October 8/9 2012 - Discussion of potential proposals for NIS production workshops and training

Home > October 8/9 2012 - Discussion of potential proposals for NIS production workshops and training

October 8/9 2012 - Discussion of potential proposals for NIS production workshops and training

Wed, 09/26/2012 - 9:13am — mobrien

doodle poll: http://www.doodle.com/ygeagtz32bqiddin#table [1]
note times are 1pm mountain and 10am mountain, respectively. the time settings were left on EDT

notes below for Monday and Tuesday

Monday Notes:

8 Oct 2012 IMC Water Cooler
Discussion of potential proposals for NIS production workshops and training

attending:
Corinna, Wade, Jason, Suzanne R, Margaret, John P, James C, Gastil, Duane, Mark

Margaret moderating. No presentation. Early feedback on proposals for 2013 workshops. A list started at ASM, in the ppt, posted on ASM IMC agenda page.

Announcements from IMExec (at end of this vtc)

Wade
Ideas for trying to determine need for and scope of an EML BP version 3 doc, or the work to produce that. Discussed scope, what to cover. In addn to updating BP to be more useful to use EML to support loading data into pasta and with workflow applications, also need to look at more theoretical treatments in v 1 and 2. Update for real-world use cases, actual applications. Started realizing we are merging into notion of data pkg mgmt, data archiving best practices. Good to get opinions whether useful to start pursuing that. Data pkg mgmt (data and metadata formats) optimized for synthetic activities. Beyond what and how in EML to how to best factor the data to deal with table arrangements, missing value codes, layouts. Is there enough community interest in participating and using the information. Would be more valuable. If just a revision of BP doc, may be able to do with just vtc and writing assignments. If expand to packages then beneficial to have in person working group meeting, with prep work, and broader recommendations.

About half the people on this call are in this WG: Margaret, Gastil ,Duane, Servilla, Corinna, Wade, John P, Sven, Ken, Dan. One third of network.

Margaret
Problem of widening the scope, unless focus on particular uses.

John P
Could be linked to (Wade's) EML BP: get together small group of scientists where attempt to use tools to actually try to do some synthesis activity. Actual use will illuminate issues. Semi-automated processing of EML-described data.

Wade
That is worthwhile. If we corral that process, go them to focus on themselves using the tools but on helping clarify what are target output types for such projects. What format would they like to pull all LTEr data into? For example a data frame in R, or a consistent spreadsheet file with certain characteristics. Clipping data, met data, species data. . . what tools, actual examples. May be premature to have them try the technology now. Not quite ready for real users. Translate their needs into practices and workflow tools.
Particularly with attributes. Concerns with EML for synthesis in units. Scientists have had trouble with existing unit represenations. Use the (camel-case) EML units just as unique identifiers, not as titles on plots.

Margaret
Agree re units. We are using the unit id's as if they were labels. That has been a problem.

We are looking for what the Use Cases are to learn how to build. Maybe timing right now to try to "infiltrate" a working group.

Corinna
Agree. We will not get much out of workshop just asking what they want. They want everything or they dont know what they want. Better to infiltrate science working groups. I have tried with a WG at NTL run by a couple students. By being friendly, they might invite me to their workshop or talk to me afterwards with results in hand. Have to hear about these working groups and talk to them one-on-one.

John P
There are some things out there. At ASM we talked to Mark Harmon about VegDB. I sent him an R workflow to merge two biomass datasets in pasta. Biggest challenge working with existing WG is we need to learn from the process. That can slow down what they are trying to do. 5 hours to get into a workflow vs 15 minutes into spreadsheet, losing all provenance and workflow capture. StreamChemDB, VegDB, there are groups intersted. There are also the Coastal Water Chemistry (coastal version of StreamDB) could try to partner with. The best way to learn what is required to make one of these synthesis activities work is for them to try the tools.

Corinna
Not intending to make workflows for them at their workshop. After the fact, using their spreadsheets, we coudl model the process with a workflow. One product at a time rather than everything possible.

John P
2 workshops: one for IMs to make questions for scientists to work on, then one for scientists to work in collaboration with scientists. Also think great to do WG related to units to find preferred forms. See all options currently but no vetting.

Corinna
Enough if we do what we proposed with first workflow workshop, go thru datasets that need ClimDB format. Enough to start with. A dataset from every site which needs to go into ClimDB. Eye-opening and attainable and useful for data management practices.

Margaret
ClimDB example has come up before. Would be productive. Several of us have questions, places where current recommendation of ClimDB may not be good practice. ie timezone for aggregation. We sort of mapped a path forward for ClimDB. Have an EML template for ClimDB input.

Suzanne
With ClimDB, we have an issue that the Climate Committee is defunct. there are decisions that the IMC cannot make for them. Julia Jones went to the climate meeting.

JohnP

3

Bruce Hayden at VCR has been involved with the Climate Committee. We could ask that group questions. ie could ask if should extend to hourly data. Could identify things in EcoTrends that are not being updated that maybe could be automated with workflows.

Wade
I like that idea because ClimDB there are decision points to work forward. Discussed replacing some data submission ends with workflows. But EcoTrends does get us back to one of the long-standing goals of PASTA. Moving towards rationale for developing pasta in the first place. Take a few high-profile datasets in EcoTrends and figure out what is needed to standardize data and develop workflows to do that, that could have some appeal. Since ClimDB is a functioning system, less novel.

Margaret
How do we identify the high-profile datasets in EcoTrends? It is not at LNO?

MarkS
EcoTrends has been located at JRN for 2 years. Dont know if they are doing download stats. Thought there was some progress thru Science Council to identify high profile datasets in EcoTrends.

James B
Is there a wish list for ecotrends to make dataset from EML in metacat? ANPP and all precursors, biomass, etc.

Is there tracking of use of EcoTrends datasets? Can we get access to that to determine which are important.

John P
Would like to see more sophisticated data preview tools.

James B
Are we tracking data downloads?

Duane
There is a database table dt_audit. Yes we do track.

James B
ANPP was pointed out as one potential high profile dataset, according to network chair.

margaret
who is the ecotrends admin? Ken Ramsey.

Wade
Another option is to jump in on ground floor with VegDB because no legacy baggage.

James B
will get productivity there with VegDB.

Suzanne R
Expect that Mark (Harmon?) would appreciate help getting into pasta, either earlier or later. There was nobody representing ocean sites but there were grasslands. Biomass.

Wade
They (VegDB) wanted workflows and support to calculate biomass from allometrics such as clipping studies from varieties of biomes. Then take datasets of stem counts or dbh measurements, plot level data, apply these workflows to generate biomass. Broad applicability.

John P
Appealing: a range of difficulties. Grassland studies relatively easy. Allometrics may require additional database of site characteristics such as slope, requiring additional data gathering. Could see how far workflows could take us toward that.

Interfaces. If we produce value-added dataset, does that go into pasta? Or would there be a specific interface for just this kind of data product. Is ClimDB the only one with specialized displays? There are tools to make a kit to use EML to produce graphs.

Margaret
Two workshops: one where build workflows for VegDB. One where we gather lessons learned and put them into a special case of best practices.

John P
That might be 3 workshops. Useful effort to guide BP for EML. Gather people working on tools to make tools more useable, more robust, before present them to PIs. (Shocked to find Nitrate+Nitrite as a measurement name)

Wade
Effort to push workflow technology forward. Informally capture ideas for EML and data pkg mgmt. Follow up later to work on BP documents. Do another workflows workshop. Use EML-based tools to help develop concepts for what are best practices: data formatting and archiving practices.

Time horizon is August 2013. We have a WG. Can accumulate input for BP under normal IMC wg activities. The more important work is actually pushing-forward the workflow implementation.

Margaret
is VegDB putting in a proposal?
Suzanne
Yes

Margaret
Units. The Units WG does not really have a chair. Linda would like to not have that role. Next task is to start the units vetting process, cleanup, and map to reasonable representation. That group needs a parent.

MGB
Workflow training?

Wade
Did discuss training on workflows. Would like to see more dev work first with more tame audience. Woudl be frustrating for end users. Still dealing with fundamental issues, strange incompatabilities. Shaking out the details. Not helpful to have a scientist trying to get real work done. More helpful to first develop workflows for use cases ourselves, then later train scientists.

John P
After ASM was thinking a workshop of IM and non-IM might be useful but based on Wade's points, might want to focus that on the IM component. Perhaps a training and other workshop for just IMs.

Automated tools only take you so far, data ingestion process. But up to you to know enough about that sw to do useful things to get further. The workflow will not teach you R, matlab or Kepler. Are there routine tasks that need to be done. Example: standardize date formats. There may be some re-useable code.

_____

Duane
Working Group scoping question:
EML Metrics & Congruency Reporting Group (Philip, Adam, . . .)
EML Data Package and Reporting

Rolling out new sets of quality checks, congruency checks, over 3 or 6 months. Which group will drive the suggestions for controlling that set of checks. If EML Metrics and Congruency Reporting group is complete, would it now be reasonable to ask Reporting group to take over?

Margaret
Good question
Initally I thought (B) that Reporting group is concerned with audiences and reports. Then . . .
Checker group would still do that (what?) but not actually build reports from the result.
These groups definitely have to communicate.

Duane
The report is the last milestone of the Congruency group and now becoming inactive. Asked Philip if scope of reporting group could be expanded to include development of new checks.

Margaret
Just wishing that Congruncy group was done.
Maybe the report group tells the Congruency group to do more work.
Or Duane asks when he tackles more checks on the list.

Suzanne
On this new group. Moving fast. Data pkg reporting is very different from EML Metrics. Not this group's task. There is an issue that there will be questions as pasta goes into production. Duane will need guidance from the metrics group. So that group cannot become totally inactive.

Margaret
Metrics is not the title. We're the Checker group. We discuss the mechanics of the checker. The reports will report metrics. The checker is defining bits of logic: expectations what should be found, defining what is an error. Good to know what the current pattern is in existing eml docs. "Checker" for mechanical group, "Report" for report group.

Margaret
IMExec items:
Began last summer, IMExec members phoned IMC members to discuss activity. Message received was that people wanted to be better informed. Did not like elitism. Changes have been made. In past years, IMExec previewed all the proposals. Instead this year, we will all post proposals where anyone in IMC can comment on them. There is no ranking, no pre-review. IMExec will not serve as a review panel.

In the past, each person on IMExec acted as a liaison with an active WG. Now instead, when WG wants to inform IMExec, best way is for WG to hold a vtc. In the month prior, someone from the WG would join into the IMExec call, use as sounding board, get feedback on items to highlight.

Suzanne
Totally getting away from liaison? Might need a whip. Might still want an occasional visit to group.

Margaret
IMExec might prompt a WG.

Corinna
Elaborate on "egalitarian" and "e?"

Margaret
Some comments last summer was that IMExec was not getting enough information out to IMC about the goings on last spring. In May we had about a week to decide that we did know how to add up what was in datasets. IMExec just decided to do that.

Corinna
That is why we have IMExec. It has to be that way.

margaret
In general, the IMC is ok with giving IMExec a lot of power until they're not.

Corinna
IMExec posts minutes and anyone can read those and call up if there's a problem.

Corinna

Another WG proposal suggested at ASM:

SensorNIS

may want to put in a WG that explores best practices for keeping information about sensors. Looking at OGC Sensor Web Enablement (SWE) and other places where a lot of sensors are manageed. Sensor metadata, calibration schedule and such information. Perhaps extend into recomendations for flagging data QA/QC from sensors.

Margaret

Sounds familiar. Please tell me what I should be doing. This would address that question.

Corinna

Would help sites evaluate the different approaches.

Wade

The time is right for it. An emerging issue. Few are committed to a strategy or system. Good time to offer recommendations. A WG to explore and share systems would be useful. The SensorNIS WG at ASM, a lot was already put together about . . . best practices document. I would favor being involved with a proposal for that.

Corinna

What is general opinion about training? Do we have enough training proposals going in?

Margaret

I thought 'train the trainer' was next. Does that have a person to propose it?

Corinna

Introductory training should be done at each site for their own people. John P suggested another site's IM present at other site.

Advanced db, advanced Kepler, advanced XML, offered to IMs first then filled out with scientists.

May be too late now to find trainers to do these things.

Wade

The GCE Toolbox training is a separate track.

JamesB

There is some training for scientists being put in.

John P

I started a data intensive science lunch time informal group. There is value in having a strange face show up to give advice. If scheduled then higher priority. We could be a train the trainer workshop. One day? One hour?

Jason

Hybridize. On-site, invite addl IMs, raise level of training. Would like to be

involved.

JohnP
Have used remote vtc training successfully.

Who is interested in the writing of particular proposals?

Workflows - was Wade & Corinna last year
Wade could take lead this year. (Corinna wrote it last year.)
Mark (Servilla or Harmon?) could provide input
Could leverage the Coastal Water Quality suggested supplement.

EML BP - defer in-person
First capture feedback from real use cases

Corinna
have started to move EML BP onto web pages
would be good to start updating that as living doc (remotely). That could be
the draft version.

Train the trainer. 5 sites.
2 hours with grad students, prep for site IM.

Suzanne
We do training every year at our site. Visiting IM may not enhance.
John P
Could make it a condition of coming that a signif portion of grad students will
attend.

Wade
Video training modules that could be used by sites. More lasting value than
face-to-face. Long term more impact. Would allow IMs to expand expertise
beyond their own.

John P
Best way to develop online presentations is to do live presentation, refine, try
out, then ready to go online. Repeating a presentation lets you refine it more.

Corinna
That is what we got funded last time. EB made it clear that funding depended
on that. The videos from the August training is online now. Next need a
professional polish these materials, use pedagogy.

John P
Production values. Those need to be coupled with (hands on) exercises. Cannot
just hand it in a workbook. There are no exams associated. Some portions
would need to be re-videoed.

Corinna
This is a serious investment of time to professionally complete those materials.

Margaret

We're not teachers. It is a lot of work to develop curriculum.

John P

Scope. List of things to learn to manage grad data. Not in-depth training of how to do it. Just a start.

Corinna

Should I propose to involve a graphic designer or someone who knows pedagogy?

John P

Could propose primary funds go into materials, then use materials for trial training at 3 to 5 sites to test using them.

Tuesday Notes

October 9th LTER IM Watercooler:

1. Attendees: Sven Bohm (KBS), Don Henshaw (AND), Hap Garrett (PIE), Yang Xia (LNO), John Chamblee (CWT), Inigo San Gil (MCM), Theresa Valentine (AND)

2. Decision Making Changes on IM proposals: John C. reiterated the new process for decisions on IM proposals. New changes reflect what is stated in the IM Exec meeting notes: http://im.lternet.edu/imexec/meeting_notes/2012_09_13 [2] This makes LNO the final reviewer of proposals and matches what is stated in the Operational Plan. IMC comments/suggestions will be on an internal website and be from October 15-Nov 1. See James Brunt's email for details.

3. Proposed Workshops/proposals:

a. Theresa: Working on 3 proposals (1 training, 1 post-doc, and 1 production oriented workshop). The production oriented workshop will be for a meeting of key GeoNIS IM folks, GeoNIS contractor, and LNO @ ABQ in January/February timeframe 2013. Purpose is to work with contractor and LNO to make sure systems are operational, and scope of work is progressing to work with PASTA.

b. Inigo: DEIMS group: training proposal, widening the reach to everyone. Make it more basic, wider audience. Post-Doc, Eric Sokol has approached some IM's about a synthesis proposal on biodiversity: wants a data manager: Corinna, John Porter, Inigo. Tasks: prepare datasets for pasta, create workflows, R scripts that will prepare data for analysis, and wants a workflow to automate searches in PASTA, create EML for derived data sets. Inigo will be working with him. John C. can also assist. Question on DEIMS? Any effort towards documentation? Inigo said that there was insignificant at this time, with DataBits and Peer review articles, however asking contractor to help documentation in RFP. IM Videos: People are watching them, generating questions at unrelated talks.

c. Don: StreamChemDB, VegDB: Working with Mark Harmon on VegDB, Science proposal. He wasn't planning a production workshop for work flows. Not sure what was decided yesterday.

4. Comments on working Groups:

a. Need to coordinate the science and IM workshops. Would help move things forward. Involve IM's early and often in synthesis products. IM and VegeDB did good job of integrating at ASM. Not clear to Don who's writing what proposal.

b. SensorNIS: how to manage metadata about sensors, how they are deployed,

calibrated, etc, how to flag the data. Don not sure if he or Corinna or Wade is writing. Don will follow up with them.

c. Sensor middleware training: Don doing the writing.

d. ARRA funded workshop @Georgia, documentation for using the toolbox for sensors. Wade is key player. Don will be in contact with Wade and Corinna.

e. List was made at the meeting of proposed working groups/training, however it was not available for VTC.

f. Best Practices Document: covered yesterday, lots of e-mails going on about it.

g. Web Services for personnel database from ARRA funds (not part of this proposal?)

h. Proposals should be posted on the page that James sent out, and we'll be able to review and give comments. Production workshops to be decided by December 1, as noted in email from James Brunt. Deadline Nov 1for final.

5. Communications: Change in IMEXEC, bottleneck between IMC and IMEXEC. Doing away of liaisons, any working group is encouraged to schedule time with IMEXEC, set up to do VTC's. Meet with IMEXEC the month before. More as a communication, helping mechanism. In ASM final IMEXEC meeting notes. Not in terms of reference, don't have to do it, etc. Helps us figure out the schedule for VTCs, might keep from being so re-active and pro-active. Hope to have outcomes from VTC's, help working groups.

6. DataBits: looking for articles by next Monday. Little more flexible in the fall. Little extension of time, might be good to have each of the sites that participated in SensorNIS, do a write up of what they are doing. The spring issue included something similar, so Adam should be up for it.

- Virtual Updates [3]

Source URL: http://im.lternet.edu/node/1082

Links:
[1] http://www.doodle.com/ygeagtz32bqiddin#table
[2] http://im.lternet.edu/imexec/meeting_notes/2012_09_13
[3] http://im.lternet.edu/taxonomy/term/169