Multi-Scale Maps of Scholarly Activity

Katy Börner
Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, The Netherlands & Cyberinfrastructure for Network Science Center, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

Research Talk at the Organisation for Economic Co-Operation and Development (OECD)
Paris, France
http://www.oecd.org

April 12, 2012

Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th>Statistical Analysis/Profiling</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>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of science, all of US, all of states</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temporal Analysis (When)</th>
<th>Funding portfolio of one individual</th>
<th>PNAS, bursts of PNAS, 113 Years of PNAS Research</th>
<th>PNAS, grants, VxOrd/Topic, NIH funding</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Geospatial Analysis (Where)</th>
<th>Career trajectory of one individual</th>
<th>PNAS, publications, intellectual landscape</th>
<th>NIH’s, geographical distribution</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Topical Analysis (What)</th>
<th>Knowledge flows in Chemistry research</th>
<th>VxOrd/Topic in NIH research</th>
<th>NIH’s, publication distribution</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Network Analysis (With Whom?)</th>
<th>NSF’s, co-author network of one individual</th>
<th>NIH’s, co-author network</th>
<th>NIH’s, co-author network</th>
</tr>
</thead>
</table>
**Modeling Science Dynamics**

- multi-level,
- mixed methods, and
- multi-perspective models


---

**Descriptive Models of Science**

- Detect advances of scientific knowledge via "longitudinal mapping" (Garfield, 1994).
- Synthesis of specialty narratives from co-citation clusters (Small, 1986).
- Identify cross-disciplinary fertilization via "passages through science" (Small, 1999, 2000).
- Understand scholarly information foraging (Sandstrom, 2001).
- Knowledge discovery in un-connected terms (Swanson & Smalheiser, 1997).
- Determine areas of expertise for specific researcher, research group via "invisible colleges" (note that researchers self definition might differ from how field defines him/her) (Crane, 1972).
- Identify profiles of authors, also called CAMEOS, to be used to for document retrieval or to map an author’s subject matter and studying his/her publishing career, or to map the social and intellectual networks evident in citations to and from authors and in co-authorships (White, 2001).
Descriptive Models of Science cont.

- Identification of scientific frontiers [http://www.science-frontiers.com/].
- *ISI's Essential Science Indicators* [http://essentialscience.com/].
- Import-export studies (Stigler, 1994).
- Evaluation of 'big science' facilities using 'converging partial indicators' (Martin, 1996; Martin & Irvine, 1983).
- Input (levels of funding, expertise of scientists, facilities used) - output (publications, patents, Nobel prices, improved health, reduced environment insults, etc. - influenced by political, economic, financial, and legal factors studies (Kostroff & DelRío, 2001).
- Determine influence of funding on research output (Boyack & Borner, 2002).
- How to write highly influential paper (van Dalen & Henkens, 2001).

Process Models of Science

Can be used to predict the effects of

- Large collaborations vs. single author research on information diffusion.
- Different publishing mechanisms, e.g., E-journals vs. books on co-authorship, speed of publication, etc.
- Supporting disciplinary vs. interdisciplinary collaborations.
- Many small vs. one large grant on # publications, Ph.D. students, etc.
- Resource distribution on research output.
- …

In general, process model provide a means to analyze the structure and dynamics of science -- to study science using the scientific methods of science as suggested by Derek J. deSolla Price about 40 years ago.
Chemical Research & Development
Powers the U.S. Innovation Engine

INVESTMENT IN CHEMICAL SCIENCE R&D

$1 Billion
FEDERAL GOVERNMENT

$5 Billion
INDUSTRY FUNDING

$8 Billion
TAXES

$1B + $5B = $10 Billion

$40 Billion
GROWTH IN GNP

+ 600,000
JOBS CREATED

20 YEARS

TIMELINE FROM CONCEPTION TO COMMERCIALIZATION

The Council for Chemical Research (CCR) conducted a study to assess the macroeconomic implications of public and private R&D investments in the chemical sciences. The study found that investment in chemical science R&D yields significant economic benefits through innovation, job creation, and growth in the chemical industry and the U.S. economy.


Clickstream Map of Science


Adrian White and the National Geographic EarthPulse Team. 2008. A Global Projection of Subjective Well-being
Science of (team) science research and practice requires an interdisciplinary, multi-level, mixed-methods approach. Expertise, theories, methods, data, and tools from diverse research fields need to be applied and advanced to arrive at a holistic understanding of the science system.

**Mixed-Methods, Multi-Level Science of Science**
(or Team Science or SciSIP) studies need:

**Expertise** – identify and access it at the perfect moment using, e.g., Facebook, LinkedIn, Academia, VIVO, Harvard Profiles, Elsevier's Collexis, Loki, Stanford's CAP, or other systems.

**Data** – find, interlink, unify, merge, reformat, share them, e.g., using web sites analogous to [http://www.diggingintodata.org/Repositories/tabid/167/Default.aspx](http://www.diggingintodata.org/Repositories/tabid/167/Default.aspx), SDB, or LOD.

**Tools** – identify, learn, advance, share code, e.g., via Plug-and-Play Macroscopes, to arrive at a holistic understanding of the science system.
Expertise – identify and access it at the perfect moment using, e.g., Facebook, LinkedIn, Academia, VIVO, Harvard Profiles, Elsevier’s Collexis, Loki, Stanford’s CAP, or other systems.

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

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

Funded by $12 million NIH award.

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?

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

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.
Data – find, access, interlink, unify, merge, reformat, share them, e.g., using web sites analogous to http://www.diggingintodata.org/Repositories/tabid/167/Default.aspx, SDB, or LOD.
Supports federated search of 25 million publication, patent, grant records.
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

Register for free access at http://sdb.cns.iu.edu
Since March 2009:
Users can download networks:
- Co-author
- Co-investigator
- Co-inventor
- Patent citation
and tables for burst analysis in NWB.

Semantic Web: 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

Save Data.gov, sign the petition at
http://om.ly/BRPRE

Twitter
#savethedata

August 2007
Tools – *continuously* identify, learn, advance, share code, e.g., via Plug-and-Play Macroscopes


Video and paper are at [http://www.scivee.tv/node/27704](http://www.scivee.tv/node/27704)
Designing “Dream Tools”

Many of the best micro-, tele-, and macrosopes are designed by scientists keen to observe and comprehend what no one has seen or understood before. Galileo Galilei (1564–1642) recognized the potential of a spyglass for the study of the heavens, ground and polished his own lenses, and used the improved optical instruments to make discoveries like the moons of Jupiter, providing quantitative evidence for the Copernican theory.

Today, scientists repurpose, extend, and invent new hardware and software to create “macrosopes” that may solve both local and global challenges.

Plug-and-play macrosopes empower me, my students, colleagues, and 100,000 others that downloaded them.

Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of data sets representing the structure and dynamics of complex systems. Analysis, navigation, and management of these continuously evolving data sets require a new kind of data-analysis and visualization tool we call a macroscope (from the Greek macros, or “great,” and skopein, or “to observe”) inspired by de Rosnay’s futurist science writings. Macrosopes provide a “vision of the whole,” helping us “synthesize” the related elements and enabling us to detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macrosopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend.
While microscopes and telescopes are physical instruments, *macrosopes resemble continuously changing bundles of software plug-ins*. Macrosopes make it easy to select and combine algorithm and tool plug-ins but also interface plug-ins, workflow support, logging, scheduling, and other plug-ins needed for scientifically rigorous yet effective work.

They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. **Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.**

**Macrosopes Serve the Changing Scientific Landscape**

- Different datasets/formats.
- Diverse algorithms/tools written in many programming languages.
- Health
- Finance
- Smart Cities
- Epidemiology
- Crime
Related Work

Google Code and SourceForge.net provide special means for developing and distributing software
- In August 2009, SourceForge.net hosted more than 230,000 software projects by two million registered users (285,957 in January 2011);
- In August 2009 ProgrammableWeb.com hosted 1,366 application programming interfaces (APIs) and 4,092 mashups (2,699 APIs and 5,493 mashups in January 2011)

Cyberinfrastructures serving large biomedical communities
- Cancer Biomedical Informatics Grid (caBIG) (http://cabig.nci.nih.gov)
- Biomedical Informatics Research Network (BIRN) (http://nbirn.net)
- Informatics for Integrating Biology and the Bedside (i2b2) (https://www.i2b2.org)
- HUBzero (http://hubzero.org) platform for scientific collaboration uses
- myExperiment (http://myexperiment.org) supports the sharing of scientific workflows and other research objects.

Missing so far is a common standard for
- the design of modular, compatible algorithm and tool plug-ins (also called “modules” or “components”)
- that can be easily combined into scientific workflows (“pipeline” or “composition”),
- and packaged as custom tools.

OSGi & CIShell

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

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

CIShell is the basis of Network Workbench, TexTrend, Sci2, and the upcoming Epic tool.

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

CIShell Features

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

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

Learn More...

- CIShell Papers
- CIShell Powered Tools
- Algorithms
- Plugins (coming soon)
- Misc Tool Documentation
- CIShell Web Services (coming soon)
- Screenshots

Getting Started...

- Documentation & Developer Resources
- Download

Getting Involved...

- Contact Us
Sci2 Tool – “Open Code for S&T Assessment”

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

Sci² Tool

Geo Maps

Circular Hierarchy

EpiC Tool

Welcome to the EpiC tool, which supports modeling, analysis, and visualization of epidemic processes.

EpiC uses the Cyberinfrastructure Shell (https://cnsi.indiana.edu) developed at the Cyberinfrastructure for Network Science Center (https://cnsi.indiana.edu).

Please cite as follows:
OSGi/CIShell Adoption

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

- **Cytoscape** [http://cytoscape.org](http://cytoscape.org) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

- **MAEviz** [https://wiki.ncsa.uiuc.edu/display/MAE/Home](https://wiki.ncsa.uiuc.edu/display/MAE/Home) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

- **Taverna Workbench** [http://taverna.org.uk](http://taverna.org.uk) Developed by the myGrid team led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.

- **TEXTrend** [http://textrend.org](http://textrend.org) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

- **DynaNets** [http://www.dynanets.org](http://www.dynanets.org) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.


As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.
Few have access to or time to visit “Visualization Domes”

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

Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org/
Read about and zoom into maps at [http://scimaps.org/exhibit_info](http://scimaps.org/exhibit_info)

Call for Maps for the 8th Iteration of the Places & Spaces: Mapping Science Exhibit on "Science Maps for Kids" (2012)

[http://scimaps.org/call](http://scimaps.org/call)
Mapping Science Exhibit at MEDIA X was on May 18, 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 Opening was on April 23rd, 2009 by German Chancellor Merkel [http://www.expedition-zukunft.de](http://www.expedition-zukunft.de)
Illuminated Diagram Display soon on display at the Smithsonian in DC.

http://scimaps.org/exhibit_info/#ID
About

This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Elinor Ostrom - Nobel Prize in Economic Sciences 2009

Born: 7 August 1933, New York, NY, USA

Affiliation at the time of the award: Indiana University, Bloomington, IN, USA

Arizona State University, Tempe, AZ, USA

Prize motivation: "For her analysis of economic governance, especially the commons"

Field: Economic governance

Contribution: Challenged the conventional wisdom by demonstrating how local property can be successfully managed by local commons without any regulation by central authorities or privatization.

Interact

Select any location on the Geographic Map location (by breaching your finger over an area on the lectern's touch screen) and topics studied in that area will highlight on the Science Map, the brighter a topic grows, the more papers on that topic originated in the selected area. Conversely, lacking a scientific area in the Science Map illuminates places on the Geographic Map where that topic is studied. People and topic buttons support the exploration of publication output by selected Nobel laureates and particular lines of research using MEDLINE data from 2000-2009.

Keywords Search

2000-2009

About

This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Top Five Continents

North America - 13,000 records
South & East Asia - 3,889
Australia - 2,041
Africa - 2,046
South America - 1,542

Top Five Scientific Disciplines

Math & Physics - 1,000 records
Health Professionals - 5,589
Social Sciences - 8,419
Agricultural, Chemical, Mechanical & Civil Engineering - 2,036
Humanities - 1,542

Search

The keyboard supports retrieval and display of papers based on Medical Subject Headings (MeSH) and MeSH qualifier terms. If multiple terms are entered in a field, they are automatically combined using "AND." So, "breast cancer" matches any record with "breast" or "cancer" in that field. You can pull AND between terms to combine with "AND." Thus "breast AND cancer" would only match records that contain both terms. Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer," and not records where "breast" and "cancer" are both present, but the exact phrase.
Ingo Gunther's Worldprocessor globe design now on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan
References


---

Mixed-indicators model for identifying emerging research areas

Hanning Guo · Scott Weingart · Katy Börner

Abstract This study presents a mixed model that combines different indicators to describe and predict key structural and dynamic features of emerging research areas. Three indicators are combined: sudden increases in the frequency of specific words; the number and speed by which new authors are attracted to an emerging research area, and changes in the interdisciplinarity of cited references. The mixed model is applied to four emerging research areas: RNAi, Nano, h-Index, and Impact Factor research using papers published in the Proceedings of the National Academy of Sciences of the United States of America (1982–2009) and in Scientometrics (1978–2009). Results are compared in terms of strengths and temporal dynamics. Results show that the indicators are indicative of emerging areas and they exhibit interesting temporal correlations: new authors enter the area first, then the interdisciplinarity of paper references increases, then word bursts occur. All workflows are reported in a manner that supports replication and extension by others.

Keywords Burst detection · Prediction · Emerging trend · Temporal dynamics · Science of science (Sci²) tool

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

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