

NISAC Meeting Report

LTER Network Office, Albuquerque, NM March 24-26, 2010

Attendees: Will Pockman (Co-chair, SEV), Wade Sheldon (Co-chair, GCE), Barbara Bond (AND), Libe Washburn (SBC/MCR), Kristen Vanderbilt (SEV), Karen Baker (CCE/PAL), John Porter (VCR), Bob Waide (LNO), James Brunt (LNO), Mark Servilla (LNO), John Vande Castle (LNO), Chuck Hopkinson via VTC (PIE).

Summary of discussion and action items:

- 1) The Committee reviewed and approved CY 2010 milestones for its January 2011 evaluation of progress in NIS development.
- 2) The committee discussed the progress of working groups developing the unit registry and a controlled keyword vocabulary for LTER metadata, both critical pieces required to improve EML metadata consistency in the LTER data catalog.
 - a. NISAC supports the continued development of the unit dictionary and registry to facilitate automated unit conversions in PASTA workflows.
 - b. The controlled vocabulary group has submitted a list of 640 keywords to IMExec and IMExec has recommended this product to EB. NISAC will review this topic at a future meeting if so requested by EB.
- 3) The committee heard an update from an ad hoc IMC web services working group and recommends that LNO support their activities as a product-oriented working group to enable sites to leverage existing network databases without creating standalone solutions.
- 4) The committee discussed the status of the LNO operational plan submitted to NSF for review and approval after a lengthy review and revision in Spring 2010. Important aspects of this plan that were the topic of extended discussion include:
 - a. Plan for development of PASTA, including the milestones noted above.
 - b. A communication plan to engage IM and domain scientists in PASTA development. NISAC supports this effort because it offers an important opportunity to make sure that PASTA addresses the needs of network scientists and working groups.
 - c. Utilization of Tiger Teams, groups of 2-5 individuals from the network to provide rapid, in-depth feedback to LNO development team on specific aspects of PASTA development. NISAC recommended that LNO develop approaches to recruit broadly from the LTER community.
- 5) The committee reviewed the current draft of the Cyber-infrastructure (CI) implementation Plan and revised it to reflect progress since the document was last reviewed. A revised version will be forwarded to EB for review.
- 6) The committee reviewed current policies for committee composition and terms of service, and recommends the following:
 - a. Composition to include 6 domain scientists (3 yr terms, approved by SC), 4 information managers (2 yr terms voted by IMC) and LNO staff. Terms for those serving as co-chair will extend to 4 years.
 - b. Greater transparency regarding discussion and voting on issues with real or perceived conflict of interest for some group of committee members.
- 7) NISAC activity in coming months will include VTCs in late April and early June.

LNO report to NISAC on Operational Plan Approval. This spring, LNO submitted an operational plan to NSF to meet the requirements of ARRA funding of its program. By agreement with the EB, the operational plan addresses all aspects of LNO operations, including NIS development. A component that was discussed extensively during the review process is a communication plan that will define mechanisms of engaging the community in the process of NIS development. After review by NISAC, IMExec and a panel of external advisors, and completion of a reverse site visit with NSF program staff in Arlington, NSF is expected to approve the plan without major changes. NISAC considered several aspects of the operational plan as detailed below and will present an overview of the operational plan at the upcoming Science Council meeting May 12-14. One of the goals of this presentation is to encourage members of Science Council, and others from their sites, to participate in Tiger Teams and other efforts in support of NIS development (see below).

LNO milestones for CY 2010. The operational plan calls for yearly NISAC review of milestones measuring progress on NIS development. At this meeting LNO proposed milestones for completion during CY 2010 and, after minor revisions, NISAC approved a motion to accept them (see Attachment A). An evaluation of progress on these milestones will be conducted in January 2011 and submitted to the EB.

Discussion of the use of Tiger Teams to engage LTER community. Tiger Teams represent a significant component of the LNO operational plan and a principal conduit for feedback from the community on different aspects of NIS development. The goal is to identify 2-5 individuals who agree to be very responsive to requests for discussion, review and testing of NIS components over a several month period at critical points in NIS development. LNO will develop guidelines for period of service and characteristics for participants, and identify activities that may require participant compensation (e.g. honoraria). The success of NIS development, as measured by the functionality of NIS/PASTA in the long term, will depend upon engagement from all areas of LTER. This is an opportunity for scientists to make sure that NIS/PASTA becomes a powerful tool for addressing science questions using synthetic data from across the LTER network. LNO is considering whether they need to hire an individual to manage community engagement in this process. If hired, this individual would participate in recruiting Tiger Teams and facilitating their interaction with NIS developers.

Increasing access and use of LTER data. The committee discussed the need to increase the utilization of LTER data in cross-site synthetic research. It was widely agreed that one strategy to achieve this is through increasing the availability of high quality derived datasets with complete metadata. Maintaining access to primary data is important, but synthesizing primary data requires a significant upfront investment to produce comparable data from multiple sites. NISAC recommends seeking input from the Science Council to determine which value-added datasets should be made available. Possibilities range from summary data sets intended for research to summary data published as graphs in presentation format with accompanying explanatory materials

suitable for educators. Also, because the nature of synthetic research is iterative with evolving questions, NISAC recommends that the development of added-value datasets occur in tight communication with the working groups that requested them.

Increasing access to legacy data. NISAC discussed possible approaches to incorporating legacy data in the LTER network (i.e. generation of EML and solving data formatting problems to make them available to PASTA and NEON). Our discussion was limited because this topic is currently being considered by NSF. Several models for accomplishing this data incorporation were mentioned. The committee agreed that there must be a mechanism for including local LTER personnel in the process since they (responsible PIs and IM) are the ones with the knowledge required to work with these data. As this issue develops, NISAC stands prepared to provide input to the EB. NISAC could play a primary role in reviewing the plan for legacy data incorporation and advising the EB on the implementation of this activity. The process of assigning priorities, however, should not be tasked directly to NISAC but rather should seek the broadest participation of scientists (SC) to make the most informed choices. Moreover, sites will need to play a leading role in identifying legacy data, although there should also be consideration of network level priorities to provide broad data on topics that are relevant to NEON (the possible funding source for this effort) and the types of data that would be most useful. NISAC recommends collaboration with NEON personnel to discuss goals and data priorities if this opportunity progresses.

CI Implementation Plan. NISAC reviewed the draft CI implementation plan that was under development before committee activity was diverted to the LNO operational plan evaluation. We noted that many items in the draft CIIP were addressed in the LNO operational plan or are being addressed by IMC working groups (e.g. web services, unit registry, controlled vocabulary). The draft is currently being revised to reflect these new activities, and will be submitted to IM-Exec and the EB for consideration in the near future.

Unit Registry Working Group. The committee reviewed plans of the LTER Information Management Committee Unit Working Group (IMC UWG). NISAC endorses the development of an LTER unit dictionary and registry by the IMC UWG in coordination with LNO as a central element of the LTER NIS development. LTER datasets all require reporting of units; the unit dictionary objective to develop a searchable database of standard units ensures that units from one site are comparable with those of another and sets the stage for automating conversion of units. Also, the units registry avoids comingling of concepts in defining units (e.g. mg N liter⁻¹. where the unit is mg liter⁻¹, and the element (N) involved isn't part of the unit). Poor unit naming practices make data integration and interconversion between units more difficult, and use of the unit registry will reduce this problem. The unit registry transforms the EML unit dictionary from a static list of units into a dynamic process by providing for site submission of units as well as for the designation of different levels of scope (e.g. LTER site or investigator) as part of a unit vetting process. Plans for two activities - a site-to-site meeting in the Spring and a sites-network meeting in early summer - are moving forward with support from the UWG post ASM proposal.

Controlled vocabulary working group. NISAC reviewed the controlled vocabulary project that has been underway for some time through the efforts of a group of IMs. Searching for datasets is difficult without a standardized vocabulary, leading to problems finding comparable data. Only 3% of keywords in existing LTER EML documents are used at 5 or more sites with synonyms used at a broader group of sites for many core datasets. This working group has examined existing keyword sets (e.g. NBII, GCMD) but found them insufficient because of differences between LTER research and the goals of the projects that created these sets. The current recommended LTER list has 640 words (148 synonyms), but just 200 words in common with NBII and GCMD. IMexec has recommended the current product to EB. Based on our discussion, NISAC

- Recommends adoption of an approved LTER keyword list (with some reservations considering the effort required to achieve it)
- Suggests that the list be managed (e.g. additions considered) on a regular but infrequent basis by a subcommittee of the IMC
- Acknowledges that few resources are explicitly allocated to this process at the moment but would be required to see it through to completion
- Acknowledges that the keyword list is not meant to supercede existing site keyword lists, and sites will not have to retroactively keyword their resources with the LTER Controlled Vocabulary. Rather, the Controlled Vocabulary will be a preferred list of terms for sites to use when generating metadata.

Next steps after a product is approved should be decided at a later time, but might include development of other keyword lists (geographic locations, taxonomic names, etc) and tools for automatically adding approved terms to documents. Also, the current keyword list resides in a spreadsheet and needs to be moved into a database. NISAC will provide evaluation and feedback to the EB if requested at a future meeting.

Web services working group. Wade Sheldon reported on the activities of an ad hoc web services working group that was organized at the 2009 IMC meeting. This group will collaborate with LNO personnel in developing application programming interfaces (APIs) for network level databases (e.g. personnel, bibliography, SiteDB) to facilitate leveraging of these databases by sites. They will also develop best practice guidelines, prototype applications, and provide assistance to sites developing web services for their information systems. NISAC recommends that IMExec formally establishes this group as a product-oriented working group, and that LNO provide funds to organize a face-to-face meeting in the near term. This group is ideally suited to addressing key elements of the CI Implementation Plan (particularly developing a service-oriented architecture for the network) and its activities should be encouraged at all levels.

ChemDB update. NISAC received an update from Don Henshaw concerning progress on development of ChemDB with independent funding from USFS. A postdoc has been hired for two years, and a programmer will be hired in fall. These efforts will build on other issues discussed by NISAC (such as repurposing SiteDB to make it a gateway to synthetic databases such as ClimDB and HydroDB) and will need EML adequate to allow PASTA to incorporate data. It is unclear how databases such as ChemDB will be maintained into the future, particularly when short term funding is exhausted. The LTER

network should carefully consider whether it can contribute to supporting the database products into the future and how USFS can contribute, particularly since a large proportion of the data in the database may be non-LTER data.

Processing and incorporating sensor network data. NISAC briefly discussed the need for an SC sub-committee or working group to address approaches for handling increasingly frequent data streams from sensor networks at LTER sites. NISAC recommends requesting that IMExec identify a way to develop shared approaches for managing high volume streaming data, including QA/QC on these data. This effort should involve some collaboration with a complementary group in NEON seeking to address this problem at their intensively measured sites.

NISAC structure and length of appointment

The committee discussed at length its own role in the network governance structure and the possibility of changes to its membership as suggested by a group of IMs following the 2009 ASM. NISAC was created to meet the need for informed evaluations of issues affecting the Network Information System by bringing together domain scientists, information managers and LNO personnel to discuss the interaction of science and information management topics. After the 2009 All Scientist's meeting, a group of IMs met with the EB to request consideration of changes in committee structure in light of the increasing reliance upon NISAC for input on a variety of issues. This group suggested making the committee smaller and changing the representation of the three types of members to address the potential for conflicts of interest within the committee (e.g. LNO members offering input on tasks assigned to LNO, or domain and IM members asked for input that may influence allocation of resources or workload assigned to their sites).

After a wide-ranging discussion, the committee concluded the following:

- The recommended changes in committee structure do not favor the best performance of NISAC. COI issues can and should be dealt with through greater transparency, identifying members with COI on a particular issue and by having them sit out of votes on that topic.
- The demands on the three different elements of the committee are different and our service terms should reflect these differences. NISAC recommends including **6 domain scientists** (an increase of 1 from the present structure) with a 3 yr commitment and a possible fourth year for those who agree to serve as chair. The previous 2 yr commitment is very short for domain scientists who face a steep learning curve to understand the issues that are relevant to NIS/PASTA development and activity. In addition, it is often difficult to arrange attendance of all domain scientists at each meeting (especially in person meetings but even for VTCs) because of scheduling conflicts with their other activities. We recommend including **4 Information Managers** with a 2 yr term and an additional 2 yr commitment for those willing to serve as co-chair. These members are currently selected by vote of the IM committee so this policy, if approved by EB, would also require input from IMC. We recommend that members from **LNO** (Executive Director, CIO and NIS developer) remain as permanently appointed

members of the committee and that other LNO staff participate in meeting as necessary to facilitate discussion of particular issues.

NISAC Membership Rotations

Our rotation of members in recent years has been slower than planned, with the result that many of the current membership have exceeded or met their planned terms.

Domain science: currently 4 members (Pockman – 4 yrs in 2010, Bond – 3 yrs in 2012, Hopkinson – 2 yrs in 2010 agreed to extend to 2011, Washburn – 2 yrs in 2011). Our new policy would require that we immediately recruit two domain science members for 3 year terms. In the absence of willing candidates and to maintain consistency in the chair position as Sheldon leaves the committee, Pockman agreed to extend for one year – serving as co-chair until 2011.

IM: currently 4 IMs (Vanderbilt – completing second 2 yr term in 2010, Sheldon completing 4 yrs in 2010 – 2 yrs as co-chair, Baker 2 yrs in 2011 and Porter – 2 years in 2010). Wade Sheldon will leave the committee. IM Co-Chair will be identified by vote of IMC from among the three remaining current members and one new member should be approved by IMC. The committee thanks Wade Sheldon for his service over the last four years.

The committee also briefly discussed the issue of populating the committee with qualified and active members in the face of increasing requests for input and competing demands upon committee members time. The committee considered, and rejected as inappropriate, the possibility of seeking compensation for service on NISAC. This issue will continue to be a challenge that will influence the response time of NISAC on the tasks brought before it. Solutions to this problem will include: better distribution of workload among committee members rather than relying upon co-chairs to handle the bulk of the workload between meetings and tasking LNO members for meeting planning, agenda preparation etc.

Future meetings: NISAC activities for the remainder of the year will include VTCs in April and June and year-end activity for annual review of LNO milestones. Our next in person meeting will be in March 2011, in advance of SC meeting in April/May.

Attachment A. Milestones for LNO NIS Development Activity – CY 2010
Approved by NISAC, April 24, 2010.

Milestone 1 - By the end of the 2nd quarter, identify requirements and demonstrate a functional prototype (Phase 1) of the Data Catalog component (see section 3.2.3.10.1 of the LNO Operational Plan 2009-2014) of the PASTA framework. Specifically, engage NIS Tiger Team to evaluate necessary web-service interface requirements and install and configure a Metacat server instance. Criteria: Done or not.

Milestone 2 - By the end of the calendar year, identify requirements and demonstrate a functional prototype (Phase 1) of the Data Loader component (see section 3.2.3.10.1 of the LNO Operational Plan 2009-2014) of the PASTA framework. Specifically, engage NIS Tiger Team and EML Metrics working group identify requirements for quality checks and reporting of quality of metadata and deploying an instance of the Data Manager Library for testing metadata and data congruency and loading data into the Data Cache. Criteria: Done or not.

Milestone 3 - By the end of the 3rd quarter, identify requirements and demonstrate a functional prototype (Phase 1) of the Data Cache component (see section 3.2.3.10.1 of the LNO Operational Plan 2009-2014) of the PASTA framework. Specifically, engage NIS Tiger Team to evaluate necessary web-service interface requirements and install and configure a RDBMS server instance. Criteria: Done or not.

Milestone 4 - By the end of the 3rd quarter, engage NIS Tiger Team to complete a draft specification (Phase 1) of the Discovery/Access API (see section 3.2.3.10.5 of the LNO Operational Plan 2009-2014) of the PASTA framework. Criteria: Done or not.

Milestone 5 - By the end of the calendar year, engage NIS Tiger Team to complete a draft specification (Phase 1) of the Identity Management Services component (see section 3.2.3.10.6 of the LNO Operational Plan 2009-2014) of the PASTA framework. Criteria: Done or not.

Milestone 6 - By the end of the calendar year, complete a draft plan for integration of the EcoTrends and ClimDB/HydroDB databases into the Network Information System (see section 3.2.3.11.1 of the LNO Operational Plan 2009-2014). Criteria: Done or not.

Milestone 7 - By the end of the 3rd quarter, complete and deploy services that support HIVE-LTER interoperability. Specifically, a) format the LTER controlled vocabulary in Simple Knowledge Organization System (SKOS) format, b) install and test the HIVE vocabulary repository, c) develop a functional prototype of a metadata generation vocabulary tool, and d) develop a functional prototype of a metadata discovery vocabulary tool. Criteria: Done or not.