11-10-2009

DataONE: Protecting the future of environmental and ecological data.

Suzie Allard
University of Tennessee - Knoxville, sallard@utk.edu

P. Cruse

Robert J. Sandusky
University of Illinois at Chicago, sandusky@uic.edu

Carol Tenopir
University of Tennessee - Knoxville, ctenopir@utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_dataone

Part of the Library and Information Science Commons

Recommended Citation

DataNetONE: Protecting the Future of Environmental and Ecological Data

Suzie Allard
University of Tennessee   Email: sallard@utk.edu

Patricia Cruse
California Digital Library   Email: Patricia.Cruse@ucop.edu

Robert Sandusky
University of Illinois, Chicago Email: Sandusky@uic.edu

Carol Tenopir
University of Tennessee   Email: ctenopir@utk.edu

This interactive panel addresses the theme of the conference "Thriving on Diversity- Information Opportunities in a Pluralistic World" by exploring the role of librarians and information scientists in DataNetONE (Observation Network for Earth), an emerging cyberinfrastructure project that supports the full data lifecycle for scientists in the diverse domains that are embodied in environmental and ecological science. Collaborations between scientists and information professionals are increasingly important, which is reflected by the soon to be announced "Science and Metadata Community" within the Dublin Core Metadata Initiative.

The research activities of environmental and ecological scientists produce diverse multi-scale, multi-discipline, and multi-national observational data. These data provide insights to address new environmental, social and technological challenges caused by climate variability, altered land use, population shifts, and changes in resource availability (e.g., food, water, and oil). Therefore scientists, educators, librarians, land managers, and the public need open, persistent, and secure access to well-described and easily discovered Earth observational data. These data are critical because they form the basis for good scientific decisions, wise management and use of resources, and informed decision-making.

Securing our global scientific knowledge base by preserving these documents and datasets requires active management of the resources and the supporting technology, as well as, an awareness of current research.

DataNetONE is one of two National Science Foundation DataNet Partners. The DataNet Partners serve as exemplars for national and global data research infrastructure organizations, and provide unique opportunities to communities of researchers to advance science and engineering research and learning. Eventually there will be five DataNet Partners.

DataNetONE focuses on multi-disciplinary observational data collected by biological (genome to ecosystem) and environmental (atmospheric, ecological, hydrological, and oceanographic) scientists, national and international research networks, and environmental observatories. However, the DataNetONE structure is designed to be domain-agnostic, so that it can be extended to serve a broader range of science domains both directly and through interoperability with other DataNet Partners.
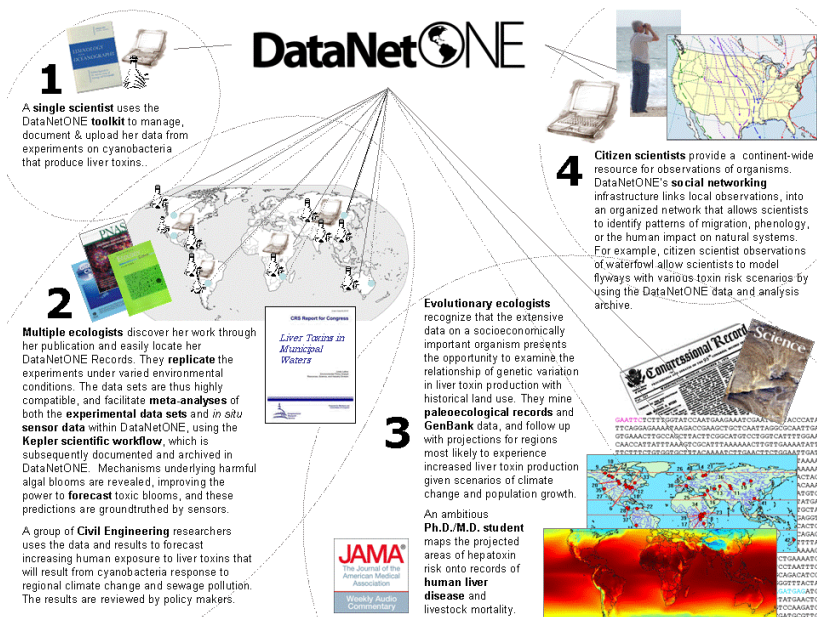
This panel of DataNetONE investigators focuses on four areas related to the challenges of the preservation of digital scientific data. The panel will be facilitated to encourage the audience to provide feedback and exchange ideas related to the role of information science, libraries and librarians in the process of creation, discovery, access, and manipulation of electronic scientific data.

About DataNetONE

A Case Study: DataNetONE's potential is best illustrated through case study that helps illustrate what Seeing the DataNetONE process helps illustrate the relationship between scientists and librarians/information professionals. This describes the figure below.

STEP 1: The scientist needs efficient and effective tools to manage and upload her work to a database that can be accessed through the DataNetONE network. Usability design and testing will create tools that can be used easily by practicing scientists while only minimally impacting their normal workflow. These tools will also help tag data so that it can be more easily discovered. Additionally the tools provide the foundation for better preservation practices. The scientist takes the time to learn these tools because she has learned through various media, the importance of best practices for electronic ecological data.

STEP 2: Proper labeling of the data through standardized metadata and interoperability of the system allow other ecologists to discover and access her records.

1 A **single scientist** uses the DataNetONE **toolkit** to manage, document & upload her data from experiments on cyanobacteria that produce liver toxins..

2 **Multiple ecologists** discover her work through her publication and easily locate her DataNetONE Records. They **replicate** the experiments under varied environmental conditions. The data sets are thus highly compatible, and facilitate **meta-analyses** of both the **experimental data sets** and *in situ* **sensor data** within DataNetONE, using the **Kepler scientific workflow**, which is subsequently documented and archived in DataNetONE. Mechanisms underlying harmful algal blooms are revealed, improving the power to **forecast** toxic blooms, and these predictions are groundtruthed by sensors.

A group of **Civil Engineering** researchers uses the data and results to forecast increasing human exposure to liver toxins that will result from cyanobacteria response to regional climate change and sewage pollution. The results are reviewed by policy makers.

3 **Evolutionary ecologists** recognize that the extensive data on a socioeconomically important organism presents the opportunity to examine the relationship of genetic variation in liver toxin production with historical land use. They mine **paleoecological records** and **GenBank** data, and follow up with projections for regions most likely to experience increased liver toxin production given scenarios of climate change and population growth.

An ambitious **Ph.D./M.D. student** maps the projected areas of hepatoxin risk onto records of **human liver disease** and livestock mortality.

4 **Citizen scientists** provide a continent-wide resource for observations of organisms. DataNetONE's **social networking** infrastructure links local observations, into an organized network that allows scientists to identify patterns of migration, phenology, or the human impact on natural systems. For example, citizen scientist observations of waterfowl allow scientists to model flyways with various toxin risk scenarios by using the DataNetONE data and analysis archive.

STEP 3: Metadata and interoperability also provide the foundation for these data to become available to researchers outside the original domain.

STEP 4: Communication with citizen scientists is enhanced with best communication practices.

DataNetONE is enhanced by expertise in communication and information. From a socio-cultural perspective, research areas include studying the creation, use and communication of information between scientists and with the non-scientific community; the management and sustainability of virtual data organizations; the creation of a data preservation best practices in the scientific community; the role of libraries in the scientific process in regards to datasets; and the influence of social networks within and between scientific domains. From a usability and assessment perspective, research areas include the usability and effectiveness of DataNetONE services and tools, and assessment of the effectiveness of DataNetONE across the many sectors interested in the data sets. From a cyberinfrastructure perspective, DataNetONE will focus on capabilities such as architecture of portals; distributed approaches to preservation and access; replication; secure, controlled access; authentication methods; tools deployed, and supported; and data discovery and interoperability methods.

Areas the panel will discuss

Preservation challenges and opportunities. DataNetONE faces myriad preservation challenges, but there are also exciting opportunities presented by leveraging the expertise of academic and digital librarians. This discussion will be lead by **Patricia Cruse,** who is the 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.

Infrastructure primer: The architectural underpinnings of DataNetONE aim to provide agility and sustainability. Libraries and librarians play pivotal roles in building and maintaining this infrastructure. This discussion will be led by **Robert Sandusky,** Assistant University Librarian for Information Technology and Clinical Associate Professor at the University of Illinois at Chicago's Richard J. Daley Library. He has experience designing, building, and operating highly-secure and reliable national-scale data communications and networks.

Creating a preservation culture: Digital preservation can be hindered by socio-cultural barriers such as: resistance to change; unwillingness to learn new technologies; lack of incentives for adopting new processes; and lack of supporting policies. Librarians and libraries can be essential in helping to establish a preservation culture at their organization, and among scientists. This discussion will be led by **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. Allard will also serve as the session facilitator.

Helping scientists use DataNet ONE – Usability and Assessment: Project success relies on tools and processes that are easy for scientists to use. Considering usability during the design phase, and conducting usability tests throughout development can help assure success. Additionally, ongoing assessment is essential to determining if DataNetONE is successfully meeting its goals. This discussion will be led by **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 the library with digital resources. She will discuss usability and assessment.

This session is sponsored by SIG-DL, SIG-STI, and SIG-KM