**Since then - Margaret**

Since 2008, many of the ideas of the Cyberinfrastructure documents of the previous decade have been implemented, with a focus on coordinated Information Management Systems (IMS) and practices at sites, plus refined data publication practices that are codified in the PASTA data repository.

[[ TO DO: Sentences about cross-site sharing of IMS components (DEIMS and Metabase, local catalog code, GCE Toolbox, \_\_), ]]

Data publication in the LTER has had two major transitions since 2008. First, the exclusive LTER-focused system migrated from a metadata-only catalog to a data archive with the advent of the PASTA system (2009 - 2013). Secondly, in 2016, that exclusive system transitioned to one in which the LTER Network is one of many members, during the closing of the LTER Network Office project and the beginning of the Environmental Data Initiative project. Each transition was a remarkable feat of sustainability, and reflects the IMC’s resilience and ability to adapt to new environments. Both transitions were expertly facilitated by the UNM software team’s active engagement with their user community, cognizance of their needs, and responsiveness.

[[ TO DO Sentences about Pasta - online in 2013, with ECC

How the IMC was involved - Tiger teams, IMC influence on pasta-dev ]]

Figure %bubble\_plot% shows a meta-analysis of the LTER contribution history that covers both these transitions. It compares two major aspects: the number of contributions per year (both new and updates), and the completeness of metadata relative to 25 general metadata concepts that can be traced to “FAIR” principles (“Findable, Accessible, Interoperable, Reusable”, Wilkerson 2017). Some contribution gaps early and late in the time series reflect LTER sites entering or leaving the network, however, the patchiness in frequency in early years also reflects the lack of fully operational systems at LTER sites. With regard to metadata completeness (circle fill), within each site, generally, there is an improvement, to approx 75-80% completeness in most sites in 2018. The difference in package granularity is also evident (i.e., whether a site chooses many smaller packages or fewer larger ones), and the issues arising from these sorts of packaging choices is likely to be the basis of some LTER guidelines in coming years.

Further, analysis of pasta quality reports from the first few years (2013-2015) showed that a system of reporting increased quality at a faster rate than did simple narrative recommendations, and that LTER sites were more than able to respond to a reporting scheme (Fig 5 in O’Brien et al). Generally then, when compared to a decade ago, more data are online today, and can be found centrally. Data are of higher quality, and demonstrably more FAIR. Today, LTER is a benchmark. Gordon and Habermann (paper in prep) compared these same metadata concepts across the entire corpus of DataONE holdings (from over 40 distinct repositories); LTER regularly scored in the highest percentile. The main drivers of improvement are the PASTA system, plus supplemental funds from NSF to sites that were used to upgrade local IMS (years 2009 - 2011).

The LTER IMC has held regularly annual meetings since the late 1990s, which were typically inclusive, welcoming guests and collaborators from other networks. As the Network grew, the size and number of guests grew as well, and reached a point where accommodating them became difficult. In 2008 and 2011, the IMC co-sponsored (with NCEAS) the “Environmental Information Management Conference”, as a forum for information managers, scientists, and informatics researchers to present and discuss advances in environmental information management and analysis. The meetings were a success (convening over 100 participants each time), but the requests for more frequent, larger meetings made it clear that another more sustainable option was needed. In 2014 LTER joined the Earth Science Information Partners (ESIP), which acts as the data management hub for many federally funded data-centric agencies such as NASA, NOAA, USGS and many independent research groups. In addition to freeing the IMC from meeting planning, joining ESIP afforded the LTER IM community broad exposure, and a host of new colleagues.

Figure %bubble\_plot%: Data packages from 26 LTER sites from 2005 to 2018. Sites (not identified) are in columns. Circle size and color reflect the number of packages contributed (both new and updates) to the LTER catalog that year, and indexed by DataONE (binned, R: 1-25, O: 25-50, Y: 51-75, G: 76-100, B: 101-200, V: >200). Circle fill reflects that year’s average completeness relative to 25 general metadata concepts. Replotted from Gordon & Habermann, 2019.

Cruft below here

Meta-analysis graphic showing LTER contribution history (sites to central catalog) number/year, completeness of metadata during each year (FIGURE \_). Chose general metadata concepts which could be traced to “FAIR” principles (wilkerson).

Upshot: Compared to a decade ago, more data are online, and can be found centrally. data are of higher quality, and demonstrably more FAIR. Main drivers: pasta system, plus supplemental funds from NSF to sites that were used to upgrade local IMS (years 2009-2011). Further, analysis of pasta quality reports from the first few years (2013-2015) showed that a system of reporting increased quality at a faster rate than did simple narrative recommendations, and that LTER sites were more than able to respond to a reporting scheme (Fig 5 in O’Brien et al)

Other: meetings so popular that we experimented with running a conference! twice - EIMC. But discovered a better option - joined ESIP which freed us from meeting-organization, and gave the LTER IM community broad exposure, and a host of new colleagues. Set up working groups for sensor networks, …

Figure here

Figure \_: Data packages from 26 LTER sites from 2005 to 2018. Sites (not identified) are in columns. Circle size and color reflect the number of packages contributed (both new and updates) to the LTER catalog that year, and indexed by DataONE (binned, R: 1-25, O: 25-50, Y: 51-75, G: 76-100, B: 101-200, V: >200). Circle fill reflects that year’s average completeness relative to 25 general metadata concepts. Replotted from Gordon & Habermann, 2019.

Meta-analysis graphic showing LTER contribution history (sites to central catalog) number/year, completeness of metadata during each year (FIGURE \_). Chose general metadata concepts which could be traced to “FAIR” principles (wilkerson). Gordon/habermann compared these same metadata concepts across the entire corpus of DataONE holdings (from over 40 distinct repositories); LTER regularly scored in the highest percentile (pers comm, paper in prep).

Upshot: Compared to a decade ago, more data are online, and can be found centrally, data are of higher quality, and demonstrably more FAIR. Main drivers: pasta system, plus supplemental funds from NSF to sites that were used to upgrade local IMS (years 2009-2011). Further, analysis of pasta quality reports from the first few years (2013-2015) showed that a system of reporting increased quality at a faster rate than did simple narrative recommendations, and that LTER sites were more than able to respond to a reporting scheme (Fig 5 in O’Brien et al)

Other: meetings so popular that we experimented with running a conference! twice - EIMC. But discovered a better option - joined ESIP which freed us from meeting-organization, and gave the LTER IM community broad exposure, and a host of new colleagues. Set up working groups for sensor networks, …

---------- notes, frags below -------

2009 - DEB 0936498

ARRA funding

 Co-development of standards for ecological information management, development of new cyberinfrastructure tools, increased access to data for educators, and collaboration with other environmental observing networks will lead to a more robust and efficient national cyberinfrastructure for ecological research. Expanded collaborations and broad-scale transdisciplinary activities will be encouraged through working groups.

2010 - site supplements

to help position LTER for future funding similarly to other environmental observatories, and to provide resources to sites in parallel with the development of the Network Information System (NIS) at the LNO.

several common themes wereidentified that could become the basis for future collaboration, production workshops, or topics for our annual meetings. Some common themes:

* The use of a common data model to manage and store metadata to allow common tools to uniformly deliver metadata as EML or generate queries of the data (e.g., using the Drupal content management system)
* The need for tools to facilitate the capture of metadata at the front end of experiments such as web user interfaces
* The use of best practices as a guide for improving and completing metadata elements
* The need to develop or generate EML for geodatabases
* The installation of sensor networks and the need for tools to facilitate the quality assurance of data, and generally the need to improve the flow and processing of collected data into managed information systems
* The incorporation of standardized approaches such as the controlled vocabulary or unit dictionary into site information systems

These themes represent the spectrum of necessary site work required to begin paving the way for more seamless and interoperable access of site data throughout the LTER network. Our community must now address the best use of funding opportunities to meet this daunting

The operational plan! What an energy sink!

Relational data model comparison: The benefits of this comparison are 2-fold: it allows sites to understand the most successful features of existing systems as they consider adapting one of the models, and second, allows sites to compare their systems features and dependencies to other sites within the network.

2011

EML usage is not yet standardized across the network. Full standardization is the long-term goal, and is best addressed by upgrading site IM systems in general. However, some short-term remediation may be helpful as well. Non-standardization also inc

The first checking!! 5 essential features of data packages:

Title (quality check has a recommended length range, not in V1)

Abstract is meaningful (quality check has a recommended length range, not in V1)

Keywords from lter vocab (quality check exists, not in V1)

Entity description (attributes?, if yes, required by schema, and PASTA, V1)

Entity URL at the entity level (PASTA requirement, V1)

2013:

PASTA!

Some sites are still thinking through their approaches to PASTA submission and others are putting recent supplement funds to use adapting their EML-generating systems. In the process, many sites are re-evaluating their inventories, re-designing data packages, and generally improving the quality of the data going into PASTA. These efforts will all contribute to greater overall availability for LTER data. The accelerated timeline for PASTA’s development also brings some significant challenges to both policy and site-level implementation. The adoption of PASTA as the basis of the LTER data cataloging system means that we will switch from the current metadata-only catalog to one designed for both data and metadata. This step will improve availability and accessibility because PASTA provides DOI-based identification for data packages, the capacity for synthesis data provenance tracking, and more fully exploits the Ecological Metadata Language schema.

IMC workshop will design documented workflows and provide practical, real-world experience that will inform best practices for EML metadata and PASTA development (Sheldon, et al 2013).

rapid progress. In many ways this is a watershed moment in LTER, when we are poised to capitalize on our long-term investments. We focused here on the challenges, but the fact remains that we should all be very proud of PASTA’s existence as a production framework, and that IMC has pulled together to do what it takes to make it our “one-stop shop” for LTER data.

Web of science selected LTER as a key resource :

Data Citation Index (DCI), which will support discovery, access, citation and attribution of digital data. Initially their work has focused on data sets deposited in repositories and which are used in the research published in the traditional scientific literature. Their system is planned to enable the discovery of data studies and data sets, and to help track the full impact of research and institutional output. After reviewing LTER data catalogs, Thomson Reuters has selected our Network as a key resource for this new initiative. Inclusion in the new /Data Citation Index/^SM will provide LTER with increased visibility

LTER CI External Review Committee. 2010. Cyberinfrastructure Operational Plan External Review Report

<https://lternet.edu/wp-content/uploads/2010/12/LTER_CI_Plan_External_Review_Final_2010-02-07.pdf>

<https://eim.ecoinformatics.org/eim2008>

<https://eim.ecoinformatics.org/eim2011>