Session 5

Prioritizing Future Member Nodes

Bob Cook
*Oak Ridge National Laboratory*

Suzie Allard
*University of Tennessee*
Member Node Subcommittee

Members
Mike Frame, *USGS* (lead)
Suzie Allard, *University of Tennessee*
Paul Allen, *Cornell University*
Bruce Wilson, *University of Tennessee & ORNL*
Bob Cook, *ORNL*, presenting
Roadmap

1. Introduction
2. Becoming a Member Node
3. Approaches for identifying and choosing next member nodes
4. Discussion
Member Nodes and DataONE

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)
  - Replica services

- **Investigator Toolkit**
  - Mendeley
  - VEGBANK
  - Specify 6
  - MyExperiment
  - Excel
  - Zotero
  - MATLAB

---

Introduction

Becoming a Member Node

Prioritizing next Member Nodes

Discussion
Member Node Functions

• Data storage
• Data access
• Access control
• Metadata quality
• Primary user interaction

If needed
• Replication
Metadata for discovery and access

Science metadata
• EML, FGDC, DC, ISO, DIF, ...

System metadata
• Globally unique IDs for data & metadata (DOI, GUID, Hdl, ...)
• Checksums of objects
• Object policies

Three original member nodes:
• demonstrate different ways of integration
• diverse range of technical and business processes.
Implementation Tiers

**Tier 1** Supports publicly readable content without authentication or more specific access control rules

**Tier 2** Tier 1 plus access control support

**Tier 3** Tier 2 plus ability to add content through the DataONE service interfaces and provides full support for interaction with DataONE Investigator Toolkit

**Tier 4** Support the full set of DataONE APIs and can operate as replication targets, accepting content from compatible (technical and policy) Member Nodes and fully supporting the DataONE content access control rules

Member Node API Definitions are available:

- [http://mule1.dataone.org/ArchitectureDocs-current/apis/MN_APIs.html](http://mule1.dataone.org/ArchitectureDocs-current/apis/MN_APIs.html)
## Current Member Node Characteristics

<table>
<thead>
<tr>
<th></th>
<th>ORNL DAAC</th>
<th>Dryad</th>
<th>KNB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Community</strong></td>
<td>Agency repository</td>
<td>Journal consortium</td>
<td>Research network</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>Ecology and biogeochemical dynamics</td>
<td>Biosciences</td>
<td>Biodiversity, ecology, environment</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>~ 1,000 data products, ~ 1 TB</td>
<td>~ 1,000 data products, ~ 5 GB</td>
<td>20,000 data products, 100s GBs</td>
</tr>
<tr>
<td><strong>Services</strong></td>
<td>Tools for data preservation, replication, discovery, access, subsetting, and visualization</td>
<td>Tools for data preservation, replication, discovery and access</td>
<td>Tools for data preservation, replication, discovery, access, management, and visualization</td>
</tr>
<tr>
<td><strong>Metadata standards</strong></td>
<td>FGDC subset</td>
<td>Dublin Core application profile</td>
<td>EML, FGDC</td>
</tr>
<tr>
<td><strong>Degree of curation</strong></td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Data submission</strong></td>
<td>Staff-assisted submission and curation of final data product</td>
<td>Web-based data submission at time of journal article submission</td>
<td>Self-submission via desktop tool at any time</td>
</tr>
<tr>
<td><strong>Sponsor</strong></td>
<td>NASA</td>
<td>NSF/JISC, societies, publishers</td>
<td>NSF</td>
</tr>
</tbody>
</table>
ORNL DAAC

**Existing Capabilities**
- Data Ingest
- Metadata preparation
- Search and Order System
- Tools and Services
- Data citation
- User metrics
- User Help Desk
ORNL DAAC
As a DataONE Member Node
Benefits of Becoming a DataONE Member Node

• Increases the exposure of your data products and tools to a larger user base
• Provides your user community with access to broader DataONE data and tools
• Enables leveraging, networking, and coordinating with others to address similar data management and cyberinfrastructure issues
Projected Growth of Member Nodes

Cumulative number of member nodes

Year 1: 3 in test
Year 2: 6
Year 3: 10
Year 4: 20
Year 5: 40
Next New Member Nodes

USGS Earth Science Data Clearinghouse

AKN Avian Knowledge Network

CUAHSI HIS Sharing hydrologic data

Compute and Storage Node

Next.....
Steps to become a Member Node*

**Step 1:** Expression of interest
- Complete Member Node Description document and sign letter of interest
- Join DataONE Users Group (at any time)

**Step 2:** Technical and strategic review
- “Criteria for Evaluating / Prioritizing DataONE Member Node”
- DUG input today

**Step 3:** DataONE Response

Affirmative: Next slide

Not at this time:
- Link up with appropriate Member Node
- Obtain advice about becoming a Member Node in the future.

*Additional information available on-line:
http://www.dataone.org/content/dataone-users-group
Implementation Steps

• Agree to DataONE service guidelines and partnership agreement
• Operational implementation by DataONE and Member Node

Participate in the following:
• DataONE User Group
• Development of:
  • Overall MN strategic plan
  • 5-yr plan for MNs
  • Roles, responsibilities, value proposition
Approaches for choosing next Member Nodes

1. Why prioritize?
   - sustainable, technological, strategic

2. Share our ideas for a prioritization approach

3. Solicit your ideas for a prioritization approach

- Year 1: 3 in test
- Year 2: 6
- Year 3: 10
- Year 4: 20
- Year 5: 40

*cumulative number of member nodes*
Member Node Subcommittee

Goal: Prepare a draft set of new Member Node evaluation criteria for refinement by DUG

Members
Mike Frame (lead)
Suzie Allard
Paul Allen
Bruce Wilson
Bob Cook (presenting)
Minimum set of MN requirements

- The metadata format used by the candidate Member Node is supported by DataONE
- Most data in the collection is public or can be shared upon request
- Basic level of physical and cyber-security is in place
- The candidate MN intends to implement at least the Tier 1 DataONE Member Node API
Key Characteristics

**Data**
- Quality assurance
- Scientific value
- Extent of collection
- Risk

**Strategic Partnerships**
- Community
- Funding
- Technical expertise
- Data management expertise
- Synergy

**Diversity**
- Geographic
- Underrepresented groups
- Linguistic or cultural diversity
- Institution type

**Leadership/Management**

**Technical**
- Human resources
- Technical resources
- Technical compatibility
- Technical stability
Discussion: Prioritization

What criteria should be used to prioritize the selection of future Member Nodes?

What approach should be used to prioritize the selection of future Member Nodes?
Discussion: Speeding addition of MN

Given limited resources, what approaches could be used to facilitate addition of Member Nodes?

• Package of information
  • “How to prepare your data center to become a DataONE Member Node”
• Train the trainer / incubator approach
For more information

DataONE
http://www.dataone.org/

DUG and Member Node Information
http://www.dataone.org/content/dataone-users-group

Robert Cook
cookrb@ornl.gov

Suzie Allard
sallard@utk.edu