Afternoon Session  
Wednesday, May 1 2013  

Rama: Annual Survey:  

Miriam comment: is documentation useful, from the end user point of view.  
Cobb: some things you can automatically quantify, others only via survey  
Rama: number of publications from member nodes  
Cobb: number of science highlights (slightly different measures)  
Miriam: sort of a core - start rippling out. Everything DataONE does is related to member nodes. Where is the boundary? Number of publications resulting from use of member nodes - how does that reflect back to membership of the DataONE institution with member nodes. What can the member node group of people associated with DataONE grab.  
Cobb: Impacts.  
Rama: Impacts of DataONE on Member Nodes (If MN’s were operating before they became members of DataONE, what incremental improvements resulted as a result of the membership)?  

Ranjeet and Giri comment that Mercury can collect statistics about queries but ONEmercury does not have this turned on. It collects the query requested, where it came from (IP) and results. It could be used for building up information about common search terms, etc.  

Thursday, 2 May 2013 - 11:00a - MN Scalability  

Present: John, AmberO, Holly, Line, Miriam, Todd, Robert, Suzie, Tanner, Laura  
Virtual: Lisa, Bruce Grant  

See John’s slides <get link>  

Everyone answer the question: How many MNs can DAtaONE support, given the current DataONE org structure, staff, resources, etc.  

What limits the scalability of D1 MNs?  
- physical CI (storage, network latency and bandwidth)
- human resources (CI staff to assist w/deployment, outreach activity, ability of D1 to "act" on MN requests - process/policy friction
- target market - can we exhaust the pool of potential MNs, can we exhaust pool of potential MNs in specific areas, discipline/subdiscipline areas, "types" of MNs (observational, simulation, remote-sensing)

Brainstorming: list 2-3 things which might limit scalability

Robert: lack of available and widespread stacks that implement the D1 API (Dspace, iRODs)
Tanner: may be popular and/or proprietary s/w (i.e. ESRI ARCGIS) that is locked on proprietary data silos
Miriam: running out of funding
Holly: failure to provide a compelling reason to be a MN
Amber: our expectations may be limiting (we might be able to have more than 40 MNs by the end of Y5)
Line: if we don't have a compelling story in terms of data discovery, potential MNs will become discouraged from becoming a MN
Tanner: this is one of my favorite "stories" about DataONE and the utility of the data: http://cci.utk.edu/can-big-data-save-planet
Suzie: lack of support from the MN home organizations (i.e. mission alignment)
Todd: lack of peer support from MNs from things like DUG, etc.
Laura: failure to make "why would you want to be a MN" case
Bruce: outreach to MN's user base
Robert: lack of diversity of ITK tools may limit the end-user base (if all we have is R, non-R users may be turned off)
Tanner: more money, more problems, complexity of the organization increases as we add MNs, complexity of CI increases as well; who's watching
Laura: might need to broaden scope of MNs beyond eco/enviro/bio, etc.
Miriam: sustainability concerns (guarantee D1 availability)
Amber Owens: to add on to the complexity idea - information systems are like biological systems, they must become more complex - causes developers to rise to occasion. Lack of innovative or creative minds - Tanner interjects "employed" innovative or creative minds.
Line: Raising expectations too high for the member node.
Allard: new approaches to data warehousing - what might the future hold that our member node model is not flexible enough to hold
Suzie: inability to provide governance for an expanding network (if D1 doesn't deliver, MNs may pull out and start their own network)
Lisa: competing initiatives (i.e. EarthCube) may pull potential MNs from D1; need to be clear in our message and value of participating in D1
Miriam: efficiency (or lack thereof) of the process of bringing MNs online; if inefficient/arduous process, ppl won't want to be MNs
Suzie: model might change where D1 becomes more of a repository; if so, there won't be MNs as repositories because D1 is the repository
Tanner: energy costs - price/gigabyte as a limiting factor
Next topic: what could we do in response to these potentially limiting factors?

- increased funding: easy answer but not necessarily "smartest" - push the frontier envelope
- increase efficiency - in what areas?
- change how we do things
  - focus on deploying s/w reference implementations (metacat, DSpace, Fedora, IRODs...)
  - develop and promote deployment of "MN in a box" as a turnkey implementation
  - develop MN retreat - short workshop with goal of deploying MN in a week

Brainstorming:

Bruce: DUG is important organization to develop further; help with support base, help make the case to become a MN
Lisa: need compelling MN stories to point to that, after becoming a D1 MN they’re getting more hits, did someone find a dataset via D1 that they couldn’t find elsewhere; demonstrate ROI
Amber: increase communication between all members of the project; communication to MNs can be enhanced to demonstrate action on their behalf, etc.; MNs can communicate improvement areas to the D1 team
Holly: sustainability, the need to involve MN parent organizations through membership, governance, maturing organizations
Line: capturing MN metadata, increase visibility of their data,
Miriam: study and verify importance of discoverability
Suzie: identify potential next communities (beyond initial targeted community of eco/bio/enviro), look at complimentary disciplines (i.e. health)
Laura: survey and learn from our competition and deploy times that would of benfit
Holly: co-develop tools with competitors; this would distribute resource demands
Tanner: at AHM, no one self-identified as geospacial; need more outreach to that community and proprietary data silos; engagement of data publishers to ensure accessibility of their data (even if not freely available, make it available via D1)
Robert: diversity of revenue streams; specifically non-profit model, selling our expertise as a service, requires D1 to incorporate as a non-profit.
Bruce: echo Robert’s idea for non-profit, courses at a university, perhaps a certificate program
Miriam: continue and expand data management tutorials/education efforts on CE side, then the data in the MNs will be "better" and the MNs will be better prepared to come on board; right now holdup is lack of metadata standards - need to improve
Suzie: strategic planning w/3 year horizon - concerns with over and under expectations, continual review (strategic plan); 3 year is more realistic in terms of keeping a handle on technological changes and being responsive
Miriam: Develop the value proposition to MNs discussed above in "limitations"
Miriam: Develop the efficient process discussed re bringing MNs on board. (discussed Wed.)
Todd: further development through the DUG,
Amber: SWOT-honest look at operation, assessing effectiveness, where there are gaps (relating to emergent needs of MN's) try to arrive at solutions; be flexible too
John: currently in process: outreach to potential MNs is improving; hoping to prioritize potential MNs to get them on board in a timely fashion.

==============================================
Report outs, Thursday, 2 May 2013,
==============================================
(see http://epad.dataone.org/Sp13-SCUAwg-jtmt for full details)

- Working Group Surveys (Kevin, Alison, and Carol H.)
- Assessments Subgroup (Ben et al) (academic libraries/librarians, federal libraries/librarians FU surveys)
- FAQs (AmberB, LynnB, KimberlyD, BruceW)
- Member Nodes (John, AmberO, Laura, Suzie, Robert, Tanner, Holly, Line, Miriam