**IM VTC--Guidelines for LTER Information Management Systems**

13-Feb-2017

E. Boose

This is an update of a document created in 2005 and updated in 2009. IMC and NISAC were commissioned to create a document which was used in NSF site reviews. It became out of date and no longer used during Saran’s tenure at NSF.

Renamed “guidelines”. No effort to prescribe technology. Focus on functionality.

Subcommittee began with original document. Final version to be shared with LTER EB. Serve as tool for self-assessment and planning. Perhaps also for site reviews.

Items highlighted where committee not sure or looking for input. “Should” indicates an expectation for all LTER IM systems. “Is encouraged” indicated desirable but not essential items (Kristin).

**Section A1e**. Evidence of data use.

Our ability to track external use is limited through PASTA (just downloads). There are some operational constrictions now that should be acknowledged. Tracking internal use is easier (Wade).

Our data are now available through PASTA and DataONE, which provide different functionality. NSF is very interested in getting these guidelines and will use them in reviews. We can’t be responsible for decisions made by DataONE (Corinna).

Peter McCartney would like to see a clear metric of how LTER data are managed and used. But we were not asked to create a review document (Kristin).

It would be good if LTER had more input into how DataONE implements search capability, etc (Mark).

It’s the scientific community that needs to be engaged in finding data. Number of downloads is not the best measure of a dataset’s value. Publications are a better measure. A dataset may sit unused for a long time and then become important (Jim).

Data citations would be better than number of downloads. Data downloads from DataONE are actually made through PASTA, so these downloads are counted. See Perl programs from John that will compile download numbers. It would be best not to prescribe a particular solution here (John).

Linking datasets to publications would be valuable (Ken).

Can bots and web crawlers be separated from actual downloads? (Don).

Still a problem. PASTA filters out known bots but can’t get them all (Mark).

Don’t include tasks that aren’t worth doing (e.g. counting downloads) (Jim).

It might be useful to have numbers crunched at the network level (e.g. by EDI) (Ken).

This document may persist for many years. Move from downloads to citations (Gastil).

Need to define terms “should” and “is encouraged” in introduction. Putting “should” before “is encouraged” is good (Margaret).

Who is the audience here? More details would be helpful for new IMs (Mark).

**Section A3b**. Guidelines for LTER web sites.

A new document will be needed. Maybe list as “in preparation”? (Gastil).

Difficult to keep up to date because of rapid changes to websites (Wade).

Maybe use a more permanent URL? (Margaret).

Recall that in the past the EB has made changes to the website document (Ken).

The idea for website guidelines was originally proposed by NSF because sites varied so much (Jim).

Peter McCartney would like to see users go to DataONE for data rather than to site web pages (Kristin).

Site scientists have more particular needs than other users. Site web page and DataONE serve different audiences. How to resolve this tension? (Mark).

For how many sites will this document require major changes? None. But the original document did not require major changes either (Corinna).

The document captures good practice for LTER. It affords sites some protection if IMs are asked why they are or are not doing particular things (John).

Peter indicated he is interested in seeing the document and may use it to compile review criteria (Kristin).

NSF management team lost track of this and related documents (Wade).

In the past, site IMs have used this document to prepare for reviews and to keep activities on track (Gastil).

NSF typically gives broad review criteria (e.g. intellectual merit, broader impacts). In the past, some non-LTER reviewers had unanticipated expectations. This is our chance to speak (John).

Because of the ~10 year revision cycle, the document should be big picture and long term (Jason).

Check language in current RFP regarding availability of data in PASTA (Wade).

**Section A4b**. Current list of research projects.

Should we drop this item? (Kristin)

How to interpret “projects”? (Jim)

Including this item sets an expectation. Better to omit. Sites can do if they want (Wade).

Some sites have been encouraged to track data submissions. This is somewhat different and is now part of the RFP. Maybe add this as an item (Wade).

**Section B5a**. Site contributions to network databases.

Are these network resources (e.g. SiteDB) still active? If not, they should not be listed here (Ken).

Many of these issues are still in flux (Mark).

**Wrap up**. Please send along any editing suggestions, esp. for introduction (Kristin).