Metadata Zoo

Rebecca Koskela

CAMP-4-DATA
September 6, 2013
Promoting data discovery

Provide universal access to data about life on earth and the environment

1. Building community

2. Developing sustainable data discovery and interoperability solutions

3. Enabling science through tools and services
Discovery via DataONE

Three major components for a flexible, scalable, sustainable network

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

Three major components for a flexible, scalable, sustainable network

Coordinating Nodes

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

![Network Diagram]
Discovery via DataONE

Three major components for a flexible, scalable, sustainable network

Coordinating Nodes

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

![Map with coordinating nodes](image)
Discovery via DataONE

Three major components for a flexible, scalable, sustainable network

<table>
<thead>
<tr>
<th>Coordinating Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member Nodes</td>
</tr>
</tbody>
</table>

Investigator Toolkit

- command line interface
- python
- Java
- ONE Drive
- DMP Tool
- ONE Mercury
- DATA UP
- Kepler
- ONE R
Scientists want to share data

- Use other researchers’ datasets if easily accessible: 84%
- Willing to share data across a broad group of researchers: 81%
- Appropriate to create new datasets from shared data: 76%
- Currently share all of their data: 6%

Metadata standards:

- DIF: 12
- DwC: 21
- DC: 26
- EML: 95
- FGDC: 95
- Open GIS: 96
- ISO: 97
- My Lab: 266
- none: 676
# Libraries not yet providing data services

<table>
<thead>
<tr>
<th>Service</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata creation</td>
<td>67%</td>
</tr>
<tr>
<td>Conversion of data/datasets for ingest</td>
<td>75%</td>
</tr>
<tr>
<td>Selection of data/datasets for ingest</td>
<td>70%</td>
</tr>
<tr>
<td>Selection of data/datasets for repository</td>
<td>66%</td>
</tr>
</tbody>
</table>

- **never**
- **occasionally**
- **monthly**
- **weekly**
- **daily**
DataONE Metadata Interoperability

**Member Nodes**

- **KNB**
  - EML, ISO, FGDC

- **LTER**
  - EML

- **ORNL DAAC**
  - FGDC, ISO

- **CDL**
  - FGDC

- **USGS CSAS**
  - FGDC, ISO

- **DRYAD***
  - METS

**Coordinating Nodes**

- Internal Metadata Index
- Metadata Extraction

- **Virtual Portals**
- **Numerous search capabilities**
- **Metadata has link to data, which reside at Member Nodes**

*Future