OK - I plan on joining Spring 2013 Usability & Assessment / Sociocultural Working Group Joint Meeting
Member Nodes Subgroup

Twitter sharing during the meeting @DataONEorg #SCUAwg
To Contact Rama for inclusion into this group - dial him directly at 301-614-5356.
He is available from 1:30 - 2:30, then potentially after 4pm ET. He will use the Etherpad.
Rama: You can contact me at 865.924.9661

Current Meeting Documents

MN Policy DRAFT

MN Checklist (this is higher level for public use)
http://mule1.dataone.org/OperationDocs/member_node_deployment/mn_checklist.html

and MN Procedure DRAFT (this is detailed for internal use)
https://repository.dataone.org/documents/Committees/MNcoord/Coordination%20Work%20Area/SC%20UA%20Joint%20WG%20mtg%2030Apr-2May/DRAFT%20Member%20Node%20procedures.docx

MN Persona Template
https://repository.dataone.org/documents/Committees/MNcoord/Coordination%20Work%20Area/SC%20UA%20Joint%20WG%20mtg%2030Apr-2May/DataONE%20Member%20Node%20Persona%20Template%20-%20DRAFT.docx

MN Persona DRAFT (for PPSR-Public Participation in Scientific Research)
https://repository.dataone.org/documents/Committees/MNcoord/Coordination%20Work%20Area/SC%20UA%20Joint%20WG%20mtg%2030Apr-2May/DataONE%20Member%20Node%20Persona%20-%20PPSR%20-%20DRAFT.docx

Past Meeting Documents

Presentations from February NSF Review of DataONE (Reverse Site visit)
https://docs.dataone.org/member-area/documents/management/nsf-reviews/nsf-reverse-site-visit-february-2013/presentations_final_versions

Working Group Repository
Background and Related Work


PREVIOUS WORK From Joint UASC WG 2012” https://docs.dataone.org/member-area/working-groups/usability-and-assessment/meetings-usability-and-assessments-working-group/joint-u-a-sc-wg

Tier 1 member node: public data repository; public data only; no access controls; moderate availability level; uses Tier 1 API
Tier 2 member node: entry of authenticated user; log presence, use, responses of member node (including authentication), delivery of data
Tier 3 member node: authenticated user with write access to contribute datasets; public and restricted data
Tier 4 member node: responsibility to back up their own and possibly other Tiers’ data CCIT WG will review”


Subgroup work products:

Task lists:

Meeting Agenda: (MN portions)

Block 3: (Tue: 1:30-3:00)
Block 4: (Tue: 3:30-5:00) Persona Discussion start: introduction; brainstorming; selection; writing assignments
Block 5: (Wed: 9:00-10:30) Document discussion: Ext web: procedure; policy
Block 6: (Wed 11:00-12:30) initial report out: Continued Persona development
Block 9: (Th 9:00-10:30) MN process assessments - what are important metrics? (analytical, anecdotal, general)

background:


Block 10: (Th 11:00-12:30) Discussion of Member Node Scaling limits; how to increase MN count; "MN lite"

Block 12: (2:00-3:00) Subgroup report out

Meeting Notes: (Free to all to edit and add notes - this is the main reporting for the meeting)

Block 3: (Tue: 1:30-3:00)

Block 4: (Tue: 3:30-5:00) Persona Discussion start: introduction; brainstorming; selection; writing assignments

Block 5: (Wed: 9:00-10:30) Document discussion: Ext web: procedure; policy

Block 6: (Wed 11:00-12:30) initial report out: Continued Persona development
Block 9: (Th 9:00-10:30) MN process assessments - what are important metrics? (analytical, anecdotal, general)

Background Info:
DataONE PMP: https://docs.dataone.org/member-area/documents/management/project-management-plans-pmp
Goals yr 5 (yr 4)
40 (20) MN's
60 TB Storage
1M Metadata records


Block 10: (Th 11:00-12:30) Discussion of Member Node Scaling limits; how to increase MN count; "MN lite"

Block 12: (2:00-3:00) Subgroup report out

Issues not discussed elsewhere (i.e. "Parking lot")

Rama's question during Tuesday Morning: Don't forget to use prior outputs of previous SC&UA WG meetings.

BEGIN CURRENT E-PAD

Member node breakout

Who is present:
Todd Suomela
Suzie Allard
Kevin Crowston
Tanner Jessell
Robert Waltz
Amber Budden
Holly Mercer
John Cobb
Rama - virtual for parts of meeting
Ranjeet - virtual for parts of meeting

Thursday morning - reconvene
Usability and assessment metrics; scalability

Policy Documents we will be Reviewing - check under current meeting documents (scroll up on e-pad)

Not rigid policy; guidelines is a good word

Checklist; External Web presence (Tomorrow Morning)

Today is the notion of "Personas"
This group has done much in generating user personas
When we say "Member node persona" it is the description of an organization - an "Org-Sona"

Amber Owens will present on what Laura Creekmore and John Cobb have been working on - we will generate a group of candidate personas to think about, vote, then spend the bulk of the time writing draft personas.

Had the notion that there’d be more people, 8 or 10 - 12 personas. This is an arduous process, so we will be satisfied to have drafts for a couple.

Pick your collective brains!

Questions are welcome during Amber's Presentation.

Start-up Question: Tiers that we have talked about in the past - Tier 1 through Tier 4 - both implemented in what we are doing operationally, a big part of how we will organize the candidate member nodes. Different member nodes will want to participate differently.

**Member Node Personas - Slides**
Amber Owens is a masters candidate in the UT School of Information Science. Part of the SciData IMLS program scidata.sis.utk.edu Interests: User-Centered Design and Sociocultural impact on cognition and interfacing potential

**Slide 1**
MN Personas - help define / refine MN coordination activities
Personas were informed by Stakeholder network input on potential MN personas
Consistency of personas at DataONE

**Slide: Templates**
DataONE usage scenarios
Data Conservancy Scenarios
Data Conservancy profiles from Illinois and Purdue

**Slide: Features**
Positive feedback on member nodes
What a good persona should have:
Background: name, education, age; socioeconomic class and desires; life or career goals, fears, hopes, attitudes; Reasons for using DataONE to share and reuse data; needs and expectations, skills... (more).

**Slide: Process**
Goals - are we answering the needs of the users, how do we expand the user base
Methods - who are we servicing? Spreadsheet of potential member nodes, attempt to categorize (Identifier, URL for external web site) categorize by type
Guided scenarios from the template

**Slide: MN URL and Classification**
Name, Description, Location, URL
Just a way to lay out who the constituency / stakeholder might be
Who are we trying to reach out to?

**Slide: Classification Schema**
Different kinds of repositories, organizations, location and classification along with name, description.

**Background**
Informing the process: developing the persona in the exercise
2009 - Scott Ambler - Intro to Persona

**Model**
Data Curation Profile (Online in DataONE docs)
Talks about user needs, but mainly focusing on the data, a good template but taken a step further in trying to figure out what to do with the data and how to facilitate best practices, growing each organization

A user scenario on the DataONE Docs site, this pinpoints the different kinds of users, research scientists, librarians, activity within the DataONE project.

Sufficient for the needs of the scenario, but taking a step further with informing the practice and meeting the needs of the user.

Alan Cooper - The Inmates are Running the Asylum - What makes a really good persona?

**What are the Categories Established?**
Institutional Repository
Discipline Science Repository for Researchers
Government Repository
Individual Investigator Repository
Replication Node
Public Participation in Science Repository
Remote Sensing Large-Scale Data Repository

Q Are these the categories of Member Nodes that we are to consider? Also, these are categories dreamed up, we may add more.

Kevin: Maybe pick a few that are prototypical instead of trying to abstract.

Personas need to be specific. Comparison is individual researcher that graduated from Cornell in the late 1990s.

Robert Waltz: With personas, there is a level of specificity that is encouraged, but they are also useful in a general way to talk about general characteristics. If we get too specific in a description, maybe a potential member node will exclude themselves from the dialog. Might be taken too literally.

Kevin: Point of personas generally speaking is more for the design team. If you are thinking about a researcher interacting with the system, think of this person who has these specific skills.

Here's why some repository that hasn't thought of it.

Robert: Also, personas are for advertising

Suzie Allard: If you were thinking about being a member node, it is easy to think of this and this and this in a way that tech people are looking at in a different way.

Robert: Personas are fictitious, based off of user data, imaginary friends in the design world. Trying to create the same type of thing, imaginary best friend member node.

Fictionalized e-bird versus real e-bird?

Freedom to fictionalize something about e-bird

Profiles of typical member nodes.

We may have a small repository versus a large repository.

Kevin: Did that with researchers, early career, late career
John: might have a funding change - characteristics of member node might be captured within the entire persona

Kevin: there are many more ways that organizations can change.

John: Person might have an spectrum of opinion

Kevin - organizations might be merged from several

John: a few features should rise to the surface and be prominent

Set of things to ask for organizational things?

A few more slides will cover that.

Salient questions or characteristics:
Who are my users
What do I value
What do I accept
Curation Level - OAIS model
Where do I fit in the data life cycle?
Fixity - data and the state of change - object permanence
Tier Placement
Funding

Tanner: What is fixity?
A: permanence. Assurance that retrieved data is the same as stored data (including retrieved replicas and retrieval across time)

Template components:
Description
Users
Data
Resources / Funding

Expectations - Create 8 templates and narratives
Look at components and sub-components

Comments:

Tanner: What about explicit Geo-spatial oriented repositories.
The concept of a data package was interesting for me.
For GIS, the package may include many files in a package.

Consensus: Yes important.
It might be a characteristic of a repository
It may have special features in terms of searching and retrieval

Geospatial cuts across many issues

Examples:
USGS national Hydrology dataset
Cornell University Geospatial information systems (CUGIR)
EDAC (a prospective MN)

Amber: do we need to add industrial

Kevin: Looking over the human personas, the characteristics that we try to collect are name, age, education, socioeconomic class, desires, life or career goals, reasons to use... etc.

From there we went to usage scenarios.

Strikes me that almost all of these could carry over to the organizational one (name, age, history where it came from, maybe class and desires (mission statement), funding levels, goals that a repository is trying to accomplish) fears, hopes and attitudes.

Fears is an important one: what are they afraid of?

Smaller might be losing funding, unable to sustain; conversely a large one might have a different set of fears.

Understanding those might inform DataONE on how to support.

Tools that researchers are going to use: very different set for the repositories; what might they use from DataONE?

What skills do they have? Technically sophisticated repository, equal partner, versus a group without a lot of resources, need a lot of help from DataONE where DataONE essentially helps professionalize what they are doing.

List of categories might re-think some of the dimensions, repository that has a steady source of funding than one that has been scraping by on overhead, technology is not that sophisticated.

John Cobb: Agrees. Social parts about integration of project, member node growth plan, organizational ranking to lead to success or failure.
Kevin: Think of as market segmentation - which go after immediately, which are the longer-term targets. "Hand Holding" to get up and running, versus someone who has more data, more technology, and more staff. A hypothesis.

John Cobb - good to characterize, but difficult to approach.

Kevin: Useful data divided by hours of integration work

Suzie Allard: Is it a leader member node? Is it worth the investment if they will be a leader of other member nodes?

Vis-a-vis other organizations.

Suzie Allard: D-space has a huge reach within the library community - may have 100 extra people.

Kevin: something the organizations have that people don't is the base system. If we can find someone who is really visible in genetics / oceanography / hydrology what have you then we suddenly have a higher profile.

More of a marketing use of personas than a systems development use of personas.

Allard:
Cross over - do they need to be standalone or can they have some crossover - important if looking for buy-in from a community of libraries, that would be a different kind of organization, an institutional repository type.

There are different fears and challenges at an organizational level for those challenges.

Could have value for if we know their fear is "not having a tool add-on to do x" because that is what their subject group needs, we don't have the tool we need.

Robert:
Ability to transfer their user base could be a fear or impediment. All XML is marked up with the users specific credentials, if a MN will not join InCommon, will users lose ability to access information? Losing control of the data.

Kevin:
Suggests making a list of fears that organizations have - losing control of information, losing privileged connection to users - concern that DataONE will come between you and users.

Don't think the service being provided, newspapers are concerned about Google News; "we send you links" but they never know the headlines came to us."
Allard: Holly Mercer is connected to the Academic Library

Holly Mercer: Investment in infrastructure that people don’t use.

John Cobb: Territorial issues.

Amber Owens: Fears and Value - looking at external website. Ask them?

Kevin: user persona, a number of people had been in that role and could write it, for others like data librarian, you would go off of the interview.

A first draft which sticks pretty closely to repositories. Closely on E-Bird, case study on e-bird is 30 pages. Has a fair amount of goals and technology.

Write something that you try to generalize across, identify what the population is, interview some sample, then try to abstract.

Not convinced this is going to be the most illuminating for the purposes - variation is more important, well set-up technically, good resources versus not, large collection of legacy data versus more flexibility about the data.

Productive:
1) What do we need to find out? Organizational versus individual persona
2) Thinking about points of variation - representation of the different kinds

Amber: Small versus large?
Kevin: Perhaps.

Life History would be interesting.

How it came to be, how it is currently run.
Who am I.
Hopes and fears - what would the organization like to become? What is it afraid of?
In general and around DataONE.

Smaller repositories might be more at-risk from funding.

John Cobb:
People have mined their database and sold without knowledge.
Influences interactions with DataONE.

Tanner: are the institutions happy with their current system.
BePress versus DSpace- front end for other places that they put data. UT Libraries has discussion.

Kevin: Reasons for using DataONE?
Skills or internal capabilities - what is this organization good at.
Could be an entire section on "What will this organization "get" out of DataONE?

Gets at the fears and hopes - what is it they are afraid of that DataONE is going to help them do better.

John Cobb: Bullet on "how do I perceive my role within DataONE"

Kevin: Organizational Capabilities - where do they have a lot of capabilities, where are they missing capabilities.

Individual investigators - "implement the stack" how / why should we do that?

I understand why that would help, but it’s just me, not sure what you are asking me to do?

Should be highly specific, point value on some distribution, but it really is the description of some individual.

Suzie Allard: We have the data to do a persona on libraries - interviewed librarians, had a separate questionairre at the organizational level - we do have hard data.

Kevin: Libraries are likely to become member nodes?

John Cobb: Yes, institutional repositories.

Suzie Allard: Most researchers are interacting with an "information organization" along the way (data center is different from library) ORNL does have a library... does not handle data.

We do have hard data on academic libraries, has a huge potential for DataONE in multiple ways - lots of researchers are seated there. Housing most of them. Way of keeping their institutional capital, intellectual capital. Huge opportunity, and a potential funding stream from libraries, they are used to paying subscriber fees for other people’s data, if there is something there to help them expose their intellectual capital, best practices for exposing their own data, that could be lucrative - not to say that we are taking money, but we can be inexpensive next to other solutions.

John Cobb: Another fear and hope - undermining of funding. Would not say it out loud, but "joining dataONE leads to a diminished recognition of what we do?"

Tanner: purpose of each repository - type of data offered by member nodes should match up with how DataONE connects users to Data.

Robert Waltz: Fictional representatives, framework in our mind for who our customer base is, discuss with others, how they work with us. Useful in promoting DataONE

Allard: Marketing obviously has value; were there far before computer scientists had them.
John Cobb: Look at member nodes already on our Radar.

This is a redmine.dataone.org/rb/master_backlog/mns

List of operational member nodes.

Things about the reverse site visit - end of year four, first half end of year 2. In process laundry, deploy by end of year four. Other things: not written in stone. Just a shared.

As member nodes, this could inspire us.

Ticketing System - backlogs of things to get done.

This would be things you deploy to the incremental rev.

Turned in to kind of a scheduling tool for targeting.

20 by end of year 4, 40 by end of year 5.

Q5 - person to help, waiting on something to happen.

Member node by some resarcher - none in that list.

Closest is ONeShare - As you take a spreadsheet, do DataUP.

Kevin Crowston has a Terabyte of data - currently use Google Code

Robert - we know these people exist because "we know them."

The amount of work to bring them into the fold - not realistic.

Member node in a box.

Some standardized way, queries that could be written in some standardized form read by a harvester and slurpped into a DataONE member node (Github, Drupal, Google Code, Figshare)

John Cobb: DataONE is not in the business of providing direct archive collection services to users of data - we don't scale correctly; we don't want to poach on our partners' scope.

We find ourselves in a situation saying, thanks for coming to us, we have a member node set up just for what you do.

IN one instance, we did not, that's how OneShare came about.
How will the query tools expose the member node data.

Kevin: Operational List: fairly large, fairly well institutionalized, technically savvy,

John Cobb: Opportunistic because they are metacat?

Robert Waltz: SANParks is not necessarily well-funded.

Pathfinder is an African Member node.

Dr. Allard: from personal experience concern seen in South Africa is keeping up with the rest of the world, and researchers do not have support to collect data. Not sensitive about exposing data.

Kevin: Biological samples - Brazil (why are american drug companies coming down, discovering neat stuff, turning it into a drug, selling it)

Suzie Allard: when dealing with an actual sample, has only a specific amount of slices that you can get out of it, there is a whole lot more territoriality on physical samples.

Robert Waltz: Should we export data to another continent and not have it on our continent

Reconvening After the Brea
Choosing some:

Template- draft a member node persona -

Example from what Amber handed out - Laura and John Cobb assisted.

Persona Number 1 - eBird
Amber had a set of questions; she answered the questions for E-bird as a chrononical EPSR
Abstracted that into an attempt to make concrete personas.

Persona Number 2: What should that be?

What category, how can we answer questions

EDAC - Like (http://edac.unm.edu/) - a discipline science repository, GIS oriented or aware

Including Search and Query

III. How is it Funded
Multiple funding streams; single archive

IV. What is the user community?
Federal agencies; state, local and tribal governments; professional societies, organizations; and advisory bodies nationally and internationally

V. What do they support?
Resource management, scientific research; GIS oriented but specific to New Mexico.

VI. Access Method
GIS Clearinghouse, File type, how best to connect the user to data
Raster Data, Vector Data, metadata files that come with it

What are we not capturing in the description form that we are also needing for a persona?

a. Hopes / Fears
b. Operational Age
c. Skills
d. Size of Collection (800 objects, well used)
e. staff size

Suze Allard Suggest change in direction: Prioritize categories by where we can recruit 20 more - should we discuss categories without agreeing that these are the right categories; are we missing any; should these be taken out.

Pulling the Categories from Earlier in the Document:

- Institutional Repository
- Discipline Science Repository for Researchers
- Government Repository
- Individual Investigator Repository
- Replication Node
- Public Participation in Science Repository
- Remote Sensing Large-Scale Data Repository

Rama's initial reaction is that these may overlap.

Example: Large repository / Government repository
Most Remote Sensing are Institutional;
Is institution really an educational institutional?

NGO vs. Academic Repository?

Is there a difference between a government repository at state, federal level and an academic repository.

Rama: Specialized repositories that hold only one scientific discipline (e.g., only Shakespeare's works)

Remote Sensing: Continuous Data Flow, mostly government institutional kind of repositories
(Or at least funded by government)

DataONE will not be a repository for everything, but to the extent possible we would like DataONE to be able to interoperate to the extent possible, but not everyone who interoperates needs to be a member node.

Who is funding a lot of this data? Do we look at where the Data is coming from?

Suzie Allard: Problem with categories from different dimensions
- source of data
- type of data
- funding of data acquisition
- administer of the repository
- infrastructure node. Service provider, e.g. member nodes,

Rama: look at tiers, capabilities that they are going to have.

Robert Waltz: Justifying categories that we have

How do organizations represent themselves to the public? How do they describe themselves?
Select a few examples

Allard: Discipline science can be the organization in charge, public participation.

Waltz: Replication node is certainly a data one defined type of thing.

Cobb: Series of entities (besides DataONE) that provide cyberinfrastructure as a service

The type of resource available is really along a completely different dimension - warehousing or bridging service

Rama: Sounds more like a technical skill or capability.

Cobb: Has a lot of sociocultural differences - concerned with Dataset created or provided. Infrastructure node is concerned with the service provided.

May be providing "brokering service" for nodes providing datasets to node providing datasets.

Waltz" Organizational affiliations - organizationally how do we define DataONE.

John Cobb: Size, Quality, and a Type Quality could be geospatial.

Rama: is there a summary?

A: The word file on the WebEx is the place where this is being included.

Suzie Allard: NGO might be a characteristic to describe some of these categories. We can hit some / talk about (if you thought of this as a matrix, different characteristics, funding streams, NGO, but look at commonalities and difference)

Rama: Phrases for categories; attributes for entities.

Change the word "public participation in Science repository" change to something like a "noun phrase." (Changed Public Participation to Citizen Science).

Cobb: Kevin used the "Weirdo" category in that it would stretch our notions - museums, iDigBio (not citizen science, but somewhat contributory, high metadata to data ratios, might be a photo or description of a sample at a particular collection location).

Waltz: Museum or other public works?

Rama: Individual PI repository

Waltz: we had an individual repository but lost it to editing -
Cobb: To take away: Assignments to answer the template or extend the template for a particular example in one of these categories, get back here tomorrow.

Allard: take categories, try to put down some of those attributes, see if we defined those categories so we see commonalities.

Cobb: Amber began doing that - the rest of her slides talk about these abilities or qualities.

Kevin’s comment was a lot of the exact questions we were asking of the user personas - equally applied. We should include that original template as well.

Presentation gave the URL to a lot of them.

Hanging tasks off of people in hopes of getting something by Thursday. Amber laid out a whole list of stuff to do - travel the whole thing back and forth and then re-group.

Tanner and Amber can do large geo spatial data repository. (Rama can advise)
Suzie and Holly can work together - Academic Institutional Repository

Robert Waltz and John Cobb - will work on Replication Node.
Todd will work on something like the "Cultural Heritage Repository"

Ever have a private corporation as a member node?
If I am a private company doing overhead information, my funding model is I gather information and sell it to people. Overhead information systems are third party payor, where agency or institution collects and makes it freely available, it is funded out of another mission.

Spreadsheet document added to e-pad.
https://docs.google.com/spreadsheet/ccc?key=0AmuOOMpSMNgLdDRNMERjOVVPUnlKcXhuSHZQNTLUEE#gid=0

candidate member node URL's and classification

Things that did not fit the mold... some time Thursday but doing other topics.

Going to Amber's Extensive Research (2013)

Rama Re-joined the break-out at 4:11 and the WebEx

About to be looking at <http://mule1.dataone.org/OperationDocs/member_node_deployment/mn_checklist.html>
Couple of personas selected to start trying to draft.

Wednesday Morning:

Policy Discussion:
https://docs.dataone.org/member-area/member-nodes/coordination-work-area/scua_wg_20130430_materials/mn-policy-draft-v0.8.0

External Web study:

Process/Procedure

What should be on this list?

**MN Checklist (this is higher level for public use)**
http://mule1.dataone.org/OperationDocs/member_node_deployment/mn_checklist.html

Governing that:

**and MN Procedure DRAFT (this is detailed for internal use)**
https://repository.dataone.org/documents/Committees/MNcoord/Coordination%20Work%20Area/SC%20UA%20Joint%20WG%20mtg%2030Apr-2May/DRAFT%20Member%20Node%20procedures.docx

Road Map to a Member Node
20121217_MemberNodeImplementationPlan.pptx

https://docs.dataone.org/member-area/committees/external-advisory-board/2012_17-18dec_eab_meeting_washingtondc/ppt-for-december-2012-eab-meeting-member-node-implementation-plans/?searchterm=20121217_MemberNodeImplementationPlan.pptx

Decision Points along the chart.

Guidelines, Suggestion might be weak, but something firm might be giving boundary conditions external to DataONE.

Talk about Policy Based on Graph from Above.
Solicit Comments
Go through Exercise to look at external web space

From a perspective member node's point of view, what do you need to know about the process.
Allard: Promoter of best practices and not enforcer of standards. Coordinator seems like a "big guy who knows everything and runs roughshod over everyone." Caution on enforcement role.

Crowston: implications - if a member node wants to do something dataONE does not believe is best practice, we will not say "you can't do that, we're not going to let you in if you don't do it that way. Is that the implication?

Allard. Yes, the issue to tease out is between the technology, political issue, from technology is to say this is the best version to make things work smoothly.

Crowston: large amount of resources, may not happen in the short run.

Some policy that has been adopted, in the opinion of application evaluators, that's not a best practice for data management. e.g. update files whenever, if they want to update an item in the repository, they should just do it.

DataONE would say - that's not really what an archive means.

Cobb: Looking at the e-bird guys - they take observations from anyone - including joe six pack who just saw everything on his life list.

Crowston- e-bird does vett things that would seem unlikely. They don’t delete, it is your record, but they may just flag it as unreliable.

Cobb: opportunity w/in dataONE ecosystem for external annotator. Likely, unlikely, credible but unverified.

Miriam: Recognizing member nodes that were following in certain ways - credit or a stamp for doing that.

Waltz: argument against minimal, if you don't have well formed XML we are going to throw it out. So far, if you are creating well-formed XML, for some instances on the backend, valid EML, DataONE can accept technologically what is provided. May not be worth indexing for discovery purposes, but we can hold on to it just in case in the future you get back and want to provide an update to make it more usable.

As a repository we have the capacity to be a part of someone's workflow and create data, not just preserving the product for the long term.

Might want to put in policy - we are validating XML, not only that but you better have these 10 fields in your record or else...

Crowston: Application, assessment of readiness, technologically we don't think you are ready but hey, whatever. As opposed to come back when you are ready.
Waltz: level of tech - must be able to talk to each other.

Crowston: you don’t have good metadata, we can’t figure out what’s in the repository. Come back, or sure, whatever.

Cobb: 1 decision process becomes - when dataone reaches agreement - how should we be making that decision? Need many people to weigh in.

CCIT is "running the show" because of a vacuum. See things from on high, strategic needs, but CCIT says this will not work unless you have well-formed data packages.

From an evaluation standpoint, as we get to 40 member nodes, there needs to be a process in place. Part is quality issues. Couple questions. How do we make that decision?

Allard: Discussion of this before at the second DUG. Afterwards, came to an agreement, this is in the leadership team notes somewhere. Suggest that we should carefully form a research question that we are talking about. Where have we stood before? Where do we need to go - talking about scalability, basically the same thing, we have made decisions to help us hit our marks - strategic decisions, talk about different qualities - high-value, high-risk data, ability to bring in a large amount of data, one node, or with associated groups that come in with it. Prioritized somewhere.

Goal should be to find that or talk about in the same way that we talked about it before - recapture the investment. Talked about opportunistic: low hanging fruit. Just "who we were associated with" then metacat.

High Risk, technological readiness. Then there was a shift that the most recent discussions would be ones that would grow the network in records or number of nodes, most quickly.

Brought up earlier - terminology issue. Careful about technological quality in all aspects: content, tech. Weigh against the ability to be welcoming and inclusive.

Mriam: Document what the process has been?

Allard: if not document, take into account that we have talked of reasons for bringing people in.

Implication that they have well-formed data.a

DAAC does have a lot of good things, but tech issues.

Cobb: have a discussion, reconnect with all stakeholders, present to DUG, once and for all say this is our control document so in the future we don’t say "remember back in 2011 we talked about."
Miriam: essentially a decision tree.

Allard: We have specifically avoided a decision tree.

Cobb: you should consider, you should strive to - one reason the language is kind of "wimpy"

Allard: allows greater flexibility, technology may be ready but comes down to a personnel issue, a decision tree does not account for that. E.g., making decisions that were not going to be implemented.

Cobb: project talks about this a lot before, brainstorm of the different discussions - DUG meeting linked by Robert W.

Miriam: helpful to list criteria to be considered? Topic, quality, quantity etc.

Cobb: Three things: make contact with institutional history 2. issues just mentioned 3. what is the process for reaching a decision for moving forward with a mn.

Allard:

- data
- diversity
- st. part
- leadership
- management
- Technical

From the document "MN_Prioritization_2011.07.09.pptx"

N0t taking lab-created metadata, must be some standard.

Miriam: from slide 17, is that publicly available out there on the Web site - minimum set of MN requirements.

Allard: When mike made some of those early handouts, that’s what this was based on.
Waltz: metadata format used by the candidate is used by DataONE. We have mn’s coming in that does not have it - 2011. Thinking about this at the time. Can be supported may be a better way to discuss it. Negotiation between dataone and member node.

Introduce: Chelsea Williamson Barnwell - working on a paper for a class - STEM communication, current member nodes, help classify, set parameters for doing personas.

Crowston: proprietary, but important data is of high value and public.

Mriam: not a quantity issue

Allard: Public data in collection of value are available and shared upon request. People will share it, public in that sense, you must request to get it.

Crowston - project adopting that - physical access could be you made publicly available, or that you have to send an e-mail and ask.

Mriam: does it describe somewhere what that basic level is?
Cobb: we can point them to repo mule
Mriam: maybe create something that describes that, define what basic level is

Robert Waltz: at last CCIT meeting, even the last point, can kind of debate, the idea of tier 1 member node implementation might be too high a level for some mns, talked of creating a member node light. Have implementations of some other member nodes. Although it is good to keep that in, that is even subject to change - this "lite" version of a member node.

Cobb: important, unsure how to re-word to capture that.
Waltz: may be a
Crowston: might use a different "member node lite" if you don't have tier 1
Allard: if you go to tier 1 you go to being an actual member node
Crowston: "silver member"

Next slide; Key Characteristics - technical, resources, match the topic, could be that the member nodes are not in our sphere. If it matches DataONE’s earth and environmental...

Crowston: another characteristic of data that is not captured there - you can imagine the domains coverage would expand. Might not have been a priority initially but's in scope enough.

Mriam: Then talked about high priority - many member nodes competing for resources - some kind of prioritization of quantity versus priority

Cobb - we are going to quickly get to a place where there are more takers than resources to bring on MNs. We will need a way or prioritizing who comes first.
Crowston: isn’t idea anyone who wants to be a member node can be... do now vs. do eventually. Imagine a scenario in the future, common repository software could implement the member node software. Download DataONE D-space plugin. There, whenever ready to start harvesting

Cobb: question to pose to group. Important archive, requires 2 person years of dataONE. Implement member node, pay DataONE for quarter of FTE, carry their own freight. Marginally important to Dataset. Take one or the other, or both?

Waltz: that amount of effort comes down to “do we have the funding stream to support that.”

Cobb: Find some company which you are not aligned to, decide they want to fund that. Crowston: hard to imagine DataONE saying “we don’t want your data” from Elsevier.

Cobb: DataUP was a bit of a trade - Microsoft carried the freight.

Crowston: last thing said about Elsevier saying ‘btw we’re doing this for all journals, some earth science data, but we can’t exactly tell you which files are which.”

Miriam: finding again to make sure we go back over, criteria is good/important. Question posed: what do we do in this situation - that situation, process is important. Do we need insight from CCIT, here, there...

Kevin: proposal: resource allocation problem, finite pool of resources, potentially large demand. Grant competition, you would essentially treat each member node as applying for a grant. Might have multiple pools of money. Some might say they need no money, they are self funding, others might need large amounts. Could send real amounts, send to a foundation or funding agency. Basically, you would have a review committee, here are the 30 grant requests, let’s prioritize, allocate this year’s budget for bringing up member nodes, if there are deserving ones, they can go to the next year’s funding allocation. - Ooh, I like that! - LM

Miriam: modification, assuming when talking about funding, really talking about resources. Dual process, initial ranking along criteria, of all the suite that comes in, those that rank highly are reviewed by people. There is a big political capital issue impossible to rank. Once candidates have gone through a ranking process, still needs to be some subjective looking.

Crowston: believes it is all subjective. Has review committee. NSF has “intellectual merit” and broader impacts. Intellectual merit of the data - opens up to a new community of interest, diversity of data, achieves an infrastructure for the next step, spent a lot of effort bringing this one online.

Waltz: objective criteria - software platform that already conforms to dataone API that gives a prioritization.

Crowston: some would basically say, “we’re done, accept us”

Waltz: give us a certificate so we can talk to you. That is an objective criteria.

Crowston: internal administrative review, but not peer review.

Miriam: question about originality of this discussion

Crowston: probably rolling, but for easy ones, does not require any resources, resource allocation decisions. "on this date we will prioritize" episodically analyze submissions in terms of resource allocation for the project. If it turns out you want external input, all ones will have meeting at DuG, set priority for next 6 mo
Cobb: absent other input CCIT has been doing this, x many dev, what should we focus on. Prep mth. for reverse site visit santa fe, put folks on redmine ticket system. Look and say, what about SEAD - what about Taiwan, Data Fed consortium, have not freed up resources, process occurs in ad hoc fashion.
Cobb: not exactly "defensible" I don't have 5 principles, etc.
Miriam: ultimately needs to be transparent to these potential nodes. huge question
Allard: needs to be set up in a way that is agile. does not block out someone who just came on the horizon and really is someone who should be online.
Cobb: when terrapin existed, group 10 centers working together. Inside, wrote a policy on what it takes to join. Inflexible. Expensive (tech requirement, 1 million network requirementt). Outside, this was a "walled garden" problem. DataONE is arbitrarily deciding, playing favorites - was a pitfall for Terrapin. People were advocating for defunding it.
Miriam: you need this process for making decisions, but also principles for guiding transparency
agility
defensibility
is the process "just" and "useful."
Suggestion of submission process.
Allard: talk about what the system is of addressing different characteristics. People have some idea of what's important. Some point depending on if it's transparent, what's going on with DataONE key personnel. May shift between diversity in terms of data at risk, may weigh heavier than other points depending on how many people are engaged.
Miriam: these are the criteria that which we consider - still transparent, these are the criteria this is the process...
Crowston: easier way to say is here are things that get taken into account, up to the committee to decide how they want to decide...
Cobb: reference in policy doc, there is a process - will take months to ratify
Crowston: app process should focus on letting proponent make case for why this is a deserving cast. Tech info on what is the kind of data, why is this data which is of value to the dataone community. might be an initial screening phase. CCIT- guess as to what kind of resources.
Cobb: NSF - point of contact is the PI. not much more than advisory, helps the PI decide.
Crowston: governance - who should be involved. DataONE user's group? can say we recommend d1 users', pi, giving suggestions as to what we think a committee would look like - that requires so little resources it would take more time to discuss when one should just do it. >>>who maintains the list of requests, bugs to submit information that is missing. <<<If you have the NSF analogy, who is the "program manager"
Cobb: multiple tracks.
Single form that people sbumit, low demand, request human resources, have a standard stack already implemented. NO value-added way for the committee to implement.
Is it a 2 page or a 15 page proposal How long should it take us to act.
How long should it take us to act, if less than 15 pages?
Allard: ballpark of 2 - 4 weeks is reasonable. if we take too long to act, tired of waiting, whatever level of excitement is there is gone. may not be offering. people will figure it out. not asking for any resources, expect response.
Crowston: Committee meets on the following dates. get submission in on following deadline. Get a response after committee meets to review. How easy to upgrade software.
Cobb: consensus - couple of weeks to respond.
Allard: at least a couple of weeks, seems totally reasonable.
Crowston: empowered to say if what you are talking about, metacat, implement software, empowered to say "you're in"
Cobb: one thing in draft, in process, member node coordination, it is a DataONE form, kind of help them get that in, Laura went through process, cleaning up member node descriptions.
www.dataone.org/current-member-nodes
http://www.dataone.org/sites/all/documents/DataONEMNDescription_USGSClearinghouse.pdf

OneShare would have a different answer.

DataONE generated this document to showcase member nodes based on site. Serves a lot of other purposes. Document used to have URI designators and stuff. Need for implementation.

Crowston: Could be like 1040 ez, 1040 a and 1040 where if all you want to do is fill out this form. Or if you want a standard implementation, already have metacat, then download the software, get it running, let us know, and you are done. IF you have weird metadata, weird software stack, fill out the 1040. In fact that is just the opening part of the conversation.

Most people would say, you get all the easy ones out of the queue, of the ones that are left, whatever the unit of scheduling is. What’s the timeframe for scheduling of programmer resources
Cobb: agile -
Waltz: usually at CCIT meeting once per year we go through backlog, see what effort/focus will be. Overall kind of focus is.
Crowston: Question is, if it is once of year, answer for larger scale implementation, get your application meeting august, every september at the all-hands meeting, move the discussion on.
Waltz: in terms of making more routine, room for enhancement.
Cobb: operational acceptance. CCIT are really ready to roll. at start of production, almost an outreach. can be something you open for grand opening. have some communication highlighting. new member node coordinated with the member node. like to see associated nuggets. Set a date for public release. Coordinate with CEO. Put in the news letter.
Crowston: on the form, ask for the mailing lists for the communities of interests - can send to the communities "hey we just came live" or they will know...
Cobb: one that came up, we will lose a member node at some point. What do we want to do.
A unique process. Reach a decision like signing a contract with a dying person. You may
have the member node come in, has data at-risk. If we talk about maintaining a scholarly
record, DataONE seems to make that data available in the future. Who would inherit the
data.
Cobb: situation can be different than one planned for
Vieglas (on break): dataone can be an insurance policy
Crowston: best practices
Waltz: Assign to anyone: wipe from repository?
Crowston: should be in check-list. if not willing to agee that in the event that you go out of
business, you will assign the data to someone else. Deletable 50 years, 25 years. Seems
incompatible with the notion that these things are permanent parts of the scholarly record.
ONly until the year 2065?
Cobb: edvage, gravitational constant. Important paper. Data is not relevant at this point.
Crowston: very different than most ecological data; in principle anyone could re-do the
experiment. Observational data about the ecosystem at a point in time, you either have it
or you do not.
Cobb: in data set point, important, bytes - ... - if we can get the community to do lifetime
planning process. Some are really hard decisions. not high priority. Absent that - letting
projects cease to exist.
Crowston: If you go out of business, what will happen with data. DataONE is "pointing" at
something.
Waltz: easy checkbox, if you go out of business, will you assign dataONE preservation
rights.
Crowston: in the unlikely event that your project goes out, who will you assign with a big
blank box. Then a check-box with "i am willing to assign to dataONE."
Cobb: opportunity versus responsibility. Insurance means assuming risk. As a project we
can't assume risk.
Cobb: all i need to do to preserve data is just buy disk drives.
Vieglais: keyword: there is a protocol/process called Drambora - goes through process of

Breaking at 11:00 for 11:15 reconvene.

Discussions on doing DataONE web page - as a propspective or operational member node -
what do you want to see?

Cobb: Speaking personally - I don't go to DataONE's external Web site.
The collaboration is so large, most on the DataONE mailing list are probably the same
thing. As we get more and more high profile externally, in the context of member nodes,
what would I hope to find?

Where would you find member nodes.
If you participate- you find membernodes
Under resources - you don't.
www.dataone.org/member-nodes

Revise, put into place.

Question becomes: what would you like to see about member nodes.
See something over here that would list the set of member nodes that are active, and what their status is. One as a user - get information about status of nodes.

Take a screen shot of this page - this many member nodes running this dataset.
Looking at the left: you see the same menu as the drop-down at the top.

Who are the current member nodes. Could go for some higher res logos

Guidance - really needs to be revised. Idea is people who might express an interest.
Putting on "prospective member node hat"

Going around the room - collect and discuss for a few minutes thoughts on content on the page:
Robert Waltz:
Researcher coming to site: might want to see, in terms of a widget
Number of items included in the repository
Types of Data (formats, so if I have tools I might want to know what types of information I am receiving from the membernode).
Chelsea Williamson:
Criteria for the member node.
More of an interactive map - little nodes where Nasa has data centers, you can click a brief bio, visit the site for more information.
(2) none

Little Data Description documents.
Tanner (1): PDF averse, suggests that PDF takes too long to download, would prefer a short vimeo video with someone friendly. 1.5 to 2 minutes.
Tanner (2) watch text-heavy design

Allard: Should have reasonable production values, not necessary.

Some PDF documents describing member nodes, could do something else to grab people's attention better, be more effective.

Kevin Crowston:
In a couple of places, "for more information, contact DataONE" but does not say who to contact - e.g., the "Member Node Coordinator"

Allard: with a picture

Kevin: should definitely be on the 'spash" page, prominent up in the corner.
Kevin (2): up at the top - download member node fact sheet. Who is it intended for? And they are written for multiple audiences. Fact sheet includes detail about the API along with some more high-level information. Worth thinking really about personas - write documents targeted at specific kinds of individuals.

Amber: follow up - how do you articulate - short one pager, here's a more comprehensive ..
Crowston: executive summary.
Amber: concern if you do not fit, the level of capability of the audience is concerned.

Technical document for the PI.
Amber: tend to point to mule a lot for technical.
Allard: how would we shorthand that - easy to parse tech document?
Amber: consideration of audience
Cobb: jumbling in is "ask" website - can be part of the Website.

Amber Owens (1):
Talking about sustainability, long-term vision, potential costs of becoming a member node.

Allard: some level of expectation. what did it look like, idea of how to look at it.
Amber: workflow of the steps associated with becoming a member node.

Allard: what do I benefit, and what would it cost me.
Cobb: what do my users get?

Amber: how hard will it be - all part of benefit and cost.
Todd: either complete list of current member nodes or an example of highlighted ones.
Miriam: There is a list of things that relate to this that came into Rebecca through the current contact DataONE. May also want to look at what people really are asking.

Point is it is a way to crowd-source relevant questions and accurate answers. Some things will end up - ask responses. How do we integrate ask.dataone.org.
(Todd 2): list of Institutions who have member nodes. have that question

Amber (2): tier question tabulation
Allard (2): what does dataONE mean to me - subsuming, augmenting - member node vs. coordinating node. Excellent technically but troublesome politically
Cobb: what is the power structure
Amber Owens: pre-requisites - if you can't support these kinds of technology, then you will have to wait or bring better metadata. Pre-requ checklist.
Cobb: data collections that they want archived. If you have data that you need to preserve, you should go to a Member node - not DataONE itself.
Answer data owners question about where to archive.
Amber Budden: Clear statement of what a member node is, what a member node is for.
Cobb: maybe put something in Ask, then reference it. If I had the need, and came to DataONE.org, where would I go. I have two excel files I want to save..
Amber: You go to contribute, then you are told to go elsewhere.

Crowston: who is responsible for what?
More visible.
Chelseae - should the form go up?
Amber: there is some concern, there is tension around how much do you expect to do up front?
Tanner: MN interactive for readiness, contact
Redmine list:
Capture some of that in RedMine.
Cobb: prospective member nodes up front.
Allard: Keep transparency of things to be considered
Amber Budden - notes in the record.
Redmine: multiple platforms using, cumbersome, prone to error
Cobb: selection of working group members - we do not publicly make available results of the discussion.

Miriam: regardless of waitlist, public or private, effective vehicle. Public list on the web.
Cobb: variation of mushroom process. Teach staff not to create embarassing comments - "clueless user"

Dryad used: trello.com/board/dryad-development/4f9563b87072e3eb5f06951a

Sent to track.
Waltz: was on the short list of ones to evaluae, and redmine was the other one.
You can look at redmine not logged in, can’t add anything.
Prioritization is one field. They make a point, new release notes, positive development plan. Documentation, dryad. So, in 15 minutes.

Cobb:
Is there any organization that comes through. Page of text is not that useful. Some sort of structure. Is there a natural structure that is emerging?

Land on member nodes - have the option of become and current
There will be some reluctance changing the buckets - some URLS already posted. If it is critical then it would be accepted, the preference would be not to change.

Allard: Very fast, very glaring: why is it important; gain eyes, gain status, be discoverable, "people will like you"
Amber: Web site does not have it. Very verbal.
Allard: expect people to explore the page and figure it out. Mind enabled shortening that gets people hooked - fave way to think of it - tapestry you see pictures. NOT stiching and thread - seem to be suggesting that people examine the stitching and thread. For folks on the edge, must find a way to sell it to them. Better way to sell on the front.
Something that is needed - put some serious divide conversation into pieces - technical piece, things to be ready, interactive features that help someone see how close they are to even thinking about member node-ship, best practices versus Sheriff.

CONversation about recruitment - whole different conversation - don't even get to the point where they take a self-test to see if they are ready. Taking low-hanging fruit, have a bit more "Pizzaz" recruitment should be simple and sexy but in technology discussion you must be technical and precise.

Recruitment has come from word-of-mouth, maybe two people express interest via web

Tanner: even people knowing from word-of-mouth need some info to back up their idea to become a member node.
Allard: people within a long-term organization, in order to get the funding, go after and use their funding to become a member. Something to point them to and explain it.
Allard: not necessarily recruiting - should do information that fits the top two. Kinds of things hearing we keep coming back to.
Cobb: speaking to organizations, not users. Not convincing people to use Paypal. Example to point to -
Budden: not entirely true, we do talk to researchers.
Allard: ETDS - getting instances around the world - 3 forms of growth
Administrator - told people it would happen
Librarian - have to sell to administration (most successful)
Domains - students saying to advisors - grassroots up - a department might start something that got contributed as an institutional repository.
Institutional set: identified
University admin
Library admin
Faculty (tough, engrained behavior)
IT (we do what we are told/paid to do)
Students (easy)

In australia, they went national almost immediately. Big country on the edges of the continent. Germany - national quickly. Doing Info Rep. work for years.

Cobb: Don't "browse by viewing" think of as "had a business card"
Kevin Talking about audiences - think about fact we have multiple audiences. Prioritize audience we care most about reaching - address.
For member nodes, individual researcher is not top audience. Here are all the technical specs, down in the weeds, here is the path where someone who is a level above, not having direct responsibility, needing to understand because they are in charge of resources.

HOlly: not necessarily the technology person, but the person who has to think in terms of funding, strategic priorities, and...
Allard: allocation of limited resources
Cobb: rolling out, there will be some communications. Reasonable to think as we gain momentum are ther people willing to do that
Allard: once creating awareness, where do we drop them? Awareness is different than informing or leading to a decision.
Technology piece: details on types of data, repository, other piece is "they don't care about that detail" talk to Greg at ORE - happy you have that technological thing, but what is the answer, how do you make it work, whose responsibility, and how much money do you need.
Talking about Nat'l Climate data center - still decision makers there that have to be addressed even if the hard core data folks thing this is a good idea.
Right now we are dealing with people who make decisions. Outgrowing. Subsuming issue - territory issue. ETDs - when math / engineers had their own ETD, there were issues about whether the ETD should live within Math or integrated. IN that case, we gave librarian’s argument for why it should be put in.
Gave really good ideas, for what would and would not work, did a lot of incubating. Those are all things that as we move forward, number two recruitment, if we stay just on number one, tech and people who already want to come on, that's a different set of materials.
HOlly: For two or so hours, what is the plan in terms of activity. List of things that should be on the site? Content written?

Break for Lunch

Afternoon Sessions:
Review of Personas
  revising guiding infro: Categories "ility" dimensions
  Review of Assignments

  Progress?
  Issues/probleasm?

MN related metrics yr5 (yr 4)
  Metadata records 1,000,000 (400,0000)
  Number of data sets 360,000 (180,000)
  Number of supported metadata schemas 8 (8)
  MN count 40 (20)
  Total Storage 60 TB (40 TB)
  MN countries 10 (5)

What should we do?

Break
Scalability

Wrap

A. Metrics in Project Management Plan
Line 54 has a link to stuff related to metrics.

- Reference statistics above
- These are fairly constant - goals from year five since review of the project
- These metrics go back to right after the award. Questions from the review committee.
- There was a tweak in the project management plan.
- Idea was here for about 5 years. Scale things rather rapidly in years 4 and 5.
- Now facing years four and five.
- Looking at member nodes, doing things that we can measure, and doing things that are softer (member node candidates)

What are the important things to do and to measure, even if not in the project management plan?

Not new ground for this group; a lot of material is already there. Rama wrote a paper. If you look at Line 54 - in that line, there is a paper by Rama.

"Rama's Paper" may 02, 2012 - UA "Role and Utility of Metrics in Data Systems"

Assignment; come up with three top metrics that are good measures of performance by a member node.

Tanner Suggestion: Downloads from member node site
Cobb: is this a metric of DataONE success? or Member Node Success?
Member node wrangler success

Chelsea Williamson Suggestion: user activity - who contributes, who uses, if you have multiple people contributing it makes the data richer. For a single member node, the number of people who contribute to that member node.
Robert Waltz Suggestion: if a tier 4 member node, how many objects do you have replicated. How many replicated objects are there.

Miriam Suggestion: How long to respective member nodes wait to come online (activation time)

Suzie Allard Suggestion: Diversity of member nodes.

Diversity of data that they have. Multiple aspects. Characteristics.

Amber Budden Suggestion: Number of byte amount of data objects, not metadata available.

Cobb: you would not count people who have data available, but that you could not get back.

Allard: You can get to the metadata set, but it only points you to the source.

Ratio of two of those things (dataset download versus views). Number of searches could be meaningless,

Allard: in past, static, member nodes themselves, and dynamic measures (services, how much used, usage - transmission).

Holly Suggestion: How would you demonstrate a transdisciplinary.

Cobb: how many science studies retrieve data from more than one member node, via dataONE or dataONE facilitator.

Amber Budden: Something advocated before, providing some guidance on the methods section, language, easier to search for papers.

Cobb: in cyber infrastructure, journal, computing experience. in a lot of journals, looks like a methods table. Experimental role. Is there a place to publish what Amber Budden was talking about.

Recommended language. Write down as a recommendation.

Cobb: Comment on that: publish a paper on DataONE and encourage that as a citation, you get that as a publication. Not a metric.

Amber Owens Suggestion: Capture something more qualitative - how is your repository treating you - have a list of potential - asking potential member nodes. Survey of some group of member nodes.

Crowston Suggestion: a metric that would capture success or failure - referrals - how much are people finding stuff through DataONE, and how many more deposits would you get before people start contributing.
Miriam Suggestion: how many people want to become member nodes; how long is the waitlist.

Robert Waltz:
Measure of replication multiplicity for data objects.

Chelsea Suggestion 2: curation level, how active are they in DataONE, some are more active than others, using or promoting. Engagement.

Tanner Suggestion 2: Persistent Identifier - Document ID, citation counts from known documents from member

Dr. Allard Suggestion 2: number of data sets
Rama Suggestion: User satisfaction

Note: due to concern about the stability of the current e-pad, the discussion is being continued on a second E-pad available here: http://epad.dataone.org/Sp13-SCUAwg-membernodes2

New pad started at around 2:45 pm on Wednesday.

**Process needs to be transparent, defensible, and agile. not arbitrary.**

**************************************************************************************************************************************************************************************************************

Jump to (or near) line 1066
**************************************************************************************************************************************************************************************************************

Member Node policy draft - version 0_8_0
**************************************************************************************************************************************************************************************************************

**** Purpose: This policy is designed to give guidance for the processes of how DataONE handles Member Node (MN) issues including prospect identification, response to inquiries, entrainment, development, deployment, operations, and removal.

**** Background: An important part of the DataONE as an operational infrastructure is the engagement of Member Nodes as project partners. Member Nodes are somewhat unique within DataONE in that they are project partners instead of internal DataONE resourced activities. Member Nodes, inherently, are collaborating partners. One of the core principle of DataONE is to enable access and interoperability to a large, varied, important set of data collections residing within the associated Member Nodes.

Thus the process of recruiting, entraining, and coordinating with MN's is one key aspect of DataONE's overall success. However it is not advisable to pursue production deployment of every conceivable Member Node prospect for various reasons. Consequently, DataONE needs guidance and clarification in the formulation of processes and procedures for various MN activities. In many instances, the course of action is clear, but there are a few decision
points where DataONE will benefit from prior, agreed-upon, processes to avoid situational decisions that may not be consistent with long-term, strategic goals of DataONE. This policy is an attempt to collect, organize, and reach a project-wide consensus on the principles that will guide the steps in the process that require decisions.

A high-level summary of the proposed DataONE workflow for Member Node activities from identification to production operations is outlined graphically at <https://docs.dataone.org/member-area/member-nodes/coordination-work-area/scua_wg_20130430_materials/MN_Workflow.pptx/view>. The key identified decisions points from DataONE’s perspective are:
- Outreach strategy: How to focus on potential Member Nodes to recruit and how to respond to inquiries from potential Member Nodes
- Evaluation of initial proposal to initiate a new Member Node development
- Operational Acceptance of successful development completion
- Coordination of activities at start of production
- Possible termination of Member Node status at some future time.

Some overarching considerations are that these decision points should be guided by DataONE’s mission and vision statements:
Mission: Enable new science and knowledge creation through universal access to data about life on earth and the environment that sustains it.
Vision: DataONE will be commonly used by researchers, educators, and the public to better understand and conserve life on earth and the environment that sustains it.

In addition, the DataONE sociocultural working group has articulated five summary principals at <https://docs.dataone.org/member-area/working-groups/usability-and-assessment/meetings-usability-and-assessments-working-group/joint-u-a-sc-wg/faqs-documentation-environmental-scan/D1%20Principles%205.2.12.docx/view?searchterm=five%20principles> for data contained within the collections included in DataONE:

Include by reference not inclusion
Which version? - WE should go with RSV principles (where stored in docs?)...

What is DataONE's role within the community?
service to MN's ; promoter of best practices; or enforcer?

A: coordinator and promoter - not enforcer

Do we include suspect data?

perhaps support both but try to annotate with quality cues

c.f. Bilder's prior comments about indicating or "approving" certain practices and practitioners
Robert: We can hold onto records that are "well formed" orthogonal to data quality.

1. Data science is transforming environmental science. [??? jwc: Still include ??]  
2. Data should be part of the permanent scholarly record and requires long-term stewardship.  
3. Sharing and reuse maximize the value of data to environmental science.  
4. Environmental science is best served by an open and inclusive global community.  
5. The data environment is dynamic and requires evidence-based decision-making about practice and governance.

Finally, the frameworks described in this policy are not absolute and inflexible, but rather are designed to provide guidance than can help to maintain constancy and coherency of DataONE objectives across many different activities and interacting with different communities.

**** Outreach Strategy and Targeting Criteria:  
The goal is to create a collection of MN's that advances science, enhances MN's, and sustains DataONE. First we hope to enhance community and content. Second we want to enhance, diversify, and simplify underlying infrastructure and interoperability. In addition, it is valid to consider pragmatic considerations such as potential member node willingness and eagerness to participate in DataONE and the technological implementation feasibility (or difficulty.)

Is a MN-lite a MN or some other type of creature?

**** Evaluation of Initial Member Node proposal:  
There is an associated process and procedures document that outlines needed activities along the workflow path before review. The character of many of these activities are in terms of information gathering, planning, and scoping. After that discussion proceeds to maturity, the potential Member Node will propose moving forward to development. That proposal will be considered by DataONE and the Member Node with the goal of reaching a mutually beneficial understanding of the prospective Member Node. From the DataONE perspective, the DataONE Member Node coordination group will assist the prospective Member Node in providing information and answering questions. (The coordination group will also facilitate connections throughout DataONE and not act as the sole resource for prospective Member Nodes). Once a complete, but brief, proposal is prepared, DataONE will undergo and rapid but sufficiently detailed consideration of the Member Node. The goal of this review is to have a sufficiently complete understanding of important dimensions of the Member Node including: the community it serves; the characteristics of its data collection(s); the implications for DataONE of needed physical cyberinfrastructure, software cyberinfrastructure, effort hours required to support Member Node development/deployment and continuing operations; any unique or new cyberinfrastructure needs that DataONE will support beyond its currently developed and operational infrastructure; and the required resources and their availability to the Member
Node to undertake activities during the Member Node development, deployment and operations phases. The exact form of such a proposal is not specified here, but will be provided as a template by the Member Node coordination group who will also assist prospective Member Nodes in completing the proposal. The template itself is not a controlled document, but will evolve as needed.

Suzie: NM DUG discussed this- find and re-confirm our position.

Qualities:
High Value data
High risk data
Volume of data
Technological readiness

(Now) scalability

We need to touch base with those discussions

Document where we have been:
NM DUG https://docs.dataone.org/member-area/planning-for-dug/dug-meetings/dug-2011-meeting-planning-folder/dug-2011-discussion-notes/DUG_member_node_prioritization_notes.docx
Also see Powerpoint presentation https://docs.dataone.org/member-area/planning-for-dug/dug-meetings/dug-2011-meeting-planning-folder/dug-2011-presentations/MN_Prioritization_2011.07.09.pptx

Public facing pages on member nodes
https://www.dataone.org/become-member-node
https://ask.dataone.org/question/10/where-can-i-find-more-information-about-becoming-a-member-node/

S&G WG discussions:

Miriam: items we listed
Technology
resources
alignment with DataONE

Process?
Kevin: review committe
Once submitted, the proposal for a Member Node to move to development will be available for review and comment. Specifically, comment will be solicited from the core cyberinfrastructure team (CCIT) as conveyed by the director of development and operations, director of community engagement and outreach, and the DataONE project manager. The proposal and comments will be reviewed by the DataONE leadership team who will be asked to make a recommendation. The DataONE principle investigator will make the final decision. The purpose of a gate and release process at this point is to endeavor to understand the importance, impact, and needed resources to move this Member Node through development and into continuing operations.

Process:
- needs to be transparent
- agile defensible

Dave: (up in discussion: DRAMBORA (ref????) ) risk analysis for consideration/evaluation of data repositories

**** Operational Acceptance of successful development completion:
At the end of the development period, it will be necessary to insure that the Member Node is functioning in collaboration with the DataONE infrastructure. This is needed to insure that users of DataONE will be able to access the collections of the Member Node effectively and efficiently. It is also necessary to insure that adding the Member Node to operations will not create operational problems for DataONE or other Member Nodes. The vast majority of issues that need to be examined will concern cyberinfrastructure unit-tests and dev-ops readiness issues. Consequently, the DataONE director(s) of development and operations will make this decision with a recommendation from the CCIT with notification given to the DataONE leadership team and DataONE Users Group. If during the course of development, the characteristics of the Member Node change significantly, then the director of development and operations can request a re-evaluation of the initial Member Node proposal.

**** Coordination at start of production:
As the Member Node moves to production, the Member Node will coordinate with the DataONE director of community engagement, education, and outreach in order to develop communication plans for the start of production. Such items might include: revision of internal and external project documentation and communications release to the general public, funding sponsors such as NSF, and social media.

And put it in the DataONE NEwsletter

send to mailing lists, including MN-=related mailinglists

**** Possible termination of Member Node status at some future time. [??? jwc: "terminating" is a double-plus ungood term - alternatives? ??] It is quite possible that at some point a Member Node or DataONE may no longer wish to continue to be a DataONE member node. A likely example might be the end of a
overarching project that developed and supported the Member Node. Often the nature of such events are poorly resourced efforts at the end of projects. To the extent possible, DataONE and the Member Node should seek to understand and foresee this type of event with enough foresight to allow an orderly and reasonable termination. DataONE and the Member Node should reach an understanding about data legacy issues including the preservation of data already published; continued availability of data where needed, for example data referenced by persistent identifiers; continued access to collections; management of identities created by the Member Nodes for access controlled data; and other issues as appropriate. Also, there needs to be a mutual understanding about the continued access and access control of data that is not publicly available without restriction. In addition, there may be a discussion about a new agreement to host and maintain data at DataONE or elsewhere for Member Nodes that cannot (or do not wish to) continue to manage the data in their collections moving forward.

[??] jwc this is a bit unclear to me at this point. Have I captured the issues? What is the decision tree?

In a sense it's a will.
Add a data will question to initial evaluation process.
is "Data Estate PLanning" part of DataONE's sustainability plan.
wording: "legacy" "succession planning" - this is the term in land ownership

Vieglais: also consider change of authoritative MN from original MN.

Kevin: Ask for these items at evaluation. Especially preservation and a data will -- realizing that we may not get complete answers.

Add a checkbox on evaluation form "will you allow DataONE to maintain rights to the data? Kevin: reword ask how they plan to make their data avaialble in the long-term and let them choose DataONE.

Cobb: Also, keep in mind that this permits DataONE to have the data but does not constitute an obligation by DataONE (i.e. don’t go into the insurnace business unbonded)

Dave: (up in discussion: DRAMBORA (ref????) ) risk analysis for consideration/evaluation of data repositories

Policy Review and Ratification:
The proposed route for approval as a DataONE project policy is for this document to be review and discussed with various DataONE working groups and the DataONE user group. Then it will be submitted for discussion and recommendation by the DataONE leadership team and approval (or rejection) by the DataONE PI.

This policy will be in effect until revised or revoked. The policy should be reviewed and if necessary revised annually, but if that review fails to occur, it will remain in effect unmodified. (i.e. not automatic unseating.)
External Web process

Robert: Types of Data (formats)
   What types of information from MN.

Chelsea: Criteria for the MN-

Chelsea - interactive clickable list of MN's that provide similr to NASA Earth Data Centers.

Tanner: Factsheet and partnership guidelines are PDF. Send me a Vimeo short description. slideshare embeeded, short-ish - not more than 3 minutes. (Suzie say 1.5- 2 min)

Kevin: Form mroe information contact DataONE. Who do I contact.
   instead have a link to MN coordinator
   Add a picture

   note not on main MN page but is on the become a MN page. should be on the first page

Amber: When talking about sustainability - talk about costs of beoming a MN. "'What will this cost me?"
Difficult to anser decisvely but this is a question
C.F. Mule work by Suzie

Side comment Miriam: do we include this question in the Persona develoment.

Amber B. I would like to see the workflow/process.

Suzie: What is the value proposition to the MN> Why would I want to? What do I get?
What do my users get? How hard will it be? (cost)

Todd: Complete list of current members and/or highlight examples

Miriam: Exmaples of good MN evaluation proposals

Miriam: Who are Member nodes?
Miriam: examine current questions coming from incoming. In the FAQ’s in the SC mtg.

Tanner: I like the second paragraph is too much text fro the web. Break it up into groups and bottom one.

Kevin: Kevin describe the documents to download - explain why one would want to look at these documents.

Kevin: The backing documents are uneven in their aimed audience. Write documents targeting different personas (decision-maker, technical person, ...)

Amber: How do I articulate that without excessive length?
A: Kevin: executive Summary

(discussion on how to pick apart the content)

Todd: List of institutions that run MN’s - this is in the ASK FAQ’s

Amber: Tabulation of the Tier structure and SW alternatives.

Suzie: What is the structural and "power" relationship between me and my MN and DataONE.

Amber O: Pre-requisites to being a MN. Easy Pre-req.

John: Answer DataOWNer’s question about where to archive. (Maybe in ASK)

Kevin: Who’s responsible for what? (It’s buried in the documents)

Chelsea: Where do I find forms I need to fill out?
Amber: We didn’t have it on the page initially, but we gave it out. We are not sure. Consider again. Don’t scare people off.

Tanner: Have an interactive checklist. Congratulations you scored 95%. Do you want to be a MN?

Amber: Can and should MN’s see the redmine prioritization. Some items should not be public. But this works against desire to have everything in Redmine. Think about it and figure it out.

Miriam: There is potential embarrassment.
Amber: Example TRELLO tracking in Dryad.

Shift: What are the buckets to organiz data>

current: MN
Become a MN
Current MN's

Amber: I alittle difficult to change.

Suzie: Add something quicj and fast about why being a MN is helpful!
See the picture on the tapestry, not the stiching.

Maybe add this before the "What is a MN"

Suzie: Divide this conversation into a couple of peices
1. The technical piece documenting
2. Breaking the discussion into digestable bites (Tanner)
3. Recruitment conversation. (we haven’t recruited yet)
   How is recruitment occuring via website or word of mouth?
   We have two website inquires

Suzie show alliance with MN goals

Give prospects ammunition for internal discussions and decision process.


Example: Electronic THeses and disserations. (ETD's)
   two approaches:
     Talking to University Admins
     Talking to Libraries
     Metrics of three kinds: static, dynamic, transmission. see

     Coming from grassroots moving up.

     Identified 5 audience

     s
Library
University level administrators
Faculty
Students
IT

Suzie: DataONE is missing the boat currently in having a place to understand why this is important.

We have multiple audiences and they will come in different ways.

Have two paths when getting to MN pages:
  technical
  level above in terms of responsibility (Decision maker)

Awareness: Once we have created awareness where do we drop them to give them information to inform them.

Operationally nodes may want "just the facts Ma'am" to "Git-R-dun"

Personas - timeline draft 3 weeks; do phone call Laura will doodle us Tanner & Amber O & Chelsea - geospatial large govt repository