EML Best Practices Working Group Report

Date: 20-Feb-2013
From: Wade Sheldon (GCE), working group chair
To: LTER IM-Exec

Working Group Goals –

Review and update the EML Best Practices version 2 document (http://im.lternet.edu/node/910) to reflect LTER developments since 2011, including:

- The EML congruency checker
- LTER NIS changes (e.g. revised data catalog, PASTA)
- EB metadata recommendations
- New use cases for EML (e.g. programmatic workflows in R, MATLAB, Kepler, SAS, etc)
- New thinking about specific content recommendations (e.g. taxonomicCoverage, keywords)
- Produce "exemplar" documents for LTER sites to reference

Working Group Status –

In September 2012, members of the EML BP 2 working group were polled about their interest in continuing their involvement, and then a call for new members was circulated to the IMC email list. All but two original members of the EML BP 2 working group were interested in remaining active, and three new members were recruited. The current composition of the working group is:

Dan Bahauddin (CDR)
Emery Boose (HFR)
James Brunt (LNO)
Duane Costa (LNO)
M. Gastil-Buhl, (MCR)`
Corinna Gries (NTL)
Margaret O'Brien (SBC)
Ken Ramsey (JRN)
Mark Servilla (LNO)
Wade Sheldon (GCE) - chair
Aaron Stephenson (NTL)
Theresa Valentine (AND)

The LTER email list for the working group (emlbestpractices@lternet.edu) was updated to reflect the membership changes.

Working Group Activity –

The scope and approach to creating an EML BP version 3 document was discussed via email in October 2012, then the collective ideas were presented to the IMC for comment during a Virtual Update on October 8-9, 2012 to determine if a NIS production workshop proposal was warranted. Based on the working group discussion and IMC feedback, we concluded that work on the EML BP 3 document should be deferred until later in 2013 to allow several critical activities that will inform the document to make additional progress, including:

- NIS Data Portal production release
- EML Metrics and Data Reporting working group
- Controlled Vocabulary

We decided to focus instead on writing a NIS Production Workshop proposal focused on leveraging EML-described data sets in PASTA for data synthesis workflows (i.e. to follow up on a successful February 2012 PASTA workflows workshop). We felt that additional testing of PASTA and EML for performing real-world data synthesis for LTER cross-site projects (targeting VegDB, ClimDB and coastal water quality synthesis activities) would provide critical insights into how EML implementations and data formats used by LTER sites (and their congruency) affect usability in different workflow environments (Kepler, R, MATLAB, SAS). EML and data format recommendations would be a key product of this workshop.

Data Synthesis Workshop Proposal Status –

The workshop proposal "Data Synthesis Workflow Development and Testing" was submitted by Wade Sheldon, John Porter and Corinna Gries in November 2012, and was selected for full funding by the EB in December 2012. The abstract is listed below as a summary:

> The NIS PASTA framework is nearing its first production release in December 2012, and LTER sites have made great progress over the past several years in getting EML-described core data online. LTER site Information Managers and others in the eco-informatics community have also made progress in developing or adapting software to perform analytical workflows using EML-described data. In many ways this is a watershed moment in LTER, when we are poised to capitalize on these long-term investments and realize the vision of computer-mediated data synthesis based on structured metadata.

> We therefore propose to convene a workshop to develop and test workflows that leverage the PASTA framework, EML-described data and EML-based software to generate derived data products for one or more network synthesis projects (e.g. ClimDB, Veg-DB and Cross-Site Coastal Water Quality). This workshop will not only

provide documented workflows that can be used by these synthesis efforts, but will also provide crucial feedback to workflow software developers, EML providers and NIS developers on any problems encountered to improve these resources. This workshop will also provide practical, real world experience that will inform views on best practices for EML metadata content, EML-data congruency and site data package management, helping to guide site practices as we move from the era of human-readable to machine-readable metadata and data in LTER.

The proposal includes funding to convene a group of 10 IMs, workflow developers and scientists for a week-long workshop at LNO. However, due to scheduling conflicts with many other workshops and training events in spring 2013, the dates for this workshop are still not finalized. We are currently exploring options for holding this workshop in mid-June to early-July in Madison, WI to avoid conflicts with the Data Management Training course being taught at LNO in June.

Working Group Plans for the 2013 IMC Meeting –

Plans for meeting at the 2013 IMC meeting have not yet been discussed with the other group members; however, based on the timing of the meeting this would be an ideal opportunity to review the lessons learned from the initial Data Portal production release and progress reports from the EML Metrics and Data Reporting and Controlled Vocabulary activities and start organizing the EML BP version 3 document.

Possible IMC meeting themes or break-outs that would contribute to or follow-up on our activities include:
- EML-data congruency and workflow readiness
- Optimizing EML for the LTER Data Portal and DataONE
- Generating EML for synthetic data sets