July 2/3 2012 – ASM IM-related working groups

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

Home > July 2/3 2012 – ASM IM-related working groups

July 2/3 2012 – ASM IM-related working groups

Wed, 06/27/2012 - 10:49am — mobrien

Already proposed:

1. **Adam Skibbe:** GIS Data and Tools for the LTER: resources for site research and synthesis [1].
2. **John Porter:** A Researcher's Guide to Automating Analysis of LTER Data [2].
3. **Inigo:** DEIMS
4. **Don Henshaw:**SensorNIS: Building a sensor network resource guide through community participation [3]

Likely to be proposed:

1.**Theresa Valentine:** GeoSpatial Coordinating Committee, maybe focused on using LiDAR to calculate biomass.

2. **person TBA:** workflows (based on imexec june notes)

3.**John Chamblee:** white paper outlining current state and potential future directions for LTER Information Management.

5. **O'Brien (with a co-proposer TBD, she hopes):**Data inventories.
The 11 proposals submitted in 2012 required a data inventory. The 3 sites in the synthesis data pilot-project have also produced inventories. We could compare how we built these, identify the common components, and/or draft a general inventory template. There is a difference between an inventory used for data package management and one for a proposal. The Tuesday-group thought this should be in 2 - 2 hour sessions. Margaret will ask NSF for feedback on on the usefulness of the inventories that were submitted with proposals.
Draft title: Developing data package inventories - I and II:
I. Design and Management of Data Package Inventories – include an overview by Yang on the inventories the LNO received in response to Scott's request
II. Strategies for finding the 'Dark Data' – focused more on the experience gained by the synthesis data pilot sites as they add more data to their inventories

6. **Corinna (or TBD):**What could a Central IM system look like?
A scenario building exercise. Everyone can go thru the idea of what could be done centrally and how it would affect individual sites. Get the discussion rolling on possible alternative ideas for certain aspects of information management that might increase efficiency (eg, not everyone has to know every detail of EML) Looking to the future.

7. **Emery & Phil:** Conceptualize what the future LNO should look like.
Emery and Phil are chairing the network group for this. They may propose an activity for the ASM, IMC should also discuss at our meeting, and focus on technical aspects.

9. **Sven?:** standardized attributes
In order to compare data (e.g., in a workflow) you need to know full details about table attributes. One way is to use standardized attributes where possible, or develop synonyms that sites can map to their local attribute-definitions. In 2010 we started this discussion. Probably in 2 or 4 hours we could review old discussions, and outline a possible trajectory for an attribute-dictionary project. For example, how do we accommodate data table attributes into a workflow? Does EML have the structures we need? What systems do we need, and what is the timeline?

Other ideas - ASM:

3. **Eda CMC:** Discuss ways to enhance the relationship between a site's Information Management (IM) and its Science components.
How can IM collaborate with Science in its "plan for synthesis of long-term data to answer important scientific questions"; How can IM collaborate with Science in "demonstrating how each data set is to be used in the proposed research."

How can IM assist Science in the Knowledge Discovery of their own databases (KDD [4]). Are these issues IM-related, Science-only-related or a well-thought collaboration and system design is needed to achieve syntheses of a site's pool of data?

When fully developed, the NIS will foster data discovery to do cross-site syntheses…the issue I am presenting here has to do with ways the information management staff of a site can collaborate with the site's scientists to facilitate the site's own data knowledge discovery and ways to design an information management system that fosters this.

4. **person TBD:** data package design practices
several people have expressed an interest in having guidelines for dataset design, e.g, another set of 'best practices' recommendations. These might be particularly useful for packages we expect to become input to workflows, or also could include guidelines so that certain data could be presented in a similar manner. Look for emergent patterns.

5. **person TBD:** Metabase users?

6. **person TBD:** Integrating all those other DBs,
like VegDB, SiteDB, ClimDB, projectDB, personnelDB, untsDB, streamChemDB, myFavoriteDB

Add to IMC 1-day agenda and assess need for ad hoc WG then:

2. **IMC chairs? or TBD:** Supplement funding coordination
As of July 1, we don't know how much supplement money will be going out to sites, but we should know by late September. Even if funded at ~60%, the total amount could be about $2 million. We know that there is some overlap among sites' needs, and that some funded projects could have cross-site use. So some degree of additional cross-site coordination could be necessary. A working group composed of sites with supplement funding could a) summarize their funded projects and b) determine any needs for additional coordination, e.g., a general workshop which is not covered by the supplement projects themselves. Address in one day meeting.

3. **person TBD:** Compare NSF reviews of IM sections in 2012 site proposals. NSF has ramped up it's interested in data availability. Depending on how soon we receive the proposal reviews, it could be interesting to ask each of the sites that were recently renewed to comment on the reviews they received addressing information management .

7. **TBD** Conceptualize what the future LNO should look like.
Discussion at the IMC meeting and maybe have an ad-hoc group. Focus on technical aspects. Coordinate with Emery & Phil.

Dropped:

7. someone from dataONE? or an LTER IM who did some prep first:
An introduction to using the dataONE API for LTER Site catalogs.
the OneMercury interface seems fast and easy, although not publicly introduced
yet (as of early July). How can the DataOne API be used by sites in their own
catalogs? With a little background, a working group might be able to develop
some demo forms pretty quickly and also promote a common look-and-feel to
site data catalogs.

Virtual Water Cooler notes from Monday, 2 July 2012

attending
video: Corinna Gries (NTL), Don Henshaw (AND), M. Gastil-Buhl (MCR),
James Conners (CCE/PAL), Jason Downing (BNZ), John Porter (VCR), Wade
Sheldon (GCE), Hap Garret (PIE), Kristin Vanderbilt (SEV), Suzanne Remillard
(AND), Yang Xia (LNO)
phone: Ken Ramsey (JRN), John Chamblee (CWT)

Don
Use ideas to build sessions for IMC afternoon and also wg during ASM.

John P
We proposed a workshop to bring in researchers to teach them about workflow
tools. Idea is to try to train researchers in use of workflow tools: Wade's matlab
web service, toolbox; Corinna's Kepler workflows, Sheng-Shen Lu's tools from
Taiwan to pull in eml and walk thru analysis. Web services for R and spss
(statprog).
Requesting two 2-hour sessions. Possibly try for just one. One session of
presentations, basic familiarity. 2nd session more detail, hands-on, bring laptops.
Push benefits of EML out to research community. Aimed at researchers, graduate
students.

Suzanne R
I recently gave a talk for Judy Cushing's class on stat methods for ecology.
Wants to put together workshop on data analysis.

John P
visualization and data analysis can be 2 hours in itself. The wg is aimed more
at learning the tools. Perhaps she (Judy) would like to see statprog, to get thru
the loading phase quickly. Could show the statprog urls for R, SPSS, SAS which
load an EML dataset. Those could save her some time, the rote work for loading
data.

Wade
Would be good to have both IMs and researchers to open lines of communication,

4

this is a partnership. Put effort behind tools that can deal with nitty gritty to get to synthesis more efficiently. Would be nice to make this a meet & greet.

JohnP
Any IM can come as long as they bring a scientist or grad student with them. Really want to speak to the researcher community.

Corinna
Proposing to do a workshop called Central Information Mgmt for LTER Sites? A scenario building exercise. Everyone can go thru the idea of what it would it look like, what would it do for your site? Is it completely out of the question? Get the discussion rolling on possible alternative ideas for certain aspects of information management. Notes on why does everyone have to know every detail of EML. Why not do this more efficiently? an imaginary scenario building. Just an idea.

Wade
We owe it to ourselves and everyone envisioning next iteration of the network office to go thru that exercise. If we go in with open minds and discuss as objectively as possible... with in mind future construction of the next iteration of LNO. Interested in participating.

John P
Would you be happy to talk about LTER information mgmt 2.0 or 2020. While there are some solutions centralized, also happy with idea of web services, so really central but appears distributed. If just talk about centralization, may not bring in all topics, such as web services. All advantages of both central and distributed.

Corinna
Not a technical discussion. PIs might be interested.

Wade
I agree. Emphasize "New models for Information Management". Centralized presupposes a specific structural approach. Balance centrally-provided along with site-supported services. Need to re-envision supervisory structures. What would be a new model for more efficient IM that would be sustainable? We have been told our current model is not sustainable. This should be an "out of the weeds" discussion... By the time this comes in we may be in different planes of technology anyway.

JohnP
PIs may say sure just do it central.

Corinna
Then they could go thru this mental exercise of what would this mean for your site.

JohnP
I think if you take the IM out of the site then you take away the connection,

"your person".

JohnC
I agree with JohnP.
White paper. One shot. Have not circulated this yet: an outline of what we talked about in IMExec. 6 presentation workshop conference presentations. Then hour discussion. Then what distills goes into white paper.

Corinna
Getting out of our comfort zone. Draw the PIs into the discussion. No consequences, no paper. Completely creative.

JohnC
One concern is my perception is there is not an understanding of IM, everything that is involved in site level IM, even as it is now. How will a discussion like that go if folks do not have the frame of reference of what we do now.

John P
One thought... concerned that with no presentations, group has no direction. Maybe 3 or 4 short presentations of Bold Visions of what LTER IM could look like 10 years from now. One might be linked data, the concept that dissolves the dataset. Find data point by point. Search. Taiwan group does stuff with that, RDF sparkle. Another might be a centralized model of IM. Make these bold and stark. then use those to drive the discussion. Round table, half group move. Hold certain outline for the scenario. As much as I do like the linked ata idea, not sure that will attract or get thru to the scientists.

Don
Two 2-hour sequence, one focused on white paper, current view of IM, somewhat what Wade mentioned for future model of LNO. Then second 2-hour session for the bolder central view. Include what LNO might look like because NSF is interested in looking at new ideas.

Wade
The big moving piece on the chessboard is LNO potentially changing. The impact on network-level IM, how it functions, could be large. The better we can help steer that process (LNO changing) the better off we would be.

SuzanneR
That format would not bring in the scientists that Corinna is after. Her idea PIs would be interested in, or Exec comm from sites, because they know what the system is. There are people at each site that DO know what is in their IM system.

JohnCh
The main thing with the white papers would do is to bring in the PIs. We have one shot at bringing the PIs in. ASM will be very busy. We can get them for 2 or 3 hours. Whatever approach will get them engaged and productive. After the 30 year review, we felt like the white paper was necessary so we could put together a veiw of IM as it currently stands. Need to let people know what we

are doing now and where we stand. But thinking about the future might be a better way to go.

There were serious misconceptions about what we do. Perhaps that is past.

JohnP
At the ASM, it is catch as catch can. Those you most want are likely to be running another workshop. Perhaps put in a formal workshop proposal to NSF. 2 or 3 day workshop. Bring in top people.

JohnCh
idea of the product: one reason was thinking of presentation format is because if everyone came with completed presentations, then discussion at end, then would have a product. 15 minute papers. supporting material. distill down to WP. NSF workshop might pull in a few PIs interested in this.

Don
Sounds like there are two separate workshops.

Hap
I like Corinna's idea because what we need to do as IMs is to get scientists to be more active stakeholders. How do scientists feel about network system, what can they do with it. We should be at network level. What can we do with IM at network level to do work yet. Not the details. As Corinna explained, the feedback you get from people is the product. Need more engagement to know direction to go.

JohnP
At the last ASM 2009 we brought scientists and IM together to discuss what the needs were. Scott Collins primarily. What they want is integrated, polished data products. They do not want to do unit conversions. They want data that has already been through extra QC. What we call value-added or derived datasets. Right now, we do not have good methods for producing those. We are mired in low-level work, leaving no time for integration. Nearest thing has been EcoTrends. One idea is to talk about the EcoTrends experience from data user experience.

Those are science products. Not just take this file merged to that file. All sorts of scientific decisions.

———

Don
SensorNIS workshops last fall and this spring. Talked about a followup DataBits this next issue. Conducted training. Is it worth doing anything re sensors at this point and what can be done? We had 45 people at Sensors workshop last time. Expect interest is high.

Corinna
Could discuss what is the way forward from here. Left that open at end of training workshop. Where do we go from here. What kind of documentation

would help and where should it go? Should we start a group? Is ESIP a good forum? Need to continue ongoing discussion.

JohnP
One thing that brought in 45 people in 2009 was to see the neat things sites were doing with sensors. Still a role for a workshop for experienced users to show their successes. Then round table to discuss setup and management. More the scientific side.

Don
At IMC meeting half-day could have short forum on sensorNIs workshop, think about QA/QC there, then leaving the ASM workshop for something - not hands-on training - but perhaps reprise some talks to show what tools are out there. What next steps could be.

Wade
My biggest concern is the amount htat can be accomplished in that time. All would like to do more with sensors. Some kind of RCN or workshop proposal, bigger, to identify tools, standards, documentation. Help people craft the idea where work is needed. Had good discussions with Derik Barsegian and Michael Nekrasov on ways we could try to leverage the sensor description standards to handle the QC stuff. Good to scope those things out in a working group format.

Don
Probably worth submitting for a workshop for SensorNIS.

Ken
The five essential features such as title, abstract and keywords: how we can improve our titles and abstract so they can be more useful without talking to other sites?

Don
That might be something to spend time on at open meeting of IMC, improvement of EML.

JohnCh
Need to talk to others offline re PersonnelDB because put that in a production proposal.

Don
We have several of those types of projects that we have not been able to move forward. Part of that is folks are waiting to hear what comes back on mid-term review of LNO before they can commit funds. Waiting a few more weeks.

Don
We might want to hear from DEIMS or Metabase groups to hear why folks might want to participate in those activities.

Wade
Waiting to see if Metabase activity is funded. Will not move forward unless get release time. A couple parts of that. Pieces could be useful to sites in different

ways. A workshop toward end of year. Tentatively yes, depending if that moves forward.

Hap
Regarding resources, do we know when the supplement proposals' funds may be coming in? Will we find out right at the ASM?

Don
We expect to find out in the next couple of weeks. Not clear. Margaret suggested we could discuss coordination between sites of using our supplement funds, as proposed.

Corinna
Even if we don't get all the funds for DEIMS, Im sure there will be some DEIMS workshops and get-togethers at ASM. We will definitely have to coordinate. I

Don
Not clear if that is an ASM workshop or something informal.

Wade
Nothing planned independently. Informal get-togethers planned. Would like to leave open some time commitments to make GCE Toolbox more useful to more sites.

Jason
One idea. . . personnelDB, SiteDB. Might be able to draw in a few more people. Could be an offline time or a formal session.

Don
We are trying to figure out how we can take a system like ClimDB and have it functional within PASTA. At short SiteDB meeting we discussed how we might use that sort of framework to house site information. Could use a lot more feedback on direction. Could be an informal time.

Suzanne R
The SiteDB thing came up at the VegDB workshop a couple weeks ago at AND. Some had not realized that their climate stations had been added to SiteDB. They were interested in knowing. . . what characteristics should be included in SiteDB. There is interest out there in expanding SiteDB.

Don
How can we assemble the extended metadata to describe those sites, ie for veg.

Corinna
You should just schedule this as a workshop. Not all ASM workshops need scientists. If you get different people, you get different ideas. Do not schedule so much for the off time. Make it an official workshop.

Don
if we can get people to commit to come.

JamesC
Interested in a lot of these topics.

===
end of Monday's VTC
===

======
Tuesday's VTC
======

Attending:
VTC: James, Margaret, Emery, Jason,
(Phone) Sven

- Virtual Updates [5]

Source URL: http://im.lternet.edu/node/1030

Links:
[1] http://asm2012.lternet.edu/working-groups/gis-data-and-tools-lter-resources-
site-research-and-synthesis
[2] http://asm2012.lternet.edu/working-groups/researcher%E2%80%99s-guide-
automating-analysis-lter-data
[3]   http://asm2012.lternet.edu/working-groups/sensornis-building-sensor-
network-resource-guide-through-community-participation
[4] http://resources.metapress.com/pdf-preview.axd?code=n156r64152288441&size=largest
[5] http://im.lternet.edu/taxonomy/term/169