook.com/pages/Cyberinfrastructure-for-Network-Science-Center/144339535612571

Exhibit Facebook: http://www.facebook.com/mappingscience