Information Manager Buy-Out Proposal
EML Diagnosis and Best Practice Implementation Mentor
November 2011
M. Gastil-Buhl (MCR) and Sven Bohm (KBS)

Summary:
Of the EML datasets being contributed to the LTER NIS, a significant fraction is not of sufficient quality for automated use, and a variety of approaches will be necessary to remedy this. The EML Congruence Checker (ECC) is being developed to report on dataset usability, but experience indicates that simply informing sites of dataset issues is insufficient and in many cases, intermittent, individualized consultation between site information managers and a mentor will be necessary. We propose that one or two site information managers apply for funds to offset time during 2012 so that they are able to act as mentors to other sites, advising them on issues related to "PASTA-readiness."

Products:

1. An increase in the number of valid datasets contributed to the NIS from LTER sites
2. Site information managers who are better trained in Ecological Metadata Language content, use and expectations
3. Training and example material archived for future use
4. Early feedback to development of EML-generating tools at sites.

Participants:
The work will require one or two persons (mentors) with the following qualifications:

- The ability to run the ECC web services as these mature during 2012
- Close familiarity with the EML 2.1 schema
- Demonstrated proficiency with the generation and evaluation of EML 2.1 documents

The mentor(s) will respond to requests from site IMs for advice and evaluation of EML metadata relative to EML Best Practices and the developing ECC and metrics. To date, two site information managers have expressed an interest in taking on this role, Sven Bohm (KBS) and M. Gastil-Buhl (MCR).

Background:
The preliminary results from the early version of the EML Congruency Checker (ECC) indicated that some sites' EML would need upgrades to be compliant with anticipated NIS requirements for data packages. Sites have varying degrees of familiarity with EML form and content. The ability to evaluate one's own EML is essential; the SIM is often the gate keeper for submission of datasets to the Network, and as such, the person who, if not engaged in programming, must hire and then evaluate the work of others. Even with sophisticated EML generation tools, SIMs must be able to identify the requirements for improvements. The 2011 IMC annual meeting provided all IMC members with a chance to comment on both the Congruency Checker and EML Best Practices, and many expressed a desire to become better equipped to evaluate their EML.

The ECC code is too immature to serve this educational need alone; interpreting results is not yet straightforward. Shortcomings in the code are being addressed; however sites are anticipating future higher expectations for data packages and are already engaged in rewriting code for EML generation. They require usable feedback on their EML products in the very near term. This

mentor position will serve that need. The concept of an EML mentor supported by NIS funds has received verbal support from several SIMs during VTC conversations among the IMC.

Tasks:

1. Train site IMs how to evaluate EML generated at their sites so that they can identify requirements for improvement of their own site's IMS
2. Compare site-EML to Best Practice using examples from individual sites with recommendations for convergence
3. Train site IMs on Network tools used to evaluate EML and dataset congruence

Implementation:

The IMC and the mentor(s) will decide the most effect implementation. However, these estimates offer an example. 0.083 FTE (i.e., 1 month, or the maximum buyout allotted to one person) represents approximately 160 hours over the course of a year. We anticipate that the mentor(s) will spend approximately half of that time in preparation (either of site material or reports) and in refinement of examples and training material, with the remaining time for available for consultation. Those 80 hours could serve 10 sites with an average of four 2-hour consultations each. One suggested format would be established "office hours" where the mentor was available for drop-in consultation. Possibly, a combination of office hours and scheduled appointments may be the best combination. Scheduling should take into account the PASTA milestones.

Expected outcome and resulting activities:

ECC reports are likely to indicate a need for training modules to advance site EML proficiency; however, the scope is yet to be determined. The time described here for intermittent, individualized consultation with site information managers will provide invaluable information to that process. Additionally, the materials assembled for individual consultations will serve as examples for the larger SIM community. Often, all that is needed to identify an issue is an example, and SIMs, while willing to share their products, often have no time to do so. Support of the mentor position will ensure that this material is made available to a large audience.

Exclusions:

Certain activities are specifically not covered by this proposal. As a member of the LTER IMC, a mentor already is expected to participate on working groups and teams that coordinate the development of LTER NIS with IMC members. The mentor also will not create new code for sites. Apparent fixes to simple tools may be suggested but this is not the primary focus.

References:

[1] EML Congruence Checker reports from V0.1 Data Manager code are available from LNO.

[2] EML Best Practices V2: http://intranet.lternet.edu/documents/eml-best-practices-2011

[3] Notes from IMC Breakout (26 Sep 2011) are available as an upload to the IMC website: http://im.lternet.edu/meetings/2011/breakout1

[4] LNO Operational Plan: http://intranet2.lternet.edu/content/lno-operational-plan-draft-2

Budget:

$11,958:   1/12 FTE for one person, billable in 2012, which will cover ~4 hours/week for 10 months (May 2012 through February 2013). This example budget is the fully burdened rate for M. Gastil-Buhl using the University of California's Productive Time Calculation for recharge rates.

Justification: Sven Bohm wrote code to generate EML for KBS LTER, which is congruent and has demonstrated skill with XML and related tools. During his involvement with the Data Package Manager tiger team he gained familiarity with the developing NIS and its requirements for well-constructed EML. M. Gastil-Buhl has constructed EML manually for 2.5 years using oXygen and has used XSLT to query and display EML. She contributed to EML Best Practices and became familiar with the Data Manager Library for EML during her ex-officio involvement with the Data Manager tiger team testing. By running the Data Manager component of PASTA on her mac in debug mode she has become intimately familiar with the ECC. She has compiled examples of EML from most sites as they relate to congruence or best practice.