30 May 2003

REPORT FROM THE NETWORK INFORMATION SYSTEM ADVISORY GROUP: TOWARD A STRATEGY FOR THE DEVELOPMENT OF NETWORK INFORMATION INFRASTRUCTURE

Introduction

The 20-year review challenged the LTER network to enhance its inter-site research activities by adopting a strategy for network-based research. The network's response shows a strong emphasis on strengthening existing cross-site integrative research and encourages new research initiatives conducted from the outset as collaborative efforts among multiple sites. Along with the continuing mandate for quality site-based research, we are now faced with the need to develop an information infrastructure that serves a single investigator as well as an entire network while maintaining involvement with global Information Technology developments.

The primary goal for the Network Information System (NIS) is to support synthetic research efforts at the LTER Network level. NIS development would facilitate the synthesis process, improve the quality of resultant synthetic research products and databases, and increase the amount of synthesis work conducted.

The LTER Network has identified six areas of emphasis for the next ten years. The NIS directly supports the network's strategic goals in the areas of synthesis, education, and legacies, and collaterally benefits the areas of education and outreach. These areas of emphasis are described in the Mission of the LTER Network as follows:

The central, organizing intellectual aim of the LTER program is to understand long-term patterns and processes of ecological systems at multiple spatial scales. The Mission of the LTER Network is to achieve this aim in six, interrelated ways:

- ***Understanding***: Gaining ecological understanding of a diverse array of ecosystems at multiple spatial and temporal scales
- ***Synthesis***: Using the network of sites to create general ecological knowledge through the synthesis of information gained from long-term research and development of theory
- ***Information***: Creating well designed, documented databases that are accessible to the broader scientific community
- ***Legacies***: Leaving a legacy of well designed and documented long-term observations, experiments, and archives of samples and specimens
- ***Education***: Using the uniqueness of the LTER programs and network to promote training, teaching, and learning about long-term ecological research and the earth's ecosystems
- ***Outreach***: Providing knowledge to the broader ecological community, general public, resource managers, and policy makers to address complex environmental challenges

Development of Goals and Policies to Enable Synthesis Through Networked Information Infrastructure

Development of a Network Information System must balance efforts in two major areas. First, the NIS should foster development of network-level datasets describing basic information on site characteristics, climate, and hydrology. These and other network-level data sets identified on an ad hoc basis through annual Coordinating Committee (CC) science themes or other mechanisms are more immediate products that will facilitate synthesis activities. Second, the NIS should involve longer-term development of generic

solutions or tools that will make discovery, access, aggregation, and visualization of data across multiple sites easier and more efficient. Reliance on either of these approaches alone will not be sufficient. Creation and maintenance of specific datasets for specific purposes using existing technology will lead to short-term, immediate successes but will be inefficient over the long-term. Similarly, development of predominantly generic solutions or tools will prolong the development of specific datasets needed for synthetic activities over the short-term. Finding the appropriate balance will be a continuing challenge for our scientist/information manager partnership.

Network-based research in LTER involves a diverse array of activities requiring multiple solutions for supporting its information needs. Although referred to as the “Network Information System”, NIS is really a collection of technical solutions rather than a single comprehensive database application. These solutions include the provision of persistent, online archives of electronic data, documentation (metadata) describing the organization and meaning of those data, aids to facilitate discovery of the data, and standardized protocols and formats for retrieving the data. Processing applications are being developed to transform, import, and direct data into larger synthetic databases and analytic and modeling activities. An information infrastructure is needed to support all levels of network research from comparisons of aggregate properties of entire sites such as net primary productivity, to syntheses of primary data such as in the ClimDB project, to the creation of entirely new shared, cross-site databases such as in the LIDET project.

The goal of a scalable information infrastructure cannot be accomplished without close coordination between IM development and the strategic plan for network-based research in LTER. There are several points where this can be defined: (1) annual science themes can identify and initiate research activities that include data products; (2) competitions for cross site research activities should engage information management personnel to ensure that the infrastructure is used to its best potential; and (3) ad hoc efforts by an individual or smaller groups of researchers should be able to make use of the infrastructure to support efforts to discover, evaluate and access data.

In examining the constituencies the NIS is designed to serve, the network of LTER sites comes first with the broader ecological community a close second. The main concern of LTER scientists is in the development of products that further LTER science goals. Education, policy makers, and general public inquiry would also benefit collaterally from these synthetic database products.

Impediments to Synthesis Science

There are several approaches to cross-site synthetic research, including modeling, cross-site comparison through specific hypothesis testing or empirical induction, and new cross-site experimentation. Each of these approaches requires: 1) significant time commitments of site scientists; 2) adequate and predictable funding opportunities; and 3) appropriate access to data from multiple sites. Although each of these three factors are currently limiting cross-site synthetic research, the strategic plan described here for the Network Information System focuses on the last of these impediments.

Our current information management systems at site and network levels are limiting synthetic research in several different ways:

- Data discovery: It is difficult to identify easily and efficiently what data are available across multiple sites. Moreover, common information about individual sites that one might expect to find at all sites is not easily found. Information content on site web pages is highly variable, metadata and data are presented in diverse formats, and metadata content is inconsistent across sites. The general requirement for sites to put data online does not resolve this issue.
- Data accessibility: Data are not easily accessible from all sites including critical basic data such as meteorological data, physical site descriptions, and study site locations necessary for interpreting

other research study data. Poor accessibility limits the development of science theme and other synthetic research databases. In some cases tools have been provided to automate this collection of data (e.g., web harvesters), but complete or even near-complete accessibility for all sites is rare. Proprietary restrictions placed on some data sets are another limiting factor.

- Data comparability: Differences in sampling protocols and methods frequently hinder synthetic efforts. Making data available in an appropriate format with comparable time and space resolution can be difficult given the range of ecosystems in the LTER network. Aggregating data in meaningful ways across disparate sites is challenging from both a scientific and an information management perspective.
- Community citizenship: The practice of good citizenship by the sites to provide data and information is necessary, but has not been defined adequately.

Identification of Roles in the Development of the Network Information System

Network information system development requires a partnership among LTER scientists, Information Managers, and the Network Office. LTER science drives the development of infrastructure to enable synthetic research. The science community takes responsibility for defining cross-site initiatives and identifying associated data needs. LTER scientists in conjunction with the LTER Coordinating Committee will recommend science theme proposals, identify database products, and identify a lead scientist for these tasks. While science theme proposals are the primary mechanism for presenting these initiatives, the LTER All-Scientist Meeting and other supplemental funding competitions also provide opportunities for synthesis proposals. These proposals should also designate and involve an Information Manager in the planning and implementation of the effort.

Additionally, the science community will identify necessary basic data required for all or many LTER sites. Current examples of existing basic data are climate (ClimDB), hydrology (HydroDB), site characteristics (SiteDB), and a Data Catalog. LTER scientists will recommend other potential basic data sets (e.g., land cover, sampling location maps, instrument locations, etc.), secure development funding, and cooperate with IMs to accomplish database construction.

The Information Management Committee plays a combined role of providing both technical expertise and providing a wealth of site scientific data through NIS participation. The IM Committee will define necessary functionality required for short-term and long-term synthetic efforts and develop recommendations for multi-tiered NIS standards. Information Managers will involve LTER scientists in the development and modification of standards via the CC. The IM Committee in conjunction with NET will articulate technical solutions to science questions proposed in science themes and other synthetic projects, assist in planning, and lead their development and implementation. Additionally, Information Managers will participate in NIS efforts through contribution of scientific data and by providing a testbed for new technological applications.

The LTER Network Office (NET) will develop and support infrastructure for synthetic databases (but not perform data collection), host production databases of basic observations (e.g., climate, hydrology), and maintain, curate, and expand the utility of the LTER Network databases (personnel, bibliography, site characteristics, data catalog, image archive, document archive, remote sensing archive) to provide framework data for other applications. In addition to supporting these databases NET will provide collaboration software services to the network (e.g., mail lists, mail aliases, portals, website and database hosting, persistent archive). The network developer dedicated to supporting synthesis information infrastructure will develop the database and web software framework necessary to include new synthetic databases easily. This development will be an iterative process beginning with the first science-theme database. This person will also provide necessary software and aid to sites for participation in a distributed data network, including development of protocol adaptors for site information systems and definition of the communication protocol to be used. Priorities for NET efforts in supporting synthesis information

infrastructure will be based on NIS Advisory Group recommendations. Finally, NET will continually apprise the NIS Advisory Group of developments on the global IT front for review and consideration.

Role of the NIS Advisory Group

The NIS Advisory Group will contribute to the Network Strategic Plan by creating a strategic plan for the NIS development. To accomplish this planning mission, the current NIS Advisory Group needs to continue as a group until the Network Strategic Plan is completed (the deadline for completion is Jan/Feb 2004). The NIS will require ongoing planning and assessment throughout its development, and the NIS Advisory Group will become a standing committee as approved by the Coordinating Committee in May 2003.

The NIS Advisory Group provides a necessary forum for discussion and planning among the three crucial sets of participants in designing and implementing an information infrastructure to support network synthetic science: the LTER science community, LTER Information Managers, and the Network Office. It is recommended that the NIS Advisory Group meet twice a year to draft goals to implement the strategic plan and to conduct assessments of NIS development. The NIS Advisory Group will make recommendations on goals and timeframes for NIS development and will prepare an assessment report for presentation to the Coordinating Committee. These presentations will be made at least annually. These recommendations will guide the activities of Network Office staff assigned to NIS development.

The meetings of the NIS Advisory Group will be scheduled to coincide with LTER Exec, Coordinating Committee or IMExec Meetings. The implementation recommendations accepted by the Coordinating Committee and the assessment report describing progress will be summarized in an annual article for LTER Network News. One recommendation to facilitate information management support for network synthetic science would be to produce a document on best practices for the development of synthetic databases with advice to scientists on the formulation of critical information management components. The Network Office will also provide advice on synthetic database development.

Adoption of NIS Standards

To facilitate inter-site collaboration the LTER network should adopt standards for site participation and a tiered functionality framework for promoting and measuring progress in site information management practices. These standards would ensure a level of participation necessary for increased, predictable functionality within the network.

The strength of the network lies in the degree of participation of sites. While not every site can participate in the development of every database, there are certain basic databases that will enhance overall synthesis efforts. All sites have committed to populate and update existing basic network databases (ClimDB, HydroDB, SiteDB), where applicable. The Coordinating Committee should determine which cross-site databases are essential for all-site participation. All sites would be expected to participate in the development and updating of these databases. Participation in these databases should become part of the site review criteria adopted by NSF. By selecting science themes, the Coordinating Committee will also be selecting other cross-site databases to be developed. Site participation for science theme databases will be variable, but at a minimum those sites participating in the workshop should be expected to contribute data.

Tiered Information Management Functionality at Sites

The NIS Advisory Group recognizes that development of tools to support network synthesis efforts should recognize that specific levels or tiers of IM functionality currently exist at individual sites. It is unreasonable to expect that, at any given time, all LTER sites will exhibit the same level of IM functionality. Nevertheless, it is possible to identify a generalized trajectory of site IM development and to articulate the costs and benefits of

site participation at various levels, with respect to synthesis efforts. A common goal agreed to by the Coordinating Committee is to improve each site's position on the trajectory. As sites improve IM infrastructure and functionality, features enhancing network-level synthesis including data discovery, access and aggregation will be more fully supported. The lowest tiers may be characterized by more manual methods and unstructured metadata, while higher tiers may represent highly assisted methods and structured, machine-readable metadata.

The NIS Advisory Group recommends that as part of the network Strategic Plan process, the LTER IM Committee, or a sub-committee thereof, draft a tiered functionality framework that represents the generalized trajectory of site IM development. The range of tiers identified should be sufficiently broad to include the minimum level of participation as well as prospective levels that enable next-generation synthesis tools for use by the ecological research community. The number of tiers should be large enough that advancement to next-higher tiers is reasonably discrete and manageable with respect to site strategic planning.

The tiered functionality framework will help to coordinate the advancement of network synthesis. Specifically:

- IM will articulate the objectives for each tier and the kinds of synthesis enabled by each successive level
- IM will identify the functionality requisite for each level including recommendations for the extent of implementation of structured metadata
- Sites will evaluate their current tier and decide how to allocate personnel and other resources to assure advancement to the next target level
- Advancement objectives can be stipulated in terms of the framework, from the perspectives both of the network and of the individual site

The framework will facilitate NSF site evaluations and help to identify technical or other problems limiting site participation

-
- CC will review, modify, and endorse as necessary, both the tiered framework and objectives tied to each tier

NIS Advisory Group will develop metrics for assessment of progress at site and network levels

Assessment Plan for the Network Information System

A periodic review needs to be conducted to assure success in meeting the goals of the NIS. The review should be performed by the NIS Advisory Group who will make a report to the LTER CC for consideration and final decisions. The first review should start within a year to document the progress of the NIS development. The review should be tied to a strategic plan that will contain a timeline with concrete goals and milestones. The assessment would also include a review of minimum standards for site participation and the tiered functionality framework, as well as the synthesis products themselves. Usability of synthetic databases can be assessed using statistics and logs relating to database use, downtime, and user distribution as well as user comments. Products such as databases incorporated in the NIS, publications resulting from the NIS and citations from use of the database would be important indicators of success. To assist the assessment process, access and use of the NIS products should include a user log that would document use and request notification of resulting publications.

Joint Vision for New Funding

Increasing the level of site participation and information management within the Network will have costs. The NSF 20-year Review of the LTER Network highlighted the need for increased funding for information management. It is also clear to the NIS Advisory Group that additional funding will be necessary to support the enhanced levels of IM functionality required for cross-site and network-wide synthesis. To reduce long-term costs, robust, generic solutions for developing and maintaining basic and science theme databases

should be developed. A mechanism, via Network Office funding, exists to assist in the development of science theme databases. NSF has provided supplemental funding for development of basic databases, such as ClimDB. We anticipate specific funding requests will be made to NSF as the Coordinating Committee recommends additional synthetic databases. Targeted budget increases on existing grants (plus-ups) and competitions for supplemental funds within the LTER program could play an important role in enhancing individual site's capabilities along the trajectory of IM functionality that will support increasingly powerful tools for synthesis activities.

Several other mechanisms may also be viable alternatives for enhancing the LTER Network information infrastructure. For example, the development of specific synthetic databases and support of the requisite infrastructure may be partially offset through research grants and instrumentation grants under NSF Biology Directorate's Division of Biological Instrumentation (BDI). Secondly, more visionary research proposals that result in generic solutions to broader community problems may be supported through NSF CISE's ITR and related programs. Such research efforts would likely result in new types of informatics functionality in support of synthesis, as well as new infrastructure that could be prototyped at a subset of LTER sites. Third, synthetic research projects may be expected to rely heavily on a sufficient information infrastructure, necessitating the inclusion of funds in research proposals for both infrastructure and informatics personnel.

The NIS Advisory Group foresees that many ideas for informatics research will emerge during LTER's Decade of Synthesis—e.g., a clear and compelling need for new classes of harmonization algorithms, "smart" quality assurance and quality control procedures, and new database structures. As these new ideas emerge, the NIS Advisory Group could play an important role in communicating these needs to the broader community by e-mailing bulletins to interested information managers and relevant parties (e.g., IMPLUS) and publishing routine needs assessments in *Databits*. Such an activity would ensure that informatics needs identified during synthesis efforts are fed back to appropriate IT researchers—hopefully resulting in new information technologies that better enable future synthesis efforts (i.e., a positive feedback loop between the ecological and information technology sciences).

Integration of NIS Strategic Plan with Other Strategic Plans

It is essential that the NIS Strategic Plan be fully integrated and compatible with both the LTER Network Strategic Plan and the Network Office Strategic Plan. We propose that the NIS Advisory Group be asked to review the draft of the Information Management section of the LTER Network Strategic Plan for coherence with the NIS Strategic Plan. We see the Network Office Strategic Plan as including all the Network Office roles identified in the NIS Strategic Plan. This coordination of strategic plans will be ensured by inclusion of Network Office participation in the NIS Advisory Group and overlapping participation in strategic plan development.