



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

[Home](#) > [IM Exec](#) > IM Exec Winter meeting 2011-03-29/30 (Albuquerque, NM)

IM Exec Winter meeting 2011-03-29/30 (Albuquerque, NM)

Wed, 03/09/2011 - 11:22am — [mobrien](#) ^[1]

Participants

IM-Exec: Don Henshaw, Margaret O'Brien (co-chairs), Dan Bahauddin, Sven Bohm, Emery Boose, Corinna Gries, Suzanne Remillard

LNO staff (occasional): James Brunt, Mark Servilla, Bob Waide, Yang Xia

Other guests (VTC): Nancy Huntly (NSF), Peter McCartney (NSF)

Joint Meeting with NISAC (3/29)

1. Introductions
2. LNO Performance Milestones
3. NIS Analytic Tools
4. NISAC and IM-Exec
5. Science WGs and IM
6. NSF Videoconference
7. SIP Deadlines
8. ClimDB / EcoTrends Migration

IM-Exec Meeting (3/30)

9. ClimDB Migration
10. Tiger Teams
11. Workshops
12. Water Coolers
13. Macrosystems Proposals
14. Meeting with Bob Waide
15. IMC Meeting
16. Report to EB
17. SIP Updates
18. Legacy Data Pilot Sites
19. Action Items

Joint Meeting with NISAC (3/29)

1. Introductions

IM-Exec has seven members, meets monthly by VTC, and makes day-to-day decisions for the IMC. Its activities include: planning monthly IMC teleconferences, planning the annual IMC meeting, and reviewing proposals for training and production workshops (new this year). IM-Exec

members serve as liaisons to IMC working groups. The IMC representative to the EB is an ex officio member of IM-Exec.

Most IMs are on multiple committees and working groups. Four site IMs are members of the new Synthesis Data Committee. Participation in network activities is encouraged by the IM review criteria. A newsletter (DataBits) is published twice a year. A working group has recently created a terms of reference document for the IMC and IM-Exec. The IMC strives for a balance of inward and outward activities.

NISAC includes both scientists and information managers. Some tasks are assigned to NISAC by the EB. NISAC is frequently consulted in the development of plans and documents such as the SIP. A terms of reference document for NISAC is in preparation. NISAC meets in person once or twice a year (usually once), with quarterly VTCs. Domain science members are appointed by the EB. NISAC minutes and reports are available online in the LTER document archive.

2. LNO Performance Milestones

Eight milestones for NIS development are proposed for 2011. Tiger team input will be needed for each of these. See the NIS website for a list of all tiger teams and current membership. There are three NIS deployment phases: development, development stable (accepted by tiger team, database not final), and production (real content). In development stable, the code is stable and mostly up.

What will be required of sites to participate in PASTA? The development team has tried to minimize impacts on sites. The minimum requirement is well-formed EML, network-accessible data, and data / metadata congruency. Data use forms are a problem which the DAS server addresses.

How much site data will be harvested into PASTA? Some datasets are not highly structured or are very large. Some datasets will be easier than others. Ingestion and integration are different tasks. Expectation management can be a challenge. NSF may be expecting more than LTER scientists.

Need to communicate better on tiger teams and their responsibilities. IM-Exec could promote tiger teams as an opportunity for IMs. The next IMC VTC will cover tiger teams and experiences to date. Tiger teams have been effective for the most part. Scheduling remains a problem. Tiger team members are asked to provide written comments (formal acceptance testing).

cURL is a web services tool for developers (see the project website for details). An alternative web form interface might be preferable for some tiger team members.

Future issues will include: (1) the LTER data access policy and (2) the use of Metacat as the backend metadata repository for PASTA.

The LTER data access policy affects download mechanisms and may conflict with a future federal mandate for anonymous access to the web. Need a committee to review how the policy applies to the NIS (preferably at the EB or SC level). EcoTrends has implemented full auditing of all clicks, downloads, etc. Registration is required even for metadata. DataONE is dealing with the same issues. Recent RFPs from NSF have required that sites track how their data are used.

How long do we want to use Metacat as our backend metadata repository? PASTA currently uses Metacat but does not require it. A sensitive topic since Metacat is a community product. Many international partners have invested in EML and (to some extent) in Metacat. DataONE is currently a mash of Metacat, Mercury, etc.

Metacat provides a rich interface and a suite of services but has performance limitations. There are also issues with authentication, searches, etc. Metacat supports replication. But performance is a real issue for users. Other possibilities include eXist and Lucene. The LTER Network might make a formal request to Metacat developers.

3. NIS Analytic Tools

One definition of the NIS is: Data + PASTA + Analysis Tools. How to develop general tools that others can use? Science WGs may create some general tools, but this is entirely a bottom-up process. Need to identify common tools and experts (or groups of experts) in the LTER community. Build on industry standard frameworks such as R and Matlab.

Scientists need tools and derived variables. But tool development is a hard sell at NSF. NSF wants to know general applicability and how tools will be hardened for general use. But it's worth talking to program officers.

Science WGs might include IMs or data plans if the data requirements are significant. NISAC was originally conceived as data modules created by science WGs, but this model has not worked very well.

Questions to ponder: How to bring new tools into some kind of framework? What are the primary bottlenecks to intersite synthesis? What are the primary uses of EcoTrends datasets? What will the EB expect from NISAC over the next three years? What will NISAC do when the NIS is completed?

4. NISAC and IM-Exec

Many IM-Exec members had a conflict of interest in reviewing proposals this year. IM-Exec decided to provide pluses and minuses for each proposal rather than rank them. Some proposals were passed on to NISAC.

According to the LNO operational plan the procedure is as follows: LNO asks IM-Exec to vet proposals. LNO has the option to ask NISAC for a second opinion as needed. LNO makes the final award.

Explore multiple avenues for improving communications. Add NISAC co-chairs to announcements for monthly IM-Exec VTCs and invite NISAC co-chairs (or other members) to attend quarterly. Or make one of the NISAC IMs an ex officio member of IM-Exec.

Currently the LNO operational plan names the IMC. The plan could be updated if the IMC voted to replace IMC with IM-Exec.

5. Science WGs and IM

Science working groups are funded by LNO but there is not much money. Recommend that IMs focus on big projects (e.g. prospectus projects). Need to develop best practices for this relationship. A topic for the SC, NISAC, and maybe IM-Exec to consider. Need a consensus-driven list of top priorities. NISAC and IMC have little connection to these WGs except where members participate.

For example, the stream chemistry database (ChemDB) still has no clear path to integration with PASTA. NISAC is well positioned to facilitate this process. But in the larger context we need the whole workflow, not just a laundry list of tools.

6. NSF Videoconference

2010 LTER Supplements. A group of six sites collaborated on developing Drupal modules that can be adopted by any Drupal site. Drupal can be used to manage a site's website, information management system, or just as an EML editor, and brings us closer to an "out of the box" IM solution.

EML congruency checker. Tests usability of datasets for the NIS. Reports will be available later this year or early next year. Use to develop PASTA-ready datasets.

EIM meeting every 3 years (initiated in 2008). The next EIM is in Santa Barbara on Sep 28-29, 2011 (IMC meeting on Sep 27). Meeting website is up. Call for papers is out. Registration soon. The new LTER program officer for LTER will likely attend. Travel budget is tight for NSF.

How to fund development of software tools for long-term use? The perception is that proposals are rejected for containing too much science (informatics programs) or too much focus on tools (science programs). SI2 (software infrastructure for sustainable innovation) program (office of CI) is intended to address this need. There were two RFPs last year. No RFP yet for center development (probably not this year).

ABI panels see proposals that are cross-cutting across all of BIO. Funded proposals have a connection to actual science research. Proposals specifically for LTER IM are not likely to be funded. But proposals with a connection to modeling, patch dynamics, etc may be successful. Avoid submitting an LTER proposal (LTER is already well funded).

NSF internal venture fund can be used to help direct future developments. E.g. PASTA funding came from this source. NSF has given a lot of thought to software reuse and sustainability. Proposals can include software sustainability as part of broader impacts.

SI2 workshops last year on how to craft the RFP for a software center. This concept is still very much alive at NSF. Institutes could provide long-term knowledge and expertise.

Data management plans for NSF proposals. Still quite new at NSF. How to negotiate with projects for inclusion of project data in LTER repositories? Bridge funding or other assistance? Revisions to the GPG are quite general in terms of data management. The new piece is asking applicants to indicate what they will do with their data.

LTER needs a business model for recovering the costs of managing data archives. Some consolidation is expected; not every site will want to do this. NSF expects this process will lead to a realistic estimate of the true cost for data management. This may lead to a realignment of reviewer comments with reality. Might be comparable to library costs at individual institutions.

Some interesting ideas from a recent meeting at NSF:

(1) We need a national infrastructure to store important data. We need investments in data mining and data analysis. Not everything needs to be cutting edge research (drivers don't care about asphalt specifications as long as the road doesn't have potholes). Investments to date have focused on advanced features. Need to shift focus to practical solutions for today's problems.

(2) The commercial value is in compiled code and support (not source code). A Source Forge for data would provide free versioning and quality control.

(3) Do we really want to save all of these data? The cost of digital archiving and release is low. An

alternative is to release early and rely on users for comment and correction. Users may be more motivated to do quality control than data producers.

(4) It is a misconception that CI (cyberinfrastructure) means HPC (high performance computing). Actually CI means high performance collaboration.

Terms of reference. Recently developed for the IMC by the governance WG. Developed by the community for itself. Anticipate a vote by the IMC at the next annual meeting. Terms of reference in draft form exist for IMC working groups and for NISAC.

7. SIP Deadlines

The SIP timeline is optimistic in some places. NISAC can recommend changes to the SC. IM-Exec should review specific items and deadlines and let NISAC know if changes are required. Include section number, recommended change, and justification. NISAC will also get recommendations from LNO. NISAC will provide a reality check to the EB.

Data presentation protocols will require a new working group of scientists and IMs. This requirement needs clarification.

Synthesis data effort may advance some items. The CI survey collects some of the information in the metrics section. Some items may fall outside the 5-year scope (2015) of the SIP.

The IMC can also address these issues in its report to the EB. Include activities and recommended changes linked to the SIP. EB reports are due 5 days before the EB meeting on May 17. Reports will be posted on the EB meeting website. Need to acknowledge accomplishments and not just the need for postponements.

8. ClimDB / EcoTrends Migration

Earlier this year James and Mark drafted an integration plan (generic but with a focus on ClimDB and EcoTrends) for migrating existing network databases into PASTA.

Proposed timeline for ClimDB: Site data are currently harvested from sites into the ClimDB database. 2011: create EML for harvested data (same format for each site). 2012: replace current harvest with ingestion into PASTA, add workflow to move data from PASTA into ClimDB database. 2013: add web services for ClimDB database.

Some of the site-related metadata might be diverted to SiteDB. Or information about the site (location, etc) might be placed in a separate table. For each station there could be one EML file, one table of observations, and one table of site of site documentation (e.g. deployment dates and serial numbers for sensors).

Should the data conversion (e.g. 15-minute to daily) be done at the site or in PASTA?

Add method variables that could be plotted to help with data interpretation. E.g. change in litter basket collection regime (10 days to 14 days) or different method for measuring nitrates.

EcoTrends presents a different set of challenges. EcoTrends is not updated annually and most data are now 4 years old.

Solution depends on network prioritization of datasets. Who will establish these priorities? Synthesis data committee, NISAC, SC?

The InCommon project provides authentication across institutions. The home institution forwards the necessary credentials. Not all universities participate. If the LTER Network participated, then individual LTER sites could also participate. The cost for the LTER Network would be \$1100 per year plus a one-time fee of \$700. LNO is looking into it.

IM-Exec Meeting (3/30)

9. ClimDB Migration

PASTA needs an EML file for every data file that is harvested. One strategy would be to harvest the original (level 1) data and use workflows in PASTA to create ClimDB compatible data. A second strategy would be to harvest processed (level 2) data into PASTA.

10. Tiger Teams

Bob has suggested looking outside of LTER if volunteers are hard to find. But outside volunteers may have a range of views that is too broad.

Tiger team contributions are not always technical. A fresh set of eyes is often helpful. Try to engage more graduate students.

The LNO operational plan required monthly reporting. Hence the monthly PASTA updates.

11. Workshops

IM training workshop. Jan 2012. Develop curricula for graduate students and young faculty. Meet with DataONE this summer to help plan.

Controlled vocabulary. The process of rekeying keywords is site specific. The new HIVE tool can read EML and help to suggest keywords. Developing uniform keywords for the LTER bibliography may be more difficult. Sites can have as many non-preferred terms as they want. Each non-preferred term maps to a preferred term. Only preferred terms will auto-complete in searches.

All-Scientists Meeting 2012. How to engage scientists? PASTA demo.

ClimDB migration. Create working group? Don, Suzanne, James, Yang. IM-Exec will ask for other volunteers.

Data models. Do we need a data model working group? E.g. ODM model (see Cuahsi website). Short and long-term issues. Revisit ClimDB data model. ODM works with WaterML but not EML.

Quality control. Start by focusing on ClimDB, ChemDB, and biodiversity. Problems are attribute specific. Could include range checks, seasonal range checks, and more complex dependencies.

Metadata quality control. Missing elements, keywords not from controlled list, etc. Mostly policy decisions.

EML best practices 2. Schedule working group VTC in next few weeks. Comments have been posted on WG website. A few sections are still sparse.

Rebuilding the data portal will be the most interesting part of PASTA. Compare iGoogle widgets. Working on functional prototype this year.

Training workshops:

- (1) Workflows
- (2) Controlled vocabularies

Production workshops:

- (1) ClimDB migration
- (2) Quality control
- (3) Metadata policies
- (4) NIS portal and data searching
- (5) Web services / network databases

Training workshops must be approved by the EB. Budget is \$48k per grant year. Budgets for training and production workshops are currently on track.

Governance. Circulate terms of reference before IMC meeting. Vote at meeting.

12. Water Coolers

April = Tiger teams (Mark)
May = ClimDB migration (Don & James)
June = Controlled vocabulary (John P.)
July = LTER Maps (Theresa)
August = LNO update (James)

13. Macrosystems Proposals

For funded proposals, NSF is requiring a signed agreement among awardees, LTER sites, and LNO regarding data management and related costs. LNO is involved because of the costs associated with harvesting and storing data in the NIS. A new project identifier may be added to the EML package ID to designate macro and other external projects.

14. Meeting with Bob Waide

Macrosystems projects are 5 years. How to cover data costs after the project is completed? Current cloud rate for storage is \$1000 per year per terabyte. Better to overestimate IM costs for these projects.

NSF budget is still unknown. No information yet on a supplement to support the 2012 ASM. EB has instructed LNO to hold on to funds.

EIM meeting is over budget. Overrun = cost of registration (\$185). Last year's meeting at KBS was under budget. OK to go ahead with current plans.

Funds for two production workshops this year. Recommend focusing on workshops that will get something done. Evaluate by critical path contribution to NIS.

Funds for compensation for two site IMs this year.

LNO site visit next year. In next round NSF may re-compete LNO nationally. What is the vision for LNO in the future? More training? More IM?

LTER 30-year review will be presented at NSF today.

Commercialization of data management. A group of site IMs might bid for the contract to carry out

the legacy data project.

NSF software institute. Delayed for another year. Not clear what form it will take.

Is the IMC addressing the needs of scientists? The IMC needs better public relations for its efforts.

A list from domain scientists of 20 analytic tools for synthesis would be really helpful.

SIP updates. Note that dates were largely set by the contractor hired to help create the SIP. We need to present a reasoned response to the EB.

Network coordination committee. Nick Brokaw is chair. Mission still vague. CZO may be first target. Other targets include ULTRA and NEON.

Communications committee. Get information out on tiger teams, NIS progress, etc. Some pushback from folks who don't want extra emails.

Legacy data project. Funding mechanism is still uncertain. NSF does not want to give money directly to sites. Separate RFPs for pilot and network projects. Pilot project (2-3 sites) will serve to estimate costs for network project. Contract through UNM. Non-disclosure process.

How about other ecological data without a clear path to LTER, NEON, or DataONE? E.g. LTREB, OBFS. Does LTER want to take this on?

15. IMC Meeting

Business meeting. Election for two IM-Exec positions (Corinna, Suzanne). No open positions on NISAC (but check with NISAC on member rotation). Vote on IMC terms of reference.

Site bytes. How did sites use their 2010 supplements?

Topics for discussion: training and production workshops, IM buyouts.

Breakout groups at IMC meeting: web services / network databases, ClimDB migration, attribute standardization. Two 2-hour slots.

Birds of a feather sessions at EIM meeting: controlled vocabulary, congruency checker, Drupal, spatial data, sensor networks.

Demos: DIEMS Drupal, congruency checker.

Deadline for paper submission to EIM meeting is May 27.

Agenda Monday Sep 26 (optional):

4:00 field trip to SBC

7:00 informal dinner on beach

Agenda Tuesday Sep 27:

8:00 introduction / reports

9:00 breakout groups

11:00 NSF program officer

12:00 lunch

1:00 business meeting

2:30 breakout groups
 4:00 wrap up
 6:00 catered dinner
 7:00 demos

16. Report to EB

Include IMC membership and changes, meetings, activities and accomplishments since May 2010. Describe goals for next year and their connections to SIP objectives.

For example: EIM meeting continues communication with other ecological observatories (SIP 4C). Controlled vocabulary (2A). EML best practices (2A, 4C). Unit dictionary (2A).

17. SIP Updates

1Ab. Change status to underway.
 1B. Change dates to 2012, status to planned, not funded. Procedures in place at sites. Use Scott Collin's project as first synthesis project.
 2Ac. Change status to underway. Begun with external funding.
 2Ba. Change status to underway. Prototype is new URI.
 2B. Change dates to match NIS development plan.
 4B. Change date to 2012.
 4C. Change status to underway.
 4Cc. Change status to underway.
 5B. Change by whom to LNO and site IMs, status to underway.
 5D. Change status to underway.

18. Legacy Data Pilot Sites

SGS is wrapping up. CWT is an original and augmented site with Forest Service connections, IM turnover, social data. Other possibilities include: LUQ, KBS, HBR, VCR, PAL.

Pilot project will serve to develop a cost model.

Include different data types (e.g. an aquatic site). Include GIS and other spatial data?

19. Action Items

Request volunteers for ClimDB working group (Don).
 Schedule IM-Exec VTCs for April through June (Sven). X
 Solicit nominations for IM-Exec, in august (Don, Margaret).
 Contact NISAC regarding its schedule and membership (Suzanne).
 Solicit site bytes for 2011 -- update on how you used your 2010 supplement, August (Corinna).
 Remind IMC working groups to create their terms of reference (each WG liaison).
 Solicit working group ideas for EIM meeting (Don, Margaret).
 Remind IMC about EIM submission deadlines for papers and BoF (Don, Margaret).
 Notify NISAC of suggested changes to the SIP (Don). X
 Write report to EB (Margaret, with input).
 Send training workshop proposal to IM-Exec and WG members (Corinna). X
 Contact Bob regarding legacy data pilot site recommendations (Don). X
 Compile and post notes from this meeting (Emery). X
 Add 2010 IMC meeting report to IM website (Dan). X
 Announce draft agenda for IMC meeting (Don, Margaret).

Meeting Notes ^[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/news/committees/im_exec/notes/2011_03_29

Links:

[1] <http://im.lternet.edu/user/27>

[2] <http://im.lternet.edu/taxonomy/term/3>