# Integrating your repository with the DataONE network

Ryan Scherle (ryan@scherle.org), Dryad Digital Repository

DataONE is a distributed network of data repositories. Each individual repository is considered a Member Node within the network. Member Nodes make their data available to the network's Coordinating Nodes. Coordinating Nodes provide network-wide services, including centralized search and replication of content between Member Nodes. The division of effort between Member Nodes and Coordinating nodes provides for a robust and highly scalable network.

DataONE is interested in including a broader range of repositories as Member Nodes, including Institutional Repositories. Benefits of becoming a Member Node include:

- **Broader exposure of content –** All content in a Member Node is indexed by the Coordinating Nodes. This content is exposed by DataONE's central search system, ONEMercury [1].
- **Automatic replication of content –** When a Member Node elects to have its content replicated, the DataONE network automatically replicates content to other nodes in the system. Content can be replicated to nodes that are both geographically distant and technologically distinct from the original node, providing an extra layer of fault tolerance.
- **Compatibility with a rapidly growing suite of tools –** DataONE Member Nodes are instantly compatible with DataONE's suite of "investigator tools". These tools include both DataONE-centric utilities and commonly used scientific software. DataONE tools can access content in any DataONE Member Node.
- **Access to expertise –** DataONE hosts teleconferences and physical meetings of participating Member Nodes. These meetings provide support for repository managers who spend all or part of their time managing data.

To become a member node, a repository must adhere to the DataONE API specifications [2]. These specifications are designed to cover a wide variety of repository technologies while providing maximal functionality to the scientific community. Many of the basic concepts are derived from the standards developed by the Open Archives Initiative. The general data access mechanism is similar to OAI-PMH [3]. Data files are linked with their associated metadata documents using OAI-ORE [4].

Depending on the capabilities of a repository and its willingness to participate in the DataONE system, a Member Node can be implemented at one of four levels, from a read-only node that provides no access control to a fully featured node that can hold replicas of content from any node in the DataONE system.

There are several technological solutions enabling a repository to join DataONE as a Member Node. A repository may choose to implement the DataONE API "from scratch", linking DataONE API calls directly to internal functionality. Or a repository can use one of the existing technology stacks. At the time of this writing, currently available implementations include:

- **Generic Member Node (GMN)** – The GMN is a bare-bones implementation of the DataONE API with a simple backend storage mechanism.
- **MetaCat** – MetaCat [5] is a repository platform designed to manage scientific data. It is particularly useful when managing complex XML metadata documents.
- **Dryad** – Dryad [6] is an adaptation of DSpace. Dryad's Member Node implementation is currently being generalized for other DSpace users.

By making data available through the DataONE network, a repository can greatly increase the discoverability and interoperability of its contents.

[1] https://cn.dataone.org/onemercury/
[2] http://mule1.dataone.org/ArchitectureDocs-current/
[3] http://www.openarchives.org/OAI/openarchivesprotocol.html
[4] http://www.openarchives.org/ore/1.0/toc
[5] http://knb.ecoinformatics.org/knb/docs/
[6] http://datadryad.org