

**National Science Board**  
**EXPERT PANEL DISCUSSION ON DATA POLICIES**  
**March 28-29, 2011**  
**Arlington, Virginia**

**PROVISIONAL AGENDA**  
*(A final agenda will be distributed on March 28)*

---

**Monday, March 28**  
**The NSB Board Room**  
**National Science Foundation, Room 1235**

**8:00**      **Welcome, Board Processes, and Participant Introductions**

Welcome from **Dr. Ray M. Bowen**, Chairman of the National Science Board

Welcome from **Dr. Diane L. Souvaine**, Chairman of the Committee on Strategy and Budget, NSB

Participant Introductions, Workshop Overview and Goals, **Dr. José-Marie Griffiths**, Chairman, Task Force on Data Policies, Committee on Strategy and Budget, National Science Board

**8:20 – 10:00**    **Session I: The Vision of Data-Intensive Science**

Guiding questions: What are some of the defining characteristics of data-intensive science? What are the goals for enabling re-use and re-purposing of data? What new opportunities and new types of science have yet to be realized? These questions build upon the vision for a new NSF-wide program in computational and data-intensive science.

Moderator: **Dr. Diane L. Souvaine**

Panelists:

- **Roberta Balstad**, Columbia University
- **Francine Berman**, Rensselaer Polytechnic Institute
- **Michael Lesk**, Rutgers, The State University of New Jersey
- **Maryann Martone**, University of California San Diego

**10:00 – 10:15**    **Break**

**10:15 – 12:00**    **Session II: Reproducibility, First Steps and Guiding Principles**

Guiding questions: Reproducibility starts to scope the problem and drives all sorts of related issues (curation, cost, etc.). What does this mean for types of discovery that need data sharing (e.g., medical research, such as work on Alzheimer's disease)? What are the implications for data publishing and data citation? What are the implications for simulation and software? What constitutes the release of "complete" data? Would complete data release include

the original, "raw" data; cleaned-up, publication-ready data, along with the methods for clean-up; publication-ready data with the meta-data necessary to reproduce any interpretations of the data; raw data with software to make it usable to others; data organized in a way that is interoperable to some standard; etc.?

Moderator: **Mr. Arthur K. Reilly**

Panelists:

- **Timo Hannay**, Digital Science
- **Brooks Hanson**, Science Magazine
- **Randall LeVeque**, University of Washington
- **Victoria Stodden**, Columbia University

**12:00 – 1:00 Lunch**

**12:30 Lunch Presentation: *High Performance Cyberinfrastructure is Needed to Enable Data-Intensive Science and Engineering***

**Dr. Larry Smarr**, Harry E. Gruber Professor, Department of Computer Science and Engineering, University Of California, San Diego Jacobs School of Engineering; and Director, California Institute for Telecommunications and Information Technology

**1:00 – 3:00 Session III: Exemplars, Lessons Learned**

Guiding questions: What has been your experience? What types of incentives can be created? How has data publication impacted innovation? Examples include the Virtual Observatory, Interuniversity Consortium for Political and Social Research, Protein Data Bank, etc.

Moderator: **Dr. Camilla P. Benbow**

Panelists:

- **George Alter**, Interuniversity Consortium for Political and Social Research
- **David Lynn**, Wellcome Trust
- **Reagan Moore**, University of North Carolina at Chapel Hill
- **Alex Szalay**, The Johns Hopkins University

**3:00 – 3:15 Break**

**3:15 – 5:30 Session IV: Impacts**

Guiding questions: What are the measurable impacts? What is the early experience with the NSF-wide requirement for Data Management Plans? What are the impacts on research universities? What are the international complexities, particularly for large facilities with international partnerships? What are the legal complexities? What is the potential for overlap of policy when comparing the curatorship of physical specimens and the management of large, and often digital, datasets?

Moderator: **Dr. Mark R. Abbott**

Panelists:

- **Ravi Bellamkonda**, Georgia Tech
- **Tony Hey**, Microsoft Research
- **Michael Huerta**, National Institute of Mental Health
- **Michael Mabe**, International Association of Science, Technical & Medical Publishers

**5:30 End of Monday Meeting**

---

**Tuesday, March 29**  
**The NSB Board Room**  
**National Science Foundation, Room 1235**

**8:30 National Science Foundation Perspective**

Remarks from officials from the National Science Foundation

**8:45 – 10:30 Session V: Policy Issues**

Guiding questions: Frame the issues for institutions, government agencies, publishers and any other stakeholders. What are the relative merits of various types of repositories for data? How should the various repositories be funded? To what extent should NSF assist in development and adoption of standards for such efforts? To what extent should deposit in repositories be required of awardees?

Moderator: **Dr. José-Marie Griffiths**

Panelists:

- **Daniel Atkins**, University of Michigan
- **Mike Keller**, Stanford University
- **Celeste Rohlfig**, White House Office of Science and Technology Policy
- **Ann Wolpert**, Massachusetts Institute of Technology

**10:30 – 10:45 Break**

**10:45 – 11:00 Public Comment Period**

Dr. José-Marie Griffiths will take a few comments and questions from the audience present at the workshop

**11:00 – 12:30 Session V: Policy Issues (continued)**

Discussion by Task Force Members and Stakeholders

**12:30 Adjourn**

## **Confirmed Participants**

**George Alter**, Interuniversity Consortium for Political and Social Research  
**Daniel Atkins**, University of Michigan  
**Roberta Balstad**, Columbia University  
**Ravi Bellamkonda**, Georgia Tech  
**Francine Berman**, Rensselaer Polytechnic Institute  
**Adam Bly**, Seed Media Group  
**Steve Breckler**, American Psychological Association  
**Joe Bredekamp**, NASA  
**Steve Goff**, iPlant Collaborative  
**Chris Greer**, White House Office of Science and Technology Policy  
**Timo Hannay**, Digital Science  
**Brooks Hanson**, Science Magazine  
**Fred Heath**, The University of Texas at Austin  
**Tony Hey**, Microsoft Research  
**Michael Huerta**, National Institute of Mental Health  
**Mike Keller**, Stanford University  
**Michael Lesk**, Rutgers, The State University of New Jersey  
**Randall LeVeque**, University of Washington  
**David Lynn**, Wellcome Trust  
**Michael Mabe**, International Association of Science, Technical & Medical Publishers  
**Maryann Martone**, University of California San Diego  
**Kevin Marvel**, American Astronomical Society  
**Reagan Moore**, University of North Carolina at Chapel Hill  
**Sethuraman Panchanathan**, Arizona State University  
**Celeste Rohlfig**, White House Office of Science and Technology Policy  
**Bernard Schutz**, Max Planck Society  
**Larry Smarr**, University Of California, San Diego  
**Victoria Stodden**, Columbia University  
**Alex Szalay**, The Johns Hopkins University  
**Crispin Taylor**, American Society of Plant Biologists  
**John Vaughn**, Association of American Universities  
**John Wilbanks**, Creative Commons  
**Ann Wolpert**, Massachusetts Institute of Technology

**Mark Abbott**, National Science Board  
**Camilla Benbow**, National Science Board  
**Ray Bowen**, Chairman, National Science Board  
**José-Marie Griffiths**, National Science Board, Chairman of the Task Force on Data Policies  
**Louis Lanzerotti**, National Science Board  
**Art Reilly**, National Science Board  
**Diane Souvaine**, National Science Board

**Alan Blatecky**, Office of Cyberinfrastructure, NSF  
**Myron Gutmann**, Directorate for Social, Behavioral and Economic Sciences, NSF  
**Cora Marrett**, Office of the Director, National Science Foundation  
**Edward Seidel**, Directorate for Mathematical and Physical Sciences, NSF