

Position paper on the continuity of the LTER Network Information System (NIS) by the LTER Network Information System Advisory Committee (NISAC)

A network information system enables LTER to operate effectively as a science network by catalyzing collaboration on data, models and theory. These three components are inextricably linked, such that each is required to support the other two. Here we focus on the data component and provide a specific recommendation for the continuation of an LTER NIS including the Provenance Aware Synthesis Tracking Architecture (PASTA), a technology around which a paradigm of data discovery and access has been built.

The first published reference to the LTER NIS, in 2000, describes a grassroots development effort and information flow from individual sites into cross-site administrative databases such as the all-site data catalog and bibliography and research databases such as ClimDB. Formal development and direction of the LTER NIS began in 2003 with the formation of NISAC and subsequent release of the LTER NIS Strategic Plan in 2005. The defined goals were to increase 1) data quality through standard approaches, 2) data available for synthetic activities, and 3) knowledge discovery through synthesis. Network approval of NISAC and the concept of a network information system lead to increased support from the Network office and the solidification of these early database efforts into a more formal NIS.

While there are many models for information management in distributed networks, there are functional priorities within the LTER community that inform model design. We see the following concepts as those priorities:

- Access to LTER data in a manner that is convenient and consistent across the Network for the broader scientific community and the general public
- Common 'data terminology' and defined sets of processes (protocols and workflows) to improve efficiency and reliability of data access in support of science
- Leveraging the diverse expertise of the LTER community so that technology (the melding of terminology and process with compute capacity) can harmonize effort across the Network and allow for site-based initiatives to bubble up and become part of the NIS

The LTER NIS is comprised of many varied components (i.e., data catalog, user portal, site data and metadata, network databases like PersonnelDB, BiblioDB, and SiteDB, synthesis data products generated by workflows, web services). Some of these components are obvious to the user as they facilitate the access and discovery of data. Other components are behind the scene, but core to the operation and functionality of the system and the ability to synthesize data.

The LTER Network Office (LNO) described the framework of the NIS using PASTA in their Operational Plan (2009-2015), which detailed how this cyberinfrastructure would support research activities. The NIS was further clarified in the Strategic and Implementation Plan (SIP), 2010. The overall success of this framework was critically dependent on site participation, specifically site Information Managers (IMs), informing PASTA framework component development and creating well documented and quality

data products for harvesting into the NIS. Since that time, individual sites and site personnel have invested a significant amount of effort and resources to contribute to the activities outlined in the Operational Plan and SIP. The outcome has been improved conformance at the site level to a common data standard, improved understanding of the common mission, supporting processes of the NIS, and an acceleration of the flow of data from sites to the common repository of PASTA. Thus, consensus on the goal of PASTA has led to improvements in human and technological infrastructure for LTER in addition to the goal of centralization of access to LTER data.

An effective NIS is more than a technology. It is an amalgamation of human expertise, technology, and process that evolves in response to the growing needs of the Network. LTER data synthesis is promoted by an effective NIS, and PASTA is both the foundation and catalyst for Network-level data harmonization. The LTER NIS facilitates data synthesis efforts by ensuring quality data and metadata. Currently, dataset packages downloaded from PASTA come with a guarantee that a defined set of quality control checks have been passed. No other data repositories provide this type of guarantee. These properly described data function in scripts or programs used to transform, merge and synthesize data. Although support of these workflows is inherent in PASTA design and output products from synthesis can be stored in the NIS, this has yet to be fully realized. Our vision of a centrally supported NIS would offer expertise to aid scientists in crafting workflows. We see that the foundation has been set and now is the time to build upon this foundation by exploring the possibilities of cross-site synthesis that PASTA was designed to provide.

The LTER NIS requires administration and maintenance by an entity that understands the broader vision of the LTER Network and that is in a position to facilitate coordination among sites. Experience has shown that without dedicated administration, progress diminishes and sites cannot plan for implementation of the product. Site IMs are not in a position to perform this type of administration. It is important that IMC grassroots efforts towards an end product are realized on a network level and not left uncompleted. A dedicated administration also helps to build consensus around interface standards and encourages participation, cooperation and engagement.

The LTER NIS provides essential cyberinfrastructure services which include access to data and coordination of Network activities and databases. With the implementation of PASTA, Network developers have created a robust system to store site data and metadata, provide quality checks, and discover and access site data. Sites, too, have invested in improving their data collections to meet the higher standards of PASTA. With the development of PASTA, as a base platform, the potential horizon is broadened to really accommodate building value-added and cross-site synthetic databases. We expect to reap the benefits for years to come. To ensure this, any future vision of the LTER Network central office needs to include a transition, administration, support and maintenance of the NIS and continued development in conjunction with PASTA. It is critical that cyberinfrastructure needs of the LTER Network be continued and supported so that scientists can progress with transformative Network-wide research at broad scales.