

Becoming a DataONE Member Node

IDCC14

San Francisco, CA

February, 2014





Why become a member node?

- Reach a wider audience
- Leverage existing CI
- Recognition and Credit
 - For the MN
 - For data creators
- Improve data availability
- Enhance collaboration opportunities
 - Presenting relevant collections to users
 - Additional data for your services

Metadata Interoperability

Member Nodes

KNB

EML, ISO
FGDC

LTER

EML

ORNL DAAC

FGDC, ISO

CDL

FGDC

USGS CSAS

FGDC, ISO

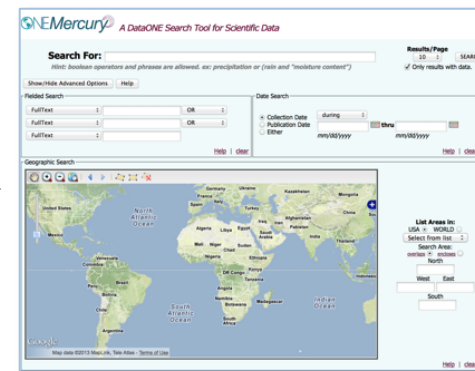
DRYAD

METS

Metadata Extraction

Coordinating Nodes

Internal
Metadata
Index



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



Tool Interoperability

Member Nodes

KNB

EML, ISO
FGDC

LTER

EML

ORNL DAAC

FGDC, ISO

CDL

FGDC

USGS CSAS

FGDC, ISO

DRYAD

METS

Metadata Extraction

Coordinating Node

Internal
Metadata
Index

Query Service Interface

VisTrails



DataONE
Library

DataONE Client
Python Library
wrapped in
VisTrails Modules



MNs: High level requirements

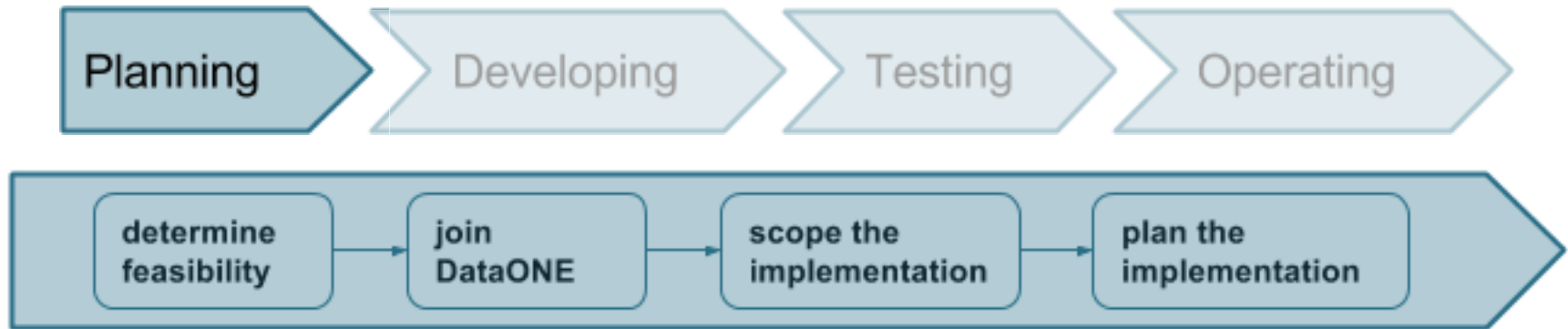
- Maintain a preservation-oriented repository
 - Use persistent identifiers for data (and metadata)
 - Ensure access to these data products over the long term
 - Ensure metadata exists with the data
- Define Data Packages using (OAI-ORE) Resource Maps
- Follow good data curation practices
 - Users need to at least know when data has changed
- Work with DataONE on practices & communication
 - Communicate downtime
 - Communicate changes in contact points
 - Help us serve you and work with us to serve science



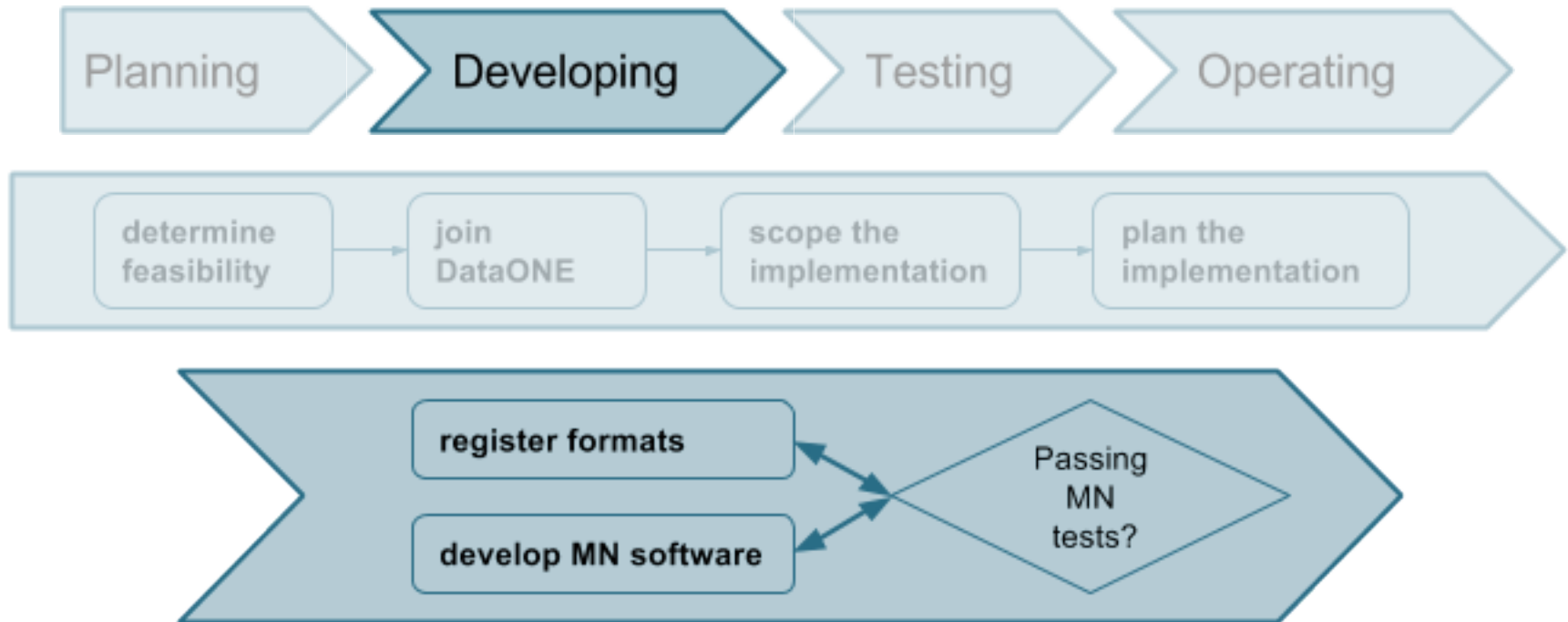
How to become a Member Node



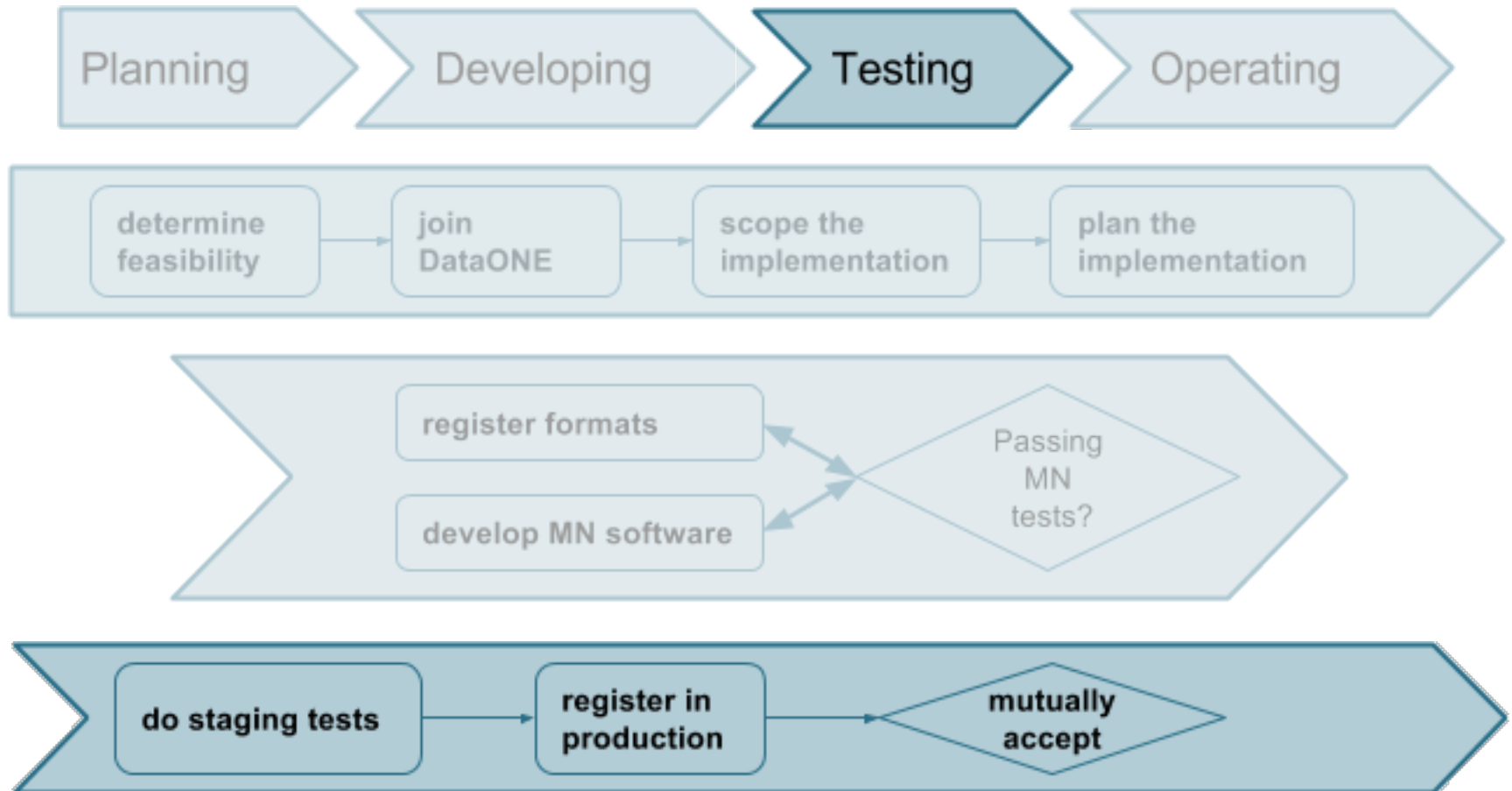
How to become a Member Node



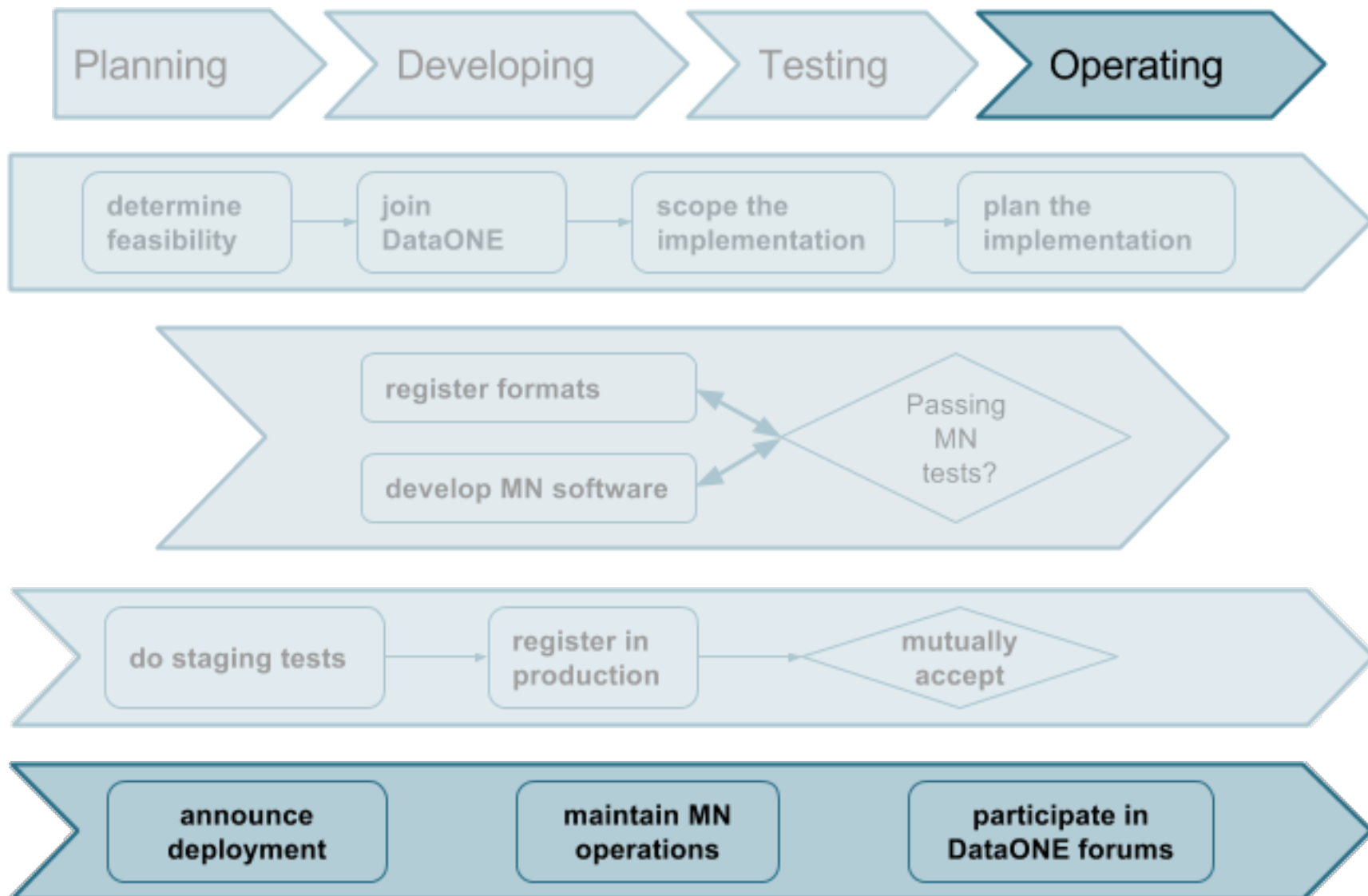
How to become a Member Node



How to become a Member Node



How to become a Member Node





Scoping: Let's talk

- Reach out to DataONE
 - Laura Moyers (Member Node Coordinator)
lmoyers1@utk.edu
 - Amber Budden (Dir. Community Engagement & Outreach)
aebudden@dataone.unm.edu
 - Contact us form on dataone.org web site
 - Any Leadership Team member
 - Any Cyberinfrastructure Team member
 - #dataone channel on irc.ecoinformatics.org
- Let's make sure:
 - You have the info needed for your decisions
 - We know about your needs (e.g. file or metadata formats)

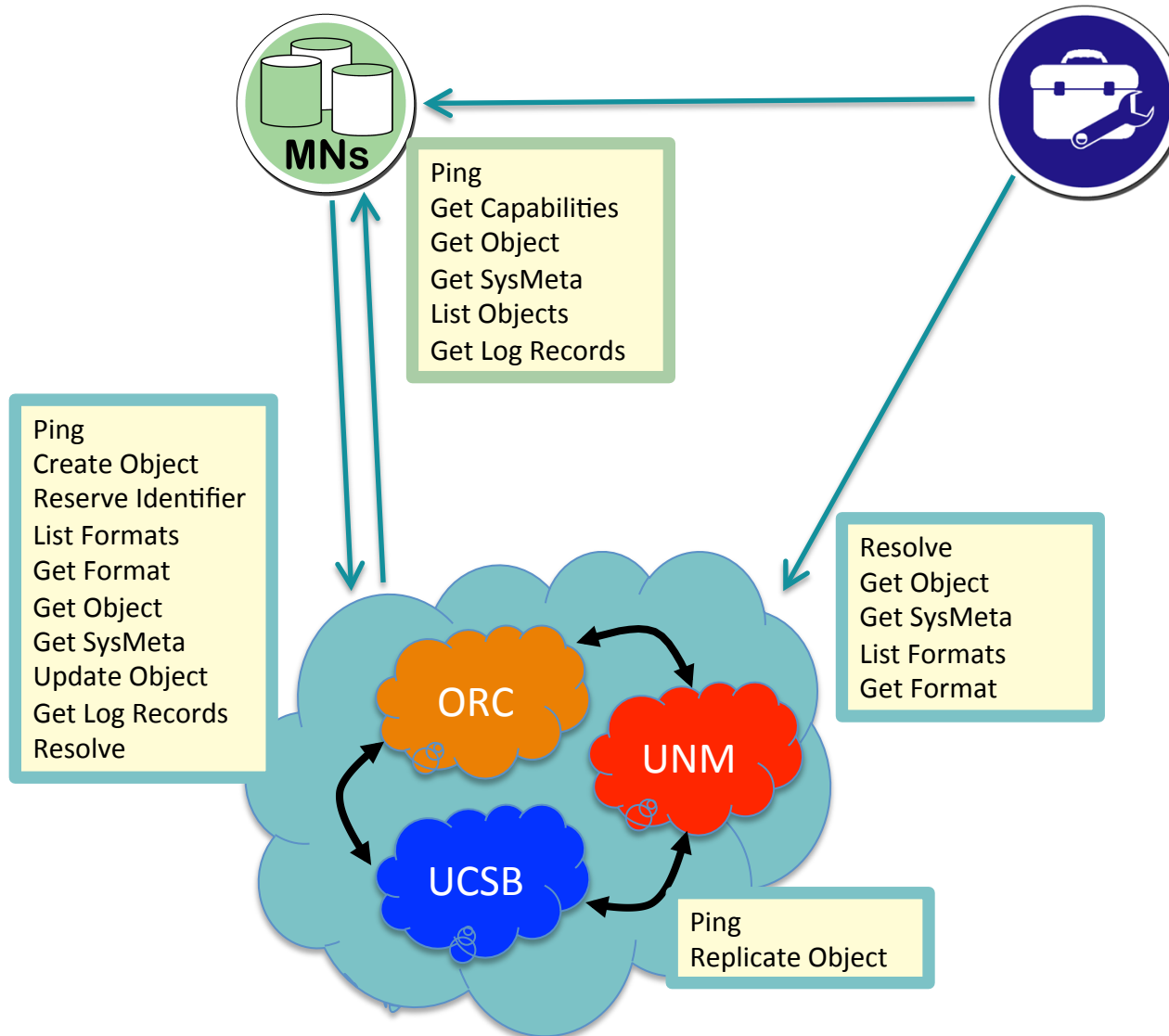


Scoping: What's your target tier?

- Tier 1: Read only, no authentication
- Tier 2: Read only, authentication (restricted content)
- Tier 3: Write enabled (you choose who)
- Tier 4: Write enabled, replication enabled

Some orgs choose to start at Tier 1 and add features later

Tier 1: What does this mean?





GETting data: the main target

Where can I get doi:10.5063/AA/mbauer.75.1?

<https://cn.dataone.org/cn/v1/resolve/doi%3A10.5063%2FAA%2Fmbauer.75.1>

```
<?xml version="1.0" encoding="UTF-8"?>
<d1:objectLocationList xmlns:d1="http://ns.dataone.org/service/types/v1">
  <identifier>doi:10.5063/AA/mbauer.75.1</identifier>
  <objectLocation>
    <nodeIdentifier>urn:node:KNB</nodeIdentifier>
    <baseURL>https://knb.ecoinformatics.org/knb/d1/mn</baseURL>
    <version>v1</version>
    <url>https://knb.ecoinformatics.org/knb/d1/mn/v1/object/doi:10.5063%2FAA%2Fmbauer.75.1</url>
  </objectLocation>
</d1:objectLocationList>
```

CN Resolve
method call

MN Get call to
retrieve data

A DataONE GET (digital object) call

<https://knb.ecoinformatics.org/knb/d1/mn/v1/object/doi:10.5063%2FAA%2Fmbauer.75.1>

MN REST Endpoint

Method

Object Identifier



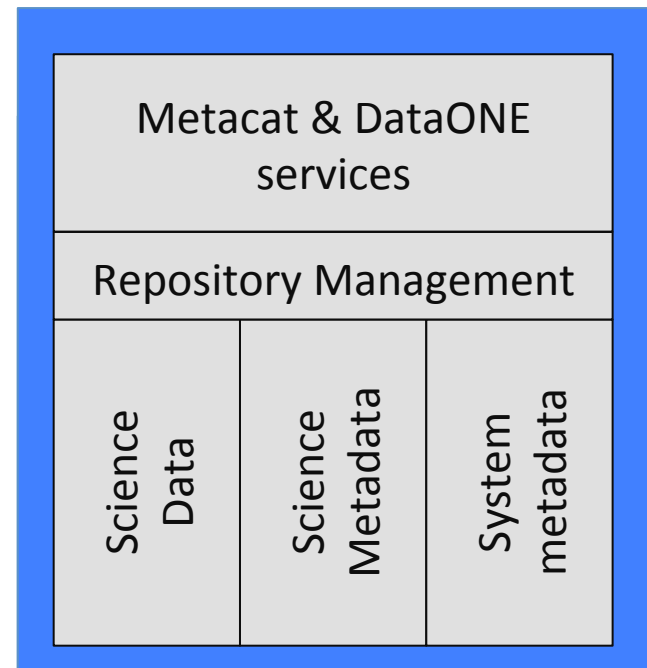
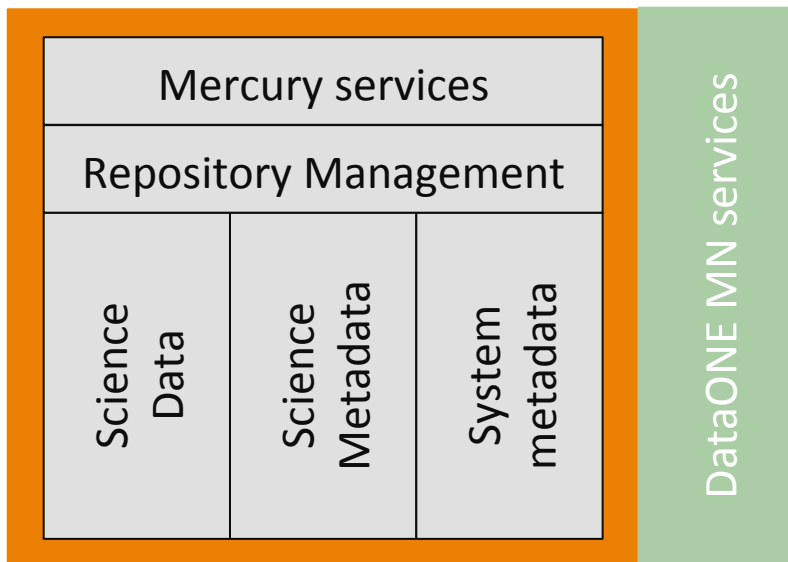
Scoping: Replication & Data Updates

- Replication is MN to MN
 - Source MN controls what (data) and where (MNs)
 - Destination MNs accept/reject replication requests
 - Peer-to-peer off-site backup
 - Log aggregation so source MN still gets credit
 - Prestage for data for HPC, service users, enable services
 - CN does the bookkeeping to track replicas
- Data Updates
 - Data can be obsoleted
 - MNs control what previous versions are kept
 - Reproducible science: all versions desirable
 - Reproducible science: clear user ability to detect change



Scoping: What are the software options?

- Use an existing “MN stack”
 - Generic MN (Python, Tier 4, reference implementation)
 - Metacat (Java, Tier 4, full source code available)
 - (future) DSpace, OPeNDAP, lightweight file based, slender node
- Adapt your existing software





Developing and Testing

- Can be simply configuration (if using existing MN Stack)

6.6. Configuring Metacat as a Member Node

Member Node Services

Enable or disable DataONE Member Node Services for this deployment

Enable DataONE Services ☒

- DataONE Staging environment
 - Check data packaging
 - Check metadata formats and parsing
 - Check software against specifications
- Support
 - Redmine.dataone.org (ticket tracking)
 - Repository.dataone.org (all DataONE source code)
 - #dataone on irc.ecoinformatics.org
 - Member Node coordinators (Laura and Bruce)



Are you certifiable?

- Certificates a common area of confusion
- Server (SSL) certificates
 - You get these for your web site, from a common CA, for https
- DataONE client certificates
 - You get these from us
 - Authenticates your servers to DataONE servers
 - Different certs for stage and production
- User authentication certificates
 - For a specific principal (person)
 - Issued by CILogon
- PGP keys (most of our developers have them)
 - For secure communication



Moving to Operations

- Register MN in production
- Harvest metadata
- DataONE and MN both check results
- DataONE and MN agree on publicity
- MN metadata visible in search interface
- Make announcement
 - Links on websites, news postings, social media
- User education -- how to use ITK elements



Sustaining Operations

- MN Operations
 - Add data, metadata, packages
 - Use MN calls to specify downtime
 - Keep MN info (contact points) up-to-date
- Staying connected
 - MN Forum (birds of a feather)
 - DataONE Users' Group
 - Ask.dataone.org community site
- More technical issues
 - #dataone on irc.ecoinformatics
 - developers@dataone.org mailing list
 - Redmine.dataone.org (ticket tracking)
 - Stage environment can be available for further testing

Questions?