Desirable Features of Large, Interdisciplinary Research Initiatives/Centers

'Large/major' funding should result in high quantity and high quality scientific products such as experts, papers, patents, databases, software tools, Cyberinfrastructures and associated

* intellectual capital
* social capital
* infrastructure capital

that is important to a research community.

They should have an interdisciplinary ‘footprint’, i.e., consume and produce scientific products from many scientific domains.

Education and Outreach are important.

Initiatives/Centers provide critical mass, physical facilities, longevity, stability, visibility that can have a major impact on the growth of a research community.
Evaluating Large, Interdisciplinary Research Initiatives/Centers

Data Acquisition
Acquire all (interdisciplinary) scientific products **consumed and produced** (experts, papers, patents, databases, software tools, cyberinfrastructures, funding) by the initiative(s)/center(s) in question as well from ‘comparison’ units.

Data Analysis
Number of consumed/produced products over time.
Geo and topic location of consumed/produced products.
Dynamic features, e.g., burst of activity, (social) network evolution, secondary effects.
Initiatives/Centers should be involved in positive (not negative) feedback cycles.

Result Communication
Top-n lists. Success stories.
Tables of major produced/consumed entities. **Profiles.**
Major produced/consumed entities in their (geo/network) context, e.g., paper citation graphs, funding-papers-PI graphs, evolving scholarly networks, impact on education/public policy.

Computational Scientometrics: Studying Science by Scientific Means

- **Places & Spaces: Mapping Science** exhibit, see also [http://scimaps.org](http://scimaps.org).
General Process of Data Acquisition, Analysis and Visualization

Analysis of Emergent Research Areas
Mapping Network Science


Data Acquisition via

Questionnaires  ISI Citation Data  Bibliography Data
HistCite paper-citation graph of publications

Components with size larger than three of the co-author network based on bibliography data.
Data Acquisition for Comprehensive Analysis
Lab/Center Management System  vs. Spacebook and MS Famulus

Designed to track, manage, and make use of data relevant for the daily operation of a medium size research team.

![Diagram of Lab/Center Management System](http://ivl.slis.indiana.edu)

Data Entities and Interlinkages

Designed for team leads, members, IT admins but also for external scholars and funding agencies.

![Diagram of Data Entities and Interlinkages](http://ivl.slis.indiana.edu)

**Not covered:**
- Queries
- Workflows
- Protocols
- Comments
- Bookmarks
- Ratings
Simplified representation of the IVL database schema

Data Entry

Tutorials - Back

<table>
<thead>
<tr>
<th>Title</th>
<th>Analyzing and Visualizing Knowledge Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link</td>
<td>http://</td>
</tr>
<tr>
<td>People</td>
<td>RESET TO ZERO</td>
</tr>
<tr>
<td></td>
<td>Aigner, Wolfgang</td>
</tr>
<tr>
<td></td>
<td>Almer, Ian</td>
</tr>
<tr>
<td></td>
<td>Alfred K.D.</td>
</tr>
<tr>
<td></td>
<td>Ambros, Suneeta</td>
</tr>
<tr>
<td></td>
<td>Andersen, Christian</td>
</tr>
<tr>
<td></td>
<td>Andersen, Fen-OleV</td>
</tr>
<tr>
<td></td>
<td>Altfors, Gernadley</td>
</tr>
<tr>
<td></td>
<td>Anda, Gunlevey</td>
</tr>
<tr>
<td>Start Date</td>
<td>01-23-2003</td>
</tr>
<tr>
<td>End Date</td>
<td>01-23-2003</td>
</tr>
<tr>
<td>Location</td>
<td>Santa Clara, CA</td>
</tr>
<tr>
<td>Venue</td>
<td>Electronic Imaging</td>
</tr>
<tr>
<td>Time</td>
<td>12:33:50</td>
</tr>
</tbody>
</table>

Submit | Clear
Demo

http://ivl.slis.indiana.edu

Time series analysis & visualization

- Grants
- Ph.D and Master students
- Publications
- Independent studies
Katy’s Travels in 2000-2006
Mapping the Evolution of Co-Authorship Networks
SEI: Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0513650 award (Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), $1,120,926) Sept. 05 - Aug. 08. [http://nwb.slis.indiana.edu](http://nwb.slis.indiana.edu)

CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, $451,000) Sept. 03-Aug. 08. [http://iv.slis.indiana.edu/](http://iv.slis.indiana.edu/)

Challenges - Interlink $ Input & Publication/Patent Citation Output

Need to interlink
- Grants and papers/patents.
- Grants/papers/patents and their PIs/authors/inventors, etc.

Use resulting networks to
- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.
Scholarly Database: Web Interface

Search across publications, patents, grants. Download records and/or (evolving) co-author, paper-citation networks.

Register for free access at https://sdb.slis.indiana.edu.
Datasets available via the Scholarly Database (* future feature)

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Records</th>
<th>Years Covered</th>
<th>Updated</th>
<th>Restricted Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>13,149,741</td>
<td>1965-2005</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PhysRev</td>
<td>398,005</td>
<td>1893-2006</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PNAS</td>
<td>16,167</td>
<td>1997-2002</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>USPTO</td>
<td>3,179,930</td>
<td>1976-2004</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td>174,835</td>
<td>1985-2003</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td>1,043,804</td>
<td>1972-2002</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18,021,560</strong></td>
<td><strong>1893-2006</strong></td>
<td><strong>4</strong></td>
<td><strong>3</strong></td>
</tr>
</tbody>
</table>

Aim for comprehensive time, geospatial, and topic coverage.
**Latest ‘Base Map’ of Science**  
*Kevin W. Boyack & Richard Klavans, unpublished work.*

- Uses combined SCI/SSCI from 2002  
  - 1.07M papers, 24.5M references, 7,300 journals  
  - Bibliographic coupling of papers, aggregated to journals  
- Initial ordination and clustering of journals gave 671 clusters  
- Coupling counts were reaggregated at the journal cluster level to calculate the  
  - (x,y) positions for each journal cluster  
  - by association, (x,y) positions for each journal
Science map applications: Identifying core competency
Kevin W. Boyack & Richard Klavans, unpublished work.

Funding patterns of the US Department of Energy (DOE)

Science map applications: Identifying core competency
Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Science Foundation (NSF)
Building Market Places not Cathedrals

➢ ‘Software glue’ has to interlink datasets and algorithms written in different languages using different data formats.
➢ The smaller the glue or ‘CI Shell’, the more likely it can be maintained.
CIShell – Serving Non-CS Algorithm Developers & Users

CIShell Wizards

CIShell

IVC Interface

NWB Interface

CIShell – Build on OSGi Industry Standard

CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework.

OSGi ([http://www.osgi.org](http://www.osgi.org)) is

- A standardized, component oriented, computing environment for networked services.
- Successfully used in the industry from high-end servers to embedded mobile devices since 7 years.
- Alliance members include IBM (Eclipse), Sun, Intel, Oracle, Motorola, NEC and many others.
- Widely adopted in open source realm, especially since Eclipse 3.0 that uses OSGi R4 for its plugin model.

Advantages of Using OSGi

- Any CIShell algorithm is a service that can be used in any OSGi-framework based system.
- Using OSGi, running CIShells/tools can connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

Ideally, CIShell becomes a standard for creating OSGi Services for algorithms.
CISShell applications can be deployed as distributed data and algorithm repositories, stand alone applications, peer-to-peer architectures, and server-client architectures.
NWB Tool: Interface Elements

http://nwb.slis.indiana.edu

Load Data | Select Preferences | List of Data Models

Console

Scheduler

Visualize Data

Open Text Files

NWB Community Wiki

Network Workbench Marketplace: An Ecology of Data Formats, Converters, and Algorithms

CiShell & OSGI

https://nwb.slis.indiana.edu/community/
Places & Spaces: Mapping Science
a science exhibit that introduces people to maps of sciences, their makers and users.
http://scimaps.org

Exhibit Curators: Dr. Katy Börner & Elisha Hardy

The Power of Maps

Four Early Maps of Our World
VERSUS
Six Early Maps of Science

(1st Iteration of Places & Spaces Exhibit - 2005)
How would a map of science look?

What metaphors would work best?
The Structure of Science

1. The Social Sciences are the study of behavior and society. A key feature of the diagram is the interconnection between disciplines, reflecting the complex interactions within and between fields. The Social Sciences are shown as a network of nodes, illustrating the interdisciplinary nature of modern research.

2. The Life Sciences focus on the study of living organisms and their processes. This includes disciplines like Biology, Biochemistry, and Genetics, which are highlighted in the diagram. The Life Sciences are connected to other fields, emphasizing the importance of interdisciplinary approaches in modern research.

3. The Natural Sciences, including Mathematics, Physics, and Chemistry, form the foundational disciplines of science. These disciplines are shown in the center of the diagram, highlighting their role as the building blocks for more applied sciences.

4. The Engineering Sciences integrate knowledge from various fields to design and develop new technologies. The Engineering Sciences are depicted as interconnected nodes, reflecting the interdependence of engineering with other sciences.

5. The Humanities and Arts explore the human condition and cultural expression. These disciplines are represented as distinct nodes, yet they also connect to other fields, showing how they contribute to the broader understanding of society and culture.

6. The Technology and Applied Sciences focus on the practical application of scientific knowledge. These disciplines are shown as nodes that connect with various other fields, illustrating the impact of technology on society.

7. The History and Philosophy of Science are shown as nodes that connect with other disciplines, reflecting the importance of understanding the origins and development of scientific thought.

8. The Ethics and Policy of Science highlight the ethical and social implications of scientific research. These nodes are connected to other fields, emphasizing the need for ethical considerations in scientific practice.

9. The Future of Science and Technology is depicted as a node that connects with other disciplines, symbolizing the expanding horizons of scientific inquiry.

The diagram serves as a visual representation of the complexity and interconnectedness of scientific knowledge, illustrating how different fields are interrelated and how new ideas and discoveries emerge at the intersections of these disciplines.
The Power of Reference Systems

Four Existing Reference Systems VERSUS
Six Potential Reference Systems of Science

(2nd Iteration of Places & Spaces Exhibit - 2006)
How would a reference system for all of science look?

What dimensions would it have?
Writing the History of Science

Evolution - Wikipedia

Evolution is any process of change in a biological population. In a metaphorical sense, evolution is also used to refer to the process of change in any system, whether biological or not.
Science can be thought of as containing themes and paradigms. Themes are areas of current research, while paradigms comprise the dominant tools and existing knowledge that are used by today's researchers. This map shows 796 major paradigms in science, along with the dominant relationships between these paradigms. Paradigms are shown as circles, and strong relationships between paradigms are indicated by the lines connecting the circles. The map was created by recently clustering the 130,000 papers referenced most often in 2003. Clustering at each level was done using VOSviewer-a force-directed graph layout routine. These papers form 10,000 clusters, 4, 100 higher-level clusters, and finally the 796 paradigms. Although each paradigm contains, on average, 1, 000 papers, some are larger and some are smaller; as shown by the different sized circles on the map.

The ring-like structure that formed by scientific paradigms is very robust. We find similar structures for different years, and for maps generated from scientific journals. "The Structure of Science," a gallery map shown in the first edition of "Maps & Spaces," is a map based on clustering of scientific journals, with superposition of papers on the journal structure, whereas this map was generated directly from highly-cited papers. The "Structure of Science" shows current science in a complementary context, whereas this map can show the breadth of disciplines that constitute a single paradigm.

Because of the robust nature of the structure of science and its paradigms, we have placed our 796 scientific paradigms under a reference system containing 17 subject areas and 1 ring. This allows the possibility of each paradigm to be identified and available for lookup, for instance "Fluid Mechanics" paradigms are in ring 01.

We have also considered and displayed the vitality of each paradigm. Vitality is a measure of the speed at which a group of researchers moves, as measured by citations. Paradigms are currently being added, but it usually takes years to reach consensus about which improvements are major. The white circles represent communities where consensus is reached relatively slowly. This is a common phenomenon in the social sciences, sociological sciences, computer sciences, and mathematics. The white circles represent communities where consensus is reached relatively quickly. This is common in physics, chemistry, biology, and many medical disciplines. Very dark circles (such as those in mathematics) may represent communities where consensus is reached extremely quickly.

The map of scientific paradigms and its reference system can be used for multiple purposes. Counting, ranking, correlation, comparison, and individual researchers can all locate themselves within the map, either by topic or by a specific collection of paradigms. Various works, such as Not Vital, can be applied as references to this reference system to highlight specific aspects. Science education and personal discovery can also be fostered by using stories and facts to the map that highlight scientific history, current advances, and relationships between scientific paradigms.

**TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE**

**Sustainability**
- The science behind our long-term hopes

**BIology & Chemistry**
- The interactions between these two vital fields

**Francis H. C. CRICK**
- Co-discovered the DNA double helix

**Albert EINSTEIN**
- Co-discovered the quark

**Michael E. FISHER**
- Co-discovered the quark

**Joshua LEDERBERG**
- Pioneers in molecular genetics

**Derek J. de Solla PRICE**
- Known as the "Father of Scienceometrics"

**Richard N. ZARE**
- Pioneers in molecular chemistry

**About this display**
- Please note: the colors are chosen to help create a thematic effect.
Nanotechnology

This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes place in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

All Topics
Sweep through all 86 scientific paradigms.

Nanotechnology
Science at the tiny scale of molecules.

Francis H. C. CRICK
Co-discovered DNA's double helix.

Albert EINSTEIN
Quantum physics, unified field theory.

Michael E. FISHER
Modeling critical phenomena, phase transitions.

Susan T. FISKE
Converting perception into Penrose tiles.

Sustainability
The science behind our long-term hopes.

Biology & Chemistry
The interface between these two vital fields.

Joshua LEDERBERG
Pioneer in bacterial genetics.

Derek J. de Solla PRICE
Known as the "Father of Scientificometrics."

Richard N. ZARE
Laser mass spectrometry, chemical kinetics.

About this display
Peek through vortices that represent organizations that helped create it.

We sweep slowly through adjacent research topics, lighting up the places in the world that study each topic. You may select a spot of the maps that deals with those three intersecting subjects by touching it.
The Power of Forecasts

Four Existing Forecasts
VERSUS
Six Potential Science ‘Weather’ Forecasts

(3rd Iteration of Places & Spaces Exhibit - 2007)

Science Maps for Economic Decision Making

Four Existing Maps
VERSUS
Six Science Maps

(4th Iteration of Places & Spaces Exhibit - 2008)
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

**Research questions:**
1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research institutions?

**Contributions:**
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.
Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.


113 Years of Physical Review

http://scimaps.org/dev/map_detail.php?map_id=171

Bruce W. Herr II and Russell Duhon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)
Mapping Indiana’s Intellectual Space

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

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry


Science Related Wikipedian Activity

http://scimaps.org/dev/map_detail.php?map_id=165

Same base map.

Overlaid are 3,599 math (blue), 6,474 science (green), and 3,164 technology relevant articles (yellow).

All other articles are given in grey.

Corners show articles size coded according to
- article edit activity (top left),
- number of major edits (top right),
- number of bursts in edit activity (bottom, right)
- indegree (bottom left).
Interested to get your own science map?
Contact the map makers!

katy@indiana.edu
The End.