Access to Scientific Data: The Social and Technical Challenges and Strategies

Sponsored by SIG/STI

John D'Ignazio (Organizer)

Syracuse University Syracuse, NY 13244 jadignaz@syr.edu

Jian Qin (Organizer)

Syracuse University Syracuse, NY 13244 jqin@syr.edu

Yale M. Braunstein

University of California Berkeley [University of California Berkeley](#)

Caroline Whitbeck

Case Western Reserve University [Case Western Reserve University](#)

Mark Parsons

University of Colorado [University of Colorado](#)

Timothy E. Eastman

NASA Goddard Space Flight Center [NASA Goddard Space Flight Center](#)

Sherry Xie

University of British Columbia [University of British Columbia](#)

The practice of science has changed in the last three decades due to the rapid development of information and communication technologies and massive increases in computing capacity, made manifest by the Internet. As the International Council for Science (ICSU) describes in its recently released five-year strategic plan, there is now more scientific data and information that is freely and openly available. This environment enables scientists around the world access to the most up-to-date data and information from his or her desktop. "Secondary analyses of data, and the combining of data from multiple sources, are opening up exciting new scientific horizons. Scientific publication practices are changing rapidly." (ICSU, 2005, 16-17) These revolutionary changes in the creation, management, and use of scientific data and information have significant economic and social implications. First among them are the economic and legal aspects provoked by open sharing versus intellectual property protection of scientific data. In addition to the impact of these issues, there are technical challenges

in managing the life cycle of scientific data. Long-term preservation strategies are evolving to ensure that the authenticity of scientific data can be verified, and to enable knowledge discovery and interoperability via metadata representations of the data collections.

To maximize the impact of scientific data, the information community needs to promote new thinking and structures in society to properly collect, preserve and distribute this resource. In response to the issues and challenges in access to scientific data and information, we have arranged for a comprehensive session that examines the topic from a holistic view. For full coverage, two panels are required: one that covers the current social and policy contexts and one that covers the system developments being driven by these broader issues. The experts on both panels will contribute their experience from conducting social, economic, and technical research of scientific data and information and invite the ASIS&T annual meeting attendees to join them in discussion.

Panel One

presents broader issues of access to scientific data from a social and policy perspective, including intellectual property concerns, ethical considerations in collaborative work, and supporting effective communication and sharing in a distributed scientific community. The speakers' topics are:

Economic Impact of Intellectual Property Protection of Non-Original Databases

Yale Braunstein is a professor at the School of Information, UC Berkeley. Most recently he served as an expert for the World Intellectual Property Organization in the area of the economic impact of intellectual property rights protection for databases. Braunstein will discuss the emerging intellectual property structures that apply to scientific data; as with scientific publishing of manuscripts, battle lines are forming over the need to protect an organization's investment in creating data versus the benefits of open sharing and communication.

Ethics of Sharing Scientific and Technological Data

As director of the Online Ethics Center for Engineering and Science located at Case Western Reserve University, Caroline Whitbeck provides engineers, scientists, and science and engineering students with resources useful for understanding and addressing ethically significant problems that arise in their work life. Her strategies include establishing or resolving the ethics of data sharing in the course of collaborative research at all levels of conduct: resolving data integrity, interpreting results, resolving conflicts of interest, and onward through authorship.

Designating User Communities for Scientific Data

Through his work at the National Snow and Ice Data Center based at the University of Colorado at Boulder, Mark Parsons has formulated best practices that support the needs of a designated user community for a data collection. As leader of data management activities for the International Polar Year, Parsons will describe hurdles he's confronted regarding data policy, international and interdisciplinary access to data, and gathering and providing access to community-based observations. Involving dataset managers in the collection process and encouraging scientists to define and then support the needs of their community has led to enhanced access to data.

Panel Two

focuses on the specific system developments that implement the broader issues on the micro-level. These developments are structures at the discipline or personal level that facilitates data archiving and promotes sharing and use of data sets among scientists. The speakers and their topics are:

Escience and Archiving for Space Science

Timothy Eastman from NASA Goddard Space Flight Center will present on scientific data archiving and preservation. Data and data systems are central to new developments in various forms of eScience or grid systems. Space science missions are developing multi-spacecraft, distributed, adaptive mission architectures that are both communication and computationally intensive and add to the data avalanche. He will present examples of space science data preservation, archiving, and access in this environment, including an application of ISO-standard Open Archive Information System (OAIS) architecture.

The Long-Term Preservation of Accurate and Authentic Digital Data

Sherry Xie from the School of Library, Archival and Information Studies at the University of British Columbia will present the latest findings of the InterPARES project (International Research on Permanent Authentic Records in Electronic Systems) with respect to the long-term preservation of scientific data existing in a dynamic digital environment. InterPARES aims at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. This knowledge allows for the formulation of model policies, strategies and standards capable of ensuring the longevity and authenticity of digital records. The next phase, InterPARES 2, develops research focusing on the issues of reliability and accuracy of digital records in addition to ways to ensure

authenticity. It also expands previous work in investigating digital records generated from administrative activities to include those generated in the course of carrying out scientific activities.

Metadata Architecture for Use of Scientific Data

John D'Ignazio from the School of Information Studies at Syracuse University will present a project that takes the ideas to encourage increased and appropriate use of scientific data sets by a broader community and embeds them in metadata architecture. Metadata provides for remote and timely appropriation of data in a standoff manner, but such characteristics suggest that the use of scientific data is only governed by convenient access. Instead, the project featured work with a research library and university-based research center to elicit user-centered information from the data producers that expands the data's potential application for research. Such information includes details of the fitness and purpose of the methodology performed by the data-producing researchers in the course of their investigation. Translating this into the metadata and metadata scheme of an information system enhances the sense of appropriateness and trust for the data-acquiring researchers to access and use the datasets in their own exploration.

References

ICSU. (2005) *Strengthening international science for the benefit of society: A strategic plan for the International Council of Science 2006-2011* Retrieved 2/2/2006 from http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/863_DD_FILE_ICSU_Strategic_Plan.pdf