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.
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
Environmental Science Challenges

- Instability of Greenland Ice Sheet?
- Atlantic Deep Water Formation
- Methane Outburst?
- Salinity Valves
- Instability of West Antarctic Ice Sheet?
- Antarctic Ozone Hole
- Southern Ocean Upwelling / Circumpolar Deep Water Formation
- Bistability of Saharan Vegetation
- Bistability of Amazonian Forest?
- Performance of Marine Carbon Pump
- Tibetan Albedo Change?
- Indian Monsoon Transformation
- ENSO Triggering
- Bodele Dust Supply Change?
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
- Analyze
- Integrate
- Discover
- Preserve
- Deposit
- Describe
- Assure

Education
1. Build on existing cyberinfrastructure

2. Create new cyberinfrastructure

3. Support new communities of practice
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
- Kepler
- Specification
- OpenOffice.org
- MyExperiment
- Excel
- MathWorks

DataONE

Pilot Data Catalog

The Pilot Catalog provides a web interface for accessing content cataloged and available through the DataONE system. The Pilot Catalog demonstrates functionality through the search interfaces being implemented on the DataONE Coordinating Nodes. It does not currently provide access to content cataloged in the DataONE system.

Search for: [Input]

Hint: Boolean operations, wildcards and phrases are allowed.

Fielded Search
- Title
- Authors
- Publication Date
- Location
-能得到的

Geographic Search

Maps available from [Google Maps]
Community Engagement Approach

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

- **Communication**
  - Newsletter
  - Web presence
  - Training
  - Outreach
Assessment-stakeholders

- Scientists
- Data
- Public Officials
- Publishers
- Libraries & Librarians
- Students & Teachers
- Citizen-scientists
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

**Librarian**
- 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
The Cloverleaf Model of DataONE Stakeholders

**Primary:**
- Discovery of new data & knowledge
- Stewardship
- Policy
- Legal regulation
- Legal compliance
- Products

**Secondary:**
- Data application
- Education
- Curricula
- Teacher Training
- Learning
- Observation
- Experimentation
- Visualization

**Stakeholders:**
- Scientists
- Citizen scientists
- Industrial researchers
- Data managers
- Observation networks
- Intergovernmental organizations
- Local, state, federal governments
- Non governmental organizations
- Commercial Sector
- Academia
- Government advisory committees
- Academic scientists
- Administrators/program managers
- Government advisory committee members
- Commercial developers
- Community
- Attorneys/judges
- Legislators

**Individuals and Organizations:**
- K-12 libraries
- Research libraries
- US Dept of Education
- Administrators – higher education
- Accrediting organizations
- Textbook companies
- Boards of education
- K-12 teachers
- K-12 librarians
- College faculty/lecturers
- College/research librarians
- College students
- K-12 students
- Publishers
- Academic
- Software companies
- Repositories
- Archives
- Museums
- Curators
- Administrators
- IT (technician) personnel
- Software developers
- Archivists
- Librarians
- Diplomats
- Academic scientists
- Administrators/program managers
- Govt advisory committee members
- Commercial developers
- Community
- Attorneys/judges
- Legislators
- Govt advisory committees
- Intergovernmental organizations
- Local, state, federal governments
- Non governmental organizations
- Commercial Sector
- Academia

**Curation**
- Stewardship
- Preservation

**K-12 teachers**
- K-12 librarians
- College faculty/lecturers
- College/research librarians
- College students
- K-12 students

**Academia**
- Publishers
- Academic
- Software companies
- Repositories
- Archives
- Museums
- Curators
- Administrators
- IT (technician) personnel
- Software developers
- Archivists
- Librarians

**Govt advisory committees**
- Intergovernmental organizations
- Local, state, federal governments
- Non governmental organizations
- Commercial Sector
- Academia

**Education**
- Curriculum
- Teacher Training
- Learning
- Observation
- Experimentation
- Visualization

**Long-term Access**
- Stewardship
- Policy
- Legal regulation
- Legal compliance
- Products

**Data application**
- Secondary:
- Education
# DataONEpedia

## Best Practices

### Search Best Practices

<table>
<thead>
<tr>
<th>Contains</th>
<th>SEARCH</th>
</tr>
</thead>
</table>

### Best Practice Categories

- **All Best Practices**
- 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
- Define quality flags or qualifying values
- Define missing values

## Software Tools

### Search Tools

<table>
<thead>
<tr>
<th>Contains</th>
<th>SEARCH</th>
</tr>
</thead>
</table>

### Tool Categories

- **All Tools**
- 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+**

---

[www.dataone.org/dataonepedia](http://www.dataone.org/dataonepedia)

Editors: Cook, Michener

Contributors: Best practices workshop participants
Engaging Citizens in Science
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>