The DataNet Partners: Sharing Science, Linking Domains, Curating Data

DataNetONE
Suzie Allard[1]
Patricia Cruse [2]
Robert J. Sandusky[3]
Carol Tenopir[1]
DataNetONE participants are from [1] University of Tennessee; [2] California Digital Library ; [3] University of Illinois at Chicago

The Data Conservancy
Melissa Cragin
Carole Palmer
Allen Renear

Data Conservancy participants are from the University of Illinois at Urbana-Champaign

Introduction and Format

This panel introduces the first two DataNet partners funded through the National Science Foundation's Sustainable Digital Data Preservation and Access Network Partners (DataNet) solicitation (National Science Foundation, 2007). The first two of an expected five projects are *The Data Conservancy: A Digital Research and Curation Virtual Organization*, based at Johns Hopkins University (Sayeed Choudhury, PI), and *DataNetONE: Observation Network for Earth*, based at the University of New Mexico (William K. Michener, PI).

Following a brief overview of NSF's DataNet vision and goals, each funded project will be introduced and positioned within the context of NSF's vision for the DataNet Partners. The next part of the panel will describe how information scientists and librarians are integrated into the projects, including research, educational, and service development objectives. The final part of the panel will discuss collaboration between the DataNet partners in order to serve as "elements of an interoperable data preservation and access network" (NSF, 2007).

Introduction (5 minutes; Sandusky)
NSF's DataNet vision (10 minutes; Allard)
DataNetONE project overview (15 minutes; Sandusky / Cruse)
Data Conservancy project overview (15 minutes; Cragin / Renear)
Roles for Library and Information Scientists in DataNet partners (20 minutes; Tenopir / Palmer)
Questions and Answers (25 minutes)

SIG Sponsorships: SIG-KM, SIG-SI, SIG-STI

NSF's DataNet Vision

NSF's vision for the DataNet program is ambitious and expansive, and the funded projects are expected to form innovative, sustainable organizations integrating

> "library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise to provide reliable digital preservation, access, integration, and analysis capabilities … over a decades-long timeline…." (NSF, 2007).

The DataNet solicitation was developed in response to a number of well-known reports on cyberinfrastructure and data preservation, including the Association of Research Libraries' 2006 report *To Stand the Test of Time: Long–Term Stewardship of Digital Data Sets in Science and Engineering*, and the National Science Board's 2005 report *Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century*.

Primary goals of the DataNet program include

- Provide reliable, long-term (over a decades-long timeframe) digital access, discovery, and preservation capabilities for science and engineering data
- Accommodate rapid technological changes in providing reliable, long-terms services
- Develop new economically and technologically sustainable organizations that are not reliant on continuous NSF funding
- Enable innovative, domain-driven data-centered inquiry for disciplinary and interdisciplinary research
- Integrate library and archival sciences, cyberinfrastructure, computer and information science, and domain science expertise

- Advance the frontiers of computer and information science and cyberinfrastructure

Improved long-term access to and preservation of scientific data can help broaden participation in science, provide a more comprehensive history of scientific inquiry, support longitudinal and synthesis studies, and improve the economics of science by encouraging data reuse instead of costly re-collection or regeneration of data.

DataNetONE Project Overview

The DataNetONE project includes information scientists, librarians, and domain scientists from leading universities, research centers, government, and non-government organizations from North America, South America, Europe, and Australia (Table 1). The DataNetONE project will build

- An open, highly federated global network of many member nodes and three coordinating nodes
- An open and transparent virtual organization that invites participation from a wide range of stakeholders through a comprehensive community engagement and outreach program
- A common, standards-based DataNetONE service interface that provides a common abstraction for all services provided by member nodes, coordinating nodes, and software tools
- A standards-based Investigator Toolkit that contains solutions for creating and managing metadata and data at the member nodes, searching and browsing DataNetONE holdings, including workflow and analytical tools
- An integrated, continuous, and multi-method assessment and evaluation process to understand current and evolving DatNetONE community practices, inform development, and create an assessment baseline and provide ongoing metrics on DataNetONE usage and impact

This section of the panel will provide a brief overview of the domains involved, an example of the new kinds of science that will be enabled, the cyberinfrastructure being developed, and an overview of the virtual organization being developed to sustain DataNetONE.

Data Conservancy Project Overview

Led by Johns Hopkins University Library, the Data Conservancy (DC) is an international network of uniquely qualified domain scientists, information and computer science researchers, librarians, and engineers (Table 2). The DC team is designing and implementing an integrated and comprehensive data curation strategy that includes infrastructure development, information science research, data curation education, and library-based sustainability. Guided by a user-centered design methodology, the infrastructure development addresses the urgent need to collect, organize, validate, and preserve data to support scientific inquiry on the grand research challenges that face society. DC information science research is building a theoretical framework that will serve scientific data curation long into the future, through development of a cross-disciplinary data model for observational data and data mining techniques for extracting and mapping diverse data to the model, as well as research on data collection description requirements, metadata granularity and relationships, and comparative analysis of data practices across the initial base of astronomy, biodiversity, earth science, and social science user communities. To strengthen the data curation workforce, DC educational initiatives support apprenticeship of data scientists and enhancement of existing data curation programs for LIS students and in-service professionals.

This section of the panel will provide a brief overview of the infrastructure design, the projected information science outcomes, and the range of educational activities. It will conclude with a discussion of the DC as a new model for research libraries in the digital age that builds on a tradition of providing services broadly and deeply for a diversity of professional and citizen scholars.

Roles for Library and Information Scientists in DataNet Partners

Representatives from the two projects will describe how members of the LIS community have been and will continue to be engaged with their projects and, the five DataNet partners collectively, and the broader LIS community. The following issues will be addressed:

- LIS community participation in proposal development, project specification, and ongoing activities
- How each project interacts with LIS education
- LIS research objectives from each project
- Planned engagement with LIS practice, such as academic library services

- Planned ongoing engagement with the broader LIS community
- Planned and potential collaborations between the DataNet Partners

References

Association of Research Libraries (2006) *To stand the test of time: Long–term stewardship of digital data sets in science and engineering*. Washington, DC: Association of Research Libraries. Retrieved January 20, 2009, from http://www.arl.org/bm~doc/digdatarpt.pdf.

National Science Board. (2005). *Long-lived digital data collections:Eenabling research and education in the 21st centuryNational Science Board*. Retrieved January 20, 2009, from http://www.nsf.gov/pubs/2005/nsb0540/.

National Science Foundation. (2007). Sustainable Digital Data Preservation and Access Network Partners (DataNet). Retrieved January 18, 2009, from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141.

Tables

Principal Investigator for DataNetONE

William K. Michener	University of New Mexico (UNM), Long Term Ecological Research Network Office

Co-Principal Investigators

Robert Cook	Oak Ridge National Laboratory (ORNL)
Mike Frame	USGS National Biological Information Infrastructure (NBII)
Stephanie Hampton	National Center for Ecological Analysis and Synthesis (NCEAS), UC-Santa Barbara
Kathleen Smith	National Evolutionary Synthesis Center (NESCent), Duke University

Co-Investigators (partial list; LIS / IS participants)

Suzie Allard, L. Normore, C. Tenopir	University of Tennessee, Knoxville
Patricia Cruse, John Kunze	California Digital Library (CDL)
Robert Sandusky	University of Illinois at Chicago (UIC)

Affiliations of other Co-Investigators

Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Cornell University
Ecological Society of America
Independent Consultant, Brazil
National Biological Information Infrastructure (NBII)
National Center for Ecological Analysis and Synthesis (NCEAS), UC-Santa Barbara
National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign (UIUC)
National Evolutionary Synthesis Center (NESCent), Duke University
Oak Ridge National Laboratory (ORNL)
The Keystone Center
U.S. Geological Survey
University of California, Davis
University of Edinburgh, U.K.
University of Kansas
University of Manchester, U.K.
University of Michigan
University of New Mexico
University of Southampton, U.K.
University of Southern California, Information Sciences Institute
Utah State University, Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI)

Table 1: DataNetONE (Observation Network for Earth); Team Members and Affiliations

Principal Investigator for the Data Conservancy

Sayeed Choudhury	Johns Hopkins University Libraries

Co-Principal Investigators

Carole Palmer	University of Illinois at Urbana-Champaign, GSLIS, Center for Informatics Research in Science & Scholarship (CIRSS).
Carl Lagoze	Computing & Information Science, Cornell University
Mary Marlino	National Center of Atmospheric Research (NCAR), Library
David Patterson	Encyclopedia of Life, Marine Biological Laboratory

Co-Investigators (partial list; LIS / IS participants)

T. DiLauro	Johns Hopkins University
M. Cragin, D. Dubin, J. MacMullen, A. Renear, M. Welge	University of Illinois at Urbana-Champaign, GSLIS and NCSA
C. Borgman, S. Traweek	University of California Los Angeles, Departments of Information Studies & History

Affiliations of other Co-Investigators

Johns Hopkins University: Departments of Molecular Biology & Genetics, Computer Science, Earth & Planetary Sciences, Mechanical Engineering, Physics & Astronomy
National Snow & Ice Data Center
Lab of Ornithology, Cornell University,
National Virtual Observatory, Space Telescope Science Institute
Institute for the Study of Society and Environment, NCAR
Fedora Commons
University of Queensland, Australia
Tessella, Inc.

Table 2: The Data Conservancy: A Digital Research and Curation Virtual Organization; Team Members and Affiliations

Panelists

Suzie Allard Assistant Professor at the University of Tennessee School of Information Sciences. Her work focuses on how scientists and engineers use and communicate information in both informal and formal channels, and how these communication processes influence the data cycle from creation to preservation.

Melissa Cragin Project Coordinator for the Data Curation Education Program and doctoral candidate at the Graduate School of Library and Information Science (GSLIS), UIUC, and co-PI on an IMLS-funded project investigating data curation needs and disciplinary variation across sciences, in conjunction with librarians' participation in university eScience initiatives. Her dissertation research concerns the use of shared data collections in neuroscience.

Patricia Cruse Founding director of the California Digital Library's Digital Preservation Program. She works collaboratively with the ten University of California libraries to develop strategies for the preservation of content that is important to the research, teaching, and learning mission of the University.

Carole L. Palmer Associate Professor and Director of the Center for Informatics Research in Science and Scholarship (CIRSS) in the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. Her work focuses on aligning digital resource development with scientific and scholarly information work and information support for interdisciplinary research. She also leads development of new educational programs at GSLIS in the areas of data curation and biological informatics.

Allen Renear Associate Professor and Associate Dean for Research in the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. His research focuses on how digital documents function as knowledge representation systems; development of models of how documents organize and structure knowledge and exploring how these models can improve document-intensive applications such as digital libraries, scientific collaboration systems, publishing systems, educational technology, and humanities textbases.

Robert J. Sandusky Assistant University Librarian for Information Technology and Clinical Associate Professor at the University of Illinois at Chicago's University Library. He also has experience designing, building, and operating highly-secure and reliable national-scale computing infrastructure. His research focuses on human and organizational interaction with distributed infrastructure. Sandusky will serve as panel facilitator.

Carol Tenopir Professor at the University of Tennessee School of Information Sciences. She has studied patterns of scientific communication and scholarly publishing, in particular the use and design of digital publications for researchers and the role of digital resources in libraries.