



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

[Home](#) > May 8, 14, 15 -- Virtual Update Water Cooler: 2012 IM Supplement ideas brainstorm

# May 8, 14, 15 -- Virtual Update Water Cooler: 2012 IM Supplement ideas brainstorm

Mon, 05/14/2012 - 11:39am — [mgastilbuhl](#) <sup>[1]</sup>

We received an RFS on May 1, for up to 150k per site for IM and education. This session is to discuss ideas, particularly for cross-site collaborations. Proposals are due Jun 1. Two to five pages.

Doodle link for participants: <http://www.doodle.com/rm2z728uyfzbtpt> <sup>[2]</sup>

Notes from

- [8 May](#)
- [14 May](#)
- [15 May](#)

are below.

Some rough notes will appear here during the vtc: <http://goo.gl/pYXtl> <sup>[3]</sup>

The text of the call for proposals:

----- The Call -----  
**Information Management:** As indicated in the recent LTER Thirty-Year Review and in NSF's response to this review, NSF places a priority on 1) timely access to all LTER core data and 2) strategic advances to an integrated and sophisticated Network Information System (NIS). The creation and sharing of long-term data were founding principles of the LTER program and the resulting data are valuable both to the LTER community and to the broader ecological community. These data are valuable, however, only if they are available, locatable, comparable, and usable. In this call, NSF invites two types of supplement proposals. Requests that implement basic expectations for data release, publication of EML Level 2, and online access to all core datasets collected at individual sites will be given the highest priority. Requests that seek to advance the functionality and capabilities of the NIS will be given secondary priority. For either type of request, sites may request additional personnel or hardware to achieve full availability of LTER data through site websites and, ideally, through a single data portal.

Basic expectations. This call for supplements provides resources to address basic issues of data availability and access following LTER-derived guidelines. NSF applauds the LTER community for establishing data accessibility as a network-wide priority. Appropriate activities for funding may include increasing the number or proportion of site data sets accessible through the LTER Data Portal or resolving barriers to making site data accessible through a single data portal. Competitive proposals should adhere to EML Best Practices for LTER Sites Version 2, and address the five

essential features of LTER EML data packages (Titles, Abstract, Keywords, Data Table Description, and Data Distribution URL).

Advancing NIS functionality. Recognizing that sites are at different stages of achieving complete data accessibility, and that some sites may have accomplished full data availability through the LTER Data Portal, NSF will also consider supplement proposals that target selected data sets to address higher level functionality of PASTA and the NIS. Targeted projects to contribute data products to NIS similar to ClimDB, HydroDB, or EcoTrends should rely on data that are currently readily accessible and tagged with appropriate metadata. They should be broadly useful across LTER sites or by non-LTER researchers, and could include, but are not restricted to, rescue of legacy data, data integration, or the creation of specific data products. Competitive proposals will be justified by their alignment with well defined research priorities and by their strategic value to the development and evaluation of the PASTA software and the NIS.

Requests must address how proposed solutions in either of these areas will be tested and validated.

---

### Notes from 8 May

Jason – vegdb, or ... sensor data mgmt. tools install from last week

Mob –

Jcham also mentioned legacy/synth project, fit with funding

JCh –

Sup avail now, legacy funds unknown. Nothing wrong with what Jason is planning.

Here:

Video: mob, jporter, cgries, inigo, Theresa, Jason, kristin V

On phone: john Chamblee

JCh – collab with gce, single proposal combining looking at strategic directed arra funding, if not that avail then will do full prop to nsf funding.

About 4 phases: fist 1 about standing up a ...

Wade is working on ways to make gce toolbox further integrated with eml.

Next steps changes to make metabase more robust in links to toolbox in way of pushing data from toolbox to metabase, and

Sw as a service so other interested sites could have a metabase instance that they could use either hosted at Georgia or elsewhere, push data in, then meta goes to metact. A matlab stage, a web interface stage.

3rd part: workshop for sites interested in particip. 2 prongs: get data in, then packaging for export in eml useful for any system. Eml and structured data come out other end.

Other portion of wkshp directed toward sw as a service.

Same things as front end in matlab then at end additional work from matlab to metabase, from web interface into metabase.

4th phase collab: address data availability. Bulk of funds directed towards pushing data into our systems, going thru proposals, and data found for legacy proposal. Some gce data wade knows is out there, time interviewing PIs to get data into the system.

Inigo –

Going to talk to PI soon. Already have draft plans. Discussed with Kristin and deims group this morning. Plan to use a portion of funds in joint proposal with deims sites to work on some of the tasks that we have, beneficial to the group to move the project forward. Have to work it out with the PIs and company. We have a laundry list of things, to give a flavor: drupal has a new version 7 with emphasis on useability. Could do the migration ourselves but would consume our tiem. Talked to

developers at DrupalCon last March. Have potential vendors to assist us.

Some site-specific items in priority list. Each diems site will allocate some of the funding to accomplish that migration. A common core that we have to 'tighten'. We have a data model with some differences site to site. Tighten data model. Having a professional to advise what team thinks best way to go forward. Will hear more from Corinna or Kristin. Have to go now to meet with PI about just this thing.

(Inigo leaves)

Kristin –

Echoing what Inigo was saying. Would like to combine funds across sites to fund developer to meet needs of most sites. Will be able to combine funds for a lot of the devel that we want. Then each site will apply for 10 to 15K to migrate their instance to drupal 7. All sites ill have functioning deims putputing high quality eml in metacat. That is what we will promise.

On top of that, selfish note ☺ need money to work with our programmers to get data out of campvll data files.

Corinna –

Can only echo. Have not had chance to think about this. Do not know if things here that are pressing. Personnally most interested in is Deims of course moving to drupal 7 high on list and all these data producsst that john was talking about. The most important part aside from publishing data is to come up with some sort of data product list that everyone could work on. Not do too many personal things.

Theresa –

Have not talked with Don and Suzanne yet. We have a developer we have worked with, who helped us with our stylesheet transform to get our geo data into eml. We ghope to get this person to devel xsl to go from arc catalog directly to eml rather than an intermediate step (fgdc) which drops information.

Then gis group is also talking about what is our next step in the geoNIS. May come up with cross-site proposal to do something with geoNIS next steps.

Mob –

The need to prioritize. When imexec discussed priority of data, did not attach funds to it. A science wk activity. Wondering how we imagine funds being used for this? Workshops? Post-Docs?

JCh –

1 thing to help prioritize data, a science priority, give them things to react to. Here are kinds of data we have. Is there a sense how much the scientists understand, across the network, willing to engage? Are they aware what data is available? How we might use IM funds to do this. As push data into system, summaries of data we have might be a way to engage the scientists to set the priorities.

Mob –

This reminds me of another hi-lited issue: we do not have a way of hi-liting certain kinds of data that we already have. Only way to find pp is to do search, and find climate data as well... vaguely related. If we had something that was more specific, some way of registering certain datasets as actually gC/m2/day, could hi-lite those. A sub-catalog to see kinds of data.

We did put that on our list for ARRA funds.

But that would... more organization... not just a web page.

JCh –

Chicken & egg problem. Get scientists to give us categories of data of priority to them, then go back using those categories to find the data, iterative process. A lot of things we don't know. We don't know the scientists priorities. Hard to orgin 2 or 3 weeks. Good thing to do at ASM. Then weave this into site-based efforts for data cataloging.

JPorter –

One problem with dataset prioritization issue is everyone stands where they sit. If you are into nutrient cycling then that is most important, if into organisms then that most critical. Try to identify champions, someone to identify at your site who cares enough that they are willing to work thru all the trouble to pull data together and integrated, motivated enough to do that. Best to start at bottom and work up, rather than network-level. Hard time doing that even in our little group here.

Mob –

One approach in past was to use current working groups, cross-site wg. Some of them presumably want data. There is a disconnect btwn proposed wg and the IMs. I do not always see what is proposed, surprised to see my site listed. Communication gap.

Do we have network-funded post-docs? Doing cross-site data collection? Where do they go for it?

Corinna –

Matt brought up that many nceas working groups have used lter data. Go thru nceas pubs to see if this is true, whether actually used lter data, then make sure these data are actually online. Possibly take those things as lead in what data products we should provide. Not get anywhere asking scientists to identify. We could put up a framework for site IMs to figure out what kinds of data products are desirable and if we have that kind of data. Not all sites have all kinds of data. If we have only 1 or 2 or 3 types of data, may not get all sites.

(off the cuff)

some way of calling all lter scientists to tell us what they would really like. Then see figure out which are feasible with data we have. Somehow communicated, published,, sites signing up for certain data products they can produce? Then IM part is to develop workflows that will actually produce data products. Products sitting in the nis with appropriate keywords. So actually discoverable. Perhaps indices of pp to compare btwn sites. Beyond the raw data. Following along neon's approach except accounting for variability of sites' data.

Mob –

Would require a dedicated person to do that initial analysis. Looking in pubs for the dark data. Fundable task, to search for dark data.

Corinna

Not just a data search, also a data product search, what the nceas groups wanted to glean from the data.

JP –

Going to try to contact sites with kinds of data we want to focus on, certain places have flux towers, have shrub expansion. Contact IM at that site and see if interested in pursuing something related to that data. A couple of pilot studies done.. how we tried to create a multi-site db that integrates a certain type of data: lessons learned.

Opportunity to also help make robust the tools we developed to aid the integration, id what tools we need.

Gastil – lessons learned from ecotrends.

Mob –

A few things common interest. Group decide some % funds propose to do something nis oriented. How to manage these? As someone to write common text for sites?

Theresa –

This is what we did for last year's proposal. Somebody takes the lead to write the boilerplate.

Mob –

This is a lot more money. More intimidating. We can handle it but adds up.

JP –

Like last time's model, people writing sections in their own proposal for a shared task. To develop integrated db, one site asking for more, that place hiring programmer or providing personnel support. Others put in travel, and site-based contributions to that database. Each site would be ... point on one db and secondary on others.

Nice to put together call for cross-site databases that they would be willing to act as point on. Circulate that list to sites. Have other sites join in. Know who to contact. Minimally prescriptive. Respond pair-wise.

Web page on im website? Quicker to send email to IM group. In 2 days send list of databases you would be willing to lead on. Then send consolidated list.

Gastil –

Is this like streamChemDB and VegDB?

JP –

Yes those would be on the list. Pre-existing groups are an opportunity.

Gastil –

Like Trophic Groups?

JP –

If you send out something too general. Too specific then too few responses. One option to say: our vision of a db. Should be hashed out during a gathering. Don't presume what scientists want.

Theresa –

Caution not to... seems free flowing. Could come up with many things. What can you really accomplish in one year. Would like to see more emphasis on things already begun, science-driven such as StreamChem and VegDB.

Too broad could siphon off time from keeping our regular data up to date. Need some kind of balance.

Mob –

JP –

Agree with Theresa. Cannot compete with putting site data online. This is why each site can choose how much they want to participate.

I can send email to IMC who interested who wants to lead. Could come out of imexec. Is this my idea or our idea?

Mob –

Useful thing to know. That kind of activity would take a lot of coordination. If we will talk about it next week then send email now. Helps people think what they might do. A good nisac thing to do.

---

## Notes from 14 May

by: Gastil and Suzanne

Participants:

Video: Aaron Stephenson (NTL), Corinna Gries (NTL), Don Henshaw (AND), Gastil Buhl (MCR),

Jason Downing (BNZ), Sven Bohm (KBS), Suzanne Sippel (KBS), Margaret O'Brien (SBC), Suzanne Remillard (AND), Theresa Valentine (AND), John Porter (VCR)  
Phone: John Campbell (HBR), Mary Martin (HBR), Ken Ramsey (JRN), Jonathan Walsh (BES), Hap Garrit (PIE)  
phone: John Campbell (HBR), Mary Martin (HBR), Ken Ramsey (JRN)

Introducing Mary Martin to the faces on the video.

Emery will represent the IMC at the EB tomorrow.

Margaret -  
Announcements

This discussion is ideas for Supplement Proposals, call came out 2 weeks ago, due June 1st. One feature of this call is this improved data availability aspect. Fact that this is specifically addressed in the call for proposals.

Metrics group thought since these 5 features came up, could make a baseline set.

At last VTC, metrics group asked to look. Starting to collect baseline metrics for our own internal use. Timing of Exec Board meeting means Scott Collins is also interested.

After this call you will receive a 4-page report with aggregated data (whole network) datasets in metacat, and a way to score these. You will also see your own scores (one row of table). Sites will want to know how they are doing with eml in metacat.

EB will see the aggregated stats tomorrow.

I (Margaret) talked to Saran this morning. We had a couple of short procedural questions, then we wanted a little feedback about potential 'advancing-nis-functionality' projects:

1. are there limits on how money can be split up between IM and Edu: answer: can all go to IM if want. Reason for including edu was that schoolyard are wrapped into renewal proposals but not all sites are renewing. Therefore included ability to write for some edu. OK to propose whole 150K for IM.

2. is there a recommendation for length (2pp does not seem like enough for large or multiple projects): Saran said keep it under 5 pages. It is important that projects are described thoroughly. (margaret's note: keep in mind that audience is the NSF-LTER committee of program managers)

Regarding 'increasing NIS functionality'. A couple of ideas have surfaced: Maybe one site would put effort into a stdized or cross-site synt product. she thought that would be a good illustration how \_\_\_\_ could work. Standardized inputs for data packages. Also a good idea.

Saran reiterated that these proposals ARE competitive. Will not be able to fund all sites at that level. Also some sites will come up to the 20% max level. Saran said they may grant part and not all, if proposals have splittable components.

Suzanne -

When AND PIs met, they were confused about the supplement being for both IM and edu, a competition between those two aspects. Leaning towards more \$ for IM since edu usually gets \$ in typical supplements.

Margaret -

Saran will arrive at EB tomorrow mid-day and will be at Science Council the rest of the week so she will talk to them then.

Don -

was she encouraging making these modular in scope so they can restrict if they need to, knock this or that out? Was she siding toward one big thing (improve all metadata) vs modular.

Marg -

got impression modular is a good idea. Our site has "two or three holes to plug".

Don -

Breakdown between improving data at your site, compliant, Best Practice, version 2 compliant (schema or BP?), list modularly. SEparate anything you would be doing internally vs across the network should be separated.

Marg -

What if my component has a dependency on a VCR component...? Maybe a site leading an effort needs to distribute "boiler plate" paragraph for other sites to reference.

Don -

that is what we did last time. ie Im working across these sites to do this thing.

phone IDs:

Jonathan Walsh (BES), Hap Garritt (PIE)

Marg -

Last Tuesday some IMs gathered re this Supplement topic, and shared ideas.

Site Supplement Ideas:

Jonathan -

We have an online chemistry DB that we would like to enhance to include all of our core data. Written a few years back in SSVR. Site-visit based thing. Take samples, enter field notes. Follows samples thru lab at UBC, Millbrook,.. to storage. We also do birds that way. Groffman said he' like to see all our core data in such as ystem. There are issues with the system, things it could use to improve. Want to extend to include all core data.

Still not, ... Metadata is still not at the attribute level. If we bring in a DB guy, he could help with that. Anyone else on GeoNIS? GeoNIS would like to see a collab project for getting spatial data online and sharable, perhaps ?XD format.

Not just a sample database. Also a data database. We put analysis results in there as well. A cradle to grave thing. I want it to me more of a data database, where it punches out attr level... ie I want to see a given year, given station, like ClimDB, to get a csv out. (A query tool.) Normally we update metacat yearly. That would be easier as csv out of a db.

Part of StreamChemDB?

Want to be; don't think we are (yet). We need to contribute more with ClimDB and StreamDB. We have a whole sensor network. I have to tie into a different data stream.

Marg -

StreamChemDB is talking about standardizing at the attribute level.

Don -

we will definitely be coordinating a stream chem component. Writing workflows, able to demo with pasta type datasets. not clear we would move entire prototype into pasta yet. Have 6 FS sites, 3 that are also LTER. Looking for more LTER sites to join in. Will provide verbiage.

Hap -

Our approach (at PIE LTER), tentatively, making more data available at quality that addresses the 5 essential features, then to address other core data not available yet: GIS, Zooplankton, Fish

graduate student data that has fallen thru cracks (want them to have time to publish then they're gone). Hire grad. Phytoplankton.

Drupal group also collab thing. Major thing: migrate from one DRupal vers to another; hiring professional person. Each site contrib.

GIS: at PIE we have 3 different GIS pkgs: Arc, RiverGIS at UNH, then EDRISI. Arc at least has FGDC. Other ones use Word Docs that are partially FGDC compliant.

Ken -

Have not yet discussed with JRN LTER leadership. Hope to collab with DEIMS version 2 as well as adding to that with JRN existing funds. Also an Edu component. Want to stay involved in any GeoNIS efforts. Not sure what level will participate since working on DRUPAL in house. Want to support FGDC and ISO and possibly EU format Inspire also planned to include in DRUPAL/DEIMS. Other xml exchange formats.

WE are trying to package up our solutions, including EcoTrends and the GeoPortal (see DataBits Spring 2012). Also hoping to integrate Deims, EcoTrends, and visualization with django. Hypervisor templates to install elsewhere all set just needs xml templates.

Mary Martin -

Not 100% on board just yet but involved with HBR LTER for last 6 months. Now thinking how our ideas fit the Supplement. EML compliance: have been talking to Marg and Gastil. Not insurmountable. Good to have resources for that.

We have data continuity issues with migration from 50 years of paper charts and weather stations to digital age. Could include resources for that in this supplement.

Core datasets: infill issues with gaps in historical record. In some cases the data is recoverable from charts. Could enhance long-term record. Early years, some manual digitizing. Now automated digitizing. Want to hear if other sites have such experience. some current proposals require data from not-yet-extracted data.

Don -

most sites planning to devote resources toward their own EML and improved EML and more data. Have not heard much at cross-site level yet. Supplement did emphasize datasets at sites.

Corinna -

We have a few legacy datasets on paper at NTL LTER from early on which we want digitized. The DEIMS, we are committed to upgrading to Drupal7. Lots of things we would like to improve in DEIMS, such as taxonomic coverage, which would mean some Web Services. More interesting programming work. We have to maintain two separate modules: drupal to eml and local NTL data query module.

We have so few GIS data and these are already available. May need effort to get those together.

StreamChemDB: Emily is interested in being part of that. But if not officially, something like that.

Aaron -

support for GeoNIS. Potentially hiring one person, hosted at one site, to do programming work. A cross-site proposal. Possibly at KNZ.

Don -

the network office does have funds right now and will devote some 80K to DEIMS. Are sites putting something in proposals saying they would be working with those developers? Are they devs or mentors?



Corinna -

Never really clear how much \$ the LNO will put into Drupal, or not. Rumors. We need somebody at each site to actually migrate the data. The big goal is that in Drupal7 the metadata data model will be the same then at all sites. An off-the-shelf distribution of DEIMS to download and use as-is, not experimental phase. That means fair amount of consolidation to re-group, converge. Had different goals. Hard to change data ways of doing things. Before we hired good php programmers but they did not know drupal. Not a drupal program. That needs to be re-integrated into the Drupal system.

Don -

if we know more what LNO is doing with ARRA funds then DEIMS sites can better align supplement with that.

Corinna -

we do not know the timing. The ARRA money will take a long time before we see it. Also some funds Kristin got. That needs to be performed by Dec 1st. Need cohesive plan for separate pots of money.

Theresa -

As Aaron said, we (AND LTER and GeoNIS) will get together this week. Pos hire programmer. Gather for another meeting. Talked with Don re someone doing the direct Arc10 to EML xslt stylesheet. We have a programmer here (at AND LTER) we could tap into.

Don -

if hire programmer, that xslt stylesheet could be used at all sites to move metadata from Arc GIS into eml.

Gastil -

Penultimate draft. Three final ideas: (1) expand breath of data in network catalog (leveraged projects, new data, other data), (2) attribute standardization, controlled vocab, protocol references, eml templates, annotations to metadata for ocean acidification dataset (cross-site), (3) implementation of SensorNIS activities to site.

Jason -

Series of 3 projects: avail of dat with EML, working with mentors, network office. improve eml generation script. in-house housecleaning, find eml pkgs, do corrections to attribute information. biggest financial one is programmer to implement kepler workflow to automate streaming sensor data into metacat and Clim/HydroDB.

Is anyone else working on workflow re SensorNIS. Possibly LNO could set up a Kepler server instance.

John P -

need for improved spatial data and rasters (but used previous supplement for this) at VCR LTER. develop value added datasets across sites (John sent out request for other sites wanting to collaborate); meeting money to hash out project ideas; have identified partners

1. Flux Towers (PIE, GCE, FCE)
2. Non-stream Water Quality (NTL?)
3. Primary Production (Mike Pace) focus on Salt Marshes  
big grassland primary productivity, forest PP, always has been tough nut
4. ShrubDB

As a group we need not have one from each site. Good to get a thru of these through the process. Kristin sent lessons learned from grassland effort.

If sites have interest in AND's or VCR's lead types of data, email us. Will send out expectation of partner sites.

Don -  
hope to get cross-site collab going on these.

JP -  
can also say will participate but not asking for funds. Being a data provider for a project takes effort. These workflows will all be set to keep on working with minimal future effort.

Margaret -  
SBC has 3 things: like JohnP, we are still working on our current supplement, Metabase implementation. We need some forms for non-experts to admin content besides IMs, ie project managers for people and project info.

About 1/2 our datasets are generated out of Matlab with EML templates. Data is stored in Matlab structures and text files. One way to integrate Matlab with Metabase is the GCE Matlab Toolbox. We have a Matlab programmer here. We will ask him to look at this code to see how to integrate. We will contribute to the Ocean Acidification data management re vocabulary and annotation and EML.

SBC scientists are eager to... continual spatial data out of Landsat-derived data products. The entire coast of California at discrete times. Currently served as text. Want query interface for users to request spatial subset. Suspect this has already been done. SBC could work on standardizing inputs and adopt tool? SBC does not use GIS for this data. Send advice.

Also adding more data to catalog and upgrading SBC's EML.

Sven -  
Dark Data - recovering grad student data. Investigating Metabase, maybe migrating to that or using parts. Have a lot of GIS data not well dealt with now. Those are the major pieces.

Don -  
At AND LTER have improvements planned for Essential 5 Features of metadata. Work with PIs to improve titles and abstracts. Re-work controlled vocab to align with standardized one. Long-term vegetation data considering how to best share in broad context, not just at our own site. Mark Harmon already funded for June workshop at HFR to consider this, incl storage. VegX type std exchange format to work with VegBank and existing repositories. Not much programming. Coordinating. If have veg data you want to share in those formats,...

AND LTER has legacy datasets in the nature of those at HBR. Charts, streamflow history. Sensor management we hope to work with rel to SensorNIS. Take advantage of tools presented at that workshop.

JohnP -  
Several mentioned CV. Have tools on website. If put \$ into VCR to help other sites. Memo to indicate.

Corinna -  
Good idea. Include NTL. Question to Don re VegDB. Community data or productivity data?

Don -  
think its productivity data. not sure they know till they meet. biomass, NPP. will clarify. follow-up at ASM.

Don -  
are people set or have questions?

Margaret -  
where overlapping, follow similar thread to other sites, look for collaborations.  
Make the point that we do not do this in a vacuum.

Margaret/Gastil-  
EML metrics; classified into excellent, good, needs help  
metrics look at 5 essential features, methods, temporal, geographic coverages  
aggregated as a % by site; email Gastil if you're interested

Hap -  
proposals must address how will test and validate.

Gastil -  
Has written verbiage to address final sentence about testing and validating EML; "all new datasets will be run through the PASTA quality engine". Useful for metacat checking, does not need to be PASTA ready. Evaluation tool doesn't look at 5 essential features.

Hap/Marg -  
the proposal specifically means the 5 essential features, which are different and less specific than the PASTA Quality Report.  
One thing you can report is how many of your sites' data packages have each of the 5 essential features.

Jason -  
want to be sure the metacat queries will be available again.

Gastil-  
All of the PATH queries are stored and SQL queries are stored. Took a lot of time to run, not automatic, but manual.

Don-  
EB will be looking at these results tomorrow. You should be hearing from your PI.

Gastil-  
Can provide some help with determining per-package numbers.  
ie we have presence/absence for these features for every dataset in metacat.

---

## Notes from 15 May

### Participants:

Video: Margaret O'Brien (SBC), Jason Downing (BNZ), M. Gastil-Buhl (MCR), Adam Skibbe (KNZ), Yang Xia (LNO), Suzanne Remillard & Don Henshaw (AND), James Brunt (LNO), Philip Tarrant (CAP)

Phone: Sven Bohm (KBS), Jim Laundre (ARC), Hope Humphries (NWT)

Pre-meeting notes sent in:

From Linda - (sent in via email prior to the vtc)

"In case I don't make the VTC, I thought I could let you know what the FCE is tentatively planning for the IM supplement. You may add this information to your VTC notes if you'd like."

Evelyn Gaiser, the FCE lead PI, wants to totally support the VCR efforts mentioned in John Porter's email: VCR: FluxDB (Fuentes) - Data from Flux Towers, WaterQualDB (McGlathery) - Non-stream Water Chemistry (Lakes, Oceans and Estuaries), and CoastalNPP (Pace) - primary production in coastal systems.

We're thinking we'll ask for supplemental funds to hire new personnel to assist with the collection and processing of these data and more specifically ensure that these data are properly QA/QCd, documented (EML) and put into appropriate formats as defined by the VCR leads mentioned above.

Additionally, some funds may be requested to supplement our ongoing sensorNIS work. An additional programmer and travel funds (for SensorNIS workshops) will facilitate our proposed sensor tool development."

Margaret -

question re Coastal NPP, is that wetlands?

Gastil -

I think so, salt marshes. [ed: Margaret's site does have a small salt marsh]

Margaret -

this is the third session. Rather large amount of money, mostly dedicated to IM. Go around the "room", 5 minutes each, describe proposed plans.

Announcements:

(1) Improved data availability. Message from Scott to PIs and mentioned in Call for Supplements. Attention on the "5 Essential Features". IMC has a metrics group. Scott asked us to start collecting some kind of metric we can report to NSF. What we did was look at current datasets in Metacat and tally up occurrence of those 5 features. A hurry because supplements due on June 1st and EB meeting now. EB is getting a report (not quite) the same thing. EB will see aggregate numbers. Site IMs will see their own numbers so we can all see where we fall.

(2) Spoke with Saran on Monday re questions about supplements. 4 questions. 1. how to split 150K between IM and Edu. She says these are intended for IM; can write for whole 150 for IM. Edu in there because network in process of moving schoolyard into renewals but that is not happening for all sites yet so some need extra opportunity.

2. length: normally 2 pages for 30K. She says up to 5 pages.

3. idea of approaching advancing NIS functionality, demonstrate capability of time-intensive datasets. Expressed idea of a cross-site database where scientists, data products designed. She said good way to demonstrate. Having a scientist is important. Or design standard inputs, try to standardize data going into catalog for given types of data.

4. she said these are competitive. May receive partial funding. Keep it modular so can carve up.

Jim L -

We are looking to use this funding to hire people to go thru our datasets. Currently only a 3rd of our datasets have dataTables in EML. We would like to consolidate to... older datasets. Also get some GIS into EML.

Margaret -

is this part of DEIMS?

JimL -

we will support them but (not?) have the funds to put into a developer for going to version 7. We are part of DEIMS and will interact with them. We have old datasets that need a look to try to find any missing data.

Hope -

We want to hire people to increase accessibility of core datasets, update EML for existing data and develop EML for spatial datasets. We want to use the GCE Toolbox to do some of this stuff. Interested in adopting Metabase.

Margaret -

do you have SQL Server or Postgres?

Hope - SQLSver

Have been in contact with John Chamblee.

Sven -

we are still trying to figure out what to do. Locate more of the grad student and one-off data that is not part of the normal sample stream. Adopt parts of the Metabase schema, esp for improving taxonomic coverage. How to catalog the Air Photos. No effective way of serving those.

Yesterday JP mentioned ShrubDB. We have a section of successional lands we have been following and wondering how to measure the biomass, more of a science question, not an IM question.

Adam -

We are in same boat as other sites. Hiring personnel to get to projects, get data online, technician for climate and sensor data to get data in cleaner. Have Aerial Photos, not even georectified yet so need GIS help. EML upgrade to EML 2.1 Adam could do or get help to do. Get some not-currently-metacat-avail data, eml written for that.

Don -

Question for Adam: planning to prioritize making EML 2.1 for this or not know if enough funds to cover that migration?

Adam -

if I dont get to it, then hire an assistant. It is a priority already. Just a matter of time.

Don -

spoke with John Blair from KNZ. Sounds like you have a good plan.

At AND, we still have datasets that could use improving. Migrated to eml 2.1 but still room for improvement with 5 essential features. Hope to engage PIs in that. Want to comply with LTER Controlled Vocabulary. Reorganizing long-term vegetation data, improving metadata, standardizing at our site to prepare ourselves for cross-site Mark Harmon HBR this june to weigh interest of other PIs to build VegDB. Spoke to Matt Jones re international exchange standard developed called VegX (sp?). We would lead only in terms of exposing, telling of that standard, how to map data into that exchange standard. Mark does not want to do further integration at this point. Eventually building VegDB. Sites interested , production data, let us know if want to contribute, and prioritize that type of data.

We also have legacy data at AND. Want to get sensor data near real time, add to provisional datasets in near real time, building up what we got trained in at SensorNIS workshop.

StreamChemDB, led by Sh. Johnson at AND. Prototype available, hope to add more sites, forest or LTER. CWT, HBR, AND, ... perhaps LUQ. Will try to write some workflows to demo utility of pasta. not promising migrating existing system.

Spatial data: not a good stylesheet between ArcGIS's xml and EML 2.1 so may hire a programmer to write that stylesheet. Bypasses FGDC which loses info. Would share that with everybody.

AND will do 135 of the 150 for IM.

Gastil -  
(add later)

Jason -  
Similar ideas. First portion into script that generates EML 2.1. Has holes to fill, better taxonomic coverage. Want to put effort into Dark Data. QA/QC into our current metadata database, rel to Synth Project. Sensor data automation, updating of EML, online. Largest funding portion. Setting up a Kepler workflow here to manage all of that.

Margaret -  
We will run up to the 20% limit because in last year of SBC II. Always have the goal of adding more data, but will have to be part of my regular job, not a line item.  
Now that we have a Postgres implementation of Metabase, need forms to enter data. Hire programmer, hopefully same as one working on 2011 funding, continue with that. Could be leveraged by anyone with pg Metabase.

Half of SBC's dataset are done with Matlab. Want to connect that to Metabase. We have a Matlab programmer wizard. Want to fund his time to connect current Matlab scripts to Metabase and GCE Toolbox.

new data type: spatial data but not GIS. Continuously spatial data. Lat, lon, temporally discrete layers. Huge text files, not easy to handle. PIs want to write a query app; suspect that already exists. Could use advice from GIS people.

Participating in OA Data Management tools project, spec: vocab and attribute std, and EML, and data in general. Not leading.

Philip -  
A lot of work currently looking at 300+ spatial datasets. Identified about 60% had some issue with the data or metadata. Already undertaking those fixes. Bulk of supplement will be programming resources to continue developing VirtualNotebook which is our metadata management system. Bulge that occurs at end of project, folks focused on next project. VN encourages incremental devel of metadata. Produces EML and data package for NIS. Lions share is programming time. Also some funds to lease a new GeoServer. Redesign includes improvements to data catalog, including display of spatial data prior to download.

Don -  
One thing mentioned yesterday: Saran said to build proposals in a modular fashion so if they cannot fully fund, it gives them ability to strike out small part of it, leaving most intact. If write all in one piece, more difficult to fund.

Margaret -  
Yes. Not like in past. Competitive.  
Hap brought up yesterday that last sentence of the Call for Proposals:  
"Requests must address how proposed solutions in either of these areas will be tested and validated."

How?

One way is there is a new interface to the pasta quality engine. It isnt perfect and not easy to understand. But we can run datasets thru it.

Philip -  
metrics. considering customer experience. are the data the right quality, properly packaged? How

useful are the data? Most downloaded data gets discarded because not what wanted or not right quality. How do we follow-up with end user. Ask them for feedback.

Don -

Wondering if this is time to re-energize the data model view of things. When spoke to John Blair (Lead PI at KNZ), they use SQL-Server, outside of the DEIMS model they're trying to standardize. Not familiar with Metabase. If sites figuring how to build systems out of their rdbms-es, what base model might look like. [Ed: he forgot about my pretty poster in 2010 so here's a link to it: [http://mcr.lternet.edu/documents/posters/Poster\\_2010\\_IMC\\_MetaDatabaseMod...](http://mcr.lternet.edu/documents/posters/Poster_2010_IMC_MetaDatabaseMod...) [4] and here's info on Metabase: <http://gce.lternet.edu/public/app/resources.asp?type=document&category=i...> [5] ]

Adam -

we are on outside looking in, with products. Do we scrap what we have and develop towards that system to integrate with network or do we keep putting funds into current system. invested in sql-server but a lot to keep going for us. not a good answer yet.  
forgot to mention: will try to collab with other GeoNIS sites. Farming out or hiring full-time at one site to push this thru. dedicated time to get product done and integrate.

Don -

enough people doing DEIMS and Metabase. Want to hear from JamesB and update of what BobW had discussed 2 weeks ago. Potential for Drupal developer hire. If those are happening, nice to cite within proposal if site working with those. to assure system will be working. standardize data model for DEIMS.

JamesB -

nothing has happened yet. trying to figure out how to cut a contract. these drupal developers want an amount just to scope up the job and bid it. outside the way UNM does things. discussion going on. last I heard, discussing last year's remaining funds with accelerated arra funds to create a larger contract. how to let the contract, how requirements put together. to make sure all DEIMS participants have agreed on the Scope of Work.  
There is some kind of Scope written but not yet comprehensive. Inigo wrote it based on what he wants to do; some response from one of the other DEIMS/drupal devs.

Jim L -

Kristin has some funds from last supplement. Went in to try to see what the scope would be. Sent it out to number of drupal developers to see what is possible. Made up a wish list, fairly reasonable list to move forward. got numbers back. All the DEIMS participants had a conference call and all saw list. All would be willing to put in more money to push this forward. There were sites going to put in funds: LUQ, SEV, NTL. All sites agree this is a good thing to do. Looking at the database model provides an opportunity to come back together with common model, now that have experimented, deviated for own site needs. Learned what works.

Margaret -

most DEIMS sites dedicating part of supp?

Jim L -

not most. all DEIMS sites agreed would be able to use/adopt/convert to new version. some may mean a site would have to come up with small amount to give to dev to put \_\_\_ into new structure.

Margaret

hard to figure where the responsibilities lie.

Don -

question re Metabase. Sites like NWT, MCR, SBC putting in funds to travel to this workshop?  
Call to bring in greater interest. Coordinated?

Hope -  
NWT is putting in funds to attend workshop.

Margaret -  
something similar, either specific funds or non-specific line item.

Don -  
Wade discussed scoping out interest in Metabase.

Margaret -  
JohnCh opened with that last week. People were more interested in Toolbox than metabase so arra for Metabase, supp

Don -  
if receiving arra funds from LNO, need to hear more about it.

Sven -  
NSF has funded these synthesis centers, at centers vs individual sites.

Margaret  
depends how dedicated those synth centers are dedicated to software.

Sven -  
NSF has put funds into Synthesis Centers. Do they expect devel will go on their for sites researchers to use? Will that be a problem with sites asking to develop locally.

Don  
concerned about putting that kind of dev to supp opposed to indiv site work?

Sven -  
KBS will likely ask for funds for devel. Concerned NSF's view re Synthesis Centers to do sw devel and lter sites using it.

Margaret -  
should we find out what the Synthesis Centers are developing? Or would they partner with us to develop things, like EML with NCEAS?.

JamesB  
no. only aware of evolutionary center in N.Carolina. They are doing HIVE and a metadata database focusing on publications, associating data with publications.

Margaret -  
is that the center that replaced NCEAS?

JamesB -  
no. Evolutionary Bio. The new one I have not seen. A meeting there in late June, bringing developers together from bio projects. Mark and Bob are going. Annapolis.

Yang -  
Preparing data package and producing EML. Attempted to put on PASTA but two files yet to debug. Hope to finish by this week.

JamesB -  
have been tweaking prototype data portal to make it look better. All legal EML but adding bits to make it show up better. Tuning the EML to the display rather than tuning the xsl. [Ed: dont do that at



home. Keep content and presentation separate. Better to correct the xsl to properly display EML than hack EML to make deficient xsl work.]

Margaret -

SESYNC ? <http://www.sesync.org/> [6] ?

#### Virtual Updates [7]

- 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/1020>

#### **Links:**

[1] <http://im.lternet.edu/user/126>

[2] <http://www.doodle.com/rm2z728uyfzbtdpt>

[3] <http://goo.gl/pYXtl>

[4] [http://mcr.lternet.edu/documents/posters/Poster\\_2010\\_IMC\\_MetaDatabaseModels\\_EMLcreation\\_at\\_LTER\\_sites.pdf](http://mcr.lternet.edu/documents/posters/Poster_2010_IMC_MetaDatabaseModels_EMLcreation_at_LTER_sites.pdf)

[5] [http://gce.lternet.edu/public/app/resources.asp?](http://gce.lternet.edu/public/app/resources.asp?type=document&category=informatics&theme=Database%20schemas)

[type=document&category=informatics&theme=Database%20schemas](http://gce.lternet.edu/public/app/resources.asp?type=document&category=informatics&theme=Database%20schemas)

[6] <http://www.sesync.org/>

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