

The following is a **first draft** of this tiered framework. Further discussion will occur at this year's IM meeting in Seattle, which may radically change this document.

Tiered Information Management Functionality at Sites

Specific levels or tiers of IM functionality currently exist at individual sites and, at any given time, LTER sites will exhibit varying levels of IM functionality. This level of functionality will determine the extent to which the site can participate with synthetic activities and tools. The NIS Advisory Committee has proposed a tiered trajectory for IM development to help coordinate the advancement of network synthesis and to provide a framework for site improvement and evaluation. 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 use/aggregation will be more fully supported. The trajectory will become part of the LTER Network strategic plan currently under preparation by the Coordinating Committee.

The IM Committee is requested to help resolve several important issues in developing this tiered framework. The IM Committee will define necessary functionality required for short-term and long-term synthetic efforts and develop recommendations for multi-tiered NIS standards. For each tier level:

- (1) Articulate objectives and the kinds of synthesis enabled
- (2) Identify evaluation criteria to measure site progress
- (3) Estimate the actual costs or resources involved in completing the work

This document is intended to provide a framework for site planning with respect to the Network Information System. Sites will be free to choose which areas of this framework are most critical, and construct a site-specific plan for improvement. 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.

In examining the development of a tiered structure, IM's originally proposed the following functional dimensions: discovery, access, and use. However, it is understood that sites develop simultaneously in all of these areas, and tiers for measuring site progress could be established within each of these functional dimensions. The tiered trajectory structure might be better thought about as a functional matrix, with the representation of a site's position called its Functionality Profile. The evaluation criteria questions listed might best be answered as a relative compliance, e.g., no/little (0-20%), some (20-50%), much/many (50-80%), most/complete (80-100%).

Functional dimension: Discovery

1) Specification

a) Tier 1

- i) Site develops a catalog of all LTER core data sets.
- ii) Minimal registry metadata is provided for all data sets in the catalog, e.g., title, creator, contact.
- iii) Catalog is available online.

- b) **Tier 2**
 - i) Site expands catalog metadata to include basic resource information, e.g., minimal registry metadata plus abstract, keywords, intellectual rights, and online availability of data and metadata.
 - ii) Site will consistently structure all basic resource information
 - iii) Site will provide a local mechanism to easily browse or search the local catalog of LTER core data sets.
 - iv) Site will provide discovery metadata (basic resource information and data URL) to a central harvester to allow a general network search for data sets over all LTER sites, e.g., Data Table of Contents (DTCOC).
 - c) **Tier 3**
 - i) Site will structure basic resource information in EML (or similar structured language)
 - ii) Site will structure geographic, temporal, and taxonomic coverage metadata in EML (or similar structured language)
 - iii) Site will allow discovery of data set through an EML registry
 - d) **Tier 4**
 - i) Site will allow semantic-based discovery through machine-based searches for data.
- 2) **Evaluation criteria**
- a) Is there a catalog of core data sets? Is the catalog online?
 - b) Is minimal registry metadata provided for all data sets? Is basic resource information provided for all data sets?
 - c) Is the basic resource metadata structured and consistent across data files?
 - d) Is coverage metadata (geographic, temporal, and taxonomic) available and structured?
 - e) Is there a searchable catalog or citation index of core data sets?
 - f) Is the site participating in a searchable LTER network data set directory such as the Data Table of Contents?
 - g) Is the site providing basic resource information in EML (or similar structured language)?
 - h) Is the site participating in an EML registry, e.g., metacat?
- 3) **Resources needed**

Functional dimension: Access

1) Specification

- a) **Tier 1**
 - i) Basic resource information and coverage metadata will be available locally to site researchers.
 - ii) Site core data sets are available locally to site researchers.
 - iii) Processes are in place to update core data sets within two years of collection.
- b) **Tier 2**
 - i) Basic resource information and coverage metadata will be accessible online.
 - ii) Methods and attribute metadata will be available online.
 - iii) All core data will be online and current within two years of data collection.
 - iv) A data access policy will be in place describing the rules governing access of site data and specific restrictions for access of certain data sets.

- c) **Tier 3**
 - i) Basic resource information, coverage, methods, and attribute metadata will be structured in EML (or similar structured language).
 - ii) Site core data sets will be queriable with a local language
 - d) **Tier 4**
 - i) Site core data sets will be queriable through a shared network language
- 2) Evaluation criteria**
- a) Are all core data sets available to local researchers?
 - b) Are all core data sets available online?
 - c) Are all core data sets current within two years?
 - d) Is all basic resource information available online for all core data sets?
 - e) Are all coverage, methods, and attribute metadata available online for all core data sets?
 - f) Is a data access policy in place?
 - g) Has existing metadata been structured in EML ((or similar structured language)?
 - h) Are data sets locally queriable (e.g., sub-sampling data set rows and columns, joining or merging multiple data sets, etc.)?
- 3) Resources needed**

Functional dimension: Use

1) Specification

- a) **Tier 1**
 - i) Site will develop human readable and comprehensive metadata (comprehensive would include the basic resource information as well as coverage, methods, entity and attribute descriptions, and physical descriptions of the data sets).
 - ii) Metadata will be of sufficient completeness and quality to allow local reuse of the data set.
 - iii) All core data sets are machine readable
 - b) **Tier 2**
 - i) Site will develop comprehensive, structured metadata
 - ii) Metadata will be of sufficient completeness and quality to allow independent reuse of the data set (e.g., independent of creator, place and time of data collection)
 - iii) All core data sets are online and interpretable from data table and physical metadata descriptions
 - c) **Tier 3**
 - i) Comprehensive metadata will be standardized to adhere to standards of use (e.g., EML, FGDC) defined by relevant communities (e.g., LTER scientists asking synthetic questions)
 - ii) Metadata will be machine-parsable in a known, shared schema or domain-defined standard to allow semi automated processing
 - d) **Tier 4**
 - i) Online metadata and data will allow fully automated knowledge extraction (e.g., KR: knowledge representation, semantic markup/ontology)
- 2) Evaluation criteria**
- a) Are the metadata comprehensive enough to allow local reuse of the data set?
 - b) Are the metadata comprehensive enough to allow future, independent reuse of the data set?

- c) Is the site participating in relevant synthetic database activities (i.e., ClimDB, HydroDB, ANPP, etc.)
- d) Is the site participating in network-level synthesis projects or theme-based activities by providing comprehensible, structured metadata related to the science question asked?

3) Resources needed