SGS-LTER 2012 Information Management Supplement Request

Introduction

The SGS-LTER site will be de-commissioned February 2014 after over 30 years of operation. The project is in year 2 of its final 3 years of support. Making data available to scientists and other end-users is an SGS-LTER site priority during its final round of funding. The aims of Information Management at the SGS LTER site are to improve data access and metadata documentation in accordance with the recommendations in the 2012 LTER revised guidelines (network communication, LTER Executive Board). We have made significant progress to this end in the past two years. However, three significant decisions made by the LTER community have occurred that will affect the quality of the work we had proposed. First, the IM community changed the ecological metadata language (EML) standard from EML 2.0.1 to EML 2.1.0. Second, the community has asked that data now be compliant with the emerging Provenance Aware Synthesis Tracking Architecture (PASTA) Framework. Third, the LTER Executive Board has asked that all datasets possess five essential features in an effort to standardize information management across the network. This request presents four goals that will allow us to meet our original aims as well as the new demands listed above.

Goal 1 - Improve Data Access: Complete the delivery of Level 5 EML 2.1.0 compliant metadata and data of SGS-LTER core datasets as data packages through the LTER Data Portal at the LTER Network Office while satisfying existing best practices and standards for the LTER Network.

Goal 2 - Preservation of SGS-LTER Legacy: Augment the data packages developed through Goal 1 with additional supporting documentation that reflect the legacy of the site and studies.

Goal 3 - PASTA Compliance: Ensure that SGS-LTER data packages are interoperable with the network infrastructure and the emerging PASTA Framework.

Goal 4 - Network Information System (NIS) Participation: Extend our involvement in the LTER NIS development initiatives by participating in the broader information management community as a member of the LTER Network.

By meeting these goals we will provide an archive of valid Level 5 EML 2.1.0 with a direct link to downloadable data of SGS-LTER science. We propose to work with local SGS-LTER scientists and staff, and LTER Network scientists and staff to improve the quality and quantity of available SGS-LTER related data, metadata and related products. This work will help preserve the legacy of the site. The resulting data archive will ensure that the research on the shortgrass steppe will be accessible to the LTER network and broader community. To meet these goals, we are requesting funds to continue support for our Information Management (IM) Team. Our IM Team is co-lead by Nicole Kaplan (Co-PI of the SGS-LTER) who serves as the information manager and database administrator, and Bob Flynn who serves as the GIS and IT manager. Both have over 15 years experience working with the SGS-LTER data management system.

Background and Justification

The SGS-LTER project has 150 datasets produced during our 30 years of research (1982-present). The proposed work will result in EML 2.1.0 and PASTA-compliant data packages (e.g., data and metadata) to be incorporated into the LTER Metadata Catalog and NIS and made available through the LTER Data Portal. Generation of the Level 5 EML for the network requires the following steps:

Step 1 - the data are vetted and entered into a digital file (e.g., spreadsheet, text file, etc.).
Step 2 - the digital form is imported in a relational database management system (RDBMS).
Step 3 - metadata documentation is obtained and submitted to the RDBMS to support the dataset with LTER best practices and standards employed.
Step 4 - scripts written in Perl and/or XSLT are developed to organize datasets and metadata into data packages and generate EML 2.1.0 from a centralized local data repository.

Table 1. Status of SGS LTER core and supplemental datasets and data packages to be served through the LTER Data Portal.			
Dataset Description	Quantity	Status	Proposed Work
Core SGS-LTER	56	In the RDBMS; Harvested EML 2.1.0 Level 5	Ensure new 2012 LTER Guidelines from the LTER EB are implemented
Core SGS-LTER	39	In the RDBMS; EML Errors	Ensure content is valid and EML generation programs are debugged
Supplemental SGS-LTER GIS Layers	19	In the RDBMS; Digital data and metadata	Generate EML
Core and Supplemental SGS-LTER	36	In the SGS-LTER IM; No EML Level 5	QAQC Data, document metadata, submit to RDBMS, and generate EML

Within each step, rigorous quality assurance and quality control (QAQC) measures are implemented. A summary of our progress on submitting valid EML 2.1.0 packages to the LTER Metadata Catalog and efforts to create new packages are provided in Table 1.

The SGS-LTER possesses 150 core datasets assembled from 1982-present. All 150 datasets are organized in digital data and metadata 'packages' within the SGS-LTER Information Management System (Step 1). In December 2011, 114 of these 150 core datasets (76%) were compliant with the EML standard at that time. With the subsequent decision to change from EML 2.0.1 to EML 2.1.0, the SGS-LTER IM staff re-evaluated each the datasets for compliance to the new standard. Currently 56 of these 150 core datasets (37%) have been successfully inserted into the LTER Metadata Catalog fully compliant with the Level 5 EML 2.1.0 standard. Of the remaining core datasets, 58 have been loaded into the RDBMS, 39 of which were found to possess errors that will require individual debugging, and 19 of which are GIS layers that will require special scripts for creating valid EML 2.1.0. The remaining 36 data sets will require addition QAQC before being loaded into the RDBMS and beyond.

We are requesting funds to retain some of our current staff so that we can complete processing of our remaining datasets and ensure compliance with the LTER Data Portal and compatibility with LTER data in the NIS. This includes Level 5 EML 2.1.0 and PASTA compliance. In addition to our datasets, we have extensive sample archives that we would like to link to the datasets in the relational database management system within the local repository. We are requesting funds to retain additional staff to complete the inventory of our sample archives so that we can link them with our long-term datasets, publications, their archive location, and the experimental site.A timeline of proposed IM activities and responsible staff members are provided in Table 2.

Plan of Operation

Goal 1 – Improving Data Access

We will deliver Level 5 EML 2.1.0 compliant metadata and data for all 150 SGS-LTER datasets through the LTER Data Portal. The datasets will satisfy existing best practices and standards for the LTER Network. Our efforts will address both the quality and quantity of data and metadata served online by inserting new data packages and appending or updating data and metadata content in our database. Each

Table 2. Timeline of proposed work and responsible staff.			
	SGS IM Work Activity	**Timeline**	**Responsible Staff**
Goal 1	**Increase Access, Quality, and Quantity of Data Packages**		
	Correct errors in core datasets	*7/2012 - 12/2013	**Kaplan, Flynn**
	Correct errors in GIS datasets	*7/2012 - 11/2012	**Flynn**
	Perform data QAQC and integration	11/2012 - 3/2013	**Kaplan, PIs**
Goal 2	**Preservation of SGS-LTER legacy**		
	Collect and curate supplemental metadata associated with flagship SGS-LTER datasets	9/2012 - 6/2013	**Kaplan**
	Organize and inventory sample archive	1/2013 - 6/2013	**Sprague, Perkins**
	Link archive inventory to SGS-LTER data products	4/2013 - 6/2013	**Sprague, Perkins, Kaplan**
	Transfer knowledge, IM tools and archive to local partner	5/2013 - 12/2013	**Kaplan, Flynn**
Goal 3	**PASTA Compliance**		
	Ensure that data sets possess the essential features defined by the LTER EB	*6/2012 - 10/2012	**Kaplan, Flynn**
	Improve EML automatic and batch generation scripts to EML for spatial and non-spatial data.	*6/2012 - 12/2012	**Kaplan, Flynn**
	Affirm Level 5.0 EML 2.1.0 compliance and valid with LTER NIS standards	4/2013 - 12/2014	**Kaplan, Flynn**
Goal 4	**Participate in the NIS**		
	Participate in NIS ASM workshops	9/2012 - 7/2013	**Kaplan**
	Contribute to NIS databases and new data integration databases	8/2012 - 10/2012	**Kaplan**
All Goals	**Oversight of Supplement Activities**		
	Ensure activities are progressing as proposed	Throughout	**PIs, Sprague**
*Ongoing activities that will be expanded and enhanced with supplement funds			

dataset will be checked for errors and corrected. We will revise XSLT scripts to ensure that the metadata and data meet the Five Essential Components as defined by the LTER Executive Board, are configured to be PASTA-ready, are validated as EML 2.1.0 packages, and reflect the full body of knowledge created by SGS-LTER researchers over the past 30 years.

Through our participation in discussions of the LTER Data Synthesis Prospectus over the past year, we have refined our inventory of site datasets with additional information on the status of EML and comments on challenges to move forward. The challenges to creating data packages for the LTER Metadata Catalog include the need for scheduling time with SGS-LTER scientists to obtain additional information to comply with EML requirements, standardization of attribute codes within or across related datasets, quality issues, or lack of digital media. With the additional information in our inventory, we are able to more precisely plan our continued work on improving data quality and metadata documentation in EML. We are requesting supplemental funds to support these efforts to ensure SGS-LTER data packages are of the highest quality for re-use.

Goal 2 - Preservation and Legacy

After 30 years of continuous data collection, the termination of the SGS-LTER project presents a unique opportunity to embody innovative pathways to ensure accessibility of data, information, samples and other products associated with the SGS legacy. Many of the founding members of the SGS-LTER are still active. These individuals, the scientists, and the support staff who followed represent an important pool of knowledge. We will collect supplemental metadata and supporting documentation to create an invaluable archive of SGS-LTER data packages for future generations. This will include well-documented data, an archived sample inventory, as well as current dynamic websites and a backend database used to manage and serve data packages and other information. Additionally, the SGS-LTER has had a long-standing relationship with the USDA-ARS and Colorado State University. We are well-positioned to work with our local partners and the Network to maintain access to data and distinctive information about the research, as well as the associated sample archive. We need to inventory this archive of samples that are associated with datasets and published findings as well as known research locations and treatments.

We are requesting funds to support experienced project staff to establish the sample inventory and link the items to our data in a way that will provide access for future researchers on the SGS or across the LTER network. We also have an opportunity to work with Dr. Helena Karasti (University of Oulu, Finland), a social scientist who has worked with LTER sites since 2002. Dr. Karasti has obtained support to attend the 2012 LTER ASM in Estes Park, where she will conduct interviews and group discussions with SGS-LTER scientists to elicit details about performing their long-term studies. This supplemental information will serve as an interpretation of research activities that shaped the creation of data packages and the knowledge of the SGS ecosystem. It will help explain changes in methodologies over time and provide information for the end user of data. This will capture underlying assumptions about the conceptual model of the ecosystem that was available at the time of data collection and fortuitous incidents that helped shape new directions in the scientific process over the long-term. We also will share the tacit know-how of conducting field research in the SGS. The SGS-LTER Information Managers will incorporate these narratives into the local data repository and link them to data packages.

Goal 3 - PASTA Compliance

We will generate valid Level 5 EML 2.1.0 data packages that can be integrated into the PASTA framework. The current structure and content of our XML needs to be improved to comply with EML best practices for Level 5 EML 2.1.0, particularly in the placement of URL links to data sets at the entity level. This will allow the Network Data Portal to better serve data packages and automate processing of site data sets with the PASTA framework. We will improve the quality and quantity of datasets available through the LTER Data Portal by ensuring the completeness of our Level 5 EML 2.1.0 metadata content as well as generating EML through Perl and XSLT scripts that are efficient and effective. We are dedicated to improving the functionality of the SGS-LTER system, availability of the SGS-LTER data, richness of the SGS-LTER metadata, and functionality of the LTER NIS as a whole. We understand that the investment we make as a site in the improvement of the quality and quantity of the data will go a long way in illustrating the wealth and accessibility of LTER data and enabling synthesis of SGS-LTER data in Network research.

Goal 4 - Network Information System Participation

The SGS-LTER is in year 2 of a final 3 years of NSF support. The requested funds will allow us to continue to participate in the development of the NIS longer than what was supported in our original budget. We identified datasets where documentation needs to be improved in response to new guidelines from the LTER EB. We will also be available to contribute high quality data packages to cross-site

science endeavors and Network databases (e.g. ClimDB, vegDB, MALS) to ensure that the SGS-LTER participates as experienced LTER scientists and information managers.

We will inspect existing metadata content to improve data set titles, abstracts, and attribute descriptions to meet guidelines that address five essential components, and add LTER-controlled vocabulary keywords where necessary. We also will adopt network standards for URL construction and location to serve SGS-LTER data well into the future. We will continue to work through issues of how to generate EML for SGS-LTER GIS data by contributing to the development of proposed LTER spatial data standards, and exploring how more specific study location data can be incorporated in our EML metadata for long-term core studies. Continued participation by an SGS-LTER GIS representative will allow SGS-LTER spatial data and metadata to be more easily used by Network projects, such as LTERmapS.

This work will enhance the SGS-LTER information management system and the availability of information through the Network. It also will provide a venue for growth and professional development through presentation of different approaches to IM related challenges. We are requesting funds to send our Information Manager to the 2012 American Society for Information Science and Technology Meeting in Maryland to present proposed models for data integration and analysis systems for LTER data and to support preparation of associated materials such as a presentation or a poster.

Budget and Justification

We are requesting $149,999 in supplemental funding. Here we provide a summary of the budget and justification. A more detailed justification is provided in the Budget Justification section. We are requesting 10 months of salary support for Nicole Kaplan to work through the SGS-LTER data inventory, create and improve Level 5 EML 2.1.0 content, create links to other supporting documentation, and work with Dr. Karasti. We also request salary support for an additional 10 months for our GIS manager Bob Flynn to work on spatial data and the EML Generation scripts for valid, PASTA ready Level 5 EML 2.1.0. We are requesting 6 months of support for our project manager, Sallie Sprague, and 8 months for experienced field crew manager, Melissa Perkins, to organize, inventory and document the SGS-LTER sample archive at the SGS Research and Interpretation Center. In addition, we are requesting funds to send our Information Manager, Nicole Kaplan, to the ASIST meeting to disseminate results obtained from analyzing and modeling a cross-site LTER data integration system. We are requesting funds for replacement of one data analysis station to support sample archive inventory and dataset QAQC efforts. We are requesting funds to replace our current server with an up-to-date version that we can transfer to our local data repository partners at the completion of the project to provide continued access to SGS information from a site with a 30-year history.

JUSTIFICATION FOR SUPPLEMENTAL FUNDING

As we decommission the SGS-LTER, we are focused on creating high integrity datasets that comply with the latest directives from NSF and the LTER Executive Board so that our data can be accessed by the LTER Network and greater ecological research community. We have processed approximately half of the datasets that are available from the site and would use the additional funds to retain experienced staff to process the remaining datasets and metadata. We are very concerned that we will lose the 15-year expertise of our data management team as we approach our final year of funding.

SUMMARY PROPOSAL BUDGET

YEAR 1

FOR NSF USE ONLY		
PROPOSAL NO.	DURATION (months) Proposed	Granted
AWARD NO.		

ORGANIZATION
Colorado State University

PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR
John C Moore

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. **John C Moore - PI**	0.00	0.00	0.00	**0**	
2.					
3.					
4.					
5.					
6. (**0**) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	**0**	
7. (**1**) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00	**0**	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (**0**) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	**0**	
2. (**4**) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	34.00	0.00	0.00	**89,005**	
3. (**0**) GRADUATE STUDENTS				**0**	
4. (**0**) UNDERGRADUATE STUDENTS				**0**	
5. (**0**) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				**0**	
6. (**0**) OTHER				**0**	
TOTAL SALARIES AND WAGES (A + B)				**89,005**	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				**21,361**	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				**110,366**	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
Server $ **5,910**					
TOTAL EQUIPMENT				**5,910**	
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				**2,000**	
2. FOREIGN				**0**	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ **0**					
2. TRAVEL **0**					
3. SUBSISTENCE **0**					
4. OTHER **0**					
TOTAL NUMBER OF PARTICIPANTS (**0**) TOTAL PARTICIPANT COSTS				**0**	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				**1,990**	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				**0**	
3. CONSULTANT SERVICES				**0**	
4. COMPUTER SERVICES				**0**	
5. SUBAWARDS				**0**	
6. OTHER				**0**	
TOTAL OTHER DIRECT COSTS				**1,990**	
H. TOTAL DIRECT COSTS (A THROUGH G)				**120,266**	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					
Modified Total Direct Cost (Rate: 26.0000, Base: 114356)					
TOTAL INDIRECT COSTS (F&A)				**29,733**	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				**149,999**	
K. RESIDUAL FUNDS				**0**	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				**149,999**	

M. COST SHARING PROPOSED LEVEL $ **0** AGREED LEVEL IF DIFFERENT $

PI/PD NAME
John C Moore

ORG. REP. NAME*
Neil Shropshire

FOR NSF USE ONLY		
INDIRECT COST RATE VERIFICATION		
Date Checked	Date Of Rate Sheet	Initials - ORG

 ***ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET**

SUMMARY PROPOSAL BUDGET

Cumulative

FOR NSF USE ONLY		
PROPOSAL NO.	DURATION (months) Proposed	Granted
AWARD NO.		

ORGANIZATION
Colorado State University

PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR
John C Moore

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. **John C Moore - PI**	0.00	0.00	0.00	**0**	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	**0**	
7. (**1**) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00	**0**	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (**0**) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	**0**	
2. (**4**) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	34.00	0.00	0.00	**89,005**	
3. (**0**) GRADUATE STUDENTS				**0**	
4. (**0**) UNDERGRADUATE STUDENTS				**0**	
5. (**0**) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				**0**	
6. (**0**) OTHER				**0**	
TOTAL SALARIES AND WAGES (A + B)				**89,005**	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				**21,361**	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				**110,366**	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.) $ **5,910**					
TOTAL EQUIPMENT				**5,910**	
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				**2,000**	
2. FOREIGN				**0**	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ **0**					
2. TRAVEL **0**					
3. SUBSISTENCE **0**					
4. OTHER **0**					
TOTAL NUMBER OF PARTICIPANTS (**0**) TOTAL PARTICIPANT COSTS				**0**	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				**1,990**	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				**0**	
3. CONSULTANT SERVICES				**0**	
4. COMPUTER SERVICES				**0**	
5. SUBAWARDS				**0**	
6. OTHER				**0**	
TOTAL OTHER DIRECT COSTS				**1,990**	
H. TOTAL DIRECT COSTS (A THROUGH G)				**120,266**	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					
TOTAL INDIRECT COSTS (F&A)				**29,733**	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				**149,999**	
K. RESIDUAL FUNDS				**0**	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				**149,999**	

M. COST SHARING PROPOSED LEVEL $ **0** AGREED LEVEL IF DIFFERENT $

PI/PD NAME
John C Moore

ORG. REP. NAME*
Neil Shropshire

FOR NSF USE ONLY		
INDIRECT COST RATE VERIFICATION		
Date Checked	Date Of Rate Sheet	Initials - ORG

C ***ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET**

Colorado State University Budget Justification

Personnel ($89,005): We are requesting funds to retain four staff members to complete data packages, sample archive inventory and linking that information. Funds would support Kaplan (FT) and Flynn (50% time) for approximately ten months, Sprague (75% time) for six months and Johnston (FT) for eight months. Kaplan and Flynn will be producing the data packages in compliance with NSF and LTER Network standards. Perkins and Sprague will be completing sample archive inventory and documentation that will be linked with the data packages.

Fringe Benefits ($21,361): Benefits are calculated at the University's federally-negotiated rates for each employment category.

Travel ($2,000): Funds would support Kaplan's participation in the 2013 ASIST meeting where she will be able to disseminate results obtained from analyzing and modeling a cross-site LTER data integration system.

Materials/Supplies ($1,990): We are requesting funds to replace an aging data analysis station to facilitate processing and QA/QC efforts on the remaining datasets, and to support the sample archive and inventory process. A small request for archive media to facilitate transfer and back up storage of long-term data files is included.

Equipment ($5,910): We are requesting funds to replace our existing server when the project ends and all data files need to be handed to a repository for long-term access. In order to help this entity, we would like to provide the 70 years of SGS LTER research information on a current design server rather than one that will then be a decade old.