An IM toolbox for the present and future to support data synthesis activities.

Bob Flynn, Nicole Kaplan, John Norman, Dianne Beer, & Adam Zeimet
Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado 80523-1170

Abstract

This poster displays four examples of topics and tools currently being implemented and developed by information management staff at the Shortgrass Steppe LTER, across the LTER Network and within USDA Agricultural Research Service (ARS). These tools are necessary in order to synthesize similar data sets from different researchers, agencies, and institutions. These tools were designed provide to support to Principal Investigators, educators, students, and policy makers that need to synthesize information to make better decisions about planning their research, coursework and land management strategies. The four areas to be presented are Geographical Information Systems (GIS) and Remote Sensing, standardization of metadata using Ecological Metadata Language (EML), integration of relational database management systems for different agencies, and the creation of useful dynamic web pages. GIS and Remote Sensing are powerful tools that allow researchers to analyze, model, and predict ecological factors and outcomes that shape the shortgrass steppe by integrating spatial and non-spatial data collected at the field site. EML consists of a number of modules that define an extensible mark-up language (XML) that creates a standard syntax for ecological metadata. This concept allows for sharing of standard metadata and data across not only the LTER Network, but throughout the broader ecological community. SGS is researching new database technologies to managing the growing amount of standard non-spatial data from the ARS and LTER as well as GIS and Remote Sensed data. This also will address the need for multi-user data access and database integration with the SGS website. An SGS Website was launched 8 years ago to provide general site information as well as detailed research information. The web site will be enhanced over the next year to improve query tools, submit metadata online, improve integration with various ecological research databases, and implement the EML standards.

GIS and Remote Sensing	Metadata Standardization (EML)	Relational Database Management System	Dynamic Web Pages

IM staff at SGS-LTER continue to development tools within these four areas that support the synthesis of ecological information.

GIS and Remote Sensing

1. What is Geographic Information Systems and Remote Sensing?
• GIS are software tools that allow for presentation and analysis of spatial data
• Remote Sensing provides a means to obtain and process source data for GIS
• Remotely sensed data include photography and multi-spectral images from aircraft, satellites or any remote source.

2. What benefits does GIS and Remote Sensing provide to Researchers at SGS?
• Can integrate spatial and non-spatial data collected at the field site
• Can analyze, model, and predict ecological factors and outcomes that shape the shortgrass steppe

3. What are example applications of GIS and Remote Sensing at SGS (Fig. 1)
• Prairie dog town temporal changes and their relations to soil and vegetation.
• Drainage and road distance to prairie dog towns and their relations to genetics
• Root depth zone maps using minirhizotron data
• Chart quadrats of grasses and forbs to show temporal changes
• Great Plains region analysis to show variability of cropping and long term C and N changes

4. What are Current/Future Plans for GIS/Remote Sensing at SGS?
• Maintain existing spatial data (prairie-dog towns, property boundary, landuse, etc.)
• Catalog all spatial data
• Compile metadata for all spatial data
• Acquisition of more imagery
• Acquisition of climate data for correlations with prairie dog towns, vegetation, etc.

Figure 1. Prairie Dog Colony Size & Location Dynamics on the Pawnee National Grassland

Metadata Standardization - EML

1. What is EML?
• Ecological Metadata Language
• It is a set of XML (Extensible Markup Language) Schema documents that allows for the structural expression of metadata
• This concept allows for sharing of standard metadata and data

2. Who's developing EML?
• The EML Project, an open source, community
• National Center for Ecological Analysis and Synthesis (NCEAS), University of California at Santa Barbara, LTER Network

3. How does EML benefit researchers at SGS?
• Ecological data are largely unorganized and inaccessible, therefore efforts to integrate or synthesize data sets are stymied
• EML allows for new opportunities for data discovery, access, integration and synthesis

4. What are the important features of EML?
• Modularity: collection of modules rather than one large standard to facilitate future growth and flexibility
• Detailed Structure: EML strives to balance the tradeoff of too much detail with enough detail to enable advanced services
• Compatibility: EML adopts much of it's syntax from the other metadata standards from Michener et. al 1997
• Strong Typing:. XML Schema provides the ability to use strong data typing within elements allowing for finer validation

5. Current/Future Plans for implementing EML at SGS
• Most SGS metadata is documented in an Access database
• Next, a specialized tool such as Xanthoria can generate XML from the database
• Then, a tool such as XML Spy or Stylus Studio can develop the XSLT script to convert the generated XML into EML
• Our metadata can then be queried by outside users using Xanthoria or a metadata server such as Metacat (Fig 2)

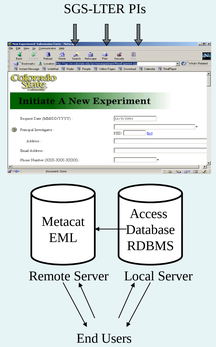

Figure 2. Mechanisms for researchers to contribute metadata

Relational Database Management System

1. Where do data collected at the SGS-LTER go?
• SGS uses data storage technology called RAID (Redundant Array of Inexpensive Disks)
• The SGS data storage system is organized into hierarchical components by identifying key attributes for categorization
• New server technology allows us to migrate older legacy data sets and new data files through a single point of entry on a centralized database that serves as a gateway to publish data sets on the web
• This system also allows flexibility for developmental data, while recognizing the need for growth and stability
• The data are entered and processed through a single point of entry on our new data storage network and assured for quality before data are archived in the SGS-LTER Relational Database Management System

2. The benefits of a Relational Database Management System (RDBMS)
• SGS-LTER utilizes Microsoft Access as the RDBMS
• Data and metadata from related projects can be stored in a single data repository
• Data managers and end users can archive, manage, and access data easily
• A RDBMS stores not only the data structure, but also how data and metadata tables are related to each other (Fig 3.)
• A RDBMS has built in tools to identify & eliminate most data inconsistency, anomalies, and dependencies on data structure

3. Future plans for the SGS-LTER RDBMS
• We must maintain an efficient and secure data flow into a centralized, safe, organized storage and data retrieval system.
• We plan to migrate to a multi-user database from Access
• We need to support our PIs by providing easy access to information, documenting and managing metadata, facilitating analyses of spatial and non-spatial data, publishing more data sets, and visualizing project information
• Growth will require the cooperation of Information Managers and PIs from ARS, SGS-LTER, and possibly PNG

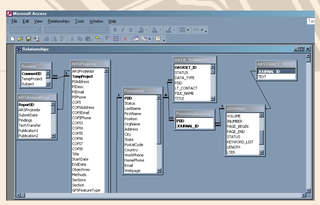

Figure 3. Relationships between metadata tables that contain information on PIs, project information, methods, and publications in the SGS-LTER RDBMS.

Dynamic Web Pages

1. How are data archived and published on the web?
• Information is transferred to the SGS-LTER Microsoft Access RDBMS
• End users may obtain project information from different tables in the SGS-LTER RDBMS from our web site (Fig. 3)
• End users may view available data and metadata in the database by category (e.g. disturbance)
• End users may query data and metadata by conducting a keyword search (Data_Summary Table in Fig. 3)
• Data and attribute definition tables are exported to archive and may be download as comma delimited text files
• Land managers and researchers may browse through information for projects conducted on the Central Plains Experimental Range

2. What tools can be developed for accessing data?
• Query tools are built using Active Server Pages
• Query tools for SGS-LTER publications by searching under keywords, an author, title, type of publication and year of publication (e.g. Journal Table in Fig. 3)
• Query data sets and metadata by searching for keywords (Data_Suammry Table in Fig. 3)
• Query species of plants and animals by searching for a species or downloading a list
• Query tools for project information and metadata by searching under a Principal Investigator (ARSProjects and Personnel Table in Fig. 3)
• Coming soon, query tools for meteorological data to summarize temperature and rainfall amounts over time

3. Web sites to check out (Please take a quick reference sheet)

•Shortgrass Steppe Long Term Ecological Research
•http://sgs.cnr.colostate.edu

•Central Plains Experimental Range
•http://sgs.cnr.colostate.edu/ars/

•United States Department of Agriculture- Agricultural Research Service
•Rangeland Resource Research Unit
•http://rrru.ars.usda.gov/

•US Forest Service – Pawnee National Grassland
•http://www.fs.fed.us/arnf/districts/png/

•LTER Network Office
•http://www.lternet.edu/

•Ecoinformatics.org (Ecological Metadata Language and other IM tools)
•http://www.ecoinformatics.org/