

EML - Bug #4393

Use datamanager for EML QA/QC

09/17/2009 01:24 PM - ben leinfelder

Status:	New	Start date:	09/17/2009
Priority:	Normal	Due date:	
Assignee:	ben leinfelder	% Done:	0%
Category:	datamanager	Estimated time:	0.00 hour
Target version:	DataManager 1.0.0	Spent time:	0.00 hour
Bugzilla-Id:	4393		

Description

As discussed at the LTER meeting this year:

Work Group: Metrics and reports for EML data package quality
The EML data manager library (contributors: Costa, Tao, Leinfelder, Servilla) was created to parse EML metadata documents and insert the described data entity into a relational database. Our experience using the library with data packages contributed to the LTER NIS indicates that a large fraction do not have metadata of sufficient quality for the data to be used in this way. The primary contribution from LTER sites to the NIS is data sets, which are intended to be used in cross-site synthesis projects. Clearly, for cross-site synthesis to make use of the NIS a certain minimum level of metadata and data quality is required.
The goals for this group:
1. establish a set of metrics for LTER EML data package quality,
2. recommend content for a report to be produced by the EML data manager library, and
3. consider implementation strategies, e.g. should the report be another choice on the EML parser page? a shell script similar to that included with the EML parser?

The quality reports can be used to
1. inform the dataset contributor about the content of the data package, and indicate whether data are of sufficient quality to be machine-readable. Our data catalog (metacat) has no quality standards beyond basic XML and EML compliance, so a data package that fails these quality metrics can still be uploaded or harvested, although its usefulness is limited.
2. in the LTER context, reports can produce a list of failure modes for LTER metadata and data entities. Such a list could provide input for the design of specific tools for data providers, or help identify gaps in a site's IM system. A site requesting supplemental funding for its IMS could use the reports as part of the proposal justification.

As a starting point for our discussion, I have started a flowchart based on my own experience with the data manager library and SBC's EML data packages.

Here is the current membership (on this cc list, and present in Estes Park):
Margaret O'Brien, SBC
Emery Boose, HFR
Dan Bahauddin, CDR
James Brunt, LNO
Mark Servilla, LNO
Duane Costa, LNO
Mark Shildhauer, NCEAS
Ben Leinfelder, NCEAS

History

#1 - 03/27/2013 02:26 PM - Redmine Admin

Original Bugzilla ID was 4393