International Scientific Data, Standards, and Digital Libraries: An NSF NSDL (U.S.) and CODATA Workshop

Laura M. Bartolo[1]*

*[*1] Kent State University, Materials Informatics Lab, United States*
Email: lbartolo@kent.edu

There is little debate that recent advances in information technology (IT) are revolutionizing the practice of science and expanding its impact on global society. It is still debatable whether the transformation can take full advantage of these advances without open, international standards, embraced by the scientific community. Accompanying the increases in computing power and bandwidth, considerable discussion has surfaced about the availability and exchange of scientific data. Common themes in the dialogue underscore the need for new visions of data organization, storage, and dissemination (*Nature* Editorial, 2005).

To add value to scientific investigation, data must be structured and organized in a consistent manner to be searchable, accessible, and exchangeable in machine-readable form. Numerous government reports have presented recommendations for redesigning the practice of science to optimize IT opportunities (Cyberinfrastructure Panel, 2003) (Digital Archiving Consultancy, 2003). These reports illustrate the benefits of integrating data acquisition, processing, storage, and access as well as the critical roles of both data organization and preservation. Shared data repositories focused on capturing, organizing, and disseminating data can ensure ongoing collection of data, provide distributed and remote access to organized collections of data, and support collaboration among scientists.

In support of innovation in education and research at all levels for science, technology, engineering, and mathematics, the National Science Foundation (NSF) created the National Science Digital Library (NSDL) to provide organized access to high-quality digital resources and tools. The organizational infrastructure of the NSDL facilitates connections between distributed users and content while offering services to sustain the NSDL enterprise (Zia, 2001). Responding to the need for authoritative content and the opportunity for repurposing data (with attention to intellectual property and digital rights management), the NSDL is built upon open and international standards for data dissemination. Its support of access to high quality scientific knowledge has resulted in the sponsorship of numerous workshops including two directly focused on the development and implementation of international standards. The first was the *NSF/NSDL Workshop on Scientific Markup Languages*, held 14-15 June 2004 at NSF in Arlington, Virginia, USA. This 2004 workshop brought together forty representatives in academe, government, and industry from the disciplines of biology, chemistry, earth sciences, mathematics, materials sciences, and physics (NSF/NSDL, 2004). Recommendations from the 2004 workshop encouraged continued discussions and expanded participation from the international data community.

In June 2005, the National Science Foundation/National Science, Technology, Engineering and Mathematics Digital Library and The Committee on Data for Science and Technology (CODATA) sponsored a workshop on International Scientific Data, Standards, and Digital Libraries held at the 5th ACM/IEEE Joint Conference on Digital Libraries.   The 2005 workshop examined successful models in the development of international standards for languages and tools in use with scientific and technical information (NSF/NSDL CODATA, 2005). This special issue brings together six papers from the 2005 NSDL/CODATA Workshop.   A common theme across the papers reiterates the enormously powerful opportunity now existing to advance scientific endeavor more rapidly through shared access to scientific data both within and across scientific domains.

- Hanisch from the Space Telescope Science Institute, United States, provides an overview of the International Virtual Observatory Alliance and its development of standards for astronomical data formatting, discovery, and delivery. His paper also presents his views for the rapid success of the project in its three years of existence.

- Brown from the Physics Department, McMaster University, Canada, describes the crystallographic data interchange format owned by the International Union of Crystallography and discusses its strengths and weaknesses.

- Burggraf from Galdos Systems, Inc., Canada, discusses geography markup language, an xml-based language to represent geographic information in a standardized way.   The paper also includes specific examples of how GML handles geographic data.

- Brase from the Research Center, University of Hannover, Germany, and Schingler from the World Data Center for Marine Environmental Sciences, University of Bremen, Germany, present information about the establishment of a registration agency for scientific primary data by the German National Library of Science and Technology.

- Cole from the Library Administration, University of Illinois at Urbana-Champaign, United States, focuses on the development of MathML, an XML-based standard for expressing mathematics, its historical background, its present state, and issues for its future.

- Erbach from the German Research Center for Artificial Intelligence, Language Intelligence Lab, Germany, reviews current research information systems (CRIS) and their transition from document centric systems to data-centric models.

- Ashino from Toyo University, Japan, defines a web ontology for design oriented materials selection. This paper has been previously published in the CODATA Data Science Journal. http://www.jstage.jst.go.jp/article/dsj/5/0/5_52/_article

Suichi Iwata, president, and Robert Chen, secretary general, of CODATA have challenged those in the scientific data communities to make a critical difference through universal access to scientific data, information, and knowledge (Iwata & Chen, 2005). The development and adoption of open international standards add support in bringing the challenge to fruition (Rumble, Hodge, Carroll, & Bartolo, 2005). By working together on the adoption of international scientific data standards, CODATA, NSDL, and e-Science centers can contribute significantly to capitalizing on the rich potential of IT for research and education in the sciences and for society.

Editorial. (2005) Let Data Speak to Data. *Nature* 438, 531.

Cyberinfrastructure Panel. (2003) Revolutionizing Science and Engineering through Cyberinfrastructure. Report of the NSF Blue-Ribbon Panel on Cyberinfrastructure. Retrieved September 13, 2006 from the World Wide Web: http://www.communitytechnology.org/nsf_ci_report/

Digital Archiving Consultancy. (2003) e-Science Curation Report prepared for the JISC Committee for the Support of Research. Retrieved September 13, 2006 from the World Wide Web: http://www.jisc.ac.uk/e-sciencecurationreport.pdf

Zia, L. (2001) Growing a National Learning Environments and Resources Network for Science, Mathematics, Engineering, and Technology Education: Current Issues and Opportunities for the NSDL Program. *D-Lib Magazine* 7, 3. Retrieved September 13, 2006 from the World Wide Web: http://www.dlib.org/dlib/march01/zia/03zia.html

NSF/NSDL Workshop on Scientific Markup Languages. (2004) National Science Foundation, Arlington, VA, USA. Retrieved September 13, 2006 from the World Wide Web: http://scimarkuplang.comm.nsdlib.org/cgi-bin/wiki.pl

NSF/NSDL CODATA Workshop on International Scientific Data, Standards, and Digital Libraries. (2005) Joint Conference on Digital Libraries, Denver, CO, USA. Retrieved September 13, 2006 from the World Wide Web: http://scimarkuplang.comm.nsdlib.org/cgi-bin/wiki.pl?SeeAgenda

Iwata, S. & Chen, R. (2005) Science and the Digital Divide. *Science* 310, 405. Retrieved September 13 2006 from the World Wide Web: http://www.sciencemag.org

Rumble, J., Hodge, G., Carroll, B., & Bartolo, L. (2005) Developing and Using Standards for Data and Information in Science and Technology. *Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data,* The Royal Society, Edinburgh, Scotland.