Collaborative Development of a Project Database for LTER Sites
Margaret O'Brien (SBC) and Corinna Gries (CAP)
15 July 2008

INTRODUCTION: Improvements to all sites' cyber infrastructure will be necessary to facilitate the research objectives of the LTER decadal plan. Although the Cyber Infrastructure Implementation Plan is not yet written and approved, the general agreement is that it will be highly modular and based on principles of Service Oriented Architecture (SOA) with the Ecological Metadata Language (EML) used as a data exchange standard. Only this approach will guarantee the highest degree of interoperability and full integration of diverse legacy systems at the sites. In addition, SOA allows for incremental implementation. Here we propose a first step towards this integrated network level cyber infrastructure, specifically the management and documentation of discrete research projects at the sites or across sites.

Most sites feel the need to document research projects for annual reporting, field site permissions, coordination of research activities in space and time etc., and various systems are in place. However, to accommodate cross-site and network level research activities these documentation systems must be more flexible and accessible from all involved sites or centrally from the LNO website, and scientists must be able to browse scientific projects in a consistent format. It will soon become necessary to link both individuals and research products (e.g. datasets and publications) to the research themes of the decadal plan.

We propose to develop infrastructure modules to manage the descriptions of scientific research projects based on the above outlined principles for 'Network Level Information Technology.' Our approach will mesh with NISAC's CI Implementation plan, will accommodate existing legacy systems and at the same time, provide a solution to local needs and a basis for extending functionality to implement the above mentioned uses of this information.

OVERVIEW: Figure 1 shows the schematic of components to be used and developed by this project: The information for each research project is stored in a database which is accessible by a web-service. This web-service returns the data in the format of an EML file, which web-clients receive. A transformation using XSLT technology and formatting according to site-specific requirements will produce a web display of project information.

DESCRIPTION: This project will use EML as the exchange standard. EML is highly regarded as a format for data descriptions and has been in use in the LTER for five years. The EML family of schemas is quite extensive, and its scope and modularity make it appropriate for many other types of information exchange as well. Many of these other uses have not yet been widely explored, although work is progressing on bibliographic databases using EML at the site level and at the LNO. By nature, all EML schemas are extensible, and so additional features can be added if necessary to accommodate the themes of the LTER decadal plan or to assist integration of this database with other collections.

Fig. 1: Schematic of Components

The EML Project files will be housed in the open source, native XML database eXist. These databases may be deployed centrally at the LNO or at sites interested in hosting them locally. eXist was chosen for its
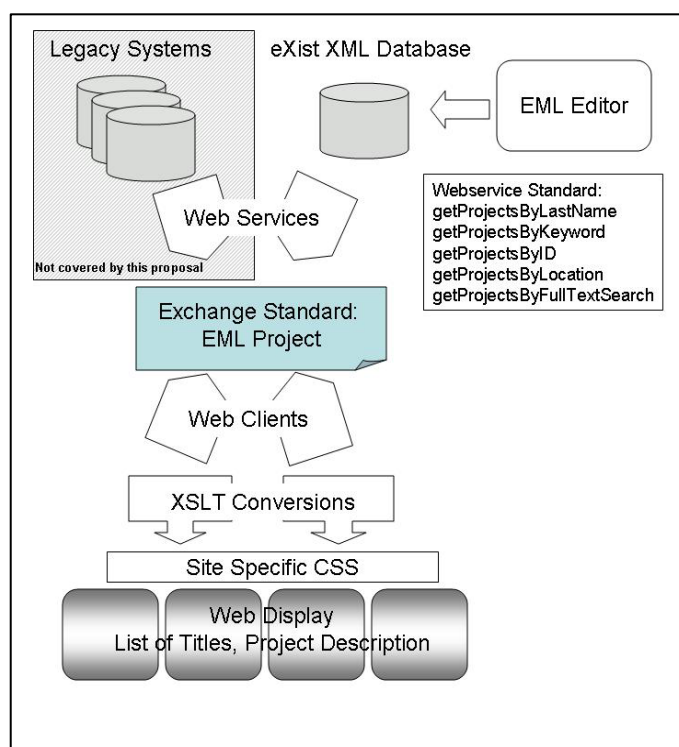
ease of installation and maintenance, and the availability of third party management tools. Information stored in eXist may be maintained via an online EML editor, which is currently being developed by CAP and LNO, providing online input forms and simple user access management.

A web-service will be developed (example methods are listed in Figure 1) which will return pertinent EML project files from eXist. As Figure 1 shows, eXist is not the only means of storing research project information. Legacy databases (i.e., sites with existing databases for projects) can participate via their individual custom web services, although programming for these is not covered in this proposal.

This project will develop a series of web-clients in a variety of programming languages to accommodate commonly used web systems (i.e. Java, ASP Dot Net, Perl). These web-clients will retrieve project descriptions when certain conditions are matched, e.g., a list of projects at a particular location. The client will employ a two part transformation - a generic XSLT transformation from XML to HTML, with individual sites' cascading style sheets (CSS) producing a human-readable web display with site-specific layout and color scheme.

SPECIFIC WORK PLAN: We have polled all site IMs as to the degree they would like to be involved in this project, either to define functionality, implement and test a final product and provide usability feedback, or to write code. 16 IMs are interested in the first two aspects and 10 chose to participate in the code writing effort.

Two workshops are planned with six or seven information managers attending each. The first workshop will be two days long and will develop use case scenarios, define the methods the web-service needs to expose, and if needed extend the EML Project schema to accommodate requirements. IMs who have indicated interest in this part of the process will be invited. The second workshop will comprise an entire work week (5 days) and will concentrate entirely on programming. We believe that writing code at a central location will be the most efficient way to accomplish this project since participants will avoid the common distractions at their home institutions. The LNO training lab would be ideal. The products we are expecting from this project are: one web-service accessing the eXist database, several web-clients retrieving the information from the web-service, one XSLT program transforming XML into HTML, and several CSS for participating sites. The workshop will also produce a functioning installation of the prototype at LNO. We expect to complete both workshops during the period Fall 08 – Spring 09.

CONCLUSION: Collaboration among so many information managers provides us with a wide variety of use cases, experiences and particularly, expertise, which will allow us to employ the different technologies required. The current list of participants (below) has a combined high level of expertise in EML, XML, XSLT and web development in Java, ASP.Net and Perl which will ensure success. The level of participation at the recent training workshop on XML technologies means that most sites are now equipped to participate in a project of this type. This project represents a new model of collaborative development by the Information Managers Committee to address a common need, as opposed to the development of individual solutions or the establishment of a central, top-down repository with mandated contributions. The design will maintain individual site identity, but the common look-and-feel will promote a network identity. Working collectively on a modular project allows each participant to further his/her expertise in a chosen area, rather than relying on the 1 IM:1 site model in which one person covers all bases – a situation known to be increasingly difficult as internet technology grows. By working in a new collaborative model, this project will help define standards and practices for developing network level cyber infrastructure.

PRELIMINARY LIST OF PARTICIPANTS: John Campbell (HBR), Barbara Benson (NTL), Wade Sheldon (GCE) Margaret O'Brien (SBC), Corinna Gries (CAP) Ken Ramsey (JRN), Kristin Vanderbilt (SEV), John Porter (VCR), Todd Ackerman (NWT), Suzanne Remilliard (AND). The two workshops will have slightly different participants and at least two IMs will participate in both.

BUDGET JUSTIFICATION: We are requesting $18,600 for a total of 14 person-trips to Albuquerque/LNO and 63 nights and per diems (workshop 1: 7 people for 3 nights each, and workshop 2: 7 people for 6 nights each). Airfare is based on July 2008 levels plus ~25%. Current lodging estimate and per diem costs were obtained from LNO travel personnel.

WORKSHOP BUDGET: Collaborative Development of a Project Database for LTER Sites

	Workshop 1 2 days, 7 participants	Workshop 2 5 days, 7 participants
Airfare and/or ground transportation	$3,600	$3,600
Hotel (workshop + 1day)	$2,771	$5,542
Per diem (workshop + 1day)	$1,029	$2,058

TOTAL REQUEST: $18,600

Support Information from NISAC:

I think this project could help define standards and practices for managing, displaying and sharing information on active and legacy research projects at LTER sites. If common data structures and web interfaces are widely adopted for managing project information, minimum effort would be required to provide cross-listing and search services on LNO and site web pages to enhance discovery of related research and data. For example, project information could be synchronized and displayed automatically on SiteDB pages, and sites could deploy community-developed web applets or code to allow visitors to search for research projects of interest at their site or across the network. Project metadata support services (e.g. EML generation) could also be developed to simplify metadata creation at sites.

Identifying common IM needs and working to develop shared solutions is critical to moving network-level CI forward in LTER, and this is an excellent opportunity to use a community development approach towards these ends.

Wade Sheldon
NISAC Co-chair