

IMC VWC VTC - Guidelines for LTER Information Management Systems
February 13, 2017

Mary Martin, Brian Herndon, Kristin Vanderbilt, Wade Sheldon, Duane Costa, Xia Yang, Corinna Gries, James Conners, John Porter, Emery Boose, Don Henshaw, Jonathan Walsh, Jason Downing, Jim Laundry, Steven Earl, Mark Schildhauer, Margaret O'Brien, Julien Brun, Ken Ramsey, Dan Bahauddin, Eda, Gastil, Hap Garritt, Suzanne Remillard (notes).

Kristin:

History of original and new documenting development. Yellow highlights identify areas that refer to documents that may no longer be relevant. Go through document by section. A.1.e. has been added. Should we be tracking data use?

Wade: ability to track data use is limited through PASTA. We've given up control over distribution.

Corinna: Document needs to include that our data are available in other places that provide different functionality. Makes searching and tracking ability different. We, LTER IM, can't be held responsible for things that DataOne don't provide.

Kristin: Peter McCartney would like to see a clear metric. The guidelines are different than a review criteria. We were asked to write guidelines and not review criteria. Peter and others can create review criteria from these guidelines.

Mark S.: DataOne would welcome additional input from LTER IMC community.

Jim L.: LTER IMs aren't the ones that need to find data; it's the scientific community that needs to guide how to find data.

John P.: Don't want to be locked into counting data downloads, but rather data citations (in the future). Downloads from DataOne come through PASTA, so the counting of downloads is there.

Ken Ramsey: thought NSF didn't want this bean counting?

Kristin: yes, this is from the committee

Ken: Yes, value in tracking publications. Can use PASTA reporting system. Wade uses this for reports. John P has R scripts.

John P.: NSF didn't want the authentication to get in the way of people downloading data.

Don: Tracking PASTA downloads; reports a lot of downloads, but no way to separate bots and actual downloads. Can these be separated?

Mark S.: Yes, that is still the case. Am trying to filter some, but can't filter out all. All are suspect, at some level. It is difficult to do this.

Jim: The sentence should be more general.

Ken: Can we automatically see who is citing data and publications

Gastil: this document needs to be general in order to be relevant in the future. State things so that when the infrastructure is available, we can do

Margaret: Fine with statement. Vocabulary (should and encouraged) needs to be defined. Likes that should are above encouraged.

Mark S.: Seems reasonable. Who is reading the document and why? Is it for current sites or new sites? Descriptive or prescriptive? How to meet these shoulds and coulds.

Kristin: describes functionality and not descriptive.

Any other comments regarding Scope?

Site Information Portal.

A.3.b. Out of date. Should we modify recommendations or eliminate this statement?

Gastil: It would be good that we update it. Could say 'in prep'.

Wade: it's challenging because they move faster than our general

Margaret: Could reference node or URL (??) Maintenance hassle.

Ken: EB changed this doc without our knowledge, which created problems at review.

Jim: Web recommendations were brought up by NSF because of frustration of different sites.

Kristin: Peter McCartney wants people to go to dataOne to get data.

Ken: scientists at our sites go to the server and not site web or nis.

Mark S.: scientists looking for data to do research versus synthesis. Ken brings up a good tension.

Corinna: Question for everyone. Does this document raise major changes for what we're doing? If not, then what are we doing it for?

Kristin: a reference

John: these aren't review criteria, but they are an example of what sites should be doing. It's a framework, good practices. It's what the network says we should be doing. It's OK if it's dull.

Corinna: Peter told her that he wants review criteria.

Kristin: got different impression from Peter.

Gastil: Should we have some shall nots?

Corinna: put us on the ??, not just the local system.

Suzanne

Gastil: a checklist of various tasks

Wade: original document was called 'review criteria'

Kristin: Henry G. came to IMExec and asked for this.

John: Broader review criteria versus specifics. This is our chance to speak.

Jason: It's hard to create a document to respond to NSF's whims. We should create a document for our community.

Ken: Both docs are useful for new PIs at the site. Very useful to prepare of site reviews and renewals.

Don: Peter may find issue with this document because there seems to be a lot of wiggle room. It doesn't specifically say that data should be in the NIS.

Wade: Agrees with Don. New language in RFPs.

Ken: Are we required to all data in PASTA.

John: we were lose in this doc because there were other appropriate places. Data should be shared in the appropriate portal.

Corinna: NSF is now saying that data should be in one of the big archives.

Kristin: A.4.b. Should we keep this in? Can take a lot of time. It is only 'encouraged' and not 'should'. It is talking about research projects.

Jim: Doesn't know how to count a project. Cut.

Others? Yes, cut. Too ambiguous.

Gastil: A new category? If time/funding allow... Even less stringent to 'encourage'.

Wade: Part of renewals is a means of tracking compliance with site data availability. Data submission. Maybe reframe; sites should track data submission.

Ken: B.5.a is a huge problem since these resources haven't been identified. It's not possible to contribute.

Gastil: Yes, these haven't been decided on. There should be some clarification about 'which resources currently exist'.

Kristin: if others want to wordsmith, please do.

Other items:

Gastil: IMExec shared a couple of docs with IMC. Last November, leadership was asked to respond to how EDI, LNCO, IMC, DataOne

Corinna was sent an email to address this issue. She started with Corinna's framework and each IMExec member won't on a section.

There will be a confab in April and want all questions to be addressed.

Mark S.: Overlapping scopes. Blah, blah, blah

Wade: IMExec represents a community that represents the broad community. This document wasn't for LNCO, but for the IMC group.

From NSF's 2016 RFP for new LTER sites:

- **Data Management Plan (limited to 2 pages).** Core data sets generated at a site must be available electronically and accompanied by metadata that meet LTER standards for the Network Information System (NIS) (<https://portal.lternet.edu/nis/home.jsp>). This section must provide a description of the data and information management system and metadata standards to be used at the site. It is expected that data derived from LTER funding will be made freely and publicly available as soon as possible, and not to exceed 2 years after collection, via the LTER NIS Data Portal. This section should include expected milestones and deliverable products from data management. NSF places high priority on the availability of site-based data to a broad research community. This section should include descriptions of how data management will be implemented in the design of research projects; how the data manager will be involved in designing research projects; and the mechanisms employed to ensure that researchers contribute their data to the NIS Data Portal. Proposers should describe the resources that will be dedicated to harvesting, documenting, archiving, managing, and making data accessible. For proposals that address ocean or coastal ocean ecosystems, data management should be consistent with the requirements described in the Division of Ocean Sciences Sample and Data Policy.