**Develop an LTER NIS Best Practices for Designing and Writing Workflow Scripts in the PASTA Framework**

Corinna Gries, Mark Servilla, Wade Sheldon, Ben Ruddell

**Summary Statement**

The PASTA Framework infrastructure of the LTER Network Information System (NIS) will support the automated execution of site provided scripts to simplify the process of creating and updating LTER synthetic and derived data products. Scripts will be required to meet specific criteria to operate successfully in the PASTA Framework environment. Such criteria will necessitate training for information managers to understand the script execution environment and how to best write a script for the PASTA Framework.

We propose to develop an LTER NIS Best Practices for writing scripts in the PASTA Framework by (1) developing a set of example scripts in a variety of supported languages, (2) documenting our experience in a best practices document, and (3) developing an agenda for a "train the trainer" approach to teach PASTA Framework script writing best practices to site Information Managers, students, and scientists. Understanding how to develop scripts for the PASTA Framework is critical for sites to participate in Network-level projects, like ClimDB, HydroDB, and EcoTrends, and future projects that are yet to be designed.

**Scope of work**

Development of the PASTA Framework began in early 2010 with the initial focus on three core services of the framework: (1) the Metadata Management Suite, (2) the Data Management Suite, and (3) the Workflow Management Suite. Together, these three services will support automated metadata and data harvesting, including the automated execution of workflow scripts by the Workflow Manager component of the Workflow Management Suite to create derived and/or synthetic products.

Software development of these three core services is rapidly progressing, and by the third quarter of 2011 real life input, testing, and training for Information Managers will be needed to verify and validate the functionality of these components of the PASTA Framework. One aspect of the PASTA Framework of particular importance to LTER sites is the Workflow Manager, which supports the contribution and execution of pre-written scripts for automating the creation of LTER derived and/or synthetic data products. The script execution environment will require that creators of scripts adhere to certain guidelines when designing and writing the actual script. As such, this working group proposal will focus on the features and capabilities of the Workflow Manager component of the Workflow Management Suite. Specifically, we propose to develop an LTER NIS Best Practices for writing scripts in the PASTA Framework that will be targeted to LTER Information Managers, students, and scientists. This best practices will address three goals:

1. Design and write multiple example scripts in a variety of supported languages to evaluate the Workflow Manager and to provide direct feedback to developers; such scripts will be exemplars for the best practices document.
2. Document the process of writing scripts, including necessary guidelines for the PASTA Framework, and any technical and/or anecdotal information to aid in the design and writing of scripts in the PASTA Framework.
3. Develop an instructional agenda in the form of a “train the trainer” module to accompany the best practices document.

Learning to design and write scripts for the Workflow Manager will require a combination of direct instruction from NIS developers, “hands-on” practice writing and debugging scripts in the PASTA environment, and providing feedback to NIS developers on issues encountered. This effort will be best performed at the LTER Network Office and by a small group of Information Managers with prior experience in programming and data analysis, and who are willing to serve as an initial set of trainers. We propose to develop example scripts with known algorithms for existing derived and/or synthetic data products, such as those in use by ClimDB, HydroDB, and EcoTrends. EcoTrends, for example, uses the R statistical analysis language to transform site-based (Level-0) data into a new derived time-scale product that shares a common data model. We will evaluate the R language scripts used in the current EcoTrends project and modify or rewrite them to meet the requirements of the script execution environment of the PASTA Framework. The use of scripts will also allow us to evaluate other PASTA Framework services, including the use of the Data Management Suite for accessing Level-1 data, the Metadata Management Suite for accessing EML metadata, and the Metadata Factory to create provenance-enabled metadata. In addition, we will replicate the R language scripts algorithms in multiple languages (e.g., Java, PERL, Python, MATLAB) to demonstrate how the same algorithm may be codified in a different language. The result of this exercise will provide direct feedback to NIS developers about their design and implementation of the Workflow Management Suite and related subsystems of the PASTA Framework.

Our experience will be coalesced into an LTER NIS Best Practices document for use by LTER Information Managers, students, and scientists. This best practices document will serve as the authoritative guide to designing and writing scripts for the PASTA Framework and will be available through multiple media (e.g., printed document, indexed website content, and/or video). Updates to this document will occur on a regular basis if changes or advances are made to the script execution environment or if new best practices are identified in script design and writing.

Finally, both the script writing and best practices document will serve as the basis for an instructional agenda to be used as a “train the trainer” module as part of the best practices document. We believe that teaching the nature and complexity of script design and writing will be more effective and efficient if taught to the LTER community by a small group of trained Information Managers. This group will be able to perform instruction at more convenient dates and venues, thereby increasing the chance of higher attendance and learning. A training proposal will be submitted to the EB for the session to be held next year as well..

**Participants and Budget Justification:**

Funds are requested for 6 people to travel to the LTER Network Office in Albuquerque. The working group will meet with the NIS developers for 5 days during which several workflow scripts will be developed. Participants will include the current Tiger Team members (Sheldon, Gries, Ruddell) plus three others with experience in workflow development and/or skills in relevant scripting languages.

**Budget:**

**Lodging:** ($ 120 6 people, 4 nights) $ 2880

**Meals:** ($60 for 6 people, 5 days) $ 1800

**Transportation:** ($ 700 for 6 people) $ 4200

**Total: $ 8880**