DataONE
(Data Observation Network for Earth)

DataONE Member Node Implementation Workshop
Amber Budden

Science Challenges
Technology Challenges

Sensors, sensor networks, and remote sensing gather observations

Data Challenges

“Building the Knowledge Pyramid”

90:10 → 10:90

Intensive science sites and experiments

Extensive science sites

Volunteer & education networks

Remote sensing

Adapted from CENR-OSTP
Individual Challenges

Plan

Collect

Assure

Describe

Preserve

Integrate

Discover

Analyze

What is DataONE?
DataONE is Cyberinfrastructure

Three major components for a flexible, scalable, sustainable network

<table>
<thead>
<tr>
<th>Member Nodes</th>
<th>Coordinating Nodes</th>
<th>Investigator Toolkit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diverse organizations</td>
<td>Retain complete metadata</td>
<td>Mendeley</td>
</tr>
<tr>
<td>Serve local community</td>
<td>Indexing for search</td>
<td>ARToolKit</td>
</tr>
<tr>
<td>Provide resources for managing their data</td>
<td>Network-wide services</td>
<td>Zotero</td>
</tr>
<tr>
<td>Retain copies of data</td>
<td>Ensure content availability (preservation)</td>
<td>MATLAB</td>
</tr>
</tbody>
</table>

DataONE is Community

Institution logos from various universities and organizations.
How will DataONE enable science?

Research Scientist
Pre DataONE

DataONE

A
B
C

A
B
C

Pre DataONE

DataONE
Data Life Cycle Tool Support
What is the DataONE process?

Our Vision and Approach:

Enabling universal access to data about life on earth and the environment that sustains it.

1. Build on existing cyberinfrastructure
2. Create new cyberinfrastructure
3. Support communities of practice
Community Engagement Approach

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

Communication
• Newsletter
• Web presence
• Training
• Outreach

Community Assessment Mechanisms

- Stakeholder surveys
- Persona and scenario development
- Usability testing
- External assessments / surveys

Importance of Access
Ease of Access
Exploration, Visualization, and Analysis

EVA Working Group
Steve Kelling (co-chair)
Daniel Fink
Kevin Webb
Bob Cook (co-chair)
John Cobb
Suresh Santhana Vannan
Theo Damoulis
Tom Dietterich
Juliana Freire
David Koop
Claudio Silva
Damian Gesler
Bill Michener
Jeff Morisette
Patrick O'Leary
Alyssa Rosemarin

Grand Challenges
- Discovery and Access
- Interoperability and Synthesis
- Processing and Analysis
- Exploration and Visualization

Process
Discovery and access
Interoperability and synthesis
Processing and analysis
Swainson’s Hawk

Exploration and Visualization

Diverse bird observations and environmental data from 300,000 locations in the US integrated and analyzed using High Performance Computing Resources

- Examine patterns of migration
- Infer how climate change may affect bird migration
Outcomes

1. Inform development of Investigator Toolkit
2. Highlight the value of DataONE to provide data and tools to conduct science
   a) search and access distributed, large, and disparate data resources
   b) integrate into one database
   c) explore, and visualize
3. Requires a close working relationship with DataONE developers, so that they can incorporate these tools and services into DataONE

Progress to date
Design and Implementation

DataONE architecture completed as a “living document”
Completion of prototype implementation

- Coordinating Nodes
- Member Nodes
- Investigator Toolkit
- Web site
Data Centers / Member Nodes

<table>
<thead>
<tr>
<th>ORNL-DAAC</th>
<th>Dryad</th>
<th>KNB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community</td>
<td>Agency repository</td>
<td>Journal consortium</td>
</tr>
<tr>
<td>Data</td>
<td>Ecology and biogeochemical dynamics</td>
<td>Biosciences</td>
</tr>
<tr>
<td>Size</td>
<td>900 data products, ~ 1 TB</td>
<td>~ 1,000 data products, ~ 3 GB</td>
</tr>
<tr>
<td>Services</td>
<td>Tools for data preservation, replication, discovery, access, subsetting, and visualization</td>
<td>Tools for data preservation, replication, discovery and access</td>
</tr>
<tr>
<td>Metadata stds.</td>
<td>FGDC subset</td>
<td>Dublin Core application profile</td>
</tr>
<tr>
<td>Degree of curation</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Data submission</td>
<td>Agency-approved, staff-assisted submission and curation of final data product</td>
<td>Web-based data submission at time of journal article submission</td>
</tr>
<tr>
<td>Sponsor</td>
<td>NASA</td>
<td>NSF/JISC, societies, publishers</td>
</tr>
</tbody>
</table>

Investigator Toolkit

The following links provide comprehensive information on tools available within the Investigator Toolkit. These tools offer additional value to DataONE and are being developed to enhance the processes of data deposition, acquisition, analysis, visualization and citation. Additional tools are being developed and will be available for download through this interface.
Community: DataONE Users Group

- Established December 2010
- Charter approved January 2011
- ‘Service Guidelines’ for Member Nodes approved January 2011
- Annual meetings co-located with ESIP summer meetings
DataONE Team and Sponsors

- 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, Richard Huffine, 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
- Ryan Scherle, Todd Vision
- Peter Honeyman
- Jeff Horsburgh
- Robert Sandusky
- Bertram Ludaescher
- Peter Buneman
- Cliff Duke
- Carole Goble
- Donald Hobern
- David DeRoure