

◆ *Interoperability in the Future—Common Data Standards and Protocols across Fields in the Digital Arena*

Moderator:

Diane Sullenberger

Proceedings of the National Academy of Sciences of the United States of America
Washington, DC

Speakers:

Clifford A Lynch

Coalition for Networked Information
Washington, DC

David Wholley

Foundation for the National Institutes of Health
Washington, DC

Peter Fox

University of Texas Health Science Center
San Antonio, Texas

Reporter:

Lynelle Korte

Elsevier
St Louis, Missouri

Attendees at this session were warned that in this world of evolving technology and increasing pressure to produce information, it can be dangerous to think solely about data without thinking about how to manage the accompanying versioning, preservation, and software issues.

Clifford A Lynch, of the Coalition for Networked Information, talked about the growing emphasis on data in the research environment. Although it is recognized that scientific disciplines have become data-intensive, he said, “this shift, which has been going on and gaining momentum over the last 2 or 3 decades, is now being much more explicitly recognized in various kinds of policy-making and funding initiatives.”

Data are being viewed as a driver of products and scientific inquiry, desirable to share and repurpose, and used to optimize investments in research. Data-management and -sharing plans are now often part of the funding, but who will be responsible for the preservation of the data, and how will they be made available over the long term? As Lynch said, “Simply preserving data is gratuitous unless you have a set of practices to make them available.”

Scholarly journals have a role in the data-sharing discussion not only because of the commitment to disseminate scientific information but also because journals codify and enforce regulations and establish norms. Many journals allow the deposit of ancillary or supplementary material with articles, but the preservation of the ancillary data may be unclear. The pressure for greater accountability and reproducibility of data means that journals will be required to provide meaningful access to the underlying data. For example, Lynch suggested that it may make sense to package underlying data with an article for researchers to use when repurposing the content.

David Wholley, of the Foundation for the National Institutes of Health, turned the discussion to the funding, policy, and organizational structures for developing and sharing data, specifically public-private partnerships. Public-private partnerships can leverage funding from and expertise of multiple stakeholders, accelerate projects, and promote efficiencies and entrepreneurial development. Of course, there are challenges in balancing the two entities, including antitrust concerns, conflicts of interest, the need for fair and productive publications, clear citation policies and methods, choices among data standards, and confidentiality, intellectual-property, and ethical issues.

Public-private partnerships must be collaborative, be built on the expertise of each partner, effectively meet clearly defined publication needs, have a formal agreement, and serve the public interest. In addition, it must be decided early how the resulting data will be accessed and published: openly available data versus controlled-access mechanisms. “These are large group efforts,” Wholley explained. “How do you recognize a large collaborative group in citations? How do readers know which version they are looking at?”

The data produced by such partnerships and others must be easily shared and accessed. Peter Fox, director of the Research Imaging Center at the University of Texas Health Science Center in San Antonio, explained that data-sharing is being promoted without a clear plan. “Maybe publication is a good trigger point of data-sharing,” he said. “Should journals encourage or enforce data-sharing?”

Examples of the availability of raw data have revealed little or no use. Fox proposed that it may be more appropriate to provide reduced data. Whereas sharing of raw data has low yield and is difficult, reduced data are more accessible, fewer, and transmitted more quickly.

Scholarly journals have a substantial interest in this discussion and can lead the way in promoting data standards and protocols by establishing policies that govern reporting standards, enabling enhanced access to data, and committing to regulation and preservation of data. 