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

A WORKSHOP SPONSORED BY THE NSF SCIENCE OF SCIENCE AND INNOVATION POLICY PROGRAM (SCISIP)

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

INFO.

A Workshop at the National Science Foundation,
4201 Wilson Boulevard, Arlington VA 22230
Stafford I Room 375
September 11-12, 2008

PAPER

Visualization of Scientific Discovery:
Conceptual and Methodological Background

OVER VIEW

The development of visualization tools made possible by advances in cyber-infrastructure offers intriguing possibilities for tracking the impact of investments in science. These possibilities include tracing the impact of basic research on innovation, examining the changing structure of scientific disciplines, studying the role of social networks in the dispersion of scientific innovations as well as making comparisons of international performance in science.

The potential contribution to federal science 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 in achieving a particular short-term outcome. They could use the information to restructure or balance their funding portfolio, or to provide information to outside queries about the value of particular investments. In addition, federal science program managers could use the tools to describe the impact of cross-cutting initiatives, such as cyber-infrastructure and ITR. Program managers might also be able to map the complex structures of multi-disciplinary collaborations and use the information to identify not only members of scientifically dispersed communities but also emerging leaders in science and innovation.

However, before such a vision is achieved, many questions remain about the robustness, validity and usability of the visualization tools. The workshop is designed to bring researchers from a broad range of disciplines to examine these key questions, and engage the federal science community in a discussion about whether and how the tools can be used in the federal context.

DAY1

Thursday, September 11th

10:00   **Coffee and mapping exhibits**

11:00   **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: Jose-Marie Griffiths (UNC), Catherine Plaisant (U Md), Bill Valdez (DOE)

2:30   Break

2:40   **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

DAY2

Friday, September 12th

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 ?˘ěC 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); Connie Della-Piana (NSF)

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?

Interdisciplinary: 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)

Tech), Silue Tampabolon (Manchester), Jeff Mulrow (NSF)

| 12:00 | **Concluding remarks** |
| 1:00 | Adjourn |