

## July 16-17 - LNO Update Water Cooler



Published on *LTER Information Management* (<http://im.lternet.edu>)

Home > July 16-17 - LNO Update Water Cooler

---

### July 16-17 - LNO Update Water Cooler

Mon, 07/09/2012 - 4:39pm — mgastilbuhl

James, Mark and Duane will update us on projects at LNO.

Doodle link: <http://www.doodle.com/zfysxnzykxn835mt> [1]

#### **Agenda:**

1. Update and status of LTER Data Portal, Metacat 2.x.y, and DataONE
2. Update and status of NIS Data Portal and PASTA

#### **Link(s):**

Monday notes here

Tuesday notes here

---

#### **Notes from Monday**

---

Attending on video: Suzanne R, Don H., Gastil, JamesB, MarkS, JasonD; on phone: none.

JamesB will update us on the LNO site review and ARRA funds allocation. Then MarkS will update on PASTA, Metacat and DataONE.

---

## LNO Update

---

JamesB

We got our site review report, dated June 17, actually just a few days ago. Trying to clarify it. We received kudos for effective promotion of network cohesion and scientific synthesis, movement of the NIS forward, strong collaboration with the IM community and development of the new website. 5 recommendations for things which will improve project success over next 3 years:

1. development of new near-term timeline for production release of pasta
2. interesting recommendation (phrased diff nsf and review team) Do more to get scientists involved in synthesis and work more toward synt across lter. Pro-active engagement (rev team). Sets priorities.
3. group of small IT... broaden audience communication. But focus only on LTER and broader eco community but not the public. Some conflicting messages.

3 things core IT support should build:

Web services connection to admin databases. They did not identify how. Will put more effort into. Some ARRA will go into that. talked to John Chamblee and James Connors re PersonnelDB. They support idea they guide actions of a programmer to complete the task then move to other DBs.  
(BW out of town so final negot. delayed.)

Increase number of trainings LNO produces. Will require IM help. Practical areas where IMs need help getting data online and available. Some conflict between review team and nsf. Basic site IM training vs NSF assumes everyone up to basic. Identify hurdles. May increase amount of trainings we need to fund. Locate trainers. (This topic is called core IT. A misnomer in the note-taker's opinion.)

More support for taking on site support tools and make avail to others. Curate site support tools. No detail. Will discuss with broader community, input from IMs re sw tools. Like DEIMS and Metabase.

Larger recommendations (these two are old):

Increase level of metrics to evaluate effectiveness and progress. Operational Plan (OP) is chalk full of metrics but looking for substantive metrics. May partner with other organizations. Looking for better ways to eval how we are doing.

Develop an external advisory structure. BW has discussed with Saran. A few external IT people to national advisory board. LNO was taken off nat adv bd. Now will add back in. IT and communication type people. Broad brush.

LNO is in process of crafting our response. BW may send out to network so you can see what NSF said. A few misinterpretations. Overall very positive, with opportunities, having to rush thru ARRA funding may actually realize some of these things.

Saran responded to BW plan for ARRA accel funds. She supported highly the support of synth data products in pasta. Also prov support to site IM expertise in EML production; support contractor for Drupal; \_\_\_\_\_ for sci questions like StreamDB, VegDB and GeoNIS. A note at end: geospatial metadata for all data is a basic requirement and not a separate activity.

Don & JamesB

Synth data products, Drupal, Needy sites. That was verbatim from NSF. Data products such as... encourages these activities. Did not mention Metabase but still well on table.

What was the “needy sites” thing? Is that mentoring? Yes. Anybody could be a needy site depending on the need. Was that modeled after the discussion of mentors, focusing on mentoring approACH FOR vehicle for getting it done.

JamesB

then we have another round where we put these into our ARRA revised funds \_\_\_\_\_.

Don

Those fit well with what IMExcec discussed in April but does not specifically cover GeoNIS.

JamesB

did already have funds allocated to specific data products. already in phase 2. NSF just putting emphasis on it in next two years. Already in plan, may not be a reallocation.

Don

So would some of that core IT, web services work,... could go beyond PersonnelDB... could that cover SiteDB?

JamesB

Yes. If we get a person to do this (coding fro PersonnelDB) then we should think of all these admin databases as one big bit. Not that big a deal to add bits. Hopefully the SiteDB working group would... dont have a BibDB working group. Have to make sure Personnel meshes well with pasta.

The way it is worded: providing web service access. does not cover restructuring the database. easier than wrapping existing database.

All still on table.

Since GeoNIS has sort of been handed to us, what can we do with GeoNIS and that side of the house, to keep that moving along. May put more funds into that area, production workshops and buy-out time.

Pull together the admin databases to think of as a whole, the three. Want to include our web people, such as Marsh, so they can use it in Drupal. Marsh has taken the ProfilesDB part of the personnel DB as far as he can now. The part left to do is to integrate with PersonneDB and ldap.

Don

You need to know from IMC what training we want.

JamesB

Have top ten list on training since last IMC. Likely changed since then.

Don

External advisory structure: separate from NAB or integrated with NAB?

JamesB

The latter.

Not a full NAB just for LNO. Just re-include LNO in the LTER NAB. Include IT and other things LNO does. That is between Bob and Saran how to do that.

---

Mark S's part: Updates on Metacat and PASTA

---

2 points.

1. metacat 1.9.5 to 2.0.1 (probably 2.0.2 later)
2. outline strawman schedule for pasta and nis for 2nd part of project, primarily to end of 2012 in better detail.

1. metacat upgrade.

The current LTER Data Portal interacts directly with Metacat. Uses underlying Metacat API. Then there is Metacat itself. In June, upgraded to Metacat 2. Integration of DataONE (D1) member node API. Briefly halted harvesting. June 30th, a Saturday, released new Metacat. Have had ups and downs, on LNO and Metacat devel sides. Stable now for everything except replication. A transition period. Trying to decide KNB old style or DataONE replication. DataONE ran into some bumps in roll-out a few weeks ago. End product will likely be thru D1.

On surface nothing has changed. Harvests and searches work as before.

New:

on ONEMercury site will begin to see LTER metadata content at that site.

We successfully synchronized with D1 but not replicated to other D1 member nodes.

Likely just to KNB at NCEAS in SB.

The content supported here at LNO is showing up on ONEMercury site.

Suzanne

What is diff btwn synchronization and replication?

MarkS

We synchronize our metadata with D1 but actual content would be replicated to other site. D1 has core metadata, like DublinCore part. But actual physical metadata should be replicated at KNB site once that begins. One is just search and discovery, other is full content. D1 designed to store data as well as meta but since we do not store the data in metacat, only meta will be synchronized.

Don

when we made that upgrade, were there some valid metadata pkgs which were valid in old metacat not valid in new version?

Did some packages not get thru the upgrade?

MarkS

Yes there was a bug fix, an interesting bug. In opening eml tag there is a declaration for the EML schema. Those attribute definitions are in quotes. In the old metacat it only looked for double quotes. (The “single quote bug”.) Assumed it was a plain xml file and was not doing any schema checks. That resulted in weird issues for searching and permissions. That allowed a lot of non schema valid EML into metacat. Now with bug fix, now can use single or double quotes so bad docs get rejected. Don’t know what happened to the content. May just not get recognised as EML or may throw an exception. Will likely cause some issue.

Change to metacat that does not affect LTER: now supports EML 2.1.1. No need to upgrade to EML 2.1.1.

We are beginning to use DOIs.

Thru LNO we now subscribe to EzID. We have our own “shoulder” or prefix. During the upgrade process, IDs were assigned to all existing content. That was synchronized to DataONE. None of those are active DOIs. Opaque. Eventually those will resolve. Those DOI s were assigned prior to the upgrade. Any eml docs harvested since Juen 30th are not assigned a DOI. Not to worry. Has no impact on sites.

Don

New eml pkgs are not assigned a DOI? Yes

Plan for future?

MarkS

Those will be picked up thru pasta. Then they will propagate into DataONE thru Pasta. Our goal is to assign those and make those active thru DataONE.

---

## PASTA Update

---

MarkS

Based on LNO review both team and nsf highly rec move to production status rather than adding features. Stabilize with current set of features. Release production PASTA asap.

Between now and end of 2012.

Long term: production version by Dec 2012. That system will have current features which have been tested and rev by Tiger Teams.

A number of existing services: id mgmt, pkg mgmt, event mgmt (includes workflows), provenance and audit services.

In addition will release the new Data Portal.

By end of this ARRA grant there are a number of things we want to be able to do.

Four items we want to incorporate into production:

1. harden sw incl how content stored within framework, able to scale without having to reinstall site data. never have to re-harvest data. May change underlying metadata or storage system but will migrate content.

2. two new features to add: 1: EML rendering (current is minimal). 2: want to integrate DOI service so users can start citing content using a DOI versus an internal resource URI.

3. Additionally,

we need to redesign our development environment for pasta. currently a 2 staged env (duane and mark) do devel on local desktops. a lot of code devel. then deploy to test env in virtual cloud such as portal and other web services. need to stabilize to 3-tier. Still devel on local systems. Also want 2nd tier but call it a “staged environment”. as stable as possible for dev testing. Site IMs can send to it to test. Then 3rd tier will be production level. Hands-off for testing. No test docs. that will be the public-facing system.

4. EML Quality Checks

We need a release schedule for the EML quality checks (QC). A series of QC that will be pass/fail on particular date. 1st date will be prod rel. A grace period leading up to milestone. Time to improve eml to pass muster. May initially put a check not at pass/fail initially, just warning with note that at next milestone will become error. A fair lead-up time.

We want to announce that at ASM.

Changing schedule. At ASM want pre-release version avail for demo. Intention... internal discussion... Two WG at ASM: one for novice user “how do I put data in and search and download via portal?”. Then an Advance PASTA session describing user API to code own apps to it, accessing pasta vi web services rather than portal. Comment period to suggest change.

Open up to sites to engage more with pasta. Set up harvest lists to test full harvest.

By end of 2012 that system will be converted to a production release.

If post-ASM sites start harvesting,... will do best to not to have to re-harvest any data.

Gastil

What about datasets w/o data? Where can those be hosted once PASTA becomes our one and only catalog?

MarkS

No plans to deprecate Metacat yet. Maybe end of 2013. Commitment to use pasta as member node to D1 by then.

Issue of dataset w/o data, a policy/social issue. How do we address the need for a metadata only repository. We do have capability in pasta to start reducing

QC so non-ready datasets can come into pasta. Would still be need to have a data object.

Don

DataONE will still be supporting Metacat? DataONE is really metadata only.

MarkS

No. Coordinating nodes (CN) and member nodes (MN). CN just provides discoverability. CN use metadata only to synchronize because not exchanging data. Supports discovery thru ONEMercury site. Coordinates thru those CN but says to MN LTER you coord with MN KNB and do direct transfer to KNB but I (CN) know about it. Data and meta are coord between member nodes. Level of metadata that is also sent to CN to support discovery process.

Don

So in those CN, Metacat is planned to still stay the main repo for metadata into the future? Metacat might go away for us but not for DataONE?

MarkS

Metacat is changing but it is behind the scenes. You never interact with that metacat on the CN. All the search metadata goes into ... a different index. Metacat will persist the way it is being used in our community. It will continue to evolve because there are 100+ installs around the world. The Metacat client API will be deprecated in favor of the DataONE API. (Ed. note: pathQuery is going away. That may be within the next year. )

---

### Notes from Tuesday

---

- Virtual Updates [2]

- Copyright © 2012 Long Term Ecological Research Network, Albuquerque, NM - This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Please contact us with questions, comments, or for technical assistance regarding this web site.

---

**Source URL:** <http://im.lternet.edu/node/1033>

**Links:**

[1] <http://www.doodle.com/zfysxnzykxn835mt>

[2] <http://im.lternet.edu/taxonomy/term/169>