**LTER National Information Management Office (NIMO) – streamlining data curation to accelerate scientific inquiry**

**I. Introduction**

Ecology is entering a new era of science (Hampton et al. 2013, Peters et al., 2013, Robertson et al., 2012, Uhlir, 2012) – one in which: (1) long-term observations are proving critical to our understanding of the drivers of environmental change, the consequences of which have local to global implications (Knapp et al., 2012, Michener & Jones, 2012); (2) synthesis by diverse teams of scientists from multiple institutions is increasingly important; (3) the tools used by scientists, especially early career, are evolving rapidly, along with the expectation for published, discoverable, and reusable data (Hernandez et al., 2012). The NSF Long Term Ecological Research (LTER) program has a special role to play in ushering in this new era by making data from a diversity of sites and domains discoverable and usable both within and outside of LTER.

The LTER program has recognized the importance of information curation, long term archival and public access to data from its inception, and LTER sites and the central LTER Network Office have made significant investments in the area. LTER Information Managers (IMs) have been instrumental in contributing to the growing knowledge base for and experience in environmental information curation (Michener et al., 2011; Porter, 2010). With the introduction of the Provenance Aware Synthesis Tracking Architecture (PASTA), as central data repository to the LTER Network Information System (NIS), data curation has been elevated to a very high standard within LTER. Most data sets produced at the sites are now securely stored, publicly accessible and documented with rich, high quality metadata in PASTA. However, now that environmental (particularly LTER) data are widely available in well designed repositories it is becoming clear that this does not automatically lead to easy reuse by synthesis projects. This is not an LTER- or PASTA-specific problem, but rather a more general problem (Hale et al., 2003, Lohr, 2014, Wilson et al. 2015) attributable to many factors. Some of these factors will have to be addressed by the larger eco-informatics community and Information Technology (IT) research projects, but some must be addressed from within the environmental data curation community (Kervin et al., 2014). The former include development and meaningful implementation of ontologies, data linking (‘linked data’), automatic indexing by larger search systems (e.g., Green 2009), and automatic metadata generation by instrumentation and during the research process. The latter, data curation related factors, include developing and following best practices and more standardized approaches to formatting and documenting data of different types and research themes before archival in a repository (Baker & Chandler, 2008), i.e., slowly moving away from the current model of one investigator, one research question, one data manager leading to unique archival formats, and uncoordinated documentation of scientific research data (metadata content).

Driven by the necessity to accommodate all LTER data, PASTA has been designed to accept datasets in any format as long as they are accompanied by a valid Ecological Metadata Language (EML) document containing pertinent metadata. Likewise, by design and necessity, EML accepts any metadata content as long as it is formatted correctly (valid XML). Advanced metadata support functions in PASTA assure completeness and data/metadata congruence, but do not prescribe a data format (data model) or standardized metadata content (e.g., controlled keywords or attribute names), which has led to a great variety of data formats and data descriptors, even for thematically very similar data. Accordingly, we are proposing a major advancement and paradigm shift in LTER information management with the new LTER National Information Management Office (NIMO) emphasizing support for accelerating scientific inquiry (Peters, 2010), a future where the process of data discovery and preparation does not require the vast majority of time of the research process (Beagrie & Houghton 2014). These technical and operational advances will be matched by initiatives that leverage both human and organizational networks to increase information management capacity and foster a spirit of experimentation and innovation without imposing major additional costs.

To achieve the paradigm shift a transition from a loosely organized federation of site based information management approach to one that can not only scale to more data but can also ease reuse of long-term data is necessary. The LTER Information Management Committee (IMC) has devised an approach to the formation of an LTER data center, termed the LTER National Information Management Organization (NIMO, Fig. @@). We envision a flexible organization for LTER Network-level data management and beyond that can respond to community-identified needs with sustainable technology, and which facilitates usability and adoption of appropriate, common tools and practices by data management personnel. NIMO does not replace site-based data management; rather it supports common practices in place at sites with broadly used centrally supported tools, community developed standards and best practices. The new organization will nurture the development of thematic specialization and task-sharing within the information management community, which will allow the system to expand quickly to other ecological research programs and small projects that currently do not have access to professional data curation services. NIMO will look inward (i.e., to the LTER community) to leverage and reinforce the power of an existing and highly effective network of data professionals, and NIMO will look outward to other data centers, federal agencies, and data aggregators to align NIMO with broader data initiatives and harness capacity that will expand and improve LTER data discoverability and reuse.

NIMO provides for innovation and extension of the LTER NIS with a foundation formed by the experience and wisdom of thirty years of information management and motivation provided by science-driven initiatives derived from LTER synthesis groups, as well as the broader LTER and ecological research communities. The approach is service-oriented and leverages and reinforces the highly effective and existing human infrastructure of LTER’s Information Managers (IMs), while engaging in new ways and with new partners to offer innovative information management solutions and opportunities for growth. The overarching goal for the three year endeavor is to be *Poised for a new era of information management while participating in defining the new era.* Specifically, goals for NIMO are to:

* Maintain, improve, and scale the LTER Network Information System functionality to ensure secure long-term availability of environmental data
* Improve efficiency of the data curation process
* Close the gap between synthesis science needs and data curation approaches
* Participate in the longer range eco-informatics visioning



Figure 1: NIMO’s Resources and Initiatives for achieving the proposed Goals

To meet these goals (Fig. 1), NIMO will adopt and stabilize existing services critical to data curation and archives (goal 1) (section II.A), harmonize site and network data curation to improve consistency (goal 3) and efficiency (goal 2) (II.B) and thus shorten the data lifecycle between site-based science and network-level synthesis, and facilitate long-term strategic planning to ensure LTER data management needs are met while maintaining LTER’s leadership role within the broader ecological community (goal 4) (II.D). This aggressive set of initiatives will require organizational governance structures that engage and empower key stakeholders both within and outside of LTER (section III).

**Definitions of Terms**

**Data curation** - the process of managing data (cleaning, quality controlling / assuring, formatting, etc.) and documenting (describing) them in EML.

**Data archiving** - the act of submitting the data and metadata to a repository.

**Data repository** - technology designed to securely store data and make them accessible into the future

**PASTA** - the LTER data repository

**IM** - the Information Manager (the person).

**NIS** - the LTER Network Information System

**II. NIMO Initiatives in Support of Goals**

**II.A Maintain, Improve, and Scale the LTER Network Information System Functionality**

Maintaining and improving the NIS functionality needs to be addressed in three major categories, and within each, priorities will be set by the NIMO Operations and Implementation Committee according to identified requirements, priorities and budget constraints. Certain technical services are designated as “core” and will be of highest priority overall because they ensure continued data security and availability, and provide a foundation for all other services. Additional services will be prioritized on a case by case basis. These include assistance to individual sites for their local needs and support of scientific synthesis.

*Core Services and Infrastructure:*

A well-documented set of core information management services is essential to support data preservation and availability, and provide a foundation for additional services.

* Central IT infrastructure (physical servers, cloud storage, administration)
* Central Data Repository with Data Catalog Service (PASTA)
* Data publication through appropriate aggregator services (currently DataONE)

Central IT infrastructure is hardware and system administration to support all network data management tasks. Infrastructure must be robust and flexible to accommodate future changes in all service layers. The PASTA software stack and associated web services comprise the central repository for data and metadata preservation and Digital Object Identifier (DOI) management. PASTA provides simple data catalog functionality, with search and display interfaces and an API, fulfilling the requirement of making all LTER data publicly available together. PASTA also has implemented the generic DataONE member node, and contributes LTER data to the DataONE federation project, making LTER data discoverable and accessible alongside environmental data from other repositories manually via online search and programmatically via an application programming interface (API). NIMO will maintain PASTA in its current functionality. In the future, however, we envision a diversification of data storage and publication solutions. The NIMO will assist LTER sites in evaluating and choosing appropriate data repositories, ensuring continued support of rich and standard metadata, secure availability into the future, and improved discoverability and reusability of data.

Central IT infrastructure will be available as testbeds for new ideas in environmental data management. NIMO will strive to support research through availability of IT infrastructure requirements that may go beyond the single investigator capacity. Of course, budget limitations apply as to what NIMO can offer, but NIMO staff will be available to develop strategies for, e.g., large scale modeling or large data set handling on a case by case basis.

Included in the category of core services are collaborative, or project management technology to facilitate the here proposed collaboration among distributed site information managers (for further detail, please see section II.C) as well as commonly used communication solutions (e.g., online conferencing, e-mail lists, etc.), which will be provided by either the National Communications Office (NCO) or NIMO as needed. Finally, NIMO will continue to support and expand upon current practices for uniform data handling through several metadata management tools and design services including the PASTA congruence checker, the unit registry, and the LTER Controlled Vocabulary.

*Scaling up to Providing Curation Services to non-LTER Projects*

Many LTER site information systems, as well as the NIS, are already accommodating data from non-LTER projects, such as LTREB and MSB, i.e., ecological projects which currently have no access to professional data management, but are expected to supply and follow through with a data management plan (e.g., Rüegg et al., 2014). However, this support is usually provided and funded informally and the supported projects frequently have some connection to an LTER site (e.g. research done at the site with additional NSF funding and LTER PIs involved). Formalizing and opening this service more widely will allow NIMO to grow and for curation services to be offered to a wider range of environmental research projects.

Accommodating non-LTER data will require scaling up of the PASTA software stack and formalizing funding for data curation services. Where possible, we will leverage work that is planned by the "ABI Development: Enabling broad-scale ecological analysis and synthesis through PASTA Plus (PASTA+), the enhanced LTER data repository" proposal submitted by Dr. Servilla, now pending in the NSF Advances in Biological Informatics (ABI) program. This project, if fully funded, will extend the PASTA data repository for use by new research communities, including the Long Term Research in Environmental Biology (LTREB) program, the Macrosystems Biology (MSB) program, and members of the Organization of Biological Field Stations (OBFS). This goal will be achieved through a combination of new software development and training workshops for scientists and students from each of the new communities. Software development activities will (1) extend user identity management in PASTA+ to effectively and securely support users from communities beyond that of LTER, (2) streamline data documentation through improved interface design and conversion of science metadata standards into the Ecological Metadata Language as necessary for PASTA+, and (3) expand the set of rules for checking the quality of uploaded data packages to the standards of these communities. The workshops will focus on teaching data management best practices as applicable within the PASTA+ framework and making data accessible through both PASTA+ and DataONE (via PASTA+'s DataONE Member Node service). This project has an expected start date of July 2016.

Data submitted to PASTA may be curated by members of the non-LTER projects, but NIMO will also offer curation services for a fee, which will require restructuring collaboration among LTER information managers, and with scientists and NIMO staff in new and innovative ways.

We propose that by the end of year two, NIMO will have a business plan established that enables such projects to contract with NIMO for their data curation and storage needs. The business plan will involve developing a fee scale for a range of curation services and archival space requirements, realistic retention policies, hardware, software, and staffing needs for various expansion scenarios (Abrams et al. 2012, Van Norden, 2013). We will work with support organizations, such as NSF’s I-Corps program. NIMO data curation staff will be hired as needed to accommodate the extra work while site IMs may continue to support local projects as has been done using the developed fee schedule as guidelines. In addition, the proposed procedures of establishing best practices and developing curated data processing script libraries will benefit ecological communities outside the LTER (see below).

**II.B Efficient Data Curation**

LTER data management, specifically, and environmental data management in general will need to streamline and automate procedures to curate and publish more environmental datasets (Abrams et al. 2011). In recent years, sites IMs have increased the level of shared data management practices and tools, and NIMO will not only enhance this process but facilitate further advancements through innovative collaboration strategies.

*Common Tools*

An initial step will be to identify technologies and approaches at the LTER sites that are highly efficient and lend themselves to wider adoption. This process will be iteratively applied to data grouped thematically (e.g., synthesis subjects) or by data type (e.g., sensor, geospatial, genomics data) with priorities being developed by the Operations and Implementation Committee. During each iteration, data curation procedures will be collaboratively evaluated, adjusted and automated as much as possible. The gained knowledge of variability, site constraints, and synthesis science needs will allow NIMO to further drive a convergence on commonly used technologies and approaches for metadata creation and data processing that are appropriate to be supported centrally, and which result in economies of scale. NIMO’s role will be to support site IMs in developing these approaches, and once it is clear which are finding adoption by several sites, to harden scripts and publish them in appropriate places (e.g., develop them into R packages on CRAN, Matlab code on Matlab Central, Python code on SciPy). There will not be one size fits all solutions, but a reasonable ecosystem of approaches needs to be supported that may also include commercial technologies and those developed by other groups in the eco-informatics community. ***Overall, NIMO’s responsibility will be to lower the barrier for adoption of any centrally-supported technology through promotion, peer-to-peer training and user-support, but not to develop any major software solution.*** For example, applicable tools already being used or evaluated by multiple sites and under active development are the GCE Data Toolbox for MATLAB (Sheldon, 2008; Sheldon, 2010) and the Drupal Ecological Information Management System (DEIMS, San Gil et al, 2010; Gries et al, 2010 and San Gil, 2011), which have high potential for streamlining processing of sensor data and for site data and information management, respectively. In addition many scripts and small programs have been developed at sites to automate data management workflows as much as possible. These scripts will be evaluated and assembled in strategic and well curated script libraries in a centrally accessible location (e.g., GitHub repository).

*Developing Centers of Expertise and Skill Exchange*

The next step for increasing efficiency in data management will be the development of a skill exchange system and centers of expertise. The proposed organizational structure for NIMO (see *Section III*) can be used to formalize sharing practices that historically have occurred informally between sites. The breadth and depth of data management tasks at each site requires a wide variety of skills within one person. On one hand, local site and science knowledge is paramount and will not be replaced by any central data management organization. However, organizational, technical, data modeling, and general information science skills are broadly applicable and may be shared among sites, especially if some specialization in one of these areas occurs.

A ‘skill exchange’ is the chosen approach for the particular situation of LTER data management, with each site hiring their own IM and each IM primarily reporting to the site PI. Informally, skill exchanges have always happened, at the very least with IMs being involved in leadership positions (IM executive committee, IMC chairs) on a rotating basis, but also direct help and advice has been offered among site IMs. We do propose that expanding on this practice will lead to the development of a more efficient data curation organization while maintaining intimate site science knowledge. To account for the special employment, reporting and decision making situation in LTER IM, NIMO will provide detailed reports of time spent by individual IMs on network wide activities and activities that benefit sites other than the IM’s home site. Formalizing the skill exchange will allow site PIs to judge time spent by their IM on network tasks versus time gained in terms of services delivered to them. An open source online task management system will be employed to accomplish this task with the long term vision of developing into a ‘skills marketplace’ of a much larger scope.

NIMO will promote two areas of specialization and skill exchange, first along the technical trajectory and second, along a science domain trajectory. We envision the development of ‘centers of expertise’ to form based on varying but overlapping and self-identified membership of site IMs. The technical centers of expertise would include specialized skills in streaming sensor or geospatial data management, knowledge in scripting approaches to automation of data management tasks (e.g., R, Python, Matlab), handling of large datasets, backup and security, etc. Single LTER sites may take advantage of these centers through training, adopting technology and reducing time spent in researching a best solution. The thematic centers of expertise would develop for information science tasks and in support of synthesis science. The former include developing best practices and solutions for handling, e.g., taxonomic species information, or understanding and resolving semantic issues of variable naming, keywords, and measurement units. The latter represent the most comprehensive change in the LTER approach to environmental data management.

Several LTER IM centers of expertise already exist, as active working groups, or have formed based on external project funding. Based on this experience, we postulate that their productivity and usefulness to the sites can be increased if supported by NIMO staff, effectively employing proven strategies of distributed collaboration, and developing a system that ensures equitable exchange of effort among IMs (see project management section). We do realize, however, that culture changes for LTER site information management will be necessary for this to succeed. Solutions serving several sites will have to be emphasized over single site approaches, and sites need to agree that providing reusable data is valuable. There is no doubt that a system of centers-of-expertise plus exchange-of-skills will be more efficient than current practices, where every LTER site IM attempts to cover all roles, and the promise of data which are handled efficiently and with the most appropriate tools makes this an exercise well worth undertaking.

**II.C Close the Gap between Synthesis Science Needs and Data Curation Approaches**

As outlined in the introduction, NIMO’s mission is to support the acceleration of scientific inquiry, a future where the process of data discovery, acquisition and preparation does not require the vast majority of time in the research process (Lohr, 2014). However, it has become very clear that having high-quality, well-documented data available in a well designed repository alone is not enough to support efficient data synthesis for answering important questions in a timely fashion. Identified obstacles to efficient data integration are clearly larger than the NIMO alone can take on. As discussed earlier, NIMO’s main emphasis will be on identifying the optimal return on investment for the data curation effort, i.e., identifying the optimal data curation effort at the LTER site that provides datasets in formats and with metadata that reduces the effort of data discovery and harmonization for synthesis research projects. NIMO will also explore the development and maintenance of harmonized data products and improvements to data descriptors that allow more efficient data discovery and reuse. Development of such data products will underlie strict scientific guidance and prioritization by the Operations and Implementation Committee, as it is impossible to support the development of meaningful data products for every synthesis question. However, it is anticipated that through collaboration between NIMO and NCO synthesis projects, a rich library of data preparation, analysis and visualization scripts (R, Python, Matlab etc.) will accumulate. NIMO will take responsibility to curate those scripts into an archival system through which they will be discoverable and reusable.

LTER already has developed several network-wide data products, such as ClimDB/HydroDB (Henshaw et. al. 2006), which will be evaluated for their scientific usefulness and accordingly maintained in their current state or archived. New synthesis data products are expected to be produced by scientific working groups formed through NCO. Each will undergo the same evaluation process and either be archived, or if found useful for future questions maintained and extended as new data are becoming available. Accumulating experience in synthesis research will allow NIMO to identify such products.

*The Role of Centers of Expertise and Skill Exchange*

We are proposing to use a process that allows for tight integration of scientific inquiry and data curation implementing stepwise, iterative change. The NIMO Operations and Implementation Committee will identify one or two NCO synthesis working groups for this process. A group of information managers with particular interest and science domain knowledge will form the ‘center of expertise’, i.e., a group that specializes in the information management for a particular science theme. A few such centers of expertise have already informally developed around themes, such as stream ecology, vegetation plot research, and the Eco-Trends project (Peters, 2013). These centers of expertise then will go through well-defined processes to(1) in collaboration with the scientists explore the problems of discoverability and reusability of data. This exercise will determine best practices for site data managers to implement locally and may include use of key phrases and concepts in the metadata that allow linking to existing ontologies; (2) explore existing data standards that may be useful for this type of data, and if no data standard is appropriate (i.e., if the data collection methods are too variable) at least develop best practices for data set structure; (3) determine the most appropriate data repository for this type of data; (4) in peer-to-peer training help local information managers automate the data curation processes as much as possible for the identified datasets.

The outcomes of this approach will be improved data discoverability and interoperability on many levels. Coordinated use of scientific terminology to describe the data will allow for more effective searching. In connection with evolving ontologies this could lead to a dramatic breakthrough in data discoverability. Determining data standards and repositories will group data of certain types to be discoverable together. We envision using currently existing repositories and possibly setting up local repositories for each center of expertise. These new repositories would rely on currently existing technology and implement either PASTA, Metacat, iRods, or other open source software (e.g., D-Space) as appropriate. This diversification of repositories will have the advantage that thematic data may be managed optimally for discoverability and interoperability, but with the obvious disadvantage that not all LTER data will be discoverable together and may have to be systematically identified by using the keyword ‘LTER’ in the metadata.

Furthermore, centers of expertise will be an important step toward integration of LTER data resources with related community efforts, institutions and agencies making LTER data more easily discovered and reused by taking advantage of sophisticated thematic data management approaches developed by those communities and agencies (Kratz & Strasser, 2014). For example, the hydrologic community has developed software stacks for managing time series (CUAHSI, Consortium of Universities for the Advancement of Hydrologic Sciences Inc.), and some LTER datasets would be well managed in this system. Similarly, some LTER data may be better managed via the Open-source Project for a Network Data Access Protocol (OPeNDAP), in an Arctic data repository, or other thematic repositories where members of those communities have better access to these data. Other examples are the USGS Office of Water Information (OWI) running the Water Quality Portal (WQP), which assembles water quality data from the USGS National Water Information System (NWIS) the EPA STOrage and RETrieval (STORET) Data Warehouse, and the USDA ARS Sustaining The Earth’s Watersheds - Agricultural Research Database System (STEWARDS). Making LTER data available through the WQP will increase their discoverability and by definition will take advantage of data integration and analytics provided by the USGS. Similarly, collaborations with other entities from, e.g., the atmospheric, oceanographic, and forestry sectors might align well with the LTER centers of expertise. Generally, these data repositories can provide more standardized and domain specific search facilities that will lead to better data discoverability and simplified data harmonization for reuse. NIMO’s role will be to ensure the highest degree of LTER data availability through collaboration with existing repositories. Data publication to these thematic repositories may be accomplished via, e.g. harvest routines accessing data in PASTA, implementing appropriate existing software stacks on the central IT infrastructure at NIMO and training site information managers in its use, or data submission to a repository provided by a different group..

For synthesis subjects where no such larger community guidance exists, a set of best practices will be developed that may be useful outside of NIMO for managing similar data. The effort proposed here will be expended on datasets that have been identified as valuable for future reuse. Maximum automation will improve efficiency of data curation procedures. In addition, this exercise will build a library of data processing scripts that may be reused and repurposed. We are expecting that with time certain automation processes will emerge as more generally applicable (one example already exists with the GCE Matlab toolbox that is used at many sites). At that point more effort will be expended centrally by NIMO staff to make the code more robust and generally usable.

**II.D Long-Term Strategic Planning**

NIMO will be an environmental data curation center and data repository in a landscape of many such centers and organizations, and the need for long-term strategic planning has been recognized by members of the community of data providers and scientific data reusers to address grand challenges ranging from capturing more data, to managing and interpreting Big Data, to preparing data useful for synthesis research (Reichman et al. 2011, Obama 2013, Hampton et al. 2015).

Long-term strategic planning needs to happen during the period of this project to ensure NIMO is well positioned to address the overall goal of being *Poised for a new era of information management while participating in defining the new era*. NIMO will have the highest probability of success through meaningful engagement of key stakeholders. Stakeholders within LTER include the network of IMs, site scientists, EB as the chief governing body of LTER operations, and the new NCO. Some stakeholders from outside of LTER have already been identified throughout this proposal and include CUAHSI, the Critical Zone Observatory (CZO), the Biological and Chemical Oceanography Data Management Office (BCO-DMO), NSF EarthCube, and federal agencies, such as USGS, NASA, NOAA, USFS, many of whom have expressed keen interest in collaborations that harmonize LTER information resources with those of respective groups and that take advantage of data discovery, integration, and delivery services either existing or under development.

This emphasis on broad community engagement pervades NIMO, from its principles through its organizational structure. By providing a culture of experimentation and innovation, NIMO will capitalize on the diversity of talent and resources both within and outside of LTER and thus realize information management advances that have the greatest possible impact.

**III. NIMO Structure, Function, and Timeline**

**III.A Project Management - Organizational Structure of NIMO**

A fully functional environmental data management organization requires expertise in many areas: project management, system and database administration, website design and implementation, content strategy, programming, technical writing, data curation and stewardship, combined with a working knowledge of environmental science. Experience has taught us that the breadth and depth required to adequately carry out all tasks are rare in one person. Therefore, we envision a system where the strengths of the diverse group of LTER IMs plus a few targeted hires can be tapped and compensated to service the requirements of NIMO over the relatively short duration of a three year award. To begin the expansion of LTER beyond its traditional informational borders, representatives from entities outside of LTER will be engaged in strategic and meaningful ways.



Figure 2. NIMO’s governance structure and interactions with LTER committees and outside partners.

The NIMO described here does not replace site-based information management; rather it relies upon that distributed talent and offers some central services and support. However, we have outlined above some changes that will be encouraged within LTER information management, emphasizing common, streamlined approaches that are more easily extended. Long-term practices cannot change overnight, and the design of the organization takes this into account. Current site-based systems reflect a long-term response to changing expectations, evaluation criteria, skills, and available technologies, and will provide the experience necessary for effective future evolution.

*NIMO will have the following organizational entities that ensure the objectives are met:* Operations and Implementation Committee; Strategic Planning Task Force; A Project Manager; A technical team; The LTER Information Management Committee. We note, however, that the initial governance structure, which is flat and streamlined, will likely evolve in complexity with hiring of key staff. We also note that policies and procedures for committees, including term lengths and rotations, have been discussed extensively by the IM community who established much of the NIMO vision. For the purposes of this proposal, we assume that all terms will last the duration of the award (three years) and that new terms will be discussed and instantiated as warranted by a future LTER data center competition that likely will occur two to three years in the future.

*Operations and Implementation Committee*

The Operations and Implementation Committee (OIP) will consist of four elected LTER IMs, three representatives from the LTER science community, the project PIs, and the project managers. A rotation schedule for the elected members will be established to provide for reasonable turnover balancing new input with established knowledge. The OIP will be responsible for developing procedures for implementing the skill exchange and centers of expertise, defining and writing operating principles and procedures, and identification of new task forces and ad hoc committees. The OIP will set priorities, make decisions, and oversee all activities of NIMO. During annual IMC meetings the OIP will report on past achievements and discuss future directions with the full IMC. Members of the OIP will coordinate with the LTER National Communications Office (NCO), the LTER Executive Board (EB) and the LTER Science Council (SC). The OIP will ensure optimal integration of science goals with information management requirements in the most democratic and transparent fashion.

*Visioning Task Force*

The purpose of the visioning task force will be to investigate how LTER data services might grow and evolve to meet the future demands of both LTER and the broader scientific community. Specifically, the task force will engage other data entities from the federal government and from other NSF-funded projects to identify strategic collaborations that leverage existing and developing services and technologies that would benefit LTER and that LTER might contribute to. Moreover, collaborations with agencies and data professionals outside of LTER will inform development of NIMO human and technological infrastructure (i.e., sound strategic planning), provide additional training and professional growth opportunities for LTER personnel, and increase the overall data and analytics capacity of LTER.

The NIMO visioning task force will stay involved with the larger eco-informatics community strategic planning effort which was initiated with separate funding (see results from prior support) and will continue with additional funding and through collaborations at the Earth Science Information Partners (ESIP).

*Administrative Function and Principal Investigators*

The NIMO principal investigators, Gries and Hanson, will be part of the administrative function and will be the liaison between NSF and NIMO as under the cooperative agreement the PIs are solely responsible for orderly execution of the here proposed tasks. The PIs will also write the annual project report to NSF. Both Gries and Hanson have extensive experience in leading large collaborative projects. They have experience in conducting training activities in the fields of information management, use of modern technologies for scientific analysis, and team science approaches. The PIs will seek additional funding to support service activities identified as desirable by the Operations and Implementation Committee, but not supported by the current budget, and are available to assist others in grant writing activities.

*Project Managers*

The two Project Managers (PMs) O’Brien and Vanderbilt are responsible for day-to-day operations and technical details of NIMO, under the guidance of the Operations and Implementation Committee and in close interaction with the PIs. The PMs will track progress and provide updates for ongoing NIMO projects, manage activities of support staff and contractors, confirm that sub-contracts are administered on schedule, and manage the request queue. The project manager will contribute significantly to annual reports, and probably act as lead for some sections.

The project managers’ tasks will differ from a traditional IT project manager’s in that they will also coordinate communications between the scientific stakeholders, the IMs and technical support staff. The tasks will be divided between O’Brien and Vanderbilt into being responsible for the details of center of expertise operations and skill exchange, respectively, i.e., supporting the distributed information managers in becoming more of a data curation organization from two different angles, but collaboratively. Obviously, this activity will require a high degree of scientific as well as technical literacy, good communications skills and overall leadership capabilities. Both, O’Brien and Vanderbilt have extensive experience as a site information managers, bringing just the right mix of science and data management knowledge as well as in leadership experience. They know the community dynamics of the LTER IMs and communicate well with all stakeholder groups.

*LTER Information Management Committee*

Currently, the Information Management Committee consists of nominated representatives from LTER sites. We propose that the LTER IMC and OIP meet annually and the long standing annual IMC meeting be augmented with strategic planning sessions for NIMO, to provide overall directions and develop decision frameworks for the OIP to evaluate single projects throughout the year. As NIMO expands into the area of non-LTER data curation it is anticipated that others will join this committee to represent community interest and outside perspectives. The IMC regularly meets in conjunction with the LTER All Scientist Meeting (every 3 years) or the Earth Science Information Partner (ESIP) annual meetings, both of which provide opportunities to gather community input and perspective organized by the Visioning Task Force.

*Technical Staff*

In addition to the project managers, NIMO will employ support staff for systems administration, PASTA maintenance, NIS component integration, and information management support. PASTA maintenance will be provided by the PASTA developers Servilla and Costa who will also be responsible to meaningfully integrate other NIS components (e.g., controlled vocabulary, unit dictionary, etc.). The position of information management support will be a new hire and is envisioned to be a person with enough science domain knowledge to be conversant with science needs, but know enough about scripting in R, Python, PHP, and/or Matlab, to be able to help sites with automation tasks, support synthesis science, refine and publish scripting packages, and provide overall technical assistance. Many of the here proposed changes to LTER data management will incur additional tasks on the information managers or require a change in procedures at the site. Therefore, it is important for NIMO to provide as much support for the additional tasks as possible within budget limitations. Another support mechanism will be to reimburse LTER IMs for particular network wide tasks that are beyond the scope of ‘skill-exchange’ based on OIC set priorities and decisions.

*Procedures of Operation*

One key to the success of NIMO will be to ensure that requirements gathering, decision making, data management automation, and administration are all supported by a robust process framework. Clearly defined and documented processes will support all NIMO activities to ensure that the organization operates both objectively, transparently and efficiently.

As outlined earlier, tasks in support of the core infrastructure will have highest priority but will be regularly evaluated and adjusted to adapt to current needs. Based on these evaluations the full IMC+OIC will decide what is part of this core infrastructure during their annual meetings and strategic planning. All other functions that NIMO will support will underlie a standard procedure of submitting a formal request including description, timeline, budget etc. as needed, the OIP (or the Project Manager for smaller issues) will make a decision based on strategic planning and priorities about resource allocation to fulfilling the request. Requests may be submitted by LTER site information managers, LTER site scientists, or non-LTER community members. The Project Manager will field these requests into categories of (1) needing immediate attention, (2) needing OIP decision, or (3) needing to be included in the strategic planning done during the annual IMC+OIC meeting. Examples for (1) are: day-to-day support questions (system malfunctioning, operational questions for centrally supported technology, etc.); examples for (2) are: assistance with cleaning, hardening and publishing analytical R code from a synthesis working group or a center of expertise, specific skill exchange request from a site, data curation request from a non-LTER project; and examples for (3) would be: which synthesis theme to give highest priority for developing a center of expertise, new proposals to NSF, staffing change/additions decisions.

As discussed earlier one of NIMO’s goals will be to move away from a system in which common tasks are duplicated at each site. We envision the evolution of a sharing and collaboration economy, similar to other peer-to-peer services (or ‘sharing economy’, Wikipedia) evolving today which enable distribution and reuse of service capacity. We fully realize that the institutional structure with overhead incurred by small, numerous subcontracts or invoices has made work-sharing across LTER sites difficult. Hence, we will explore alternative units of work-shares that are equitable and task-based, learning from the open source software development community in which contributions are based on a win-win situation (Fogel 2005). Within the context of creating services, the IMC plans to explore possible models to distribute tasks according to skills and interest, promote and manage community specialization, and mitigate the impact of staff turnover. An ‘exchange of skills’ will require a system of service or time credit that minimizes administrative overhead but allows for accounting of contributions and demonstrates benefits for participating sites to prove the win-win situation. We will collaboratively explore existing project management systems that are simple enough to use, but track time spent on tasks and allow for efficient communication between service requester and service provider, plus produce meaningful reports for skill and time exchanges among site IMs.

**III.B Timeline of Activities:**

*January - April 2016*: A supplement proposal has been submitted to enable transition of the NIS from the current LTER Network Office (LNO) to the two new organizations, NCO and NIMO. If funded NIMO PIs will determine best hardware solution for NIMO IT infrastructure (XSEDE, private cloud, or University of Wisconsin resources), collaborate with NCO on technology migration from LNO and determine respective responsibilities for different aspects of the current NIS and their interactions.

*Year 1*: Establish OIC, hire project staff, review legacy data products and other NIS components and decide on appropriate actions, evaluate which software/methods used within the LTER Network by more than one site will be supported and promoted by NIMO to make practices more uniform across sites, establish formal skill exchange procedures, establish all means of communication, progress tracking, document organization for NIMO, IMC, centers of expertise, and service requests submission and tracking, establish script repository

*Year 2*: Finalize business plan to offer data curation services to non-LTER projects and invite such projects to contract with NIMO for data curation and storage

*Annually*: Organize IMC/OIC annual meeting and planning session, participate in at least one synthesis working group and establish a center of expertise, centers of expertise develop best practice guides, data management automation and standardization, build up catalog of common tools, support LTER sites in implementing new community approved data management procedures

**III.C Outcomes and Evaluation:**

|  |  |  |
| --- | --- | --- |
| **Goal** | **Outcomes to be Evaluated** | **Responsible** |
| 1. Maintain, improve, and scale the LTER Network Information System functionality to ensure secure long-term availability of environmental data | Secure data archive - strategies are in place and implemented for all LTER site data to be securely stored and backed up. | Servilla, Costa |
| Some LTER data are archived in thematically optimized repositories | LTER IMC, O’Brien, Vanderbilt |
| Existing data products (e.g., Eco-Trends, Clim/HydroDB, StreamChemDB, BiblioDB etc.) are evaluated and either maintained or properly archived. | OIT, LTER IMC |
| LTER controlled vocabulary, unit dictionary, personnelDB, etc. are evaluated, archived, modernized and/or maintained in their current functionality | OIC, LTER IMC, Costa |
| 2. Improve efficiency of the data curation process | More uniform data handling approaches | LTER IMC, Vanderbilt, (O’Brien) |
| Better centralized support for limited number of applications | IM support person |
| Best practices documentation for data handling, quality control, data structure, metadata content | LTER IMC, Vanderbilt, (O’Brien) |
| Curated and/or formally published code libraries | LTER IMC, IM support person |
| Peer-to-peer training of LTER IMs | LTER IMC, Vanderbilt, IM support person |
| IM effort report to sites and available as basis for curation cost estimate | Vanderbilt |
| 3. Close the gap between synthesis science needs and data curation approaches | Establish science and technical centers of expertise which develop all of the above | OIC, LTER IMC, O’Brien, (Vanderbilt) |
| 4. Participate in the longer range eco-informatics visioning | Establish visioning task force and organize working groups at ESIP annual meetings | OIC, LTER IMC |
| Participate in proposal development to continue the dialog of the larger eco-informatics community | Visioning Task Force, OIC |
| Establish collaboration/data exchange with other communities and agencies | OIC |
| Determine cost for a range of research data curation services | LTER IMC, O’Brien |
| Publish datasets from non-LTER research projects curated by NIMO | LTER IMC, O’Brien |

**IV. Broader Impacts**

Several products from this project will have wider ranging impacts aside from the fact that making LTER data publicly discoverable and accessible already has and will continue to have a major impact on synthesis research well beyond LTER. The centers of expertise will contribute best practice documentation and data processing scripts which will be published and made available to the larger community well curated in appropriate repositories. Last but not least, this project will contribute in a major way to estimating the cost of curation of environmental data and suggest a path forward to capturing more data for future reuse.

**V. Sustainability Plan**

This project is aimed at streamlining LTER data management through closer collaboration of LTER IMs with the goal of supporting synthesis science and expanding data curation services to non-LTER projects while maintaining infrastructure for a central data repository. As such no large software development is proposed that needs to be sustained into the future. This project will contribute a cost schedule for curation services for non-LTER projects that will be available to be implemented into a business plan for a curation center. It is assumed that LTER will continue to fund internal data management.

**VI. Results from Prior NSF Support**

**Corinna Gries** has experiences in directing software development and large collaborations in the area of natural history collections informatics. **SYMBIOTA, A Virtual Flora Model for the Southwestern United States**. (ABI 0743827, $0.7M, 2010 - 2014) SYMBIOTA is currently widely used by Thematic Collections Networks funded by the NSF ADBC program as an open source software package for managing the national effort of mobilizing natural history collections data. She is leading **Collaborative research: North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change** (ADBC TCN 1115116, $ 4M, 2011 - 2016). This project has succeeded to digitize 2.3 million specimens of lichens and bryophytes from over 65 institutions. Information is available in a SYMBIOTA portal and through the National Collections Resource iDigBio. She is also part of other ADBC TCN and PEN projects. Most closely related to this proposal is **Collaborative Research: ABI Development: A toolbox for analysis of long-term ecological dynamics using the Kepler Workflow System** (ABI 1262458, $0.9M, 2013-2016) in which synthesis science workshops were organized resulting in the publication of the R package on CRAN, several scientific papers using this package and code to manage workflow scripts along with data sets in DataONE. Successful procedures and experiences map directly onto the here proposed approaches. The recently funded project **EAGER: Collaborative Research: Conceptualizing sustained environmental information management in the landscape of current and emerging eco-informatics infrastructure** (EAGER 1500306, $0.17M, 2015 - 2016; **O’Brien,** co-PI) has started the visioning dialog among the larger eco-informatics community on which the here proposed Task Force will build. One workshop was held in 2015 and the outcome is documented on the ESIP website.

**Paul Hanson**, Award: **CDI-Type II: New knowledge from the Global Lake Ecological**

**Observatory Network (GLEON)** (DEB-0941510; $1,381,000; 08/2009-09/2015). **Intellectual Merit:** Cyberinfrastructure in support of data transfer and storage and data interpretation in lake sensor networks has been developed and is in use by many lake observatories. Additional technologies supporting big data analyses include: numerical simulation (GLM-AED) to study phytoplankton in lakes, carbon cycling, control over mixing and gas flux, and climate change scenarios; model innovations for lake metabolism; and IT trends in ecology. Carey et al. 2012; Hanson et al. 2011; Hanson et al. 2014; Kara et al. 2012; Langman et al. 2010; Porter et al. 2011; Read et al. 2012; Read et al. 2014; Solomon et al. 2013; Staehr et al. 2010; Trolle et al. 2011; Tsai et al. 2011; plus >30 publications. **Broader Impacts:** Training workshops on data management and distributed computing, 4 graduate students, 2 postdocs supported.

**Mark Servilla,** Award: **Advancing Cyberinfrastructure for Accessing Ecological Data** ( DEB-1443175; $174,859; 2014-2015). **Intellectual Merit:** New informatics tools and resources have the potential to advance our understanding of how biological processes interact over multiple spatial and temporal scales to structure natural communities and ecosystems. This project enhanced the reuse and integration of long-term ecological data by increasing its accessibility. Two complementary activities have stabilized access to long-term ecological research data through the Provenance Aware Synthesis Tracking Architecture (PASTA) data repository. Data packages (data plus metadata) uploaded into PASTA meet high standards of quality and integrity consistent with community standards. We implemented Apache Solr as a scalable performance-based metadata search engine and deployed it into PASTA as a replacement for the aging codebase of Metacat. Both performance and stability increased by removing older technology that resulted in search 'bottlenecks' and replacing it with faster search algorithms. The second task was the completion and implementation of a new PASTA Member Node for use with DataONE. This new node improves usability of long-term ecological data by disseminating data products through a global data network. Data resources in the PASTA Member Node have been synchronized with DataONE and the DataONE Coordinating Node metadata archive tested to ensure all available data packages, including metadata and data, have been registered in DataONE. **Broader Impacts:** These improvements have increased opportunities for synthesis and improved understanding of natural ecosystems by improving access to high quality, long-term data. Data in the PASTA repository are readily accessible by researchers from a broad range of disciplines, educators, students, policy makers, and managers. Pending: **ABI Development: Enabling broad-scale ecological analysis and synthesis through PASTA Plus, the enhanced LTER data repository** (ABI-1565103; $1,072,573; 07/2016-06/2019). **Intellectual Merit:** PASTA+ will extend identity management for inclusion of the Long Term Research in Environmental Biology (LTREB), Macrosystems Biology (MSB), and Organization of Biological Field Stations (OBFS) communities, streamline data documentation through improved interface design and conversion of science metadata standards into the Ecological Metadata Language, and expand rules for quality checking uploaded data packages to the standards of these communities. **Broader Impacts:** By adding data from three new research communities to PASTA+, we will greatly expand the geographic scope of the repository and thus, its value to society. The improved analytical power resulting from expanding the pool of accessible data will increase society’s capability to predict and mitigate future changes in ecosystem services, climate, and biodiversity. Through eight data immersion workshops, we will provide support to senior and early career scientists and students to document data and make it accessible through the PASTA+ data repository.

**Margaret O’Brien**, with Matthew Jones (LPI), Mark Schildhauer, Shawn Bowers, Joshua Madin, Award: **Semantic Enhancements for Ecological Data Management** (DBI-0743429, $599,999; 08/2009 – 07/2014). **Intellectual Merit**: Produced cyberinfrastructure for semantic annotations of measurement, including three open source software tools available in public repositories: semantic indexing and search extensions for Metacat, extensions for creating annotations in Morpho, semi-automated annotation generator. **Broader Impacts**: Clarified the problems of cascading imports and the implications for knowledge modeling and ontology creation in general, and especially for modular ontologies such as OBOE. Trained nine graduate students. The Metacat semantic search system is nearing production in DataONE.

**REFERENCES:**

Abrams, S., P. Cruse, J. Kunze, and C. Minor 2011. Journal of Digital Information , Vol 12 , No 2 ( 2011 ) Curation Micro-Services : A Pipeline Metaphor for Repositories, 12(2), 2–5. Retrieved from <https://journals.tdl.org/jodi/index.php/jodi/rt/printerFriendly/1605/1766>

Abrams, S., P. Cruse, J. Kunze, and M. Mundrane 2012. Total Cost of Preservation ( TCP ): Cost Modeling for Sustainable Services, 1–20. Retrieved from <https://confluence.ucop.edu/download/attachments/163610649/TCP-cost-price-modeling-for-sustainable-services-v2_2_1.pdf?version=1&modificationDate=1397258795000>

4C Project, 2015. Collaboration to Clarify the Cost of Curation. <http://4cproject.eu/>

Baker, K. S., and C. L. Chandler 2008. Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics. Deep Sea Research Part II: Topical Studies in Oceanography, 55(18-19), 2132–2142. <http://doi.org/10.1016/j.dsr2.2008.05.009>

Beagrie, N., and J. Houghton 2014. The Value and Impact of Data Sharing and Curation A synthesis of three recent studies of UK research. Retrieved from [http://repository.jisc.ac.uk/5568/1/iDF308\_-\_Digital\_Infrastructure\_Directions\_Report,\_Jan14\_v1-04.pdf](http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf)

Brown, S., R. Bruce and D. Kernohan, 2015. Directions for Research Data Management in UK Universities, JISC. <https://www.fosteropenscience.eu/sites/default/files/pdf/1240.pdf>

Consortium of Universities for the Advancement of Hydrologic Sciences Inc. (CUAHSI) - <https://www.cuahsi.org/opensourcedevelopment>

D-Space (<http://www.dspace.org>)

DEIMS <https://www.drupal.org/project/deims>, <https://github.com/lter/deims>),

EPA STOrage and RETrieval (STORET) Data Warehouse - <http://www3.epa.gov/storet/>

Fogel, K., 2005. Producing open source software. pp. 279

Green, T. 2009. We Need Publishing Standards for Datasets and Data Tables. OECD Publishing White Paper, OECD Publishing. doi: 10.1787/603233448430

Gries, 2010. Drupal Developments at the LTER Network. Databits, Spring. Resource at: <http://databits.lternet.edu/spring-2010/drupal-developments-lter-network>

Hale, S. S., A. H. Miglagese, M. P. Bradley, T. J. Belton, L. D. Cooper, M. T. Frame, … , B. G.Peterjohn, (2003). Managing troubled data: Coastal data partnerships smooth data integration. Environmental Monitoring and Assessment, 81, 133–148.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* **11**: 156–162. <http://dx.doi.org/10.1890/120103>

Hampton, S. E., S. S. Anderson, S. C. Bagby, C. Gries, X. Han, E. M. Hart, M. B. Jones, W. C. Lenhardt, A. MacDonald, W. K. Michener, J. Mudge, A. Pourmokhtarian, M. P. Schildhauer, K. H. Woo, and N. Zimmerman. 2015. The Tao of open science for ecology. Ecosphere 6(7):120. http://dx.doi.org/10.1890/ES14-00402.1

Henshaw, D.L., W.M. Sheldon,, S.M. Remillard and K. Kotwica, 2006. CLIMDB/HYDRODB: a web harvester and data warehouse approach to building a cross-site climate and hydrology database. In: Proceedings of the 7th International Conference on Hydroscience and Engineering (ICHE-2006); Philadelphia, PA. Philadelphia, PA: Drexel University, College of Engineering: Available: <http://hdl.handle.net/1860/1434>

Hernandez, R. R., M. S Mayernik., M. L. Murphy-mariscal, and M. F. Allen, 2012. Advanced Technologies and Data Management Practices in Environmental Science: Lessons from Academia. BioScience, 62(12), 1067–1076. http://doi.org/10.1525/bio.2012.62.12.8

iRods (<http://irods.org>)

Kervin, K. E., R. B. Cook, and W. K. Michener, 2014. The Backstage Work of Data Sharing. Proceedings GROUP ‘14. Proceedings of the 18th International Conference on Supporting Group Work, pp.152-156, <http://dx.doi.org/10.1145/2660398.2660406>

Knapp, A.K., M.D. Smith, S.E. Hobbie, S.L. Collins, T.J. Fahey, G. Hansen, D.A. Landis, K. La Pierre, J. Melillo, T.R. Seastedt, G.R. Shaver and J.R. Webster. 2012. Past, present and future roles of long-term experiments in the LTER Network. BioScience 62: 377-389.

Kratz J. and C. Strasser 2014. Data publication consensus and controversies *F1000Research* 2014, **3**:94 (doi: [10.12688/f1000research.3979.3](http://dx.doi.org/10.12688/f1000research.3979.3))

Lohr, S., 2014. For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights. The New York Times. <http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-toinsights-is-janitor-work.html?_r=0>

Metacat (<https://www.dataone.org/software-tools/metacat>)

Michener, W. K., j. Porter, M. Servilla, and K. Vanderbilt 2011. Long term ecological research and information management. Ecological Informatics, 6(1), 13–24. <http://doi.org/10.1016/j.ecoinf.2010.11.005>

Michener, W. K., and M. B. Jones 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology & Evolution, 27(2), 85–93. <http://doi.org/10.1016/j.tree.2011.11.016>

Obama, B. 2013. Executive Order—Making open and machine readable the new default for government information. The White House.

Open-source Project for a Network Data Access Protocol (OPeNDAP) - <http://www.opendap.org/>

Peer-to-peer services - <https://en.wikipedia.org/wiki/Social_peer-to-peer_processes> or sharing economy - <https://en.wikipedia.org/wiki/Sharing_economy>

Peters, D. P. C. 2010. Accessible ecology: synthesis of the long, deep, and broad. Trends in Ecology & Evolution, 25(10), 592–601. <http://doi.org/10.1016/j.tree.2010.07.005>

Peters, D.P.C., C.M. Laney, A.E. Lugo, S.L. Collins, C.T. Driscoll, P.M. Groffman, J.M. Grove, A.K. Knapp, T.K. Kratz, M.D. Ohman, R.B. Waide, and J. Yao (eds). 2013. Long-term trends in ecological systems: a basis for understanding responses to global change. USDA Agricultural Research Service. Washington, D.C. 378 pages.

Porter, J. H. 2010. A Brief History of Data Sharing in the U.S. Long Term Ecological Research Network. Bulletin of the Ecological Society of America, 14–20.

Reichman, O. J., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. Science 331:703–705.

Robertson, G.P., S.L. Collins, N. Brokaw, H.W. Ducklow, D.R. Foster, T.L. Gragson, C.

Gries, S.K. Hamilton, A.D. McGuire, J.C. Moore, E.H. Stanley, R.B. Waide and M.W. Williams. 2012. Long-term Ecological Research in a human-dominated world. BioScience 62: 342-353.

Rüegg, J., C. Gries, B. Bond-Lamberty, G. J. Bowen, B. S. Felzer, N. E. McIntyre, P.A Soranno , K.L Vanderbilt, K.C. Weathers, 201). Completing the data life cycle: using information management in macrosystems ecology research. Frontiers in Ecology and the Environment, 12(1), 24–30. <http://doi.org/10.1890/120375>

Rumble, J., B. Carroll, G. Hodge, and L. Bartolo 2005. Developing and Using Standards for Data and Information in Science and Technology. In Ensuring Long-term Preservation and Adding Value to Scientific and Technical data (PV 2005) The Royal Society, Edinburgh 21-23 November 2005 (pp. 1–13). Retrieved from <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/025.pdf>

San Gil, Inigo, et al. "Case studies of ecological integrative information systems: the Luquillo and Sevilleta information management systems." *Metadata and Semantic Research*. Springer Berlin Heidelberg, 2010. 18-35. DOI: [10.1007/978-3-642-16552-8\_3](http://link.springer.com/chapter/10.1007/978-3-642-16552-8_3)

San Gil, Inigo. “The Drupal Ecological Information Management System (DEIMS): recent progress and upcoming challenges for a grassroots project. *Databits.* Fall Edition. Resource at <http://databits.lternet.edu/fall-2011/drupal-ecological-information-management-system-deims-recent-progress-and-upcoming-challen>

Sheldon, W.M. Jr. 2008. Dynamic, Rule-based Quality Control Framework for Real-time Sensor Data. Pages 145-150 in: Gries, C. and Jones, M.B. (editors). Proceedings of the Environmental Information Management Conference 2008 (EIM 2008): Sensor Networks. Albuquerque, New Mexico. <http://gce-lter.marsci.uga.edu/public/files/pubs/wsheldon_dynamic_qc_eimc2008_final.pdf>

Sheldon, W.M. Jr. 2010. GCE Data Toolbox for MATLAB website. (<https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox>)

Uhlir, P. F. 2012. The Future of Scientific Knowledge Discovery in Open Networked Environments: Summary of a Workshop. National Academy of Sciences. Retrieved from <http://www.nap.edu/catalog.php?record_id=18258>

USDA ARS Sustaining The Earth’s Watersheds - Agricultural Research Database System (STEWARDS) - <http://www.ars.usda.gov/Main/docs.htm?docid=18622>

USGS Office of Water Information (OWI) - <http://water.usgs.gov/owq/>

USGS National Water Information System (NWIS) - <http://waterdata.usgs.gov/nwis>

USGS Water Quality Portal <http://www.waterqualitydata.us/>

Van Noorden, R. 2013. Open access: The true cost of science publishing. Nature, 495(7442), 426–9. <http://doi.org/10.1038/495426a>

Wilson, A., R. Downs, C. Lenhardt, C. Meyer, W. Michener, H. Ramapriyan, and E. Robinson 2015. Realizing the Value of a National Asset: Scientific Data. Eos Trans. AGU, 95(50), 478–478. http://doi.org/10.1002/2014EO500006