Panel Discussion presentation: "DataONE: Facilitating eScience through Collaboration"

Suzie Allard
University of Tennessee

Follow this and additional works at: http://escholarship.umassmed.edu/escience_symposium
Part of the Library and Information Science Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

http://escholarship.umassmed.edu/escience_symposium/2011/program/9

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact
Lisa.Palmer@umassmed.edu.
DataONE:
Facilitating eScience through Collaboration

Suzie Allard, Ph.D.
University of Tennessee

Presented at the
Third
University of Massachusetts & New England Area Librarian
eScience Symposium
6 April 2011
A prominent feature of e-Science is the generation of immense data sets that can be rapidly disseminated to other researchers via the internet.

-- Donna Kafel, MLIS
Project Coordinator for the New England e-Science Portal for Librarians
“e-Science and Its Relevance for Research Libraries”
Environmental Science Challenges
Vision:

Provide universal access to data about life on earth and the environment that sustains it.
DataONE team and funders

- Amber Budden, Roger Dahl, Rebecca Koskela, Bill Michener, Robert Nahf, Mark Servilla
- Dave Vieglais
- Suzie Allard, Carol Tenopir, Maribeth Manoff, Robert Waltz, Bruce Wilson
- John Cobb, Bob Cook, Giri Palanisamy, Line Pouchard
- Patricia Cruse, John Kunze
- Mike Frame, Viv Hutchison, Jeff Morisette, Jake Weltzin, Lisa Zolly
- Chad Berkley, Stephanie Hampton, Matt Jones
- Paul Allen, Rick Bonney, Steve Kelling
- Ryan Scherle, Todd Vision
- Randy Butler
- Ewa Deelman
- Peter Honeyman
- Jeff Horsburgh
- Robert Sandusky
- Bertram Ludaescher
- Peter Buneman
- Cliff Duke
- Carole Goble
- Donald Hobern
- David DeRoure
Mission

Support science by:
(1) engaging the relevant science, data, and policy communities;
(2) providing easy, secure, and persistent storage of data; and
(3) disseminating integrated and user-friendly tools for data discovery, analysis, visualization, and decision-making.
Data Life Cycle

Collect
Assure
Describe
Deposit
Preserve
Discover
Integrate
Analyze

Education
1. Build on existing cyberinfrastructure

2. Create new cyberinfrastructure

3. Support new communities of practice
Infrastructure Components

Three major components for a flexible, scalable, sustainable network

Member Nodes
- diverse institutions
- serve local community
- provide resources for managing their data
- retain copies of data

Coordinating Nodes
- retain complete metadata catalog
- indexing for search
- network-wide services
- ensure content availability (preservation)
- replication services

Investigator Toolkit
- R
- Python
- Java
- OpenOffice.org
- my experiment
- Excel
- The MathWorks

Pilot Data Catalog
- The DataOne Clearhouse is a powerful resource for scientists, allowing them to share and access information about important research in natural resources.
- The catalog contains over 70,000 records from 90 data providers.

Search for: null carbon
- Result/Row: 10
- Search results: 10

Physical Search
- Publication Date: 2019
- Title: Global carbon cycle

Data Search
- Title: Global carbon cycle
- Publication Date: 2019
- Title: Global carbon cycle

Geographic Search
- U.S. or any country
- Select topic: Emissions

Department of Geography
Salt Lake State University
Candidate Nodes Around the World
Community Engagement Approach

- Engagement
  - Working Groups
  - All Hands’ Meetings
  - External Advisory Board
  - DataONE Users Group

- Communication
  - Newsletter
  - Web presence
  - Training
  - Outreach

2009 First All Hands’ Meeting
Assessment-stakeholders

Data
- Scientists
- Students & Teachers

Public Officials
- Citizen-scientists

Publishers
- Libraries & Librarians
- Publishers

Assessment-stakeholders
DataONE User Scenarios

Scientist

Project Planning
- data discovery and access
- data visualization

Research Activity
- data mgmt planning
- online training
- data download and integration
- tool discovery
- data upload

Publication
- visualization
- data upload
- best practices for citation

Project Planning
- data discovery and access

Research Activity
- software tool discovery
- spatial data referencing
- develop range visualizations

Publication
- document session activity
- archive/cite final report

Librarian
Best Practice Categories

- Content and Structure (15)
- Data Access and Discovery (5)
- Data Documentation (3)
- Data Preservation and Archives (7)
- Planning Policies and Governance (1)
- Quality Assurance and Quality Control (5)
- Vocabulary Standards and Services (2)

Featured Best Practice

**Define the contents of Data Files**

Category: Data Documentation

- Formats for dates, time, geographic coordinates, and other parameters
- Define any coded values
- Quality flags or qualifying values
- Define missing values

Software Tools

Tool Categories

- Analysis & Modeling (34)
- Data Acquisition & Modeling (30)
- Workflow (7)

Featured Tool

**S-PLUS (S+)**

Primary Category: Analysis & Modeling

S-PLUS is a commercial implementation of the S statistical programming language that was developed by Bell Labs. S+ has a cross-platform integrated development environment (IDE), provides the ability to analyze gigabyte class data sets on the desktop, and a package system for deployment of analytics.

Cost: Cost-basis

**TIBCO Spotfire S+**
Engaging Citizens in Science

Project BudBurst
Learn why phenology is important
Does climate change affect budburst?
Participate!
Report your observations online
Download free materials
Map results from around the country
www.budburst.org

eBird

Citizen Science Central
Welcome to Citizen Science Central!
A clearinghouse for ideas, news, and resources in support of citizen science--partnerships between volunteers and scientists that answer real-world questions.

IDEAS
NEWS
RESOURCES

National Phenology Network
www.CitizenScience.org
The Information Puzzle

- Discover needs of all stakeholders
- Help improve data lifecycle functions & processes
- Educate next generation of data specialists
- Test approaches, systems, & progress
- Tackle Socio-cultural issues
Linking Education and Science to Develop the Next Generation of Educators for Science Librarians and Data, Information and Communication Specialists
Digital Curation Education in Research Centers

[Photo credits: Copyright University Corporation for Atmospheric Research]
Thank You!

Suzie Allard
sallard@utk.edu
Extra slides
## Data Curation Education in Research Centers

<table>
<thead>
<tr>
<th>Data Curation Doctoral Program</th>
<th>Science Data Curation Master’s Speciality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate School of Library and Information Science</td>
<td>School of Information Sciences</td>
</tr>
<tr>
<td>University of Illinois at Urbana-Champaign</td>
<td>University of Tennessee, Knoxville</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selected Seminars &amp; Courses</th>
<th>Representative Research Areas</th>
<th>Selected Courses</th>
<th>Practice and research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Problems in Data Curation</td>
<td>Cross-disciplinary data sharing and reuse</td>
<td>Research Problems in Data Curation</td>
<td>Cross-disciplinary data sharing and reuse</td>
</tr>
<tr>
<td>Ontologies in the Sciences</td>
<td>Data units, identity, and provenance</td>
<td>Information Organization and Representation</td>
<td>Data management practices of scientists</td>
</tr>
<tr>
<td>Information Transfer &amp; Collaboration in Science</td>
<td>Collection and integration of data for scientific problem-solving</td>
<td>Sources and Services in Science &amp; Engineering</td>
<td>Participation in research teams</td>
</tr>
<tr>
<td>Metadata in Theory &amp; Practices</td>
<td>Data and literature dynamics</td>
<td>Bioinformatics</td>
<td>Shadowing scientists in laboratory environments</td>
</tr>
<tr>
<td>Foundations of Data Curation</td>
<td>Culture and practice of data communities</td>
<td>Practicum and Research Participation</td>
<td>Culture and practice of data communities</td>
</tr>
</tbody>
</table>