A DEEPER LOOK AT THE VISUALIZATION OF SCIENTIFIC DISCOVERY IN THE FEDERAL CONTEXT

Summary of a Workshop

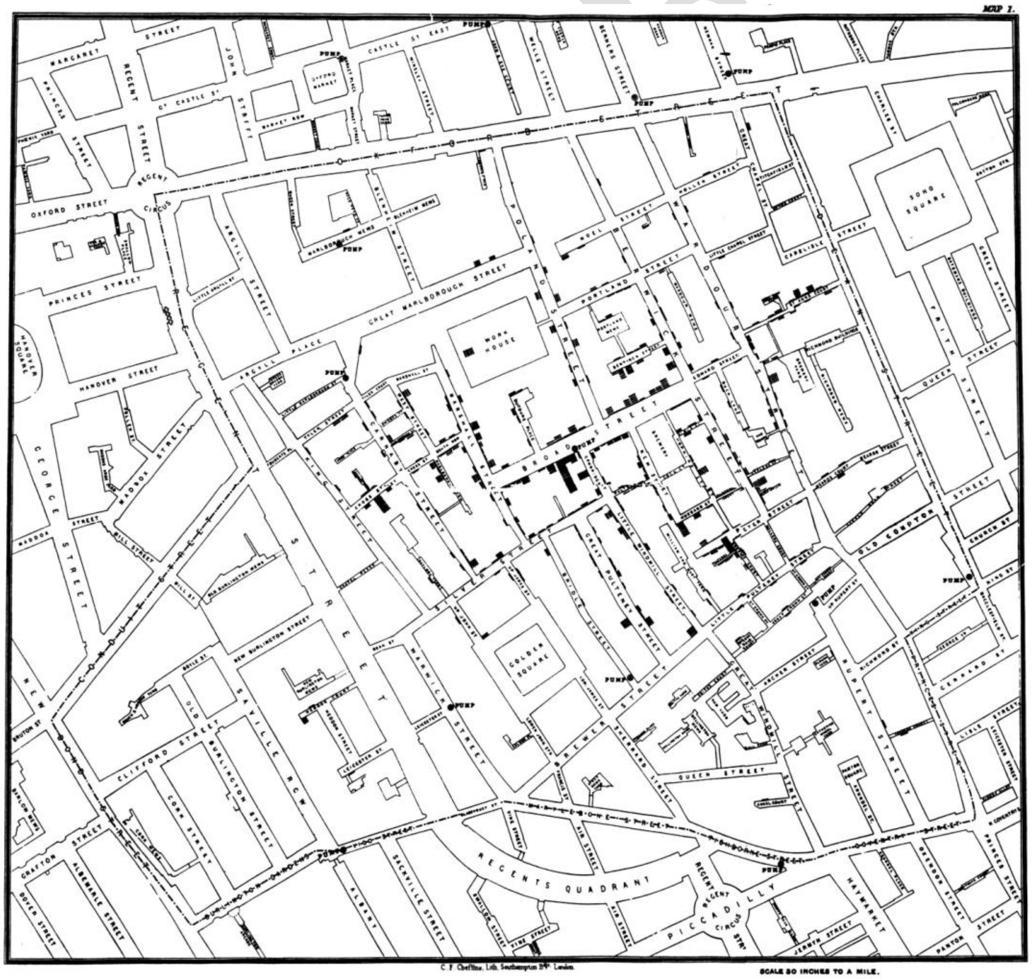
Sponsored by the NSF Program on Science of Science and Innovation Policy

In collaboration with the Divisions of Chemistry, Computing and Communication
Foundations, Industrial Innovation & Partnerships, Information and Intelligent Systems,
Science Resources Statistics and the Office of Cyberinfrastructure
as well as the Department of Energy Office of Science

Co-organizers: Susan Cozzens, Georgia Tech, and Julia Lane, NSF

National Science Foundation

September 11-12, 2008



SUMMARY AND RECOMMENDATIONS

The visualization of scientific discovery has reached an intriguing point of development. Researchers in the field are producing fascinating representations that are catching the attention of program officers and policymakers. The accomplishments are being driven by the availability of both large scale data sets and the computing power and algorithms to analyze them.

At the same time, skeptics are not sure what reality the visualizations represent and have concerns about the quality of the underlying data. Without an underlying statistical basis, it is hard to tell when a real change has occurred or when differences between teams, programs, or countries represent signal rather than noise. The almost exclusive reliance on publication, patent, and citation data is seen as too narrow and not addressing many of the crucial questions of causal relationships between federal policies and programs and outcomes for research, the economy, and society.

Nonetheless, the participants in the Workshop on Visualization of Scientific Discovery agreed that this set of tools should be developed further, especially through interdisciplinary interaction among research users, experts in visual analytics, and science map producers. Interaction with users and development in relation to their needs is crucial in this process. The workshop discussion focused on the following important steps in turning the promise of today's maps into the next generation of policy information.

1. **Data**: Federal agencies should work together to build standard data sets on federally funded projects, their outputs and outcomes. They should take the lead in making all data sets for science visualization collaborative and openly accessible.

2. **Models**. The research community should re-link visualization with models of the underlying dynamics of scientific research and technological development, as they originally were. The research community should continue to improve the models including in particular the causal relationships between federal funding and outcomes.

3. **Rigor**. The research community should develop tests of validity with regard to the models and ways to evaluate measures of relationships and changes statistically. The research community should articulate and publish quality standards and their consensus on established methods principles, while maintaining methods innovation and experimentation.

4. **Users**. The research community should develop a deeper and more systematic understanding of the variety of users for science visualization and their needs, and take this understanding into account in designing visualization tools.

AIM OF THE WORKSHOP

The development of visualization tools made possible by advances in cyber-infrastructure offer intriguing possibilities for tracking the impact of investments in science. These possibilities range from tracing the path from basic research discoveries to patents and innovation, to the changing structure of scientific disciplines, and from examining the importance of social networks to the dispersion of scientific innovations to comparators of international performance in science.

The potential contribution to federal research programs is far-reaching. Program officers could use information from such tools to examine whether one particular type or level of investment has been better than another for achieving a particular short-term outcome, use the information to restructure or balance their funding portfolios, as well as to provide information to outside queries about the value of particular investments. In addition, program staff could use the tools to describe the impact of cross-cutting initiatives, such as cyber-infrastructure and ITR. Agency staff might also be able to expand by mapping the complex structures of multi-disciplinary collaborations and using the information to identify members of scientifically dispersed communities as well as emerging leaders in science and innovation.

Federal agencies and the academic community have invested heavily in the development of mapping algorithms for a number of reasons. One important reason has been that these approaches permit the tracing of basic research innovations (such as physics and accelerator science) through the scientific literature and patents to do mini-case studies. Another important reason is that they potentially permit a cataloging of basic research into related disciplines, most notably an understanding of how disciplines interact and change. In a related fashion, these tools can potentially expand funding agencies' understanding of the previous impact of scientific inquiry as well as provide insights into the general direction of future scientific inquiry. Finally, they provide a straightforward way in which policymakers and other decision-makers can understand the dynamic interaction between funding and scientific advances. Multiple directorates within NSF have engaged in supporting research in this area.

However, before such a vision is achieved, many questions remain about the robustness, validity and usability of the visualization tools. The workshop was designed to bring researchers from a broad range of disciplines to examine the state of the art in science mapping, how it relates to program and policy needs, and what needs to be done next to translate potential into policy relevant information and analysis.

INTRODUCTION AND MOTIVATION

Julia Lane, Program Director for the Science of Science and Innovation Policy Program, opened the workshop with the image of John Snow's analysis of cholera cases in London in 1854 (see the cover of this summary). The image illustrates the use of visual representation to address a pressing problem; gathering data from existing sources, displaying the data on a recognizable grid (the street map), and demonstrating visually the clear causal connection with the Broad Street pump, which was producing contaminated water. Science visualization should aspire to produce such powerful results from such a simple tool.

Lane pointed to the origins of her program in a call by the President's Science Advisor, Dr. Jack Marburger, for a better factual base for science policy making. She raised three focal questions for the group:

- Is visualization truly scientific?
- Can visualization be useful in making scientific investment decisions?
- Can visualization be useful in explaining the impact of scientific investment decisions to policymakers?

Her hope was that the workshop would sharpen these questions and produce an agenda for further research in the area. The workshop was also designed to stimulate engagement of the federal science policy community with scientific community and identification of a research agenda for the area.

Bill Valdez, the co-chair of the interagency task force on the science of science policy, reported on the work of that group. In response to Dr. Marburger's challenge, the interagency group had produced a roadmap for a science of science and innovation policy. The three themes and ten questions that the group has formulated were undergoing interagency review at the time of the workshop. After their release, the task force will hold a gathering to plan implementation.

KEYNOTE: VISUAL ANALYTICS

Jim Thomas, Director of the National Visualization and Analytics Center at the Pacific Northwest Laboratory, reviewed the state of visual analytics and its uses in science policymaking[1]. Visual analytics is a new field of research that is developing "the science of analytical reasoning," focusing on how people interact with information to make decisions. A visual interface provides the broadest bandwidth between information and the mind. Visual data exploration provides

---

[1] See, for example, Illuminating the Path http://nvac.pnl.gov/agenda.stm

ways to make sense of the growing amounts of information available, 70% of which is from individuals.

Thomas illustrated tools that allow rapid exploration of the main themes and information in large data sets, displaying the results in representations that make them intuitively accessible to decision makers. His group has been developing applications in homeland security, but has also used the techniques to help PNL identify areas for potential collaboration with other institutions. He noted that science mapping so far has been limited in its data sources, largely to publications and citations. But many other sources of information were available now, some structured and some unstructured. Thomas made the point several times in the workshop that convergence of several types of data is a good way to address its ability to provide useful representations. His talk highlighted the necessity to *explore how users of science visualizations to absorb and use information.*

Discussion following the keynote raised a central question: Could a user of the visualizations tell a false one from a true one?

DATA: STATE OF THE ART AND VISION OF THE FUTURE

The first workshop session addressed questions about the data to be visualized. What are the different datasets available for research? What are the strengths and weaknesses of each dataset, in terms of: periods of availability; areas, fields and disciplines covered; manipulability and scalability; inclusivity of international sources? What could the future look like?

Caroline Wagner (SRI) stressed that a huge array of data is available, as Table One (drawn from Wagner's presentation) shows, so choices must be made. Those choices will be driven by the questions asked by the audience. She noted that policy audiences for visualization ask different questions from the ones scholars ask. For example, they want to know "What are we getting for the taxpayers' money?" or "Are scientists and engineers adding value to social welfare and the economy?" The primary challenge for science map-makers at this time is answering the question, "What is being visualized?"

Wagner noted that the kinds of data that are used to make maps are indicators of social relationships. If there is no understanding of the underlying social dynamics, the use of visualization does not advance metrics. However, there is the potential to understanding the social dynamics because scientific discovery is a complex communication system with certain well-established mathematical properties. In addition, communication creates social capital, one of the major concepts science metrics need to capture. She illustrated the emergence of social capital in DOE-supported nanoscience with a series of network maps. Wagner's recommendations were to

- *Propose and test theories of knowledge creation*

- *Establish a standardized data set*
- *Agree on analytic tools (or on what constitutes good practice)*
- *Move to dynamic measures and representation of scientific discovery*

Table One: Data Needs and Sources

From Caroline Wagner's presentation, "*Data Questions and Problems: State of the Art and Visions of the Future"*

Data Need	***Primary/secondary Source***
Scientific publications data (worldwide)	ISI; Scopus; Biomed; arXiv.org; Scirus; Pubmed; ScienceDirect; ulist; inasp
Scientific articles (full text)	INSPEC, Compendex
CVs	COS
Citations	ISI; Citeseer; Scitation
Patents	USPTO; LexisNexis; JPO; WIPO; EPO
Economic data 1	WDI (global); BEA, Economic Census
Science indicators 1	SEI/NSF; OECD; World Bank
Technology/engineering indicators	OECD, CORDIS, ITU
Social indicators	World Bank
Innovation surveys	OECD
Market data	BEA
Trade data	UNCTAD
Public attitudes towards science	NSB SEI
Employment, jobs, labor demand	BLS; ILO
Private sector R&D investment	IRI; SEC
Academic R&D expenditure & output	AUTM data
Joint ventures/organizational alliances	JVC, MERIT/CATI
Association membership	Association websites
Venture capital investment	VC Yearbook
Acknowledgements/informal relationships	ISI, Project data from laboratories
Entrepreneurship, establishments	PSED, ILBD, County Business Patterns
Infrastructure	World Bank; UNCTAD

Discussant Catherine Plaisant (U Maryland) reminded the group of the basic questions that need to be answered in the enterprise of visualizing scientific discovery: Who are the users? What tasks are they trying to accomplish? The answers to these questions drive the choice of data. Once we have a visual representation, how do we know whether it is right or wrong? It may appear right if we can find ourselves in it, or if it is telling us something new, but does this constitute validation? We may be able to test analyses with examples in which we know what the answer is and can explore whether the data and analytic techniques show it.

Jose Marie Griffiths (University of North Carolina) stressed that data need to be available, and urged that data sets be published so that they could be developed in open, collaborative ways. She urged investment in data management and curation. We need to define what data we need and not just analyze what we have. We need a framework and roadmap for data collection and management.

Bill Valdez (DOE) expressed the view that reliability of the data for science visualization was the number one issue facing science policy makers. He noted that we do not know the error bands around what is being used; we need indices of reliability, given that the data represent underlying social and behavioral dynamics. All science mapping results should be peer reviewed and there should be more involvement by practicing scientists for validation. The proliferation of data sets, such as online networking data, is both an opportunity and a challenge for this research area.

The discussion returned to several of the themes raised by the commentators. Visualization has so far been limited to publication and patent data, but what about the other outputs of research such as students? Another issue is the business model for data used in visualization. Several participants expressed the view that the data should be publicly accessible. There was also agreement that first the map-maker needs to know what he or she or the client wants to know; this is often more than just network information. Does mapping need to start from an understanding of the process and impact of scientific discovery or does it reveal that process? Can it answer the question of whether NSF or other agencies are fostering discovery?

Visualization experts urged the group to differentiate between data analysis and visualization itself. Visual biases may be built into the way analyses are presented, with elements like color and density.

The session ended with a plea from one agency to other agencies to make sure that the data on their activities were available, at a minimum:

- researchers supported,
- institutions supported,
- every paper attributed to the agency's support,
- a taxonomy of terms and language,
- students affiliated with their projects,
- infrastructure, and
- patents.

TOOLS: VALIDITY AND ROBUSTNESS

What are the different tools that are available? What tools are on the horizon? How robust are different taxonomies to different mapping algorithms? How robust are the apparent relationships to different distance metrics? The workshop's second session addressed these questions.

Kevin Boyack (SciTech Strategies, Inc.) pointed out that science map-makers use several different methods for calculating relatedness or similarity; grouping or clustering objects; layout; and calculating metrics. But these approaches have been sparsely evaluated. He has published several papers doing so (with Klavans and Borner) and the metrics have been compared to peer review, but the clustering and display algorithms have not been evaluated. We know little about the robustness of maps versus taxonomies (including which one if either should be the standard) and little about the relationships among metrics and between metrics and expert judgments. Indeed, there is little debate and no consensus on what validity actually means in this context.

However, there are a few things that practitioners in the field think they know, including the following. Normalized relatedness measures are preferred to raw measures. For journals, different normalized measures produce about the same results. At a very high level, different mapping methods produce similar overall structures for science. For mapping, citations are less ambiguous than text. Euclidean space produces "good enough" maps, but curved spaces are better. After these fairly well established items, however, a host of important issues remain, including how to compare maps both with each other and with expert judgments.

Johan Bollen (LANL) illustrated the issues with a new data source, "usage data," generated when users click through an information source. Maps produced from such data can be compared with standard taxonomies and show a rough equivalence. But what do the deviations mean? If maps become their own baselines, allowing for tracking of changes over time, then continuity of data collection and processing becomes crucial.

Discussant Di Cook (Iowa State) stressed the importance of making all data and algorithms used in the field public. Open source treatment would allow users to examine structure and make changes. Loet Leydesdorff (University of Amsterdam) pointed out that the interesting dynamics in the fuzzy areas of maps are treated as noise by most of the mapping algorithms, as illustrated by the data on nanotechnology. Although there is consensus on the macro-structure of the network of papers, the change is in the details and does not appear there. He urged the development of theory rather than just metrics. John Stasko (Georgia Institute of Technology) recommended moving beyond large graphs to self-organizing systems. Researchers, he said, should examine interactive systems and activities that provide new insights rather than limiting themselves to confirmatory analyses. Larry Rosenblum (NSF) noted that more work was needed on testing the validity of the visualizations; using visual communication itself has an effect. He called for more inter-disciplinarity in the analysis, for example, in teams involving both computer graphics and information science specialists.

The discussion echoed the importance of using visualization as a tool of exploration. It is hard to obtain reliability and validity when you are exploring, because you do not know what you want ahead of time. This is a challenge when federal agencies want tools not just for insight into the structure of their portfolios but also to demonstrate impact for external audiences. In the latter case, validity is essential and causality must be demonstrated. Visualization that is limited to outputs is limited in performing these tasks. The science agencies share some of the methodological challenges with other agencies, such as the Department of Homeland Security.

APPLICATIONS IN RESEARCH

The next session asked: What statistical models can be applied to visualization algorithms to validate relationships and predictability of how they are likely to evolve? How replicable and generalizable are the results? Are the data readily available and is there consensus about the approach?

Paul Gemperline (E Carolina) reported that in his field, chemometrics, visualizations have been used for quantitative comparisons and evaluations as well as in exploratory visualization for new discoveries and insights. The methods are evaluated statistically, including through meta-analysis. Chemists use visualization to present a visual comparison of properties or states in two or more systems and to present visual prediction of properties or states in the future.

Methods that can be validated include the following classic characteristics:

- Good experimental designs
- Testable hypotheses
- Valid sampling strategies
    - Representative samples, no degradation, segregation, outliers, etc.
- Reliable measurement methodologies
    - Measurements relevant to the properties of interest being studied
    - selective, sensitive precise, robust
- A sufficiently large sample to define the relationship between the measurement domain and property or state domain
    - Exemplars with properties that cover the range interest
    - Exemplars of all sources of variability expected in comparisons or future predictions

- Appropriate models
    - Statistically sufficient
    - Parsimonious
    - Rugged
- Head to head comparisons:
    - A sufficiently large validation sample to statistically test the relationship between the measurement domain and property domain.
        - Exemplars with properties that cover the range interest
        - Exemplars of all sources of variability expected in comparisons or predictions

However, exploratory methods need to have different characteristics. They should lead users to comment, “That’s funny…” There is much less agreement for exploratory methods on what counts as valid. Time to complete benchmark tasks is often used, but also often criticized. Some studies have shown time gains of 10 fold to 100 fold, but there were no accompanying measures of decision quality. Meta analysis is relatively new in this area and has not yet established itself. The theme of the session remains the million-dollar question, in Gemperline’s view: How does one quantitatively link visual perception with statistical significance?

Discussant Jane Fountain (University of Massachusetts at Amherst) highlighted the need to have construct validity, in cases where constructs are used to approximate reality. Other problems may arise that reduce validity such as citation bias, or the use of experts with insufficient knowledge of the data for validation. She also pointed to issues that can arise when pooling large datasets such as changes in variable definitions or sampling frame. In many cases it is not possible to pool many cross-country data-sets because of differences in the way the data is gathered. Science mapping is not alone in its challenges, she pointed out: Social network methods were also weak in the development of theory in their early years.

Peter Gloor (MIT) introduced an alternative method of measuring the success of knowledge workers. This involved the use of a physical device to track interactions among workers and subsequent analyses of the interactions.

Cheryl Eavey (NSF) raised some key questions for the field to answer. How do we communicate the value of science? How does the public investment in science affect the lives of U.S. citizens? A key feature for science visualization is to be able to address the causal relationships embedded in these questions.

USER APPLICATIONS AT PROGRAM LEVEL

The audience at the workshop included a number of program officers, from NSF and other federal agencies, who were interested in knowing the value of science visualization in their work. The next session addressed issues in this domain. What are the outcome measures from mapping, and how can they be used in scientific and policy analysis – particularly in the federal context? For example, can the results be included in an econometric model that analyses the outcomes of different types of investments? What types of hypotheses can be tested and how can the results be used in policy/programmatic work? How can the results be used to inform portfolio allocations?

Alan Porter (Georgia Institute of Technology) illustrated the use of science mapping for strategic intelligence by drawing on a number of examples from emerging fields. In one example, the National Academies asked for maps to help them identify who to invite for a workshop on synthetic biology. The maps helped the workshop planners identify leading institutions, leading researchers, how recent the publications were, publication impact, and topical emphases. The technique mapped a particular subset of literature against base maps of the disciplinary structure of science constructed using Web of Science data. A second example used the techniques to display the disciplinary connections in nanoscience at various points in time. A third example, kinesin research, illustrates the use of network measures to track the coherence of an emerging cluster over time and the emergence of an area of knowledge integration.

In summary, Porter said that mapping techniques could help program managers by applying at multiple levels,

- "Micro" - Characterize the research reflected in one paper, or
- "Macro" - Characterize entire research fields.

Mapping could be useful in tracking research domains by

- Benchmarking – e.g., scale of effort; US position
- Pointing out "hot" areas of special opportunity (to fund?)

It could help in research knowledge transfer, by demonstrating "seed" areas fueling other research areas. And it could help in program evaluation, by identifying loci of interdisciplinary research integration and documenting spreading impact across research fields.

Katy Börner offered her vision of how science mapping could contribute to program management. A science of science management should show reliability, simplicity, good measurement, and consistent results, and stimulate further discovery. Following a needs assessment, this science of science policy should identify the basic units of science and the concepts needed to describe it; examine data quality, coverage, and inter-linkages; identify tools for analysis; and evaluate the tools through case studies. She illustrated the display of agency funding profiles against a base map of science, much like Porter's illustrations of publication profiles displayed in terms of the base map. She stressed the greater access to scientific knowledge and expertise that mapping offers.

Discussant Arthur Ellis (University of California at San Diego) commented on possible strategic uses of science mapping in the university context. One use is in putting teams together, answering the questions, "Who else around the world is doing work like this? Are we hiring the best people in the area?" He also wanted to be able to use visualization to evaluate innovations. His university is doing a patent disclosure a day and he can't file patents on all of them. Which are most likely to pay off? Finally, he noted the democratization of information of this sort, the new assumption that everyone should have access to all the analysis.

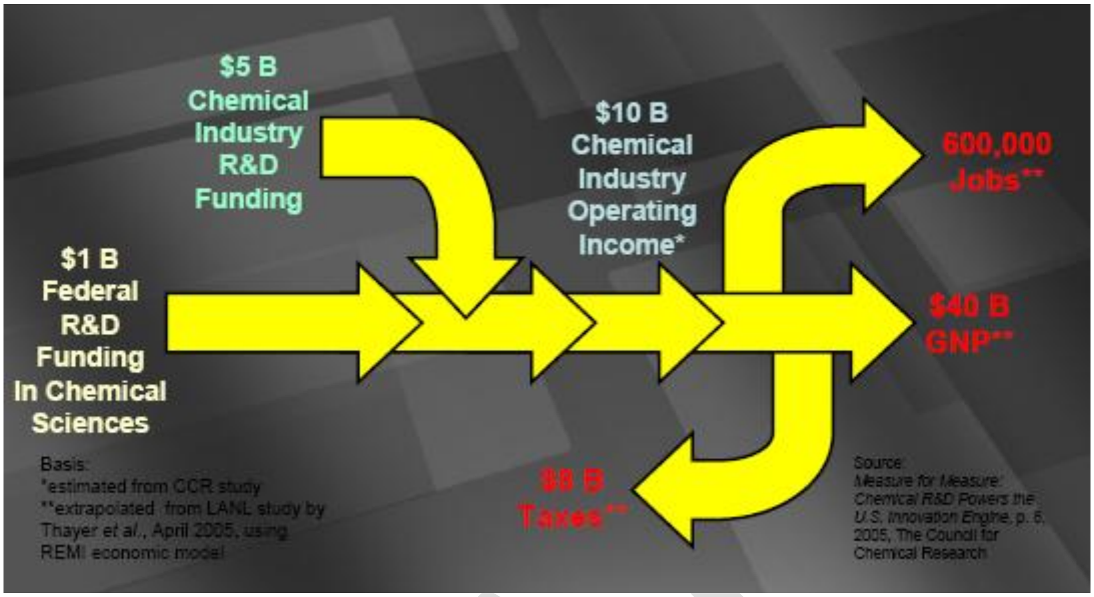
Janice Hicks (NSF) illustrated the use of the maps to inform discussions on restructuring the Chemistry Division at NSF and to streamline co-review of proposals that fall at the intersection with other programs. They were using science maps to help answer a number of key questions:

- What structure is best to catch the forefront proposals in a given area? (and how do we keep this updated?)
- What are the major nodes WITHIN a discipline, what work is outside these? How does this evolve over time?
- Which areas are US strengths, which are weaknesses, does this matter for the country (well-being, sustainability, security, economy, "competitiveness," jobs, etc.)?
- What partners do we need to involve, in what priority – other disciplines, other agencies, other countries?

The most powerful graphic they had used, however, was a pipeline diagram from the Council on Chemical Research. The yellow arrows and simplified impact measures seem to convey their message quickly and effectively.

Figure One: Impacts of Chemical Research

From Janice Hicks, discussant comments; figure drawn from Council on Chemical Research, Measure for Measure



Basis:
*estimated from CCR study
**extrapolated from LANL study by Thayer *et al.*, April 2005, using REMI economic model

Source:
Measure for Measure: Chemical R&D Powers the U.S. Innovation Engine, p. 6, 2005, The Council for Chemical Research

In addition, the CCR graphic addresses the key questions for the Chemistry Division, those that concern impacts.

- How can we link grants to output? (publications to patents to value)
- What level of investment maximizes return on investment?
- What can funding agencies do to maximize the handoff of our research?
- How can we assess the broader impacts of the research (often difficult to measure)?

Brian Zuckerman (Science and Technology Policy Institute) attempted to link the visualizations to policy questions. He noted that while some measures of outcome had been discussed thus far in the workshop (collaboration/inter-/multi-disciplinarity or changes in research trajectories; creation of new scientific fields/disciplines; simple input-output comparisons), others had not, although they could be visualized using similar techniques (traditional bibliometrics: publication/patent “quality”; training/career development; usage of physical infrastructure; leveraging additional funds). Other measures might require different or still-to-be developed visualizations, such as socioeconomic impacts; tracing translation/consequences of discovery and “high-risk,” ”transformative,” or ”innovative” research, all terms that still need definition. These tasks might require social science to move from descriptive to normative/evaluative judgments.

The presentations at the workshop, Zuckerman pointed out, had described portfolios using categorization tied to the structure of science; compared portfolios across funding organizations; and identified “strengths” and

”weaknesses” or organizational niches. These displays show the importance of commonality and comparability of portfolio data as well as the proliferation of locally-valid classification approaches. The visualization techniques might also contribute in the long run to tasks now undertaken through the peer review process, such as collaboration and inter-disciplinarity of the PI and team (especially for centers competitions); classification of incoming proposals; and “near misses” or emerging topics. The usefulness of science mapping for these tasks depends on their acceptance as decision-support tool by the scientific community.

The discussion ranged widely over the usefulness of the kinds of visualizations presented at the workshop for program management. Clearly, such data would be only one of many inputs to decisions such as restructuring in the Chemistry Division. Data on funding itself, from project information systems, could be quite useful, but past experience had shown that it was expensive to standardize across agencies. The possibility of visualizing other kinds of data, such as career trajectories, was particularly appealing.

TAXONOMIES; INTERNATIONAL COMPARISONS; POLICY APPLICATIONS

Finally the workshop turned to applications of science visualization with regard to a set of particularly strategic policy questions. How are interdisciplinary or early-stage relationships characterized? Can the validated mapping tool yield an international standard for the taxonomy of science? Can meaningful comparisons be made across countries or across agencies?

J. David Roessner (SRI) reviewed the concept of interdisciplinary research (IDR). He argued that the integration of knowledge from different fields of research is the central concept of interdisciplinarity. Interdisciplinarity is therefore a cognitive concept, not a social one. If this is so, then valid measures of IDR should not be based on team membership, organizational affiliation, or collaboration. Rather, they should be based on empirical evidence of knowledge integration during the research process and/or in the output of the research. The Keck Futures Initiative at the National Academies had developed a definition of IDR:

> “Interdisciplinary research (IDR) is a mode of research by teams or individuals that *integrates* perspectives/concepts/theories and/or tools/techniques and/or information/data, from two or more bodies of specialized knowledge or research practice. Its purpose is to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single field of research practice.”

Examples of *bodies of specialized knowledge or research practice* include: low temperature physics, molecular biology, developmental psychology, toxicology, operations research, and fluid mechanics.

Using Stirling's concept, diversity is a function of three necessary but individually insufficient properties:

- Variety (how many different types of a thing do we have?)
- Balance (how much of each type of thing do we have?)
- Disparity (how different from each other are the types of things we have?)

Roessner and Porter have operationalized Stirling's diversity measure for publications and applied it to analyze various areas of science, as Porter had already illustrated in the previous session of the workshop. The measures could be displayed as indexes or as maps.

In conclusion, Roessner claimed that to be useful to researchers, research managers, and policymakers, measures or indices of research interdisciplinarity must meet several daunting challenges:

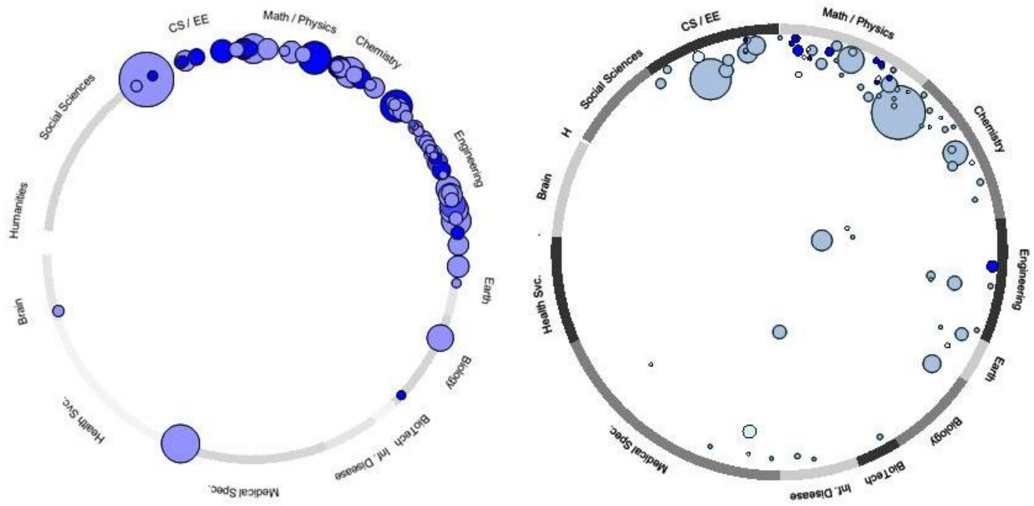
- they should be applicable to the outputs of individuals, groups, organizations, and fields of knowledge;
- they should reflect the cognitive integration of knowledge from established fields of inquiry; and
- they should reflect the diversity of the fields of inquiry drawn upon.

Richard Klavans (SciTech Strategies, Inc.) reviewed the state of international comparisons using visualization tools. There is currently no international standard for a science taxonomy, but Klavans believes that with validation, current representations can offer such a standard.

His proposed method uses bottom-up classification to identify 554 disciplines using journal inter-citation analysis. The core competence of nations or other organizations can then be measured with regard to each of these disciplines using reference co-citation analysis. His presentation illustrates the technique using data from the University of California at San Diego, Art Ellis's institution. Each faculty member and department is located from publication activity, and leadership areas are identified.

Figure Two: China's Strengths under the Old and New Methods

From Richard Klavans, *"Taxonomies; International Comparisons & Policy Applications"*



Discussant Bill Ribarsky (UNC Charlotte) reminded the group that visualization is used to provide insight on complex issues using complex data. He suggested combining visualization with other information (interviews, contexts, experts) to discover meaningful paths through knowledge and to gain insights. Interaction among data types is central. Visualization requires some subjective judgments to filter, reorganize and otherwise manipulate the data. Technology experts and the visualization community need to work together on these tasks, he noted.

Diana Hicks made another call for government supported infrastructure to produce structured data from the raw form in databases to that is cross-linked and suitable for analysis. If possible, could combine patent, publication, agency and usage information. The data management does not need to be in a university setting, but the data should be made accessible to many different users.

Gindo Tampubolon (University of Mancheter) pointed out that limitations exist at different levels, for example we cannot justify all research investments in economic terms; program managers have to deal with uncertainty or incomplete information availability that; some expressed the view that they cannot validate everything. Measures also have limitations; for instance, the diversity index may have an optimal level and negative implications should be included in the measure. Visualization must at some point be able to answer the question, "So what?"

Jeri Mulrow (NSF) was intrigued by the maps, including the patterns jumping out from them. But she noted that program managers continue to express concerns about the mapping process and implications for decision-making when output is based on incomplete data. Differences exist in how the data are represented, including data sources and types of measures. There are also challenges in collecting it: cost of data collection; unavailability or reluctance of individuals to provide data; timeliness and using the data to project to the future; data handling (tagging, coding etc.). Other challenges include resistance to change; validation of techniques/ data/ methods; and combining mapping with information from other sources. In the end, it is hard to demonstrate that what we are seeing in the maps is really happening.

A representative from Thomson ISI reported that they plan to address issues of database access and are trying to respond to user needs. They hope to integrate proceedings with existing WOS and add more journals including Chinese journals.

CONCLUSIONS

The organizers provided some closing comments. The workshop identified three potential user groups in relation to scientific mapping, who require different maps. These were

- the science visualization experts, who require detailed, interactive maps that allow users to drill down to greater levels of detail to answer interesting research questions, build theory, or improve methodological approaches;
- the NSF program managers who require maps with greater focus on disciplinary or research areas of interest and which maybe overlaid with other information such as funding allocations in order to evaluate program performance and make decisions on funding allocations; and
- policy makers require easily understood or intuitive maps or visual representation that clarify the ideas being conveyed.

The workshop had demonstrated, in the view of one organizer, that the data problems were solvable. The DOE had taken the initiative to identify researchers, papers produced and the level of attribution of knowledge created to DOE funding. The agency also invested in databases from different publishers. Other federal agencies could follow this example. It was not clear, however, if this was the right data.

Visual analytics and the complex mapping procedures are ready for use by program officers. However, these are not the tools that senior policy makers would use because they are too complex and do not provide the insights they need. Senior policy makers require tools that provide scenarios that are data rich and contribute to decision-making. Policy makers need to have the consensus; and not deal with the debates about methodology, terminology, etc. Visual analytics is also a path for creating the next generation of tools for policymakers, tools that would rest on modeling of system dynamics.

Dialogue was thought to be necessary between different groups involved in visualization, that is, the software specialists, science visualization experts, and the program managers. It would be best to keep the esoteric concepts internal to the discussion in the research community and develop a consensus on tools, so that researchers can maintain a united front on the usefulness and validity of the methods.

In general science visualization has the potential to be useful in decision-making. It provides retrospective, quantitative information. Different types of maps and information, e.g., on funding can be overlaid to increase utility. Statistical tools can contribute to validation, and some output is considered to be robust. However, more research is needed in the area to improve the quality, reproducibility, and usefulness of output. Major areas of concern relate to costly and time consuming data preparation, which can be improved through the establishment of a central data repository, where researchers can pool data collection of efforts. Precedents in this approach are federal government support in clinical biology and physics. Further, standards for/ or standard algorithms would help to improve reproducibility of mapping exercises. Researchers involved in different areas of science mapping (theory, building, computer graphics, decision makers) should also work more closely together.

WORKSHOP AGENDA

Thursday, September 11

10 am **Coffee and mapping exhibits**

11 am **Introduction and motivation/ overview**

Defining the research question from the NSF perspective (program impact, program description, new approach to understanding complex initiatives and new, complex, large and dispersed scientific communities). Unit of analysis (individual, team, community, disciplines, institutions, systems.) How do visualization measures link to NSF's research questions?

11:30 **Keynote Address: Jim Thomas, Pacific Northwest Laboratories**

12:30 **Mapping exhibits (box lunches)**

1:00 **Data: state of the art and visions of the future**

What are the different datasets available for research? What are the strengths and weaknesses of each dataset, in terms of: periods of availability; areas, fields and disciplines covered; manipulability and aggregability; inclusivity of international sources? What could the future look like?

Caroline Wagner (GWU and SRI)

Commentary from: Catherine Plaisant (U Md), Jose-Marie Griffiths (UNC), Bill Valdez (DOE)

2:30 Break

2:45 **Tools: validity and robustness**

What are the different tools that are available? What tools are on the horizon? How robust are different taxonomies to different mapping algorithms? How robust are the apparent relationships to different distance metrics?

Kevin Boyack (SciTech Strategies, Inc.), Johan Bollen (LANL)

Commentary from Di Cook (Iowa State), Loet Leydesdorff (Amsterdam), John Stasko (Georgia Tech), Larry Rosenblum (NSF)

4:15 Break

4:30 **Applications in research**

What statistical models can be applied to visualization algorithms to validate relationships and predictability of how they are likely to evolve?

How replicable and generalizable are the results? Are the data readily available and is there consensus about the approach?

Presenters: Paul Gemperline (E Carolina)

Commentary from: Jane Fountain (U Mass Amherst), Peter Gloor (MIT), Cheryl Eavey (NSF)

6:00 Adjourn

Friday, September 12

8:30 **User applications at program level**

What are the outcome measures from mapping, and how can they be used in scientific and policy analysis – particularly in the NSF context? For example, can the results be included in an econometric model that analyses the outcomes of different types of investments? What types of hypotheses can be tested and how can the results be used in policy/programmatic work? How can the results be used to inform portfolio allocations?

Prospective strategic intelligence: Alan Porter (Georgia Tech)

Retrospective information on results: Katy Borner (Indiana)

Commentary from: Arthur Ellis (UCSD), Brian Zuckerman (STPI), Janice Hicks (NSF-CHEM)

10:15 Break

10:30 **Taxonomies; international comparisons; policy applications**

How are interdisciplinary or early-stage relationships characterized? Can the validated mapping tool yield an international standard for the taxonomy of science? Can meaningful comparisons be made across countries or across agencies?

Interdisciplinarity: J. David Roessner (SRI)

Changing structure of sciences: Dick Klavans (SciTech Strategies, Inc.)

Commentary from: Bill Ribarsky (UNC Charlotte), Diana Hicks (Georgia Tech), Gindo Tampubolon (Manchester), Jeri Mulrow (NSF)

12:00 Concluding remarks

1:00 Adjourn