

### **Working Group Reports**

The remainder of the morning was taken up with IMC Working Group reports. Groups were asked to provide updates on recent activities and to submit preliminary requests regarding the kinds of activities and the amount of time they would like to have allocated for their working group during the Annual IMC Meeting. In addition, three working groups provided reports relevant to PASTA Implementation. The IMC Website and Unit Dictionary working groups did not provide reports, because they are inactive.

**GeoNIS / GIS (Adam Skibbe)** The GeoNIS group met at LNO in January and after discussions with Mark Servilla and others, the GeoNIS project has been redirected primarily toward a series of web services that will both allow users to check for the congruency of their GIS data before feeding them into PASTA and will also allow users to use GIS data in PASTA through web service calls already supported by existing GIS software systems. This will be done by a programmer and paid for via a supplement. During the question and answer session, it was suggested that detailed tutorials are needed on using the services, the spatialraster and satialvector capabilities of EML in general, and on the development of EML for GIS data. The GeoNIS working group should be able to demo the services at the IMC. Training materials will come later.

**Web Service (James Brunt)** Rick Clark, the programmer for the PersonnelDB services has worked through most of the use cases and has also worked out the authentication and authorization with Mark Servilla, as PASTA will use roles defined in Personnel DB. Authentication/auhorization will be passed in the request header using standard session protocols. Role management will be improved and there will be services for all major functions (insert, delete, etc, though in lieu of deleing information people will be inactivated. The developed materials are currently installed in a test system and the front end uses web services. Most code development will be done in April after a conference call with the web services working group supervising the project (Wade, James, and Sven). The new personnelDB services will go live in June 2013, but early adopters can begin to develop code sooner. The programmer must finish by August, since his work is tied to the ARRA funding. The web services WG will continue to work on other databases (site DB, etc), when this is completed and there should be a demo ready for the IMC.

**3. Matlab / Metabase (John Chamblee)** This workgroup held a workshop last November. Listserv for toolbox users.

- Wade is working on toolbox updates. Refactoring how data goes in, updating QA/QC tools.
- Toolbox has good support for streaming data. Need to write harvest script if a particular sensor is not supported by toolbox.
- Upcoming: refactoring metabase data model to work better with toolbox, upgrading toolbox user interface, improving knowledge base for users.

- Richard Kerry (CWT assistant site manager) is writing up detailed documentation for Campbell data loggers.
- Need to finish current development efforts by end of April. Then move to maintenance mode.
- Several LTER sites are current users of toolbox (AND, CCE, SBC, NWT, SEV).
- New metabase users: SBC, CCE, NWT, HBR. SBC and CCE ported to PostGres.
- Toolbox survey. Some sites indicated an interest in using toolbox for selected datasets. Matlab licensing is an issue for some sites.
- LNO might provide a MatLab process server if licensing is a problem.
- Currently there is a single developer for the toolbox (Wade). Code is modular and well written. No copyright issues for toolbox (Wade secured a GPL).
- Subversion repository for source code at GCE.
- But core functions are known only to the developer (Wade).
- A slow process toward a few common systems in the LTER Network.
- Some backup knowledge of these systems at the LNO would be good in the future.
- IMC meeting: reconvene group?

#### **DEIMs (Kristin Vanderbilt) </strong>**

DEIMS has contracted with an outside developer group who will finish the module. Final by May 2013 Palantir (<http://www.palantir.com/>)– really good but really expensive – will work faster. Currently in discovery phase, where they look through code. Final product will be fully featured, with most of what we need. Development process is very fast – there is one week to test and approve. Three sites are closely involved (Sevilleta, LUQ, and NTL). Additional support will need to be paid for. The code will be contributed back to the community as a module for download.

**SensorNIS (Don Henshaw) </strong> A new training event for real time sensors is being planned in early April by LTER + UCSD (Tony Fountain). Invitation was sent to wider group - ESIP and ecoinformatics mailing list (ESIP has an "Enviro-sensing cluster"). The training will also feature matlab-toolbox and CUAHSI/ODM, but not kelper this time. One goal of the workshop is a sensor management resource guide (BP). The guide is mainly for LTER mainly, but outside input will be valuable, particularly CUAHSI and NEON. One uncertainty is how real-time sensor data can be consumed by the NIS. Expected issues include whether a certain level of QC should be stipulated? Can sensor life cycle events be documented with data?**

At the IMC meeting, this group anticipates requesting feedback on draft 1 of BP doc, or a "possible discussion" difference between 'real time' data and a QC'd product.

Don also provided an update on STREAMCHEMDB. A programmer at AND (Matt) restructured the database has now a workflow that consumes data from streamchemDB in MySQL. They would like to examine how this might work with PASTA. streamchemDB might be a good candidate for a use case project for spatial

data in PASTA. It is a watershed project (e.g., nutrient flux from streams) that needs spatial descriptions of a place. Matt will do the March water cooler.

**EML Best Practices (Wade Sheldon, via VTC)** see below for his docs

next version to include additions since 2011, specifically for EML use with regard to catalog, pasta, keywording, specific text features (title, abstract), workflows, taxonomic terms.

Group plans to concentrate on real-world workflow use, and tie to current activities like climdb, vegdb, and coastal water quality. Workflows working group will meet in Madison in mid-june or early-july.

At the IMC meeting, activities could include:

pull together reports and findings from other working groups to outline the next BP doc. maybe a breakout with reps from the other WG, but probably real organization will not be till after the IMC meeting.

what specific recommendations? what changes to EML?

Wade's report includes some possible themes for discussion.

- a) EML congruency and workflow readiness
- b) EML content for dataone (a EML profile?)
- c) efficient EML generation from pasta for synthetic datasets, (also, synthesis product's publication and citation model).

**EML Metrics and Data Reporting (Philip Tarrant)**

The WG has gotten feedback from sites on mockups they shared in December, and will work that into reports. For the most part, the IMC would like reports to focus on data packages in PASTA, not those in Metacat. At the IMC meeting, they will share new mockups, discuss technicalities, like using reports to rate packages, 5-star, etc. Most likely, this will require discussions in plenary.

The reporting system is likely to need a database of its own so that previous records can be recalled.

**Controlled Vocabulary (Margaret O'Brien)** WG has scheduled a workshop in May at VCR. Before then, their goals are:

1. all current terms will have definitions and relationships (mainly synonyms) so that work can be vetted during workshop

2. additional terms have been suggested, using 2 mechanisms, either by sites (with definitions, relations and docids or with some new queries to Metacat. Both new term-sets will be vetted by group during workshop.

### 3. outline external resources for geographic and taxonomic terms

Activities for the IMC meeting:

1. The WG would like some input for how the vocabulary should be used for returning data (implications for catalog interface). E. g.
  - a) what parts of EML should be searched?
  - b) do we want to make use of keyword-only searches? More precise matches are possible, but this choice puts high demands on all packages, and will depend on implementation, and would need a timeline.
  - c) what other uses of EML-keywords could we take advantage of? (eg, using a keyword to name a synthesis project)
  - d) what kind of searches should take advantage of RTs to increase hits? (can test via tematres interface)
2. the WG could provide an introduction/tutorial on using web services to retrieve terms from tematres (e.g., for insertion as EML keywords) –probably to a subset of IMs.
3. the working group could outline of best practices (2 areas: how to suggest a new term, what to include in your EML-keywords). Also probably would appeal to only a subset of IMs.