Published on *LTER Information Management* (http://im.lternet.edu)

Site Profiles

Site specific approaches to Information Management may be accessed by subject headings in the left menu or by site, using the links below.

LTER Information Management Personnel [1].

Use the links below to see information for each site.
Andrews Forest [2]
Arctic LTER [3]
Baltimore Ecosystem [4]
Bonanza Creek [5]
California Current Ecosystem [6]
Cedar Creek [7]
Central Arizona Phoenix [8]
Coweeta [9]
Florida Coastal Everglades [10]
Georgia Coastal Ecosystems [11]
Harvard Forest [12]
Hubbard Brook [13]
Jornada Basin [14]
Kellogg Biological Station [15]
Konza Prairie [16]
Luquillo Experimental Forest [17]
McMurdo Dry Valleys [18]
Moorea Coral Reef [19]
Niwot Ridge [20]
North Temperate Lakes [21]
Palmer LTER [22]
Plum Island Ecosystems [23]
Santa Barbara Coastal [24]
Sevilleta [25]
Short Grass Steppe [26]
Virginia Coast Reserve [27]

LTER metadata systems

A half page description of the current models and systems of data and metadata management at LTER sites.

Summary

Most of the sites (15 of 26 sites) use databases to manage their metadata. Most of these use relational databases, but not all. Some of these sites (9 of the 26 sites) use a hybrid model, where part of the metadata may be stored in flat text files in an structured file system. There are 5 sites that use the file system to manage their metadata (after all, a file system can be used as a simple database).

The preferred databases are mysql [28] and SQL server [29], while some use postgres [30] and there are a few instances of Oracle [31] in the network. Two sites use eXist [32] and DB2 [33] respectively, which are XML oriented databases, especially eXist. Strictly speaking, only eXist and postgres are open source; although mysql follows an open source model, it has components that are proprietary.

The scripting languages or code use to make the data available to the users vary wildly, Cold Fusion [34], PHP [35], Perl [36], CGI [37], HTML [38], XSLT [39], Matlab [40], Java/Jsp [41], VB [42], SAS [43], Excel [44], Ajax [45], Javascript [46] are technologies used, in many instances is an amalgam of such languages.

Most of the network Ecological Metadata Standard (EML [47]) documents are very rich, with attribute level content or level 5 according to the
EML Best Practices, however, a fair amount of QA/QC needs to be conducted for these documents to be successfully used in a machine driven synthesis architecture such as PASTA [48] or Kepler [49].

As for the data, many sites use a mix of database and flat text files (whether CSV or tab separated ASCII files) to manage the data. To see details per site, follow the links!

Andrews Forest Metadata System

Andrews Forest LTER Metadata Management System Summary

The Andrews site stores their metadata using an SQL relational database. The database is the center piece of the Andrews metadata management system, and the transactions with the public and the users are coded in ColdFusion. Data are served directly from the database, although a few data files still reside in the file system (but are planned to be moved to the SQL database). The Andrews follows a metadata driven information system, and in their dynamic website offer rich metadata in both the raw XML as well as eye-friendly formats.

LTER sites with similar management systems

- clone system : KNZ
- some similarities : JRN

The Andrews Detailed

The Andrews site stores their metadata using a Windows Structured Query Language (SQL) database. The database is the center piece of the Andrews metadata management system, and the transactions with the public and the users are coded in ColdFusion. The scripts allow users to query, view and download metadata and data on the Andrew's dynamic website. The Andrews follows a metadata driven information system, and generates rich EML documents that are offered both in their website and the central repositories. Data are served directly from the database. Although a few data files may still reside in the file system, but will be migrated eventually. In addition, the Andrews uses an ArcGis server to manage their geospatial datasets, which are a significant portion of the Andrews data collection.

EML Status:

- Completion: All metadata are available in EML.
- Richness : All datasets have rich documentation, up to the attribute level (level 5)
- QA/QC : There are some mechanisms to ensure the quality of the metadata and data, mostly manual entries in the back end database, some automated mechanisms are being implemented to ease the quality assurance tasks.

Related Links

- Data Catalog (618) [50]

Arctic LTER Metadata System

Arctic LTER Metadata Management System Summary

The Arctic LTER (ARC) site stores their metadata in a structured filesystem. Metadata is gathered with an advanced Excel sheet that has evolved from an earlier Word based form sheet. The enhancements were partially inspired by the Florida Coastal LTER metadata capturing system (Excel2EML). The Excel workbook includes both metadata and data worksheets and does not use macros for entering information. A separate Excel workbook used by the Information Manager (IM) contains Excel macros (using Visual Basic for Applications) for parsing the metadata worksheet to create an EML and HTML files. The harvest list file tracks dataset metacat ID numbers. Geospacial data is served through Toolik Field Station GIS and Remote Sensing protal.

LTER sites with similar management systems

- some similarities : PIE, HBR, FCE

The Arctic LTER Detailed

The ARC site stores their data and metadata in a structured filesystem. Metadata is gathered with an advanced Excel sheet that has evolved from an earlier Word based form sheet. The enhancements were partially inspired by the Florida Coastal LTER metadata capturing system (Excel2EML). The Excel workbook includes both metadata and data worksheets and does not use macros for entering information. A separate Excel workbook used by the Information Manager (IM) contains Excel macros (using Visual Basic for Applications) for parsing the metadata worksheet to create an EML and HTML files. The harvest list file tracks dataset metacat ID numbers. Geospacial data is served through Toolik Field Station GIS and Remote Sensing protal.

EML Status:

- Completion: Nearly all metadata is available in EML.
- Richness : Ranges from discovery level to the attribute level (levels 2 to 4)
- QA/QC : Research investigators, assistants and students who collect the data are responsible for data analysis, quality control, and documentation. This insures that the data are checked and documented by those most familiar with the data.

Related Links

- Data Catalog (1372) [51]

Baltimore Ecosystem Metadata System

The Baltimore Ecosystem Metadata Management System Summary

The Baltimore Ecosystem (BES) site stores their metadata using two main systems. An open research system (ORS) in conjunction with a SQL server, Active Server Pages and Cold Fusion makes the backbone of the diverse gateways to the Baltimore data and metadata. To deal better with the enormous geospatial emphasis of the research at BES, the IMs are integrating ESRI's geospatial engines with database software. It will also eliminate the multi-entry points for storing and documenting the data. BES information and data management approach is centered around an adaptation of the Human Ecosystem Framework, called the Human Ecological System (HES). The HES interface serves as a structure by which we will integrate, or "hang", Baltimore Ecosystem Study data in a meaningful way. Google Earth services are integrated with some of the

LTER sites with similar management systems

The BES data management is quite unique, as most of BES data is geospatial. Data is served on a geospatial context with custom applications.

The Baltimore Ecosystem Detailed

The establishment of a BES Multi-User Geodatabase (BES-MUG) allows for the storage, management, and distribution of geospatial data associated with the Baltimore Ecosystem Study. At present, BES data is distributed over the internet via the Open Research System (ORS) and the BES website. While having geospatial data available for download is a vast improvement over having the data housed at individual research institutions, it still suffers from some limitations. BES-MUG overcomes these limitations; improving the quality of the geospatial data available to BES researches, thereby leading to more informed decision-making.

BES-MUG builds on Environmental Systems Research Institute's (ESRI) ArcGIS and ArcSDE technology. ESRI was selected because its geospatial software offers robust capabilities. ArcGIS is implemented agency-wide within the USDA and is the predominant geospatial software package used by collaborating institutions.

BES uses ESRI's ArcSDE (Spatial Database Engine) in conjunction with database software. Through ArcSDE and the Geodatabase model the database's capabilities are expanded, allowing for multiuser editing, intelligent feature types, and the establishment of rules and relationships. ArcSDE also allows users to connect to the database using ArcGIS software without being burdened by the intricacies of the database itself.

For an example of how BES-MUG will help improve the quality and timeless of BES geospatial data consider a census block group layer that is in need of updating. Rather than the researcher downloading the dataset, editing it, and resubmitting to through ORS, access rules will allow the authorized user to edit the dataset over the network. Metadata will automatically be updated showing who edited the dataset and when they did in the event any questions arise.

Currently, a functioning prototype Multi-User Database has been developed, using Arc SDE and IBM's DB2 Enterprise Database as a back end architecture. This database, which is currently only accessible to those on the UVM campus network, will shortly be migrated to a Linux server where it will be accessible for database connections over the Internet. Passwords can then be handed out to all interested researchers on the project, who will be able to make a database connection through the Geographic Information Systems software interface on their desktop computer.

EML Status:

- Completion: the vast majority of metadata are available in EML.
- Richness : Ranges from basic level to the discovery level (levels 1 to 3).

- QA/QC : there are mechanisms to ensure the quality of the metadata and data, mostly manual entries in the back end database.

Related Links

- Geodatabase Details (1420) [52]
- Information Management Plans (1320) [53]
- Open Res. System Portal (1333) [54]

Bonanza Creek Metadata System

The Bonanza Creek Metadata Management System Summary

The Bonanza Creek (BNZ) site stores their metadata in a database using mySQL Structured Query Language database. The database is the center piece of the BNZ site metadata management system, and the transactions with the public and the users are coded in Cold Fusion. The scripts allow users to query, view and download metadata and data.

LTER sites with similar management systems

- some similarities : HBR, GCE
-
- potential synergies : MCM, PAL, CCE, MCR, SEV

  The Bonanza Creek Detailed

  A custom mySQL relational database stores all related info to datasets. This database tracks changes in data and metadata and allows data sets to be viewed using CFM scripts. Some of the data are in raw text files and the remaining data are stored in a database. Some climate data is streaming directly from the field instruments into the database. There is an ARC-IMS server for geospatial files. Most of the metadata is managed with custom files. EML is generated with a one touch Perl script (SQL dump-»parsing-»generate EML and harvest list).

  EML Status:

  - Completion: Nearly all metadata is available in EML.
  - Richness : Ranges from discovery level to the attribute level (levels 3 to 5)
  - QA/QC : There are mechanisms to ensure the quality of the metadata and data.

Related Links

- Data Catalog (1536) [55]

California Current Ecosystem Metadata System

The California Current Ecosystem LTER data and metadata management approaches

The California Current Ecosystem LTER (CCE) site has designed a comprehensive information management environment centering on an information system called DataZoo. This system's architecture supports data aggregation, description, and interoperability. DataZoo is coded primarily in PHP, using an object-oriented design focusing on code reusability. The backend is a mySQL relational database while the frontend web interfaces for ingesting data and creating metadata as well as querying, viewing, plotting, editting, and downloading data, are enhanced through the use of JavaScript and Ajax. XLST style sheets are used to create easily readable versions of XML documents, particularly EML standard metadata.

LTER sites with similar management approaches

- clone system : PAL
- some similarities : SEV, MCM

The approach Details

CCE joined PAL in the SIO Ocean Informatics endeavor with goals of 1) creating a process for design of a community information system, 2) partnering with users and social scientists as codesigners, and 3) developing a local awareness and understanding of information infrastructure. Datazoo provides a number of services and tools to users. Users may browse and view datasets as well as perform basic plots and downloads of data (as CSV text files) and metadata (as EML standard XML files). For the local community, participants may upload and review data, perform quality assurance, and manage metadata. This system facilitates the quality control of the data using PHP scripts and code developed using the Yahoo JavaScript API serve dynamic web forms for the entry process. In addition, web interfaces provide three types of query capabilities: conditional, multi-dataset, and saved. Data integration is enabled by a trio of elements: a project-study-dataset relations architecture, a set of shared dictionaries (unit, attribute, and qualifier), and metadata description to the column level. JpGraph provides dynamic graphs as results of user driven queries of the data and metadata. CCE provides EML to the attribute level, with enhancements to support synthesis efforts. A two-tier user privilege system distinguishes community users from public users.

EML Status:

- Completion: A fraction of the metadata is available in EML. DataZoo first release occurred during the 2007 summer; population is ongoing.
- Richness : All the harvested EML is rich,attribute level (level 5)
- QA/QC : Files are quality controlled initially through data ingestion templates. Data may be further quality controlled manually and revisions reuploaded.

Related Links

- Data Catalog (Datazoo) (1344) [56]

Cedar Creek Metadata System

The Cedar Creek LTER Metadata Management System Summary

The Cedar Creek (CDR) site stores their metadata in a relational database using Oracle.

LTER sites with similar management systems

- potential synergies : MCM, PAL, CCE, MCR, SEV

The Cedar Creek Detailed

Rich EML was produced parsing the metadata info contained at the CDR project and experiment website and within the data files itself. A new model is being implemented now. During quality control checking of harvested metadata, datasets available through CDR's online data distribution system will be linked to original HTML versions of the same materials.

EML Status:

- Completion: Nearly all metadata is available in EML.
- Richness : Ranges from discovery level to the attribute level (5)
- QA/QC : currently, the harvested EML had few mechanisms to ensure the quality of the EML. All data and metadata are currently being reviewed and revised for QC.

Related Links

- Data Catalog (1384) [57]

Central Arizona Phoenix Metadata System

The Central Arizona Phoenix Metadata Management System Summary

The Central Arizona Phoenix (CAP) urban LTER site data management system follows the relational database models for the research data and site management data, while the metadata are stored as XML files in a native XML database (eXist). CAP recently transitioned from using Microsoft SQL Server and ACCESS to open source products such online PHP data entry screens, mySQL and newer technologies such as XForms (Orbeon) for editing metadata. The CAP manages data transactions on the web using Java code and the Tomcat server. The CAP pioneered many LTER Networks products, including Xanthoria, a management system that capitalized on the power of the Service Oriented Architectures, as well as XML technologies, and the EML schema for metadata. Currently, the CAP is investing heavily in the next generation of dynamic forms (Xforms) using the latest technologies.

LTER sites with similar management systems

- some similarities : JRN, AND
- potential synergies : MCM, PAL, CCE, MCR, SEV

The Central Arizona Phoenix Detailed

A custom relational database stores all research and site management data. Metadata are associated with the research databases and then reverse engineered into EML files by a Java program. Currently the workflow starts with designing the database and EML is generated after the database design as settled. The cap website provides data entry screens, some EML editing, and query systems for data discovery. Metadata are searched in the eXist native XML database, the EML files are formated for display via XSLT transformation. The predominant technology used in the website is Java and jsp for database searches, data download and general display. However, due to ease of use PHP is employed for quick generation of data entry applications for research and core monitoring data.

EML Status:

- Completion: Nearly all metadata is available in EML.

- Richness : Nearly all EML is rich, attribute level (level 5)
- QA/QC : Some quirks on EML use of attributes created some trouble in the recent past. There are mechanisms to ensure the quality of the metadata and data, mostly manual entries in the back end database.

Related Links

- The CAP site (1385) [58]

Coweeta Metadata System

The Coweeta Metadata Management System Summary

The Coweeta LTER (CWT) site data management uses a dynamically-generated searchable/indexed data catalog to keep track of the contributions to the project since its inception. All relevant data is available on the web, including the associated data files. Several formats are offered, including csv, dbf and flat text for the data, and HTML or and EML for metadata. For GIS data, there are other appropriate formats for the data files.

LTER sites with similar management systems

- some similarities : PIE, ARC

The Coweeta Detailed

The key to Coweeta's data manaegment system is that a researcher can only begin his/her work when the data management system has allocated a place for the project. Coweeta expects that the user inmediately begins documenting the project and provides rich metadata and data in several formats. This grants good documentation from the start. Technically, all the metadata is entered by the user on forms that are available on the web. All information is revised by the Coweeta Information Manager, and published for anyone to see. The data format could be in as many as three formats, microsoft excel archives, database format files (DBF files) and comma delimited files. The metadata is available both in flat text as well as EML, with about 60% of very rich and quality controlled EML.

EML Status:

- Completion: Nearly all metadata is available in EML.
- Richness : Over half of the harvested EML is rich and the quality has been throughly controlled, attribute level (level 5), and the rest is about discovery level (3)
- QA/QC : The CWT has done a tremendous effort to ensure the maximum possible quality of their data and metadata. The next release of EML will reflect that, upping to level 5

Related Links

- Data Catalog (1453) [59]

Florida Coastal Everglades Metadata System

The Florida Coastal Everglades LTER Metadata Management System Summary.

The Florida Coastal Everglades LTER (FCE) site data management collects data primarily in Excel spreadsheets. Curated ASCII data files (flat text) are offered for download in FCE's website. Thus, FCE primarily uses a file system to store their data and metadata. Rich EML documents are transformed into a readable format using XLST stylesheets and offered in FCE's website. The FCE site is working to collect some of their data in an Oracle database.

LTER sites with similar management systems

- some similarities : PIE, ARC, CWT

The Florida Coastal Everglades Detailed

The Florida Coastal Everglades LTER (FCE) site data management collects data primarily in Excel spreadsheets. These are revised and columns with corrections are added, and eventually an ASCII file (flat text) is offered for download in FCE's website. Thus, FCE primarily uses a file system to store their data and metadata. Thanks to the FCE-developed Excel2EML tool, all metadata gathered on a excel spreadsheet is transformed into EML. EML metadata is transformed into a readable format using XLST stylesheets. The site is working to collect some of the physical and chemical data in an Oracle database, and the ability to query the database real time will be opened to the user soon.

EML Status:

- Completion: All metadata is available in EML.
- Richness : All the harvested EML is rich,attribute level (level 5)
- QA/QC : Files are revised and quality controlled manually.

Related Links

- Data Catalog (1313) [60]

Georgia Coastal Ecosystems Metadata System

The Georgia Coastal Ecosystems LTER Metadata Management System Summary

The Georgia Coastal Ecosystems (GCE) LTER site manages data and metadata using a comprehensive information system that currently includes a centralized relational database management server, custom network-enabled analytical software, and dynamic web applications. All project information is stored in relational databases, which are queried dynamically to produce rich metadata for data sets as well as a searchable web-based data catalog with links to related project information (publications, study sites, species lists, and personnel contact information). Metadata-based MATLAB software developed on site (GCE Data Toolbox) is used to automate data processing, validation and quality control of primary data, and to create derived data products. Metadata and processing lineage information generated by this software during data processing are synchronized with the database whenever data sets are archived or revised to provide a complete record. Tabular data and pre-formatted text metadata are distributed to users in standard and user-customized ASCII and MATLAB formats via the web-based data catalog and GCE Toolbox search client application. EML is generated on-demand from the centralized database server via dynamic web applications. Support for GIS data will also be provided in the future.

LTER sites with similar management systems

- None. GCE has an advanced and unique management system; however, software and database designs developed at GCE have been openly shared with other LTER sites and the scientific community and can be leveraged by other sites

The Georgia Coastal Ecosystems Detailed

Most acquisition and processing of GCE data is currently done in MATLAB using the GCE Data Toolbox software. Metadata content is generated iteratively during data processing, with top-level information (title, abstract, personnel, study description, etc.) entered into a centralized SQL Server 2000 database using MS Access-based entry forms, or copied en masse from other related data sets via stored procedure and revised. Data table and attribute metadata are generated automatically during data processing, or loaded from pre-defined metadata templates for the data source, then manually edited using the GCE Data Toolbox metadata editor application. Additional metadata content, such as processing history (lineage), unit conversions, and calculations for derived fields, is automatically generated by the GCE Toolbox software during processing. Quality control is performed automatically based on detailed rules defined for each attribute (pre-defined in metadata templates or defined during analysis), with character flags assigned to individual values that meet rule criteria. Quality control can also be performed manually in a spreadsheet-like editor or on plots with the mouse, allowing automatically-assigned flags to be edited and new flags assigned visually, with all operations logged to the processing history. After data processing is complete, data sets are registered and versioned in the database programatically, at which time new and updated metadata content is synchronized back to the database server and distribution files are created and copied to the production web server. Editing and re-registering data sets automatically generates a new version of the data structure and distribution files, allowing prior versions to be retained (although only the most recent version is discoverable from the data catalog).

The GCE Data Toolbox software is publicly available online (in compiled form for non-GCE members), allowing MATLAB users to extensively customize, analyze and visualize GCE-distributed data sets and create their own derived data products with complete metadata. The entire catalog of primary and ancillary GCE data can be searched using an included search engine client, then automatically retrieved from GCE servers for local analysis. ClimDB/HydroDB and USGS NWIS data can also be queried and retrieved, then analyzed and integrated with GCE data in real time.

EML Status:

- Completion: Metadata for all primary GCE data are available in EML (ESA-FLED text metadata provided for ancillary data)
- Richness: All the harvested EML is rich, attribute level (level 5)
- QA/QC: Files are revised and quality controlled systematically with automated and manual checks, documented in the metadata

Related Links

- GCE Data Toolbox (718) [61]
- GCE EML (508) [62]
- GCE IS (475) [63]

Harvard Forest Metadata System

The Harvard Forest LTER Metadata Management System Summary

The Harvard Forest LTER (HFR) site uses a filesystem to structure and manage their data and metadata. Multi column flat text data files are offered at HFR's website to the visitor. EML files are also offered in their raw XML format for the peruse of the visitors or harvesting routines.

LTER sites with similar management systems

- some similarities : PIE, ARC, CWT, FCE

The Harvard Forest Detailed

The HFR site data management uses a structured filesystem to store data and metadata and to offer via web all details. The HFR projects titles (or their synthesized abbreviations) are the top of the hierarchy of the filesystem used to catalog data and metadata. Reviewed ASCII data files (flat text) are offered for download in HFR's website. Thus, HFR primarily uses a file system to store their data and metadata. EML documents are also offered in HFR's website, these documents enable discovery of data and metadata. GIS data are classified in similar way, however, due to the more complex nature of the data structures, files are often offered in zipped archives, and may be broken in multiple files and corresponding metadata pieces. Some data files are served by HFR's partners, that is, in external servers with their particular interface.

EML Status:

- Completion: All metadata is available in EML.
- Richness : All the harvested EML provides discovery level richness.
- QA/QC : Files are revised and quality controlled manually ??

Related Links

- Data Catalog (1370) [64]

Hubbard Brook Metadata System

The Hubbard Brook Metadata Management System Summary

The Hubbard Brook LTER site (HBR) their data and metadata in a database using mySQL Structured Query Language database. The database is the center piece of the HBR's site metadata management system, and the transactions with the public and the users are coded in PHP. The scripts allow users to query, view and download metadata and data. Rich metadata is offered both as raw EML and processed through XSLT stylesheets to make for a pleasant read.

LTER sites with similar management systems

- some similarities : BNZ, SEV
-
- potential synergies : MCM, PAL, CCE, MCR

The Hubbard Brook Detailed

At HBR, a customized mySQL relational database stores all related info to datasets. This database tracks changes in data and metadata and allows data sets to be queried, browsed and viewed using Pre Hypertext Processor language (PHP) scripts. Some of the data are in raw text files and other data are stored in a database: it depends of the data category, whether watershed studies, climate studies or more point studies. There is a tendency to sort out in a database the larger systematic

datasets. Rich EML is available for all datasets, and currently maintained manually. The site is mulling improved management options for the EML.

EML Status:

- Completion: All metadata is available in EML.
- Richness : Nearly all EML documents are very rich, to the attribute level (levels 5)
- QA/QC : There are mechanisms to ensure the quality of the metadata and data ??

Related Links

- Data Catalog (1429) [65]

Jornada Basin Metadata System

The Jornada Basin LTER Metadata Management System Summary

The Jornada Basin LTER site (JRN) manages their data and metadata in ASCII text files and relational databases. Rich metadata is viewed in a friendly text format on the JRN online data catalog (http://jornada-www.nmsu.edu/datacat.php [66]), which is generated from JRN's metadata database using web pages to administer the metadata database and JRN website content. The web pages also allow semi-automated creation of EML instance documents and the LNO Metacat harvester script. A number of applications make use of the metadata stored in the metadata database describing JRN research projects, data and metadata. Metadata is also offered in EML standard within the LTER Network and JRN data catalogs.

LTER sites with similar management systems

- some similarities : AND, KNZ
- potential synergies : MCM, HBR

The Jornada Detailed

The Jornada Information Management System (JIMS) stores data and metadata in a relational database management system (SQL Server 2005). A text editor and Excel spreadsheets are used to enter and collect metadata from JRN researchers, which are copied into the database tables with aid of PHP web pages. Data is also processed with the aid of PHP and SAS scripts to generate JRN ClimDB data harvest files. The Geographic Information System (GIS) is comprised of a separate relational database management system (ArcSDE 9.2 on SQL Server 2005) and metadata repository (entered using ArcGIS Desktop and metadata template). Jornada has applications and services to support discovery, access, and management of information stored in the JIMS. The current metadata repository consists of available Ecological Metadata Language (EML) and text documents stored and available on the JRN website. Error-checked data files are stored with associated documentation files on the file server and the publicly available data files on the Jornada website.

The Jornada geodatabase is the repository for GIS vector data layers that have a sub-meter accuracy GPS source. Jornada researchers have direct access to the geodatabase using ESRI ArcGIS or ArcView software. GIS data and metadata can also be accessed by the public via the Map Server.

EML Status:

- Completion: About half of the metadata is available in EML.
- Richness : EML content richness ranges from discovery to attribute level (levels 3 to 5)
- QA/QC : There are mechanisms to ensure the quality of the metadata and data.

Related Links

- JRN Data Catalog (1505) [67]
- JRN Data Policies (1253) [68]
- JRN Information Management (1355) [69]
- JRN Interactive Map (1250) [70]
- JRN Long-term Datasets (1208) [71]
- JRN Map Gallery (1285) [72]

Kellogg Biological Station Metadata System

The Kellogg Biological Station Metadata Management System Summary

The Kellogg Biological Station (KBS) LTER site their metadata and metadata in a relational database (DB2). Normalized relational databases are at the center piece of the KBS's site metadata management system, and the transactions with the public and the users are coded using the Ruby on Rails framework. The jsp?? scripts allow users to query, view and download metadata and data as well as to manage the KBS website in general. Rich metadata is offered both as raw EML and a text synopsis per dataset.

LTER sites with similar management systems

- some similarities : BNZ, SEV, HBR
- potential synergies : MCM, PAL, CCE, MCR

The Kellogg Biological Station Detailed

The KBS LTER site their metadata and metadata in a normalized relational database. Due to the diverse data collections, in the recent past, multiple databases managed tabular and Geospatial data. KBS is centralizing these resources to a geo-enabled, XML friendly, back end database IBM DB2. Thus, the relational database model is the center piece of the KBS's site data and metadata management system. A number of custom script aid with the data quality control process, and facilitate the investigator and the information manager to conduct and oversee the task of ensuring high quality data. The scripts allow users to query, view and download metadata and data as well as to manage the KBS website in general. Metadata is offered in conjunction with the data sets, as brief synopsis, however, the raw EML which is rich in content (level 5, attribute level) is also available in its raw state.

EML Status:

- Completion: All metadata is available in EML.
- Richness : Nearly all EML documents are very rich, to the attribute level (levels 5)
- QA/QC : There are mechanisms to ensure the quality of the metadata and data ??

Related Links

- Data Catalog (1442) [73]

Konza Prairie Metadata System

The Konza Prairie LTER Metadata Management System Summary

The Konza Prairie LTER site (KNZ) stores their metadata using a database. The database is the center piece of the KNZ metadata management system, and the transactions with the public and the users are coded in Active Server Page programming language (ASP). Some data files reside in the filesystem, while others are served from the database. The KNZ site follows a metadata driven information system, and in their dynamic website offer rich metadata in both the raw XML as well as eye-friendly formats.

LTER sites with similar management systems

- clone system : AND
- some similarities : JRN

The Konza Prairie LTER Detailed

The Andrews site stores their metadata using a Windows Structured Query Language (SQL) database. The database is the center piece of the Andrews metadata management system, and the transactions with the public and the users are coded in Active Server Page programming language (ASP). The scripts allow users to query, view and download metadata and data on the KNZ's dynamic website. The KNZ site follows a metadata driven information system, and generates rich EML documents that are offered both in their website and the central repositories. Some data is stored in flat text files, while other data are streamed directly from the database.

EML Status:

- Completion: All metadata is available in EML.
- Richness : All datasets have rich documentation, up to the attribute level (level 5)
- QA/QC : Some revision of the database content seems to be needed to fill-in some metadata gaps.

Related Links

- KNZ Site link, browse for data catalog (1302) [74]

Luquillo Experimental Forest Metadata System

The Luquillo Experimental Forest LTER Metadata Management System Summary

The Luquillo Experimental Forest LTER site (LUQ) their metadata in relational databases. Text files and HTML files are manually revised to reflect the latest updates on metadata and data. Other database (mySql) manage the site personnel and data files are kept in the structure filesystem. Discovery level EML is also offered on the Luquillo website and central repositories.

LTER sites with similar management systems

- some similarities : BNZ, SEV, HBR
- potential synergies : MCM, PAL, CCE, MCR

The Luquillo Detailed

The Luquillo Experimental Forest LTER site (LUQ) their metadata in relational databases, at this time primarily in Paradox. The data is stored in a hierarchical filesystem structured according to the Luquillo research projects. Metadata is captured in text files, and after a review and approval process, text files and HTML files are generated in the filesystem reflecting the latest updates on metadata and data. Other database (mySql) manage the site personnel and data files are kept in the structure filesystem. Discovery level EML is also offered on the Luquillo website and central repositories. Luquillo is working into transitioning into a database centric model, where all metadata will be handled and served from the database, as well as the website content. Quality controlled EML is soon to be completed to the attribute level.

EML Status:

- Completion: All metadata is available in EML.
- Richness : Nearly all EML documents are at discovery level, but plans to enrich these documents to the attribute level (level 5) are 65% implemented
- QA/QC : There are manual systematic mechanisms to ensure the quality of the metadata and data.

Related Links

- Data Catalog (1409) [75]

McMurdo Dry Valleys Metadata System

The McMurdo Dry Valleys LTER Metadata Management System Summary

The McMurdo Dry Valleys LTER site (MCM) their data and metadata in a relational database. The database is the center piece of the MCM's site metadata management system. Scripts allow users to query, view and download metadata and data. Rich metadata is offered both as raw EML and processed through XSLT stylesheets to make for a pleasant read.

LTER sites with similar management systems

- some similarities : BNZ, SEV, HBR
- potential synergies : PAL, CCE, MCR

The McMurdo Dry Valleys Creek Detailed

At MCM, a customized Oracle relational database stores all related info to datasets. This database enables tracking of changes in data and metadata and allows data sets to be queried, browsed and be viewed using procedures. Database enables provide derived data, with unit transformations, or combining measurements. Implementation of chain of custody forms which have facilitated the users the task of entering and curating their data. Some of the data are in raw text files and other data are stored in a database: it depends of the data category, whether watershed studies, climate studies or more point studies. MCM offers metadata both in the raw EML format, or processed to make up for readability. Currently, MCM is implementing plans to improve the management of their GIS data, including generating EML metadata. An interactive online map using ESRI's ArcGIS Server technology was implemented on MCM's website. This map is JSP based, and pulls layers directly from our Oracle database. The map also serves as a data portal - clicking on certain features returns actual data and metadata.

EML Status:

- Completion: All metadata is available in EML.
- Richness : All EML documents are very rich, to the attribute level (levels 5)
- QA/QC : There are mechanisms to ensure the quality of the metadata and data ??

Related Links

- Data Catalog (1395) [76]

Moorea Coral Reef Metadata System

The Moorea Coral Reef LTER Metadata Management System Summary

The Moorea Coral Reef LTER site (MCR) manages their data and metadata in a hybrid system consisting of both relational database and file system. Publicly accessible datasets are registered in the database including identifiers and data access information. A server script request the access information dynamically from the database which is returned as URL or SQL query and subsequently runs the required commands to return the data. The scripts allow users to query, view and download metadata and data. Rich metadata is offered both as raw EML and processed through XSLT stylesheets to make for a pleasant read.

LTER sites with similar management systems

- some similarities : BNZ, SEV
- potential synergies : MCM, PAL, CCE, MCR

The Moorea Detailed

At MCR, a customized PostgreSQL relational database stores all related info of datasets. This database allows data sets to be queried, browsed and viewed using Pre Hypertext Processor language (PHP) scripts. Ajax web forms and Java desktop applications serve for data entry. Ajax eases data entry by dynamically generating drop down menus and performing instant quality control. Metadata are available for all datasets, including metadata in the network standard, EML. These are currently maintained manually.

EML Status:

- Completion: All metadata is available in EML.
- Richness : Most EML documents are described to the attribute level (level 5)
- QA/QC : Mechanisms to ensure quality of metadata and data are revised frequently.

Related Links

- Data Catalog (1389) [77]
- Information Management Plans (1401) [78]
- Information Management Synopsis (1448) [79]

Niwot Ridge Metadata System

The Niwot Ridge LTER Metadata Management System Summary

The Niwot Ridge LTER site (NWT) their metadata and data are in a relational database and flat text files. An array of data entry formats merge into either of these back end systems, after a curation process. Scripts allow users to query, view and download metadata and data. Rich metadata is offered both as raw EML and also in an easy to read header section of the flat file data files.

LTER sites with similar management systems

- some similarities : LUQ, BNZ, SEV, HBR
- potential synergies : PAL, CCE, MCR

The Niwot Ridge LTER Detailed

Niwot uses a number of different entry data points, from dataloggers to hand written data forms. After initial archival and backup, data is curated, derived and passed into a final archival mode in a relational database or an structured filesystem, perl or visual basic scripts and a few unix based tools tailor the data for public consumption at the Niwot server. Perl scripts parse updated metadata from the relational system to offer the metadata in the LTER network standard. The link below could not illustrate better the data management flow.

EML Status:

- Completion: All metadata is available in EML (not all GIS data - still have issues w/ raster metadata conversion).
- Richness : All EML documents are very rich, to the attribute level (levels 5)
- QA/QC : There are systematic mechanisms to ensure the quality of the metadata and data.

Related Links

- Data Catalog (1379) [80]
- Data Management Synopsis (1314) [81]

North Temperate Lakes Metadata System

The North Temperate Lakes LTER Metadata Management System Summary

The North Temperate Lakes LTER site (NTL) manages most of their data and metadata in a relational database. Applications and scripts drive the data insertion process. Several methods allow users to query, view and download metadata and data. Rich metadata is offered both as raw EML and reading friendly formatted text on NTL's website.

LTER sites with similar management systems

- some similarities : BNZ, SEV
- potential synergies : MCM, PAL, CCE, MCR

The North Temperate Lakes Detailed

Most of the data collected by NTL reside an Oracle database. Some data are archived in text format. NTL spatial data are stored on file systems. Data are entered, updated and maintained in the Oracle database using scripts and applications on a Sun workstation as well as through use of an application installed on networked Windows workstations. Some data are entered into the Oracle

database directly through a locally installed web application. Data from sensors on lake buoys are uploaded automatically via wireless serial spread spectrum radio to a file server and then automatically uploaded to the Oracle database by a custom application. Data from the Oracle database are made available for viewing and/or downloading over a network connection by several methods.

Each NTL online data set has associated metadata online in text file and EML formats.

The metadata for the non-spatial data are stored in the Oracle database and drive the dynamic database application in the data catalog, allowing forms for querying to be generated dynamically for each data set that is maintained in the Oracle database and provided in the catalog.

A number of different quality-control mechanisms have been established. Data sets have a system of flags to indicate quality conditions such as non-standard routine or equipment used. Database triggers perform range checks on data from the automated sensors on lake buoys that are being captured in the database in near-real time. Information management staff and technicians perform visual screening of data.

EML Status:

- Completion: All metadata is available in EML.
- Richness : Most EML documents are described to the attribute level (level 5)
- QA/QC : There are mechanisms to ensure quality of metadata and data.

Related Links

- NTL Data Catalog (1436) [82]
- NTL Information Management Synopsis (1257) [83]

Palmer LTER Metadata System

The Palmer LTER data and metadata management approaches

The Palmer LTER (PAL) site has designed a comprehensive information management environment centering on an information system called DataZoo. This system's architecture supports data aggregation, description, and interoperability. DataZoo is coded primarily in PHP, using an object-oriented design focusing on code reusability. The backend is a mySQL relational database while the frontend web interfaces for ingesting data and creating metadata as well as querying, viewing, plotting, editting, and downloading data, are enhanced through the use of JavaScript and Ajax. XLST style sheets are used to create easily readable versions of XML documents, particularly EML standard metadata.

LTER sites with similar management approaches

- clone system : CCE
- some similarities : SEV, MCM

The approach Details

PAL joined CCE in the SIO Ocean Informatics endeavor with goals of 1) creating a process for design of a community information system, 2) partnering with users and social scientists as codesigners, and 3) developing a local awareness and understanding of information infrastructure. Datazoo provides a number of services and tools to users. Users may browse and view datasets as

well as perform basic plots and downloads of data (as CSV text files) and metadata (as EML standard XML files). For the local community, participants may upload and review data, perform quality assurance, and manage metadata. This system facilitates the quality control of the data using PHP scripts and code developed using the Yahoo JavaScript API serve dynamic web forms for the entry process. In addition, web interfaces provide three types of query capabilities: conditional, multi-dataset, and saved. Data integration is enabled by a trio of elements: a project-study-dataset relations architecture, a set of shared dictionaries (unit, attribute, and qualifier), and metadata description to the column level. JpGraph provides dynamic graphs as results of user driven queries of the data and metadata. PAL provides EML to the attribute level, with enhancements to support synthesis efforts. A two-tier user privilege system distinguishes community users from public users.

EML Status:

- Completion: A fraction of the metadata is available in EML. DataZoo first release occurred during the 2007 summer; population is ongoing.
- Richness : All the harvested EML is rich,attribute level (level 5)
- QA/QC : Files are quality controlled initially through data ingestion templates. Data may be further quality controlled manually and revisions reuploaded.

Related Links

- Data Catalog (Datazoo) (1344) [84]

Plum Island Ecosystems Metadata System

The Plum Island Ecosystem LTER Metadata Management System Summary

The Plum Island Ecosystems LTER (PIE) site stores their metadata in a structured file system. Data and metadata are entered by researchers into a Microsoft Excel spreadsheet designed to create EML metadata as structured XML, HTML or RTF files and data as comma delimited ascii files. Metadata and data exist as compiled flat text tabular data and also as ArcGIS, IDRISI and RiverGIS GIS file formats EML is also offered at PIE, with an attribute level of content richness.

LTER sites with similar management systems

- some similarities : ARC, HBR, FCE

The Plum Island Site Detailed

The PIE site manages the data and metadata using a structured file system that follows the PIE research program areas to recreate a file shelving structure. Data files are in tabular text format, offered unrestricted to the public but according to the data policy guidelines. Data and metadata are originally captured in an MS Excel document to an attribute level. Metadata is offered through the PIE website and in EML format.

EML Status:

- Completion: Nearly all flat file metadata is available in EML, GIS metadata is a work in progress.
- Richness : Ranges from discovery level to the attribute level (levels 3 to 4)
- QA/QC : Protocols to be added

Related Links

- Data Catalog (1410) [85]

Santa Barbara Coastal Metadata System

The Santa Barbara Coastal LTER Metadata Management System Summary

The Santa Barbara Coastal LTER (SBC) IM system is based on EML. Our metadata is displayed in a local instance of Metacat, with a customized portal for searching and browsing. Data are stored as flat text tabular files. Some GIS layers are available by request.

SBC creates EML by several methods. Where possible, EML is created as the last step in data processing using Matlab. Other data packages are drafted by scientific staff (in Morpho) and customized by IM scripts. Some packages are edited manually, and we are trying to move away from this mode. Metadata for scripts are stored in text files and as XML in eXist.

Current IM activities
IM activities for 2009-2010 are concerned with centralizing and structuring more of our metadata in systems that make it available for multiple uses. Since SBC's data system is based on EML, we promote solutions at both the site and network levels.

LTER sites with similar systems

- data and scientists in common: MCR
- similar programming languages: GCE (matlab), VCR (perl)

The Santa Barbara LTER Site Detailed
see SBC's Information management plan at:
http://sbc.lternet.edu/external/InformationManagement/documents/SBC/SBC_... [86]

EML Status:

- Completion: Nearly all site data has EML metadata.
- Richness : attribute level (level 5), except for GIS (level 3)
- SBC's bibliography is also stored in EML
- QA/QC : based in the laboratories.

Related Links

- SBC Data Index (630) [87]

Sevilleta Metadata System

The Sevilleta Metadata Management System Summary

The Sevilleta Wildlife National Refuge LTER site (SEV) stores their data and metadata in a structured filesystem. SEV also uses a relational database to manage their website and metadata at the project level. Currently SEV offers rich EML for about 70% of their total projects.

LTER sites with similar management systems

- some similarities : BNZ, HBR, LUQ
- potential synergies : MCM, PAL, CCE, MCR

The Sevilleta Detailed

At SEV, a customized mySQL relational database stores project information, mainly a description, theme, keywords, and pointers, (paths, URLs) to the related datasets. The actual details on data and metadata are stored in a project-structured filesystem on flat text files format. Projects, datasets and metadata can be queried, browsed and viewed using Pre Hypertext Processor language (PHP) scripts. Rich EML is available for about 70% datasets, and currently maintained manually. The site is working with Inigo San Gil to generate a Perl script that will transform metadata entered into a Word template into EML.

Nearly all metadata at the Sevilleta was written by the Sevilleta employees, usually the field crew who take the data on the core studies. A few grad students have submitted data and metadata, but they are the exception, not the rule. There are two PIs who have submitted metadata during the seven-year tenure of the current IM.

EML Status:

- Completion: About 60% is available in EML.
- Richness : Nearly all EML documents are very rich, to the attribute level (levels 5)
- QA/QC : There are mechanisms to ensure the quality of the metadata and data ??

Related Links

- Data Catalog (1449) [88]

Short Grass Steppe Metadata System

The Shortgrass Steppe LTER Metadata Management System Summary

The Shortgrass Steppe LTER site (SGS) their data and metadata in a relational database. Spatial data are well organized and accessible via multiple tools. The SGS offers EML for a number of their datasets, and current efforts will complete the legacy metadata conversion

LTER sites with similar management systems

- some similarities : BNZ, SEV
- potential synergies : MCM, PAL, CCE, MCR

The Shortgrass Steppe Detailed

The SGS information system includes a centralized data server for raw and in preparation data (organized as a data file management system with hierarchical project and user directory structure), an archive of finalized data and metadata organized in a relational database, and an archive of GIS products. A revised database schema that has been designed and prototyped should correct a few shortcomings and facilitate implementation of EML. Spatial data are well organized and accessible via multiple tools (web/HTML, Java applet and ESRI tools).

EML Status:

- Completion: 25% of the current and long-term data sets are available in EML.

- Richness : 10% of the current and long-term data sets have been delivered in EML at attribute level (levels 5). Work is in progress to create the rest.
- QA/QC : There are manual mechanisms to ensure the quality of the metadata and data. Automated QA/QC protocols are used for climate and vegetation cover and density data.

Related Links

- Data Catalog (1416) [89]
- Information Management Report (1301) [90]

Virginia Coast Reserve Metadata System

The Virginia Coast Reserve Metadata Management System Summary

The Virginia Coast Reserve LTER site (VCR) their data in a structured filesystem and the metadata in a database. Rich metadata is viewed in a friendly format and also as raw EML, which is generated from our metadata database and dynamically. In addition to being available for search via Metacat, EML metadata is used to generate SAS, SPSS and R programs for the underlying data and to provide a map of associated research sites for each data set. Some datasets recieve additional attention. Meteorological, tide and ground-water data are automatically updated several times each day and web-accessible text and graphical reports are generated. Web-cameras provide real-time views of selected research sites, with hourly archives that can be queried and viewed as animations or photo index pages. Mapserver and Google Maps applications provide specialized Internet maps showing time series of landscape change, research sites and data collection locations.

LTER sites with similar management systems

- some similarities : FCE, ARC, AND
- potential synergies : HBR, SEV

The Virginia Coastal Detailed

The VCR Information Management System focuses on end-to-end information management, from the planning of a project, through its inception, to the collection of data, the creation of metadata and the dissemination of data. A large number of the pages incorporate interactive features that allow users to edit or add to databases. The metadata database is populated through a set of password-protected web forms that capture information at the level of the project, the specific dataset and the variable. The metadata database is also linked to the site personnel database, so that site personnel pages list projects and datasets, and the data pages list current addresses etc.

The technologies and products that support the system include the Apache Web Server, MySQL and MiniSQL databases, the PostNuke content management system, the Mapserver online GIS tool, the Gallery image management tool, the SPSS, SAS and R statistical packages, and language support including PERL, PHP, C and JAVA.

Quality control and assurance are primarily the responsibility of the data and metadata providers. However, several widely-used datasets such as meteorological, tide and water level data are subject to graphical and statistical summaries that help spot and correct errors.

A PERL/DBI program that queries data from our metadata database and dynamically creates EML documents for all VCR datasets. Stylesheets are then used to convert the EML into SAS, SPSS and

R statistical programs and to provide a map of research locations for each dataset. VCR is actively working on streamlining the data ingestion and improving data storage to aid integration.

EML Status:

- Completion: All metadata is available in EML.
- Richness : All EML documents are very rich, to the attribute level (levels 5)
- QA/QC : There are mechanisms to ensure the quality of the metadata and data.

Related Links

- Data Catalog (1394) [91]
- Information Management Synopsis (1309) [92]

SiteBytes

The SiteBytes are short annual updates of IM related activities at each site.

2013

Site Bytes are intended as a general update on what has happened at your site during the last year. Please highlight new developments, ideas, and issues. Site bytes help us to stay informed about what is going on at other sites.

There is no theme this year. You could highlight new software or recent activities, a new IMS hire, or bring up ideas and issues. You could use it for the abstract of your demo or poster. Short is best - one or two paragraphs is plenty.

Andrews Forest SiteByte 2013

Don Henshaw, Theresa Valentine, Suzanne Remillard

The Andrews LTER has made significant headway in using the NSF supplemental funding to address data availability at our site. We proposed to make improvements to our Information Management system to assure all basic expectations for metadata and data are met and that certain data sets and workflows are developed to advance NIS functionality. Additionally, we are preparing for our next renewal proposal (LTER7) by prioritizing data updates and reviewing the databases and website for necessary improvements.

A major concentration of our improvements has been on the development of PASTA-ready data sets including the improvement of titles, abstracts, and methodology descriptions. As a means of improving these descriptive metadata elements, the Andrews IM Team conducted two “metadata parties” and participants included more than 10 PIs, professionals, and grad students. While the IM Team provided technical support and guidance, participants used our administrative interface and plugged away rewriting abstracts and improving other metadata content for two hours during each “party”. The results were superb. It turns out that while PIs will not initiate writing metadata in their free time, they will do this (and they are good at it!) if they can schedule specific time on the calendar to do this and have an IM available to guide the way.

The inclusion of data sets into PASTA has been delayed while we develop a dynamic versioning and archive system for our data sets and metadata, but we have begun uploading data into PASTA. We hope to include at least 120 data packages in the next few months. All of our EML is generated programmatically using xslt-scripts called from within a .NET framework that transform native xml from our metadata database into EML. Another program checks current data entity and metadata versions of all data packages and locally archives data as .csv files and metadata as EML whenever new versions are created. These tools have eased the burden of preparing and enhancing our data for PASTA.

Spatial data sets are being prepared into data packages compliant with best practices, and EML is generated through the esri2eml style sheets. LiDAR derived GIS layers for the Andrews are now available online and uploaded to PASTA. There has been extensive testing and revision of the esri2eml style sheet and the “Best Practices for Documenting Spatial Data” whitepaper. The LTER Landsat catalog for LTER sites is nearing completion and soon the images will be available and searchable through PASTA. Atmospherically-Corrected images are being processed and will soon be available through PASTA; thanks to a Post-Doc award given to Tom Spies, Andrews PI. Many of the Andrews study site locations have been geo-referenced, providing geo-coordinates for EML generation, and web mapping.

The IM Team has also been at work streamlining the processing of our major data collections – climate and vegetation. Streaming climate and other data collections continue to grow as phenology and carbon studies are taking advantage of the enhanced communication network within the forest. Adam Kennedy (AND) is developing a workflow for handling streaming sensor data that features Campbell Loggernet software routing real-time data into our SQLServer database and then subsequently ingesting into the GCE Data Toolbox for quality checking and data flagging. These flagged data are then pushed back into the database and the IM Team will be mapping these data into final files for online posting as provisional data, which will occur daily when this workflow is completely developed. Additionally, we are developing a new system for processing and archiving our vegetation datasets to increase efficiency, provide consistency and improve data quality. We hope that a more generic solution will accommodate our every growing use of sensor network data collections within the Forest. Mark Harmon, Andrews PI, is leading the charge to develop Veg-E and we are working towards the ability to provide the necessary datasets. We hope that Veg-E will provide a framework for other synthesis projects and that ClimDB/HydroDB and StreamChemDB can take advantage of these developments.

Arctic SiteByte 2013

content here.

Baltimore Ecosystems SiteByte 2013

At BES we have a mid-term review upcoming in October.

We should have PASTA compliant metadata by next week. It looks like we are pretty close. We have the information in a database and a script to read it and transform it. This is similar to our existing system only now the database and script have a lot more logic to include attribute data and access data. This system will be robust and in the long run, a lot less work to curate and present the metadata. Going forward we plan to integrate the attribute information with the datasets directly for some datasets. For example, our stream chemistry data are kept in a database, and exported for sharing. When this database and the metadatabase can interface each other, the attribute information for the stream chemistry data that changes with time - Such as, number of rows - will be

continually updated, and the PASTA system can be automatically updated upon any change thereof.

We have applied for Google Apps For Education. Prior to that, we began using Google Apps for various things, and continue to use them more and more as time goes on.

- We now make all of our "dynamic" web pages using Google Sites. That allows for easy editing by the author directly.
- Google Calendar has turned out to be much more useful than it appears on the surface. For instance, we have found that calendars can be associate with objects as well as people. Meeting rooms can have their own calendars, and so can vehicles, labs, equipment, and offices!
- We use Google Drive for more and more documents as time goes on. It's an elegant, straightforward way to share. For example, if someone wants to share a document, one simply has to link to it, and the notion of "attaching" the document is no longer needed.
- Online help is very good for Google Apps. It is extremely well documented.

If we are approved for Google Apps For Education, we will be allowed a great deal of storage. If that is the case, we expect to use Google Apps for pretty much everything.

Bonanza Creek SiteByte 2013

California Current Ecosystem SiteByte 2013

Cedar Creek SiteByte 2013

Cedar Creek has made a major push to identify, collect, and process lost and poorly documented historic datasets. As a part of this effort, we have added Susan Barrott to our information management team. She will also working to facilitate the interaction between the research and IM personnel, with the goal of more complete data and metadata collection, and to interpret research for distribution to various audiences. Susan has worked at Cedar Creek since 2001 managing both research and data. We are excited to have her on the team.

Many of our efforts this year have also revolved around PASTA preparedness. We have used this as an opportunity to revisit our metadata and data quality. We have also restructured our database to better reflect the needs of EML 2.1 and for general efficiency. Scripts to create EML have been rewritten. Over 225 of our datasets have been reprocessed and uploaded into PASTA. We anticipate another ~150 will be ready for upload within a month.

Through the fall and early winter of 2012, a good portion of the Cedar Creek web site was converted to Drupal. We have also replaced our internet connection with a dedicated fiber line. When this goes live in the coming weeks, we anticipate a number of changes to our IT systems.

Central Arizona Phoenix SiteByte 2013

IM activities at Central Arizona-Phoenix for the last 12 months have been quite productive, as we continue to work towards our upcoming 2013 midterm site review. Members of the Global Institute of Sustainability Informatics team have been working towards PASTA compliance by migrating our existing data inventory (55 so far!) while also improving our dataset publishing process. We have also reached out to the broader data management community at Arizona State University this year.

Our team is working with the university library's digital repository to be an authoritative source for our research data. We are also involved in a tri-university initiative for the Arizona public universities to better collaborate on many aspects of research data management.

This year has also seen some big changes in our server architecture; we have expanded our capacity, security and new features to support our websites and data management. We are currently testing a PostGIS backed Geoserver that serves our geospatial data over OGC webservice standards. This visual presentation will become an integral part of our data set previews on our website. We have also implemented an early proof of concept prototype of our Virtual Notebook and deployed it to a test group of users. We are using their feedback to guide our future development efforts.

Coweeta SiteByte 2013

Florida Coastal Everglades SiteByte 2013

FCEIII IMS Activities for 2013

The major focus of the FCE Information Management (IM) team (L. Powell and M. Rugge) has been the implementation phase for a FCE IMS physical hardware restructure and improving its network-wide standardization to facilitate increasing use of site data in synthesis projects. The the FCE IMS team has made both systematic and procedural changes to its information management system (IMS) during the first year of FCEIII and highlights include the following accomplishments:

• Completed major information migration of the FCE program’s project and research data from FCE physical servers located in the FCE office to five (5) virtual servers housed on the Florida International University Division of Information Technology’s (UTS) equipment.
• Upgraded the FCE Oracle 10g database to Oracle 11g Enterprise version.
• Upgraded FCE website content to reflect FCE III research and changes made to the FCE Information Management System.
• Upgraded FCE Excel2EML metadata tool and template from EML 2.0 to EML 2.1..
• Converted ALL existing FCE Ecological Metadata Language (EML) 2.0 metadata to EML 2.1 and enhanced metadata content by implementing the LTER controlled vocabulary list.
• Made a procedural change in the FCE IMS whereby the practice of data ‘versioning’ was discontinued to better follow the LTER community practices and to facilitate data submissions into the LTER PASTA system.
• Re-packaged 525 original FCE datasets that included versions into 138 ‘primary’ datasets.
• Working on changes to the Oracle 11g database tables to reflect recent changes to the FCE data archives.
• Submitted ALL FCE program data, with the exception of 10 dissertation research datasets, into the LTER PASTA system and made appropriate changes in the LTER Metacat database to match the FCE ‘re-packaged’ data.
• Collaborated on a custom iOS application to facilitate exploration, manipulation, and annotation of long-term ecological data signals on a mobile platform. A web-based version of the application was completed.

Future Work

• Begin work on website redesign of the FCE Data section that will include a web version of the FCE data-processing visualization tool (graphing of FCE data via the web).
• Complete populating FCE Oracle11g database data related tables with ‘repackaged’ dataset related information.

• Add newly changed FCE Database Schemas to website under ‘Information Management’ section.
• Expand FCE GIS data in the FCE IMS database and online.

Georgia Coastal Ecosystems SiteByte 2013

This has been a very atypical year for the GCE IM program that started with more staff turnover. Travis Douce left GCE for another job in 2012, and after losing 3 GIS professional in 4.5 years we decided to revise the assistant IM position to emphasize data and metadata processing rather than GIS expertise to reflect current priorities. After a lengthy search with many stops and starts, Adam Sapp joined GCE as the new assistant IM in November 2012. Adam brings extensive marine data processing expertise to GCE, from his post-undergrad work at the Louisiana Universities Marine Consortium, graduate work in coastal engineering at Georgia Tech, and post-graduate work at Savannah State University managing their field research data. Along with this transition in assistant IM role we have increased our emphasis on formally documenting IM protocols and developing training curricula for both IM and research staff to broaden expertise and responsibilities for data and metadata management across the project. Since the inception of the GCE program, the lead IM (W. Sheldon) has been responsible for the vast majority of data and metadata processing as well as software development and systems administration, but this model has become increasingly unsustainable as the IM/IT landscape and expectations have become more complex. This has been a difficult and at times tedious maturation process, but it has finally allowed us to transfer responsibilities for many routine data management to the assistant IM and other staff, giving W. Sheldon time to pursue new projects.

The other major change this year is that GCE and CWT received supplemental funding from the LTER Network ARRA grant to help other sites adopt GCE Information System technology (e.g. GCE Data Toolbox and Metabase Metadata Management System), allowing Wade Sheldon to devote 3 months to network-scope software development and support. An additional 1 month of support was provided from an NSF SI2 grant to Tony Fountain and Corinna Gries to work on improving MATLAB support in the Open Source Data Turbine streaming data middleware, and add Data Turbine support to the GCE Data Toolbox to facilitate streaming data documentation and management. This dedicated funding for software development and training led to a quantum leap in functionality and usability of GCE-developed software this year, particularly the GCE Data Toolbox for MATLAB, and greatly facilitated adoption by several other LTER sites and research programs outside of LTER. These activities will be described in a briefing at the 2013 IMC meeting.

Focus has now shifted back to core GCE IM work, where we are working diligently to catch up on archiving our backlog of monitoring and study data, as well as establishing remote communications systems to acquire and manage data from our new eddy flux tower now operating at Sapelo Island. We have already staged all accessioned GCE data packages in the PASTA staging server, and will move these to the production server by August 2013 after testing a few additional protocols as we transition from automated Metacat harvesting to a GCE-managed PASTA synchronization scheme.

Harvard Forest SiteByte 2013

Harvard Forest highlights for 2013 include:

1. PASTA. We completed a series of upgrades to our EML metadata and entered roughly two-thirds of our datasets into PASTA. We expect to enter the remaining third by early fall.

2. EML creation. We’re working with a small consulting firm (Kelly McCreary & Assoc) to develop XForms for direct creation of EML files by scientists.

3. Schoolyard. A new online database system (Apache-MySql-PHP) was created for our Schoolyard LTER program. The new system is used to enter, edit, download, and graph data.

4. Infrastructure projects. Major projects included replacement of a mile-long buried high voltage electrical line for the new NEON tower, design and siting of two new research towers supported by an FSML grant, and addition of a renovated building to the campus network.

5. Provenance. Research on data provenance was continued through ongoing collaborations with computer scientists and the Harvard Forest REU program. Recent efforts include instrumentation of R scripts to create and store provenance metadata and enhancement of associated query and visualization tool.

Funding for items 1-3 was provided by a 2012 LTER Supplement.

Emery Boose

Hubbard Brook SiteByte 2013

Jornada Basin SiteByte 2013

The Jornada has been busy for the last year updating documentation and data in preparation for populating all long-term Jornada datasets into PASTA. This has included converting our data files from fixed format to comma-separated value format and updating metadata to be congruent with the data it describes. All long-term datasets are now in the LTER Metacat with full attribute level EML and contain direct links to our dataset packages and the individual data files. We plan on processing the remaining Jornada datasets in the same manner once we have migrated to DEIMS v2 so we can take advantage of the much improved metadata entry forms.

Full attribute level EML are now being generated for all Jornada long-term datasets from DEIMS (Drupal Ecological Information Management System) metadata content. Dataset packages and files were manually created and uploaded to the DEIMS-driven data catalog after concatenating data into a single file per dataset which were then converted to comma-separated format and packaged with related GIS shapefile packages that were created by Barbara Nolen, who retired last year after ~20 years of service to the Jornada Basin LTER program. All EML metadata generated from the Drupal 6 based DEIMS has been successfully harvested to the LTER Metacat server.

Ken Ramsey created views of the DEIMS metadata to create the new Jornada data catalog (http://jornada.nmsu.edu/data-catalogs/jornada [93]) and long-term datasets (http://jornada.nmsu.edu/data-catalogs/long-term [94]) web pages. Further modifications to the data catalog and long-term dataset pages have been postponed until the metadata and data have been migrated from Drupal 6 to 7. The same goes with improving EML congruency, we decided to wait for DEIMS v2 to be released. This gave us a few months to prepare for migrating our websites and data to Drupal 7 and to upgrade our server virtualization environment and storage area network.

The current Jornada data catalog and image gallery were migrated to separate websites and integrated with the Jornada website using iFrames in preparation for migrating the remainder of the web content to Drupal 7. Ken Ramsey and Jim Lenz have also been consolidating storage arrays in our SAN to allow Jim to manage the Jornada's storage growth more efficiently. Jim upgraded our hypervisor servers to XenServer 6.1, which was not painless. A few bugs were encountered during the upgrade, which prompted the purchase of support from Citrix. We were pleasantly surprised that the pricing for Premier support has gone down significantly. We were able to license Premier support for for 6 seats of XenServer Enterprise for less than the cost of a single technical support

incident! This will be a valuable resource for Jim and help in keeping Jornada virtual servers available 24x7x365.

Ken Ramsey, Jim Lenz, and Valerie LaPlante are currently migrating the Jornada website (http://jornada.nmsu.edu [95]) from Drupal 6 to 7. After migration to the new Drupal 7 based DEIMS, we should be generating fully PASTA compliant attribute level EML that can be used to populate the LTER Data Portal with all long-term Jornada datasets. The timing of the Information Managers Meeting could not be better for Ken to learn more about deploying the new DEIMS and methods to migrate data and metadata content from Drupal 6 to 7.

Kellogg Biological Station SiteByte 2013

We had our mid-term review this year with the resulting flurry of dataset update activity. As part of the review preparation we put together a "private" score card http://lter.kbs.msu.edu/score_cards [96] to locate data tables that were incomplete or had data missing. We started to populate PASTA using a test and push approach https://github.com/kf8a/noodle [97]. Otherwise no significant new features were deployed.

Konza Prairie SiteByte 2013

Luquillo Experimental Forest SiteByte 2013

In this two years period of LUQ's LTER 5, we have dedicated most of our efforts to enhance the metadata to be PASTA ready. Using LUQ DEIMS' EML module we have been able to make about 75% of our data sets PASTA ready so far. The goal is to convert 100% of all our data sets by the end of the year. As we do this, we are also making sure that all data sets are LTER VOCAB keyed and database ready.

McMurdo Dry Valleys SiteByte 2013

Moorea Coral Reef SiteByte 2013

This past year since the 2012 ASM has seen a lot of progress. Last November, at the annual All Investigator's Meeting, the graduate students requested training in data management. We used a hands-on exercise prepared at the KBS LTER, a powerpoint from VCR LTER, and an original query activity. We opened by posting questions such as "what would you need in other's datasets to be able to use them" and let the graduate students give their answers. Then they recognized some of their ideas in the prepared materials. This gave them confidence that this IM stuff is not from outer space.

Last January we put all our unrestricted datasets into the production pasta. We have been synchronizing updates to those datasets in both pasta and metacat. The restricted datasets, although they do pass the checker, are for the present only in Metacat until the new portal offers group authentication and clarity in presentation to data users. Just this July we implemented our first pasta-subscribed workflow which deposits its derived data product back into pasta.

The ProjectDB application as implemented at SBC was installed for MCR and we began entering the project information. We like the presentation and organization of research project data and the way it mirrors ProjectDB at other sites. Although not completely populated, the site reviewer got to see it and saw this as a good direction.

The data package status inventory tracking system, although already in use a year ago, has matured and has been a significant part of how we have increased efficiency this past year. Web views have been tailored to each audience, from field tech to site PI to site reviewer. Inventories for annual reports, site renewals, and site reviews were generated as static snapshot outputs from this tool.

Part of the 2012 supplement funded an assistant, 10 to 15 hours per week, shared with SBC. This has helped us two ways. The more obvious way is the time they put in updating datasets and A fortunate side effect of training assistants is that now our written materials documenting our processes and tools are more complete and organized. Excellent advice from the CWT IM was to have the assistant write their own version of the instructions. By reading this, it becomes apparent what was not clearly explained.

Niwot Ridge SiteByte 2013

North Temperate Lakes SiteByte 2013

Palmer SiteByte 2013

Plum Island Ecosystems SiteByte 2013

Santa Barbara Coastal SiteBytes 2013

Activities at SBC during the past year have focused on the use of Metabase for datasets. Two years ago, we began using Metabase to store research theme information, and developed proof-of-concept export scripts. More recently, we have focused on population of Metabase for datasets, and on developing code for exports using object-relational mapping (ORM). SBC has collaborated with Moorea Coral Reef (MCR) LTER on all aspects of Metabase work. The benefits of this collaboration are many: we take advantage of our complementary skills; we combined supplement funds at one institution to hire a shared programmer; and a priori, all code is planned to be reusable, and by multiple LTER sites.

We are populating Metabase from our existing high-quality EML, and first pull a current snapshot with a "reverse-harvest" from Metacat. The following pattern has developed.
1) In Metabase, identify a group of tables to be populated, and in EML, the corresponding nodes.
2) Use XSLT to extract the pertinent information from the corpus of EML into text tables, and upload to ‘scratch’ tables that closely resemble the targeted Metabase tables.
3) Regularize content within the scratch table(s) as needed, and apply keys, indexes and constraints.
4) Insert from the scratch table(s) to the production Metabase table(s). We have taken advantage of the PostgreSQL XML-type to hold complete EML nodes in some cases.

The code to export EML from Metabase is being written this summer by a programmer shared by MCR and SBC. As of this writing, we are exporting fully described dataTables with dataset-level metadata that is ~80% complete (by node, when compared to the original EML). We chose Perl because it is familiar, and has full suite of ORM libraries and a flexible template system. The code is organized into modules to provide encapsulation. The “Metabase.pm” handles queries to the database, and “EML.pm” contains the classes to populate an EML document. In the database, we added views to provide an abstraction layer, which allows the code to be used with data models other than Metabase. All aspects of this process – extraction of EML content, Metabase table population, EML creation – will be further illustrated or demonstrated at the IMC meeting's demo session.

Sevilleta SiteByte 2013

Short Grass Steppe SiteByte 2013

Nicole Kaplan and Bob Flynn continue to work with SGS-LTER Researchers on information management in support of data integration, QAQC and metadata documentation to facilitate publication of scientific findings in peer-reviewed journals and data accessibility through the LTER Data Portal and the Digital Collactions of Colorado. We established objectives to improve data access and metadata documentation in accordance with the recommendations in the 2012 LTER revised guidelines (network communication, LTER Executive Board) and to preserve the local knowledge of scientists, staff and students who have worked on the SGS-LTER research site. We are statisfying data access requirements by assuring 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. The SGS-LTER data packages are interoperable with the network infrastructure and PASTA Framework. As the project nears its end, it is a good time for reflection, so we are collecting and sharing stories from people working on the SGS-LTER project. We have interviewed over 30 individuals, scientists and support staff who have worked on the shortgrass steppe with the LTER project over the course of their career. Their experiences and perspectives represent an important pool of knowledge about the site and working collaboratively. Nicole is collaborating with Helena Karasti (Finnish LTSER Network), an interdisciplinary scholar located at University of Oulu, Finland and Luleå Technical University, Sweden, who has worked with LTER networks since 2002, on incorporating these stories into publications and presentations. Different perspectives provide insight about conducting interdisciplinary science over the long-term in the shortgrass steppe and historical features of the LTER Network. It is also an opportunity to recognize formally contributions SGS-LTER has made to science and to the LTER Program.

Virginia Coast Reserve SiteByte 2013

Virginia Coast Reserve Site Byte – 2013
--John Porter

Another busy year at the Virginia Coast Reserve Long-Term Ecological Research Project. Areas of innovation include adding data to PASTA the continuing process of workflow development for additional datasets, development of a system for linking LTER publications with LTER data and integration of spatial vector and spatial raster data types into VCR/LTER EML Metadata. Needless to say, much of this would not have been possible without the addition of David L. Richardson as an Information Management Programmer (using supplemental funding).

Addition of datasets to PASTA has involved both technical and social innovations. For years we have used the LTER Network Data Access Server (DAS) to serve our online datasets. One of the features of this service is that it notifies data providers when data is downloaded. PASTA takes a different approach – providing logs of downloads, but no active notification service. Additionally, while the DAS always requires a login or authentication via “cookies,” PASTA can distribute data publically – providing no feedback on who is downloading data. To meet the needs of our researcher community, we set up web-form-based system that allows investigators to specify the level of authentication required by users downloading data. Roughly 2/3 of VCR/LTER data will be “public” – requiring no authentication. The permission-setting process is nearly completed, and mass uploading of data to PASTA will progress late this summer and early fall. Approximately 20 “test” datasets have been already uploaded to confirm metadata validity for use with PASTA.

Additionally, we added an extensive set of headers (~20 lines) to each data table. The purpose of these headers to assure that recommended citations and pointers to metadata accompany each download. They were required because we had a couple instances where data users downloaded data tables via the LNO portal without opening the metadata or used data that had been downloaded by another, and then misattributed the source of the data or failed to comply with data use policies.

We have also improved the “PASTAprog” web service (http://www.vcrlter.virginia.edu/webservice/PASTAprog [98]), which generates statistical code from EML metadata served from PASTA. The improved version for the R statistical language is much more tolerant of attribute-naming conventions (attribute names that include most mathematical operators no longer cause a failure), and data type checking is improved.
David Richardson, in conjunction with the GEONIS working group, has developed a system for producing XML stubs containing spatialRaster and spatialVector EML entities that can be incorporated into the program that generates EML metadata for VCR/LTER datasets. The entities combine the spatial metadata inherent in GIS data structures, such as shapefiles, with other EML elements. Full integration is expected in the late summer or early fall of 2013.

We have established a database to link LTER publications with related datasets. Given the large number of both publications and datasets, we expect that full population of the database will take a long time, so we are prioritizing more recent publications. Our hope is that this database will help remind investigators of the importance of promptly adding datasets to the VCR/LTER database. In the long-run, we anticipate that the use of Digital Object Identifiers (DOIs) from PASTA will allow the process of linking to be automated.

We are in the process of developing a system that couples a database of known sensor problems with code and report generators that provide problem reports, and also SAS and R code to address the issues identified in the database. The code-generators are web-accessible so that code for transforming Level 0 (original data) to Level 1 (corrected data) can be included into processing programs as simple includes. Testing with two long-term datasets of meteorological and precipitation data have been successful. A poster on this system will be presented at the meeting. Other sites interested in collaborating on development of a multi-site system are encouraged to contact John Porter (jporter@lternet.edu [99]).
It has also been a productive year for publications dealing with ecological data, including co-authorship on two BioScience publications and a book chapter on sensor networks.

2012

Site Bytes are intended as a general update on what has happened at your site during the last year. Please highlight new developments, ideas, and issues. Site bytes help us to stay informed about what is going on at other sites.

As in previous years, in 2012 we are focusing the site byte on our use of recent supplement funds. **How are you addressing data availability issues at your site; especially in light of the announcement that PASTA will be in production during 2012? Particularly, what are your plans for any supplemental funds you may have received?**

Andrews Forest SiteBytes 2012

Don Henshaw, Suzanne Remillard, Theresa Valentine (September 2012)

We plan to use supplemental funding to make improvements to our Information Management system (IMS) to assure all basic expectations for metadata and data are met and that certain data sets and workflows are developed to advance NIS functionality.

1. *Metadata development*: EML files for Andrews LTER data are generated directly from our metadata relational database through scripting programs and are regularly harvested into the LTER NIS Data catalog. Some of the necessary improvements and/or modifications to both assure completeness of metadata content and compliance with EML Best Practices (EML_BPv2) include the following:

- Site PIs have agreed to meet in a retreat-like setting to focus on metadata completeness, particularly improving data set titles, abstracts and keywords as well as data set methods and identification of related publications.
- All automatically generated EML data packages will be validated through the PASTA Quality Engine.
- Andrews theme keyword controlled vocabulary will be mapped into the new LTER standard controlled vocabulary. This mapping will assure that LTER standard keywords will be placed into EML as a means to improve the consistency of network data portal searches for LTER data.

2. *Core data development and improving data access*: In preparation for future network-wide data integration, several core data sets will be reorganized, updated, and quality assured. The goal is to improve data access to quality data through consistency of metadata descriptions, improving overall quality, standardizing attributes across data sets to facilitate application programs and analysis, and building derived products suitable for sharing with cross-site integration efforts including StreamChemDB and VegDB. Information Management System refinements will be made to assure quality and to remove physical barriers hindering fast access to site data sets.

3. *Sensor data processing*: Last year's supplemental funding resulted in a prototype for capturing meteorological data streaming from a few climate stations and performing near real-time Q/C checks. Modifications to the prototype are necessary to put this system into production for most climate and streamflow stations. The system will capture streaming data, perform preliminary quality checks and data flagging, archive into the database and post provisional data online on a daily schedule. As our hydrometeorological program collects 7 million data points per year, this system will be a huge benefit to our overall efficiency in making these data available both locally and routinely harvesting into PASTA.

4. *Spatial metadata development*: Improvements in the process for capturing metadata for spatial data sets are necessary. Stylesheet transformations (XSLT) have been used to convert FGDC metadata into EML 2.1. Changes in the ESRI metadata library (toward an ISO standard) complicated this process and require rebuilding of these scripts. The new scripts will be easily modified when the FGDC implements the ISO standard for metadata.

5. *Advancement of NIS functionality*: We will assure accessibility though network PASTA architecture (e.g., PASTA-ready) for core long-term data including stream chemistry data from eight Andrews watersheds and tree and understory vegetation data from Andrews Reference Stands. The Andrews will be coordinating data development with other participating sites and demonstrating PASTA functionality for stream chemistry (StreamChemDB). For VegDB we will be investigating exchange standards for future cross-site integration of vegetation data. Also, we will be assuring inclusion of our spatial data in the NIS through GeoNIS participation.

Arctic LTER SiteBytes 2012

SiteBytes go here

Baltimore Ecosystem SiteBytes 2012

We have been updating our website to a significant degree to reflect the changes in our program. As of our latest renewal our research questions have expanded, and therefore we have to reflect that on our public www page.

To this end, we are using Blogspot more and more since it is so easy to work with. The BES news, announcements, employment opportunities, and Director's Corner are all now facilitated with Blogspot.

My main endeavor at this time is enriching our metadata to be PASTA compliant. In order to do so I need to include file attribute information in our metadata. I am expanding the structure of our database to include tables of attribute types and descriptors. I hope to be able to identify attribute descriptions that are common to many files and be able to pull the related information based on an identifier. This should make it possible to use a single attribute description set for all similar datasets. Stream chemistry, for example, in BES, is represented by several datasets that share identical attributes – they are comma delimited lists of dates and analysis results. Enriching our metadata in this way is necessary for reasons other than PASTA compliance as well, as it allows for better synthesis with other datasets.

We received supplemental funding to expand our stream chemistry database and as part of that process we will be adopting a means to track the trips to the field with regard to equipment used, personnel, calibration and other such data. Knowing this information could be useful in the future for QC/QA purposes.

We now have a searchable online database of our publications including links, where available. This was facilitated by creating a custom export template in Endnote and writing a script to display the results.

I will add to and edit this SiteByte as the deadline approaches. See you soon!

Bonanza Creek SiteBytes 2012

SiteBytes go here

California Current Ecosystem SiteBytes 2012

This past year, Information Management at CCE LTER has been focussed primarily on reviewing the
current state of data management at our site, with a focus on identifying areas of weakness and developing plans for improvement. Recent changes have been directed by the newest guidelines set forth by the Executive Board, as well as the our readiness to work with PASTA in the approaching
new year. This September also marks the fifth month for Scott Gordon, our new Assistant Information
Manager, working in our group.

With the supplements we've recieved for the next year, we're planning to address data availability issues by hiring additional personnel for a 12-month period. We're using the results of internal data management reviews to develop plans for these individuals to assist information management in improving both data availability and quality. Our initial focus will be on improving metadata structure and content as well our NIS publishing protocols in order to provide PASTA-ready data submissions to the network.

Cedar Creek SiteBytes 2012

SiteBytes go here

Central Arizona Phoenix SiteBytes 2012

The last 12 months have proven to be another busy period at CAP LTER. The CAP Information Manager, supported by the new Informatics Team at the Global Institute of Sustainability, has been able to make good progress in growing our data management capabilities. A complete overhaul of our existing metadata inventory was submitted to Metacat improving our EML compliance and generally increasing our metadata quality.
Other improvements have been made to data entry functions as well as beginning the phased replacement of our project database that we use to keep our website (and ~21 others) up to date. This new database will eventually hold metadata as well, allowing us to create new relationships between projects, datasets and the publications that arise from them.
Over the summer we have conducted a data audit on all our spatial data and have made improvements to the general quality (standardizing projections and checking metadata) as well as repackaging to make sure tabular data and relevant spatial data are kept together.
Finally, we are developing a data entry application for our new shade tree project. This will allow citizen scientists who plant shade and fruit trees in the Phoenix metropolitan area to provide growth and health data of their prized tree specimens.

Coweeta SiteBytes 2012

SiteBytes go here

Florida Coastal Everglades SiteBytes 2012

SiteBytes go here

Georgia Coastal Ecosystems SiteBytes 2012

SiteBytes go here

Harvard Forest SiteBytes 2012

SiteBytes go here

Hubbard Brook SiteBytes 2012

SiteBytes go here

Jornada SiteBytes 2012

The Jornada LTER (JRN) participated in the SensorNIS, GeoNIS, and DEIMS cross-site workshops using last year's LTER information management supplement. This year JRN did not request supplement funds to prepare the data and metadata for the LTER NIS since our preparation efforts are already funded from existing site funds and those efforts were already underway. JRN plans on all long-term Jornada data being included in the LTER NIS with attribute level EML by November 2012. JRN also plans on supporting our participation and development efforts related to the GeoNIS and DEIMS working group using site funds.

The information management team has been preparing our spatial and tabular data and associated metadata for inclusion in DEIMS (Drupal Ecological Information Management System) and the LTER NIS. The new Jornada data catalog (http://jornada.nmsu.edu/data-catalogs/jornada [93]) is powered by DEIMS content.

We are currently loading all Jornada spatial data and metadata into an enterprise geodatabase and open source geoportal. Shapefile packages will be available for download from the geoportal initially, and from the data catalog (DEIMS) later. Concurrently, we are creating combined data files for each Jornada dataset and converting the combined data files from fixed format to comma-separated value format. We are also updating our documentation forms to PDF format and packaging dataset files (data, EML, PDF metadata, shapefile package) for download from the Jornada data catalog, geoportal, and the LTER Data Portal.

We plan to migrate from the old static Jornada data catalog by the end of November once all Jornada dataset packages have been created. We are also working to improve the quality and content of EML generated from DEIMS before November so the EML will describe the dataset package contents including direct links to the data and shapefile stored on the Jornada website.

Before November we hope to create initial map and image services using the enterprise geodatabase contents and populate the geoportal with all Jornada tabular and spatial data (packages for download, live map and data services for use). In November we plan to begin automating the Jornada information management systems processes, refining data table query capabilities of the data catalog, and to load and relate all Jornada tabular data with geographic research site locations in the enterprise geodatabase to support geospatial map and tool development by Jornada research collaborations.

It is too soon to identify information management related post-ASM activities or site supplements needed to support site information management efforts next year, but supporting DEIMS and GeoNIS working group efforts will be among them. After the LTER All-Scientists Meeting and the LTER Drupal training workshop in November, we should have a better idea of our site's information management needs for coming year.

Kellogg Biological Station SiteBytes 2012

We are using part of the supplement to orthorectify and bring online several years of air photos.

We are testing our current datasets against the pasta development server, we are trying to identify and bring 'dark data' online and improve the abstracts, protocols and descriptions in our datasets.

Konza Prairie SiteBytes 2012

IM activities at Konza for the past year have largely been spent working to respond to the 2011 midterm site review, general tweaking of pieces of the Konza IMS, and, of course, a heavy focus on updating EML and data. At the network level we have participated heavily in the conceptualization and development of the GeoNIS (including participation in the GeoNIS workshop), and attended SensorNIS as well.

Our most recent supplement was put towards personnel, with focus on assisting with EML, continued web-site maintenance (including value added projects for data access and QA/QC), and development of additional GIS data sets. Thus far we have hired a part-time student assistant to help vet/update existing EML, and a full-time (3 month position) GIS Technician to assist with the cleaning and documenting of several new GIS data sets. These data will better tie existing Konza data to specific locations of their collection... something we did not previously offer. Hiring of additional personnel is targeted for after ASM.

Post-ASM requests will likely be targeted towards GeoNIS related activities. This may include a meeting to work out final kinks in our plans / meet with a developer, or training focused on working with the product once it is finished. The ideas are there in a general form but as usual, the face to face of the meetings will help revise these into something workable, and useful to the community.

Luquillo Experimental Forest SiteBytes 2012

Automation of the rainfall and temperature manually read data . A spreadsheet that allows the technician collecting the data to enter the data and obtain, automatically, a final table of daily values for the El Verde Field Station precipitation in mm (read in inches) and temperature in Celsious (read in Fahrenheit). The precipitation data requires to be averaged among the total numer of days without reading plus one, to account for the day that the reading is actually done. The spreadsheet developed is programmed to do this average automatically, when the data is entered as read from the field. The spreadsheet renders 3 separate worksheets containing daily values for : precipitation in mm, maximum temperature in Celsius, and minimum temperature in Celsius. It also contains QA/QC that compares manually entered and averaged data to the automatically calculated one. Web Page URL: http://luq.lternet.edu/downloads/precipitationandtempdataentryand-conversiontemplatexlsx [100]) which is accessible from: http://luq.lternet.edu/IM/IMSProtocols [101]

Enhancement of Data work flow from technician gathering the data to data entry person and finally to web. In collaboration with the investigator in charge (A. Ramirez), the field technicians

gathering the data (J. Bithorn, S Matta) and the data entry person (J. Monge), we developed a new protocol that allow the investigators in charge (W. McDowell, J. Porter) to download the stream, rainfall and throughfall basic chemistry data collected at the field in a relatively real time manner. The new protocol eliminates the use of messengers and uses the LUQ Intranet (the Plone: http://rambutan.ites.upr.edu:8080/itesplone/ [102]) to share data entry sheets and spreadsheets with the data entered. Scanned data entry sheets are mailed directly to the data entry person withing a 24 hours period and data is entered and uploaded to the Plone in a 2-days period after collection. The investigator has direct access to this data and downloads it directly for inspection and further analysis.

Enhancement and publication of the LUQ Information Management System (IMS) and Web site (http://luq.lternet.edu/ [103]). New dynamic and interactive website-IMS that produces EML packages on the fly, ready to be harvested by Metacat. Metadata at the attribute level, is being updated and revised to produce PASTA ready metadata.

Collaborated with Nick Brokaw and Ignigo San Gil in standardizing LUQ's naming convention for sites and the preparation of an online map that displays the list of LUQ online data sets by Research sites. Although still under development, the LUQ GMap is in place (http://luq.lternet.edu/luquillo_plots_map). The first activity was to standarize the name of all the research sites with LUQ's PI, and rendered a standard way to name any future research sites using the following general format: research area-regional name - plot name- research use (eg. El Verde Field Area Espiritu Santo Landslide 11 (ES-11) Intensive measurements). Next and still being done was to get the georeference metadata (Lat Long). The ultimate goal is to have all the plots georeferenced as much as possible with an GPS.

Continue the training of LUQ Schoolyard teachers. In the last year IM participated in two workhops given to the LUQ Schoolyard teachers to enhance their school's data catalog and metadata.

Participated in the designed and teaching of first LTER IM Course. Since 2010 a group of 6 IMs (NTL, VCR, SEV, AND, BES and LUQ) organized course material for teaching the basic principles of managing data. In August 2012 we gave the first course to a group of 21 persons, including 3 scientists, field technicians and information managers, and graduate students. This activity included the recording and editing of the presentations.

McMurdo Dry Valleys SiteBytes 2012

What has happened during the last year at McMurdo IM?

MCM entered a new funding cycle last year (MCM-IV). The MCM PIs decided before hand that the MCM-IM will be co-located with the SEV and the LNO, with the forward looking idea of leveraging resources and expertise. That served well! Just last week I found that I needed to install a "Windows Server" yesterday. Needless to say, an unused server appeared out of nowhere, and even a license for Windows Server. I never done that before, but I had helping hands (thanks Dez) to guide me through the ordeal.

New funding cycle, new location, new IM. Those three "news" summarize last year.

The big help came from Chris Gardner (big thumb up), Sue Welch and Tina Takacs-Vesbach

1) Get familiar with the IM-related protocols at MCM (read: get ready to change the tires while the car is driving at 60mph.) MCM IM speak:

MCM <-> IT World

Database <-> Information management system
Database person <-> Sysadmin
Database person <-> Web mistress
Database person <-> Information manager
Database person <-> Database administrator
Database person <-> Data entry person
Database person <-> Houdini
You get the picture

2) Set up a Sun T3-1 server running Solaris 5.10 (I have a blog with threaded posts under the header "Surviving Solaris - thanks Dez, Renee)
The lifesaver here is opencsw.com, the closest thing to Ubuntu's repo (apt-get install blahblah -- here is pkgutil -i blablah). T3-1 motherboard blew out on a friday :)

3) Migrate Java Server Pages from a 5 to 7 year old deploy (old version of Java, old Tomcat - thanks Mark) to the new server. Make Tomcat dance
with Apache (a connector). Chris helped a lot. How the hell does he remember details from 7 years ago?

4) Catch up with season data. Find out what is that the teams collect, who owes the IM team what, and ask (awesome MCM team, with established chain of command and high expectations on data sharing -- in general -)

5) Backups. Backups of Backups. And more backups. Tera-drives, LNO new server, other Ubuntu boxes. Mainly, I backup the database, and a hard code of parts of the code, and the whole site.

6) Update the "MCM IM book" with the new changes that pertain to the UNM.

7) Break Drupal to enhance the already rockin' functionality (the site has a data query system that allows you to specify a date range -- top that!)
Introduced Drupal Bibliography module (an estelar app!) and personnel management. Teased with maps. But before revolutionizing it all, better to gain trust -- deliver on the plans that set out to do. Too much to disclose here!

8) The EML. oh, the jolly EML. Turn 189 instances of EML2.0.1 into EML2.1.x w/ further checks (it is a new era indeed!). I used Perl to do the following changes:
8.a) 2.1.0 compliant -- run stylesheet
8.b) metadataProvider, contact, publisher - from "Byrd Polar ctr" to UNM.
8.c) packageID Identifier from 000XX to XX (per Duane's guidance)
8.d) package revision from Y to Y+1
8.e) copy distribution URL at "resource" level into dataTable/physical/distro need be - watch with existing ones
8.f) add broad keywords (McMurdo, ..) to all EMLs
8.g) remove the tags surrounding the body. A workaround to a serious bug on metacat.
8.h) make sure specified the exact header of the columns of the data described.

If you can use some of that, get in touch with me.

9) Amazon Web Services. Amazing stuff.
9.a) Use the bucket (S3) to store some backups and also store static web content. Us
9.b) Use the Elastic Instances (EC2) to deploy a large Ubuntu instance with one click.
9.c) Use the Relationa Databases (RDS) to deploy an Oracle enterprise BYO to host an MCM replica.

The idea is to have an off-location instance that can pick up the traffic for MCM when the UNM's internet access is cutoff (too windy, a small flood, a electrical storm, maintenance, a human screw-up or combos of those factors). I still need to connect Apache with tomcat and the DB, move the direct connection in EC2 to the bucket (saves $$) and negotiate agile changes of DNS entries to redirect traffic as needed.

10) A new section on the History of MCM. Our sociologist has stories to tell that make a fun read. See for example his latest piece on the LTER Newsletter on "Taylor Valley's missing lake".

11) Plans for a new section on microbial diversity. It is the molecular level. That is where we see biodiversity at a scale that dwarfs our knowledge. yet, genomics is not in Kansas anymore for the casual ecologists. It presents new challenges for the information manager. If we only have a process where EML or us were more dynamic, may be LTER can actually lead a bit in this exploding field. Oh, I was saying, Plans for a new section on microbial diversity. We looked at Wade's Sapelo. NTL's clone. Now we look at the new (old) boys on the block (MG-RAST). We ask what we want to see. We plan to deliver. Many fronts: The genomics standard consortium, the Biodiversity working group, the microbial observatories, the AntaBif (an IPT incarnation of GBIF for antartica, through a Terrestrial Observatory Network championed by B. Adams

12) Sanity/Project management made simple: an IM "to do" list (at this address - deims.lternet.edu/mcmdemo)

13) Review the IM-related comments that steamed from the last site review -- address them (how fun)

14) Place some datasets on PASTA. How about 60?

14) Put out fires.
14.a) I cannot find XYZ data. Help.
14.b) Install this software. Help.
14.c) The data ABC is wrong, take it down. Help.
14.d) At Sci Council is reported that MCM EML needs help (is crap). OMG.
14.e,f) .....

...To be continued....

Moorea Coral Reef SiteBytes 2012

SiteBytes go here

Niwot Ridge SiteBytes 2012

SiteBytes go here

North Temperate Lakes SiteBytes 2012

Last year we reported on a major overhaul of the NTL information system. Data were moved from Oracle to MySQL and the website was moved from a custom jsp/Oracle system to Drupal/MySQL/PostGIS. The new system went into production by the middle of summer 2011 and was subjected to the NTL site review in September 2011. Reviewers found the new website easy to

navigate and the information well integrated. The Drupal tagging and query system allows for many different views of the information, making it approachable for a variety of people.

Overall, the experience with Drupal has been very positive. Updates are very easy and quick, we are able to post news in a timely manner, and we have integrated the Center for Limnology twitter feed and blog, both of which were easily accomplished with Drupal modules that only required configuration. In combination with a Facebook page for the LTEArts project 'Drawing Water', we are trying to make our web presence more lively and interesting to the returning as well as the first time visitor.

We are tracking data download and employed Google analytics to learn a little about our visitors' behavior. The new website saw an average of 1,200 visitors per month, with about 30% being returning visitors and the vast majority looking for data. We have supplied >500 data downloads per month to over 250 individuals during the first seven months in 2012. General information about the NTL site and its research comprises about 50% of the remaining page views. Publications and personnel pages are frequented about equally (~20% each). This leaves about 10% to specific outreach activities (for teachers, for students, for the public, etc.) with half of that being interest specifically in the LTEArts section (http://lter.limnology.wisc.edu/ltearts [104]) which now displays two projects: the NSF sponsored 'Drawing Water' (http://lter.limnology.wisc.edu/ltearts/exhibition/panel1 [105]), and 'Waterlogs', an exhibition piece by UW art professor Michael Connors based on his experience at Trout Lake station, conversations with CFL staff, and growing up at a lake in northern Wisconsin.

Although our website went into production last year, many loose ends still remained. Due to some fundamental differences and some design decisions the data transfer between the two database systems wasn't quite as smooth as we had hoped for, and many primary keys as well as indexes to improve searching needed to be added.

Between tying up those loose ends and the LTER congruence checker and early PASTA testing coming on-line, we spent most of our time outside of routine data management tasks on improving our EML output. As stated by the 30 year review and the 'Bob Robinson movie' our system already had exemplary metadata contents thanks to the efforts of Barbara Benson and Dave Balsiger. The NTL data download application was and still is based on metadata which assures a high degree of congruency between metadata and actual data tables. I.e. it is impossible to download data that have more or less columns than described in the accompanying EML file. Therefore, the LTER metrics test revealed that NTL was with 92% in the 'good' category. This test was still run on our former system before the new website went into production and obtaining NIS-ready data packages from the Drupal system (DEIMS) involved improving/bug fixing code in the DEIMS modules for EML generation and data download, plus editing our metadata content. We implemented a state of the art ASCII/csv data archive that allows us to version each dataset, maintain previous versions of a dataset, and provide unfettered machine access to the data via a direct link in the EML file. This download link implements the LTER Data Access Server for record keeping. This functionality is new to our system and is complementary to access to our live database via user friendly custom query interfaces for human interaction as was tested by Bob Robbins during the 30 year review. However, many small problems arose and NTL Information Managers Gries and Stephenson have spent long hours going through each dataset and evaluating data package congruency with the tools provided by the NIS (https://portal.lternet.edu/ [106]). During that process we also re-keyed all datasets to the LTER controlled vocabulary and classified most datasets into the five LTER core areas.

The last improvement to our metadata is adding method descriptions to all EML files. In the previous NTL IM system, methods were linked to datasets within the web application, not within each EML file. This decision was made because many datasets use the same basic methods and it was felt that it was more important to display protocols in a nicely formatted, readable fashion, which unfortunately is not supported within EML. Therefore, most of our EML files currently don't

contain methods information. However, the information is already in the DEIMS and we were able to link protocols to datasets similar to the previous system. However, retrieving this information into an EML output has not yet been encoded into the Drupa2EML module.

Progress on employing the DataTurbine server for managing NTL's high frequency streaming sensor data has not been as expeditious as we had hoped. We have uncovered a serious problem with a DataTurbine module, and rather than making this our production approach, we manually loaded the data into our database last year. However, further development of the DataTurbine applications at NTL has been funded through a SI2 grant to T. Fountain in which Gries is a co-PI. Work is starting this summer.

We were extremely fortunate in hiring a gifted young Android programmer who was able to port our fish sampling application from Microsoft PDA to the Android operating system. With a couple of improvements over the old application, it is being used this summer on 7-inch tablets which should be easier to use than 3-inch PDAs.

Regular IM tasks this year included several updates to existing software. The ChemLab application, which is being used to manage chemical water analyses and their results is now processing the data into a MySQL table, not into Oracle anymore. And an application that calculates chlorophyll parameters had a few new feature requests and bug fixes. All of which has been handled by Stephenson without hired programmer help.

Palmer Station SiteBytes 2012

This past year, Information Management at PAL LTER has been focussed primarily on reviewing the current state of data management at our site, with a focus on identifying areas of weakness and developing plans for improvement. Recent changes have been directed by the newest guidelines set forth by the Executive Board, as well as the our readiness to work with PASTA in the approaching
new year. This September also marks the fifth month for Scott Gordon, our new Assistant Information
Manager, working in our group.

With the supplements we've recieved for the next year, we're planning to address data availability issues by hiring additional personnel for a 12-month period. We're using the results of internal data management reviews to develop plans for these individuals to assist information management in improving both data availability and quality. Our initial focus will be on improving metadata structure and content as well our NIS publishing protocols in order to provide PASTA-ready data submissions to the network.

Plum Island Ecosystem SiteBytes 2012

SiteBytes go here

Santa Barbara Coastal SiteBytes 2012

SBC's data holdings grew by about 25% in 2012. Eight new ongoing time-series were added, bringing the current number of ongoing multi-station time-series to 32. We incorporated two new data types: "genomics", for kelp microsatellites and for bacterial populations, and "remote sensing" for Landsat-derived estimates of kelp biomass. All datasets now routinely receive Network

controlled vocabulary keywords and units, and all core datasets are geo-located to the level of sampling station. We routinely use the tools developed by the NIS team for evaluating data packages before publication, and all data are planned to be submitted to PASTA in early 2013.

Since EML data packages are a major product of SBC's IMS, all data available through SBC's website is identical to that found in the network data catalog. The only data packages without direct links to data entities are those contributed by students where the paper has not yet been accepted for publication, and these account for approximately 5% of our data inventory. Hence, as far as existing data are concerned, SBC's data availability is exemplary. However, we are increasing the volume of SBC data, particularly from targeted research activities and experiments, and additionally, are improving the accessibility of all data with enhanced metadata and cross links on the SBC website.

To improve accessibility through the SBC website, we replaced two major website sections with dynamic pages that include links to data. Our "research" section, was redesigned using Metabase (from GCE) and ProjectDB, the suite of software tools developed collaboratively by LTER information managers. Our project descriptions are of broad research themes that can serve as a framework for further description of specific research activities. Themes can be browsed using SBC keywords in a manner similar to that used by our data catalog, and links were built from each theme to related data collections. Additionally, our outdated "sampling sites" section was redesigned using the Google Maps API. Sampling sites are grouped according to measurement types which meant that building links to related data from the map controller was very straightforward. The map was implemented first for our ongoing time-series, and other layers can be added.

SBC has continued its collaboration with three other LTER sites (MCR, CWT, GCE), to adapt the GCE relational database schema (Metabase) to replace some of the more manual parts of our IMS. Metabase work was advanced on two fronts. SBC’s project descriptions were our first exports from Metabase, and as part of this, we began development of XSL tools to populate the database from our existing EML datasets. Secondly (with MCR), we added a schema to describe data packages for inventory management.

SBC's information manager (Margaret O'Brien) is also the webmaster and is responsible for building most code used by the IMS. To allow her time to focus on Metabase tools and new website sections, SBC used supplement funds to hire a part-time information management assistant for routine data package maintenance. The assistant started as an REU working in reef ecology, and has quickly picked up additional skills and understanding in the informatics office. The assistant is shared with MCR LTER, which has advantages: a) it distributes the supervisory load between O'Brien and M. Gastil-Buhl (of MCR), and b) highlights areas of information management at these two sites that can converge or be streamlined for efficiency. SBC also will use supplement funds for our oceanographic programmer to investigate integration of the GCE Matlab Toolbox into our IMS. SBC also has funds to hire a high-level programmer to build modular, enterprise level code for EML metadata creation from Metabase (also to be shared with MCR). We have not yet found a candidate with the right skills for this task, and have chosen to continue our search rather than to compromise.

Sevilleta SiteBytes 2012

SiteBytes go here

Short Grass Steppe SiteBytes 2012

SiteBytes go here

Virginia Coast Reserve SiteBytes 2012

The past year has seen a number of improvements in the functionalty of the VCR LTER Information System. A major task was a complete rewrite of the front-end of the data catalog. The previous version of the data catalog had been in place since 1997 and consisted of dynamically-generated HTML generated from our metadatabase. With increasingly sophisticated and attractive metadata displays proliferating through the network, although functional, it was showing its age. We replaced it with an EML-based system that uses stylesheets to generate an attractive and functional tabbed display directly from EML documents. We were able to use supplement funds to hire David Richardson, who served as assistant VCR IM during 1993-1994 to oversee the revision. He customized stylesheets that had been previously worked on by Margaret O'Brien and M. Gastil-Buhl at the SBC and MCR LTER sites that were in turn based upon stylesheets developed by Chris Jones. Truly many hands make light work!

An advantage of moving to an EML-based catalog system is that it effectively decouples the backend metadatabase system, which is used to generate the EML documents, from the display system, allowing us to modernize parts of the system without needing to revise all aspects at the same time. Also by focusing on EML the tools developed are more likely to be applicable to other sites and can be shared across the network.

We enhanced the usability of many LTER datasets by including additional ancillary data as well as the primary data tables. This ancillary data consists of earlier versions of the data, often in a variety of forms. In some cases, it includes scans of field data sheets for cross checking with digitized data and for capture of non-digital data such as marginal comments etc. Although we expect most users to make use of the primary data tables, the ancillary data helps to improve interpretability and to spot any problems in the data processing workflows.

With supplement funding, former student Bridget Long worked with VCR investigators to revise all the dataset titles to make them more useful at a national scale by assuring that spatial and temporal coverages were referenced in the titles (e.g., "Tide Data" changed to "Tide Data for the Virginia Coast Reserve 1990-2012"). We are continuing to work on making sure that dataset abstracts provide the same degree of clarity and completeness. We also upgraded the web page to include new sections on VCR/LTER Research Themes and Research Highlights. The major work on this was done by student Cat Wolner working with VCR lead-PI Karen McGlathery. Each Research Theme is connected to a list of investigators and featured datasets.
The VCR/LTER has also been active in a number of network educational activities. VCR LTER IM John Porter participated and made presentations in the SensorNIS and several training workshops aimed at training investigators in advanced IM techniques. With Paul Hanson from the NTL LTER and Chau-Chin Lin from the Taiwan Forestry Research Institute he published a review paper on information management for sensor networks as part of a special issue of Trends in Evolution and Ecology.

Continuing work on the LTER Controlled Vocabulary took an international turn with participation in an ILTER workshop in Shanghai China focusing on semantic approaches to multilingual data discovery of ILTER data. Web services created for use with the U.S. LTER controlled vocabulary were modified to interface with the RDF-based and multilingual EnvThes thesaurus to create a prototype system for multilingual searching. A poster on this work was presented at the 2012 LTER All-Scientists' Meeting.

We also have been actively working on interfacing VCR/LTER data with PASTA and have several exemplar datasets ingested, with many more ready to ingest as PASTA becomes ready to take them. That work has stimulated additional activities in the area of developing web services that facilitate analysis of LTER data by automating the routine programming tasks associated with

writing code to ingest LTER data into statistical software. With the PASTAprog web service, you can go from searching for an LTER dataset to having a working R, SAS or Matlab (thanks to Wade Sheldon of GCE) program that has ingested the data and provided a statistical summary in less than one minute (video available).

VCR IM is also have been working with Bruce Hayden and the LTER Climate Committee on extending a long-term dataset on storm frequency. Between 1885 and 1996 various US government agencies (notably NOAA) developed storm track maps that were used in 2003 to generate long-term storm frequency records for individual LTER sites. However, since 1996 such storm-track data has not been produced. Bruce Hayden is working on the laborious task of going through some 23,000 individual weather maps to construct the record from 1997 through 2012. VCR IM has been working on automating GIS processing of marked-up maps to extract storm locations and to connect them into storm tracks. Python code using the arcPy module from ArcGIS is being used to process long sequences of images. Again, a poster on this work was presented at the 2012 LTER ASM.

2011

Site Bytes are intended as a general update on what has happened at your site during the last year. Please highlight new developments, ideas, and issues. Site bytes help us to stay informed about what is going on at other sites. Please also be forward thinking and include your thoughts on the following subjects with regards to the developing Network Information System. We will discuss these question further at the annual meeting:

- Do you have outside data sets stored in your system? That is, are you managing data in your system that are produced by a project funded separately from LTER? Is that project considered to be leveraged funding and directly related to LTER?
- How did you use the 2010 IM supplement? Please tell us your success story of how you improved data availability at your site with this supplement

Andrews Forest SiteBytes 2011

LTER Site Byte: Andrews Experimental Forest
Suzanne Remillard, Don Henshaw, Theresa Valentine
September 22, 2011

1) How did you use your 2010 IM supplement and how effective was this funding in making your data more available to the Network Information System?

EML improvements for “pasta-ready” data access: We received supplemental funds for IM last year and hired a College of Forestry staff programmer to enhance and further develop our programs that generate EML from our metadata relational database (SQLServer). The original web-based program was described in a DataBits article in 2007 (http://databits.lternet.edu/spring-2007/generating-eml-relational-databa... [107]), and essentially uses a configuration file to select metadata database tables in nativeSQL for style-sheet transformation into EML. The code was greatly enhanced and revised to conform to the new EML best practices, and much more of our metadata is now mapped into EML. In particular, complete taxonomic structures for species are built into taxonomicCoverage from our hierarchical taxonomy table (we now exceed the acceptable Metacat size limits for EML files!). Datasets are now registered in the Data Access System (DAS) for programmatic access of all data from the EML. EML is constructed on demand for the EML harvester or from webpage links.

GIS workshop: The Andrews hosted a workshop at the Andrews to train LTER GIS managers and consider best practices in preparing metadata for geospatial data. Procedures for converting FGDC metadata to EML were developed and added to the LTER EML Best Practices document. Eleven LTER sites were represented at the workshop and several others participated through VTC. This was a first step in the process to assure appropriate geospatial data documentation is available to the NIS.

2) Given that LTER sites are being asked with increasing regularity to host data sets external to LTER data collections, please describe any outside data sets your site is hosting.

The Andrews Forest LTER stores data in a long-term repository called the Forest Science Data Bank (FSDB), which is supported by the LTER in partnership with the PNW and Oregon State University (OSU) College of Forestry. The FSDB was established in 1980 and has been largely funded and operated by LTER personnel since the mid-1990s. The FSDB has opportunistically acquired non-LTER data and includes well over 250 databases with more than 170 databases on-line (mostly LTER). A stable computing environment and enhanced information system, including adherence to national metadata standards, have made the FSDB a popular repository for local grant and project data. The FSDB has expanded its LTER data holdings and serves as a regional data center for key US Forest Service Research data, e.g., Research Natural Areas and Experimental Forests; USFS campaign data, e.g., Demonstration of Ecosystem Management Options (DEMO) and Mount St. Helens; National Forest System data (Young Stand Study, Blue River Landscape Plan), OSU CoF data (e.g., OSU MacDonald Forest), and the Long-Term Permanent Vegetation Plot Network (OSU, PNW, UW).

Discussions between the Andrews IMs, lead PI, and the OSU College of Forestry Computing Resources Group have tried to assess a financial value for storing (archiving) data, performing general QA/QC tasks, and producing EML documents for datasets. We have not yet implemented a fee for using the FSDB, but there is recognition that this is needed. FSDB managers are looking to partner with other projects or large grants requiring a data repository to satisfy their data management plans, and planning is underway to house Portland-Vancouver ULTRA data and NSF-funded macrosystems ecology data pending successful proposals.

3) Other Site News

Andrews is in the process of hiring a new Information Manager to manage the information infrastructure at the field site. Fred Bierlmaier, the long-time system and communications administrator, is retiring.

The Andrews developed a new website (http://andrewsforest.oregonstate.edu [108]), which includes enhanced search capability, RSS feeds, links to Facebook, and multiple pathways to information. The Andrews also published a new site map, complete with brochure (http://andrewsforest.oregonstate.edu/lter/about/site/map.cfm?topnav=219 [109]).

Arctic LTER SiteBytes 2011

SiteBytes go here

Baltimore Ecosystem SiteBytes 2011

Baltimore Ecosystem Study
2011 Site Byte

Outside data sets:

BES has been providing a feed to our stream gaging sites via the USGS website. Hopefully we can develop services to stream this data automatically into our stream chemistry database.

2010 IM supplement:

We used out 2010 supplement to purchase computers to replace the web server and create a Moodle server. Moodle is a software package for producing Internet-based courses and web sites. It is a global development project designed to support a social constructionist framework of education.

Additionally, we used the supplement to hire interns from the local college to work on enhancing our online metadata system and help with the creation and handling of metadata for GIS data.

New BES Website:

The BES website, has been redesigned. The new site is based on cascading style sheets. The site it replaces was, like most sites of that vintage, based on tables, and as a result, a real pain to modify. As of now, the new site is not live, but it will be very soon.

The publications, abstracts, data and metadata pages are now completely database driven.

The two most dynamic sections, the Director's Corner, and the BES News, are being done using Google's Blogspot.

http://besdirector.blogspot.com/ [110]

http://bes-news.blogspot.com/ [111]

This allows direct authorship and, in turn, very up-to-date content.

We have also created a Facebook group. http://beslter.org/facebook [112] The group has over 100 members and sees several posts a day. Besides BES participants, there are members of the community, including policy makers, so this might be an interesting means of outreach. As one of the BES research questions has to do with the feedback aspect of looking at an urban system as an ecosystem, linking science and policy makers is important.

Annual Meeting:

The thirteenth BES Annual Meeting takes place next month.

The Annual Meeting is designed to share the technical scientific results of the project with BES researchers, educators, local, state and federal agency representatives while also being open to the media, and interested community members. The Community Open House is an informal venue to share non-technical project information and results with teachers, students at all grade levels, community members, media, and regional decision makers.

Interviews/videos/photography:

We are working to produce more visual outreach products. We have had a difficult time producing video feeds of scientific talks at our research meetings. Lighting, media limitations, the general "bulk" of this technology, and difficulties rendering video online, along with equipment costs, have made this a challenging process. Cataloging and archiving the thousands of digital photographs has also been a challenge.

We plan to do a couple short interviews --- just 3-5 minutes each --- of selected Principal Investigators. We hope that by using two cameras, proper lighting, and proper microphones, we can produce something truly compelling to stream online.

Jonathan Walsh, September, 2011

Bonanza Creek SiteBytes 2011

BNZ used last year's supplemental funding to tackle quality issues with our legacy climate data sets. Multiple years' worth of streaming hourly data from numerous sites had produced a large archive of data that had not necessarily been fully checked for quality issues and therefore was not in a very usable state. We used this supplement to hire a research technician to systematically review, process, and organize core climate data variables from our primary weather stations. Priorities were laid out based on site and variable criteria to maximize the scientific usefulness of the products. Each variable was discussed among the BNZ executive committee and appropriate researchers for data quality, processing, and protocol issues. Data was to be assembled into clean and well documented files for importation into the BNZ database and thus harvest into the NIS. That data that was able to be processed will be useful in providing more, high quality, data to the NIS and was worth the investment. However, this project is more complex and tedious than expected or desired and thus the benefit will be less then was ideal.

BNZ uses a fairly liberal definition of what constitutes an LTER data set and we therefore host very few, if any, outside data sets. When approached with a request for such hosting services, I usually work with the individual to find a suitable host site or other options. We have tried to remove data sets from our catalog that are simply being hosted or that were only reproductions of data sets that are available elsewhere; such as some USGS stream flow data that certain researchers wanted to have available on our site as well rather that harvesting it themselves from the original source

California Current Ecosystem SiteBytes 2011

CCE LTER's primary data system, DataZoo, stores data from multiple research projects. The data system's interface presents dataset catalogs specific to project ownership. Currently, there are some datasets contained within the CCE DataZoo data catalog, as well as the data system as a whole. The type of association with LTER varies across this set of data. Some, like a long term weather timeseries from San Diego's Lindbergh field, are not connected at all with LTER, and are stored in our data system for improved access and to leverage tools built for data exploration and comparison. Other data may come from sampling done by projects within our local institution, but outside of LTER, like pier measurements. One of CCE LTER's strengths is its partnership with the CalCOFI fisheries organization, which participates in joint sampling cruises with CCE. Data from this group are non-LTER, but are included in CCE's catalog because of the coordination involved in collection for producing data products shared by both projects.

Supplement funds for CCE IM in 2010 were used in multiple ways for improving data quality and availability. Unit Registry development and coordination efforts were supported as well as opportunities for additional members of the information management team to participate in LTER network events. Funds were also used for work on improving site data quality, which included a re-factoring of the underlying data model for Palmer's primary data system and reviews and improvements to metadata content for site data. Funds also supported work to improve various data system components, such as the personnel and controlled vocabulary databases, in order to maintain system alignment and integration.

Efforts for the next year for information management include a focus on the quality of the workflow from data collection to network submission. The development of resources for use in coordinating efforts and improving communication between researchers and information management is planned, all with the goal of improving the quality and completeness of data and metadata as well as increasing the amount of available data.

Other news includes the departure of two information management team members, Karen Baker and Mason Kortz. An article co-authored by the pair and containing a fuller description of their transitions and goals is included in this Spring 2011 issue (http://databits.lternet.edu/spring-2011/lter-information-management-cont... [113]).

Cedar Creek SiteBytes 2011

SiteBytes go here

Central Arizana Phoenix SiteBytes 2011

One major development in the last year that affected CAP was that the Global Institute of Sustainability at ASU (which hosts the LTER) decided to invest additional resources in Informatics. While this decision is intended to extend support to other GIOS projects, it has the benefit of allowing us to apply the LTER standards to the majority of Institute projects. Also, because the demand for this support is growing gradually we have been able to use some of the new resources to address several longstanding areas for improvement at CAP. We now have a three person Informatics Team comprised of a part-time IM with a full-time systems programmer and a full-time data analyst.

CAP is using the 2010 IM supplement to improve data management and metadata quality. We hired a student worker to support our data analyst and systems programmer. This additional resource has made it possible to make improvements in metadata collection, work on redesigning our metadata organization/repository and collaborate with LNO to bring our network metadata contribution up to scratch.

Our work with central Arizona city governments and other agencies means that we acquire various external datasets. However, many of these are social science related and are often restricted and cannot be further distributed. Consequently, we do not currently host any external datasets on the CAP data portal.

Coweeta SiteBytes 2011

SiteBytes go here

Florida Coastal Everglades SiteBytes 2011

FCE 2010 IM Supplement Use

The major focus of the FCE Information Management team (Linda Powell, Information Manager and Mike Rugge, FCE Project Manager) this past year has been the planning phase for a FCE IMS physical hardware restructure and improving its network-wide standardization to facilitate increasing use of site data in synthesis projects. The team used their 2010 IM NSF Supplement to restructure

its Information Management System (IMS) model. They are currently in the process of moving from purchasing and maintaining their own physical servers in the FCE LTER office to leasing five (5) virtual servers housed on Florida International University Division of Information Technology's (UTS) equipment. Two of the five new virtual servers will be designated 'development' servers; a much needed addition to the FCE IMS. The 2011 IM NSF Supplement allowed for additional storage to be added to the Oracle virtual server in order to archive high-density streaming data.

The FCE program will also be adding another level of disaster recovery protection to their current plan where the Oracle 10g, FCE Web and FCE FTP virtual servers housed locally at FIU will be backed up to identical virtual servers residing at the Northwest Florida Regional Data Center (NWRDC) located on the campus of Florida State University in Tallahassee, Florida. This new 'off-site' disaster recovery plan will allow the FCE website to be continually available throughout disaster events such as hardware failures and hurricanes. This past spring, the team migrated their physical backup server from Windows 2003 to Linux, upgraded its RAID card and increased existing storage on the server from 1.2 TB to 2.2TB. They also increased the storage in each external hard drive from 500GB to 1TB.

The 2010 and 2011 NSF IM Supplements will allow the FCE IM team to focus more on FCE Data as they won't have to worry about maintaining or replacing server hardware. Having the opportunity to work on dedicated development virtual servers for both the Oracle 10g database and the FCE Website will be extremely beneficial to the team because changes can be made and tested without having to worry about interrupting service or corrupting existing system setups. The new array of virtual servers will allow the FCE IMS to receive, handle, analyze and deliver high-resolution data at the FCE site level and be comprehensive and timely contributors of sensor data and metadata to the LTER Network Information System (NIS).

Although the FCE IMS already has a very robust archive of research EML metadata, much work is needed in relating data to FCE LTER publications and adopting the emerging LTER Network standards for keyword choices, enhancing attribute descriptions and adding taxonomy vocabularies where needed. Work was started on improving the quality and availability of LTER Data and EML metadata. Existing FCE EML metadata is being modified to incorporate the LTER Controlled Vocabulary Key wording Best Practices. New keywords have been added based on results from the ‘Tagger’ tool (http://scoria.lternet.edu:8080/lter-hive-prototypes/emlTagger.jsp [114]) introduced to the LTER information managers by Duane Costa (LNO) and John Porter (VCR). The FCE IM team requested support for a programmer and once hired, the programmer’s job will be to enhance the FCE Excel2EML tool to accommodate the upcoming network keyword and taxonomy vocabulary standards.

FCE Outside Data Hosting

Now that the FCE LTER IMS team is being asked with increasing regularity to host data sets external to our LTER data collections, the team is working with the FCE LTER Lead PI, Evelyn Gaiser, to come up with strategy for archiving, delivering and sustaining these short-term data well beyond the life of their projects. One of the most difficult aspects of hosting these types of data is developing proposal budgets that cover the cost of the IM team data curating tasks and data storage well into the future. We're looking into creating a budget category under the FCE LTER account at FIU that will allow the FCE IMS team to 'bank' money awarded for information management for these short-term projects so we are able to sustain the necessary storage on the leased virtual servers mentioned earlier beyond the life of the projects.

To date the proposed FCE outside data collections include the 1) Ultra-Ex socio-ecological data, socio-demographic, socio-economic, GIS land cover/land use and water re-use data, water demand models and model generated spatial projections 2) water quality monitoring data in terrestrial settings 3) Macrosystems biology including carbon storage and nitrogen pollution data and 4) Trophic cascades in the Pacific Northwest.

Georgia Coastal Ecosystems SiteBytes 2011

Overview

Information Management at the GCE site is led by Wade Sheldon (UGA). Travis Douce (UGA) assists with general IM tasks and also serves as our GIS specialist. A major focus of Information Management effort during the past year has been redesigning our core information system to support broader classes of research data as well as new LTER initiatives (e.g. controlled keyword vocabulary, NIS work-flow execution, EML congruency checks). Notably, these changes were done in close collaboration with IM teams at three other LTER sites (CWT, SBC, MCR) that are currently leveraging or adapting GCE database designs and applications. To facilitate broader sharing of technology developed at GCE, we also began releasing all data management software developed by the GCE program under a GPL open source license and established software distribution web sites linked to our Subversion code repository for downloading code and submitting bug reports. We also developed a comprehensive research registration, permitting and tracking system for Sapelo Island, which will also be leveraged by our research partners on the island (Sapelo Island NERR, Georgia DNR and the UGA Marine Institute) over the coming year.

Overview

Information Management at the GCE site is led by Wade Sheldon (UGA). Travis Douce (UGA) assists with general IM tasks and also serves as our GIS specialist. A major focus of Information Management effort during the past year has been redesigning our core information system to support broader classes of research data as well as new LTER initiatives (e.g. controlled keyword vocabulary, NIS work-flow execution, EML congruency checks). Notably, these changes were done in close collaboration with IM teams at three other LTER sites (CWT, SBC, MCR) that are currently leveraging or adapting GCE database designs and applications. To facilitate broader sharing of technology developed at GCE, we also began releasing all data management software developed by the GCE program under a GPL open source license and established software distribution web sites linked to our Subversion code repository for downloading code and submitting bug reports. We also developed a comprehensive research registration, permitting and tracking system for Sapelo Island, which will also be leveraged by our research partners on the island (Sapelo Island NERR, Georgia DNR and the UGA Marine Institute) over the coming year.

Spatial Data Management and GIS

The primary focus of spatial data management activities during 2010 was automating routine GIS operations and improving ties between tabular environmental data and geospatial information managed in our information system. We used supplemental NSF funding received in 2010 to develop a Python-based geospatial software library that bridges between ESRI GIS tools (e.g. ArcGIS and ArcSDE), relational databases (e.g. SQL Server), and file-based data (e.g. CSV text files). This software allows us to generate distributable GIS products, such as shapefiles and Google Earth KML files, on demand for all GCE data sets and provide links to these resources through the GCE Data Catalog (http://gce-lter.marsci.uga.edu/public/app/data_catalog.asp [115]) and in EML metadata. GIS feature classes can also be registered in the ArcSDE geodatabase by parsing information from tabular data files (e.g CSV text) and Google Earth KML files, which greatly simplifies registering new data in ArcSDE for GIS analysis. We are currently developing code for bi-directional synchronization.of metadata content between ESRI GIS files and the GCE metadata database to further improve our capacity to distribute EML-described GIS products.

Website Development

This year we added new web pages describing major GCE research findings to the GCE public website (http://gce-lter.marsci.uga.edu/public/research/research_bullets.htm [116]), with prominent links on both the home page and research overview page. These vignettes were developed in concert with other LTER sites to help showcase major research findings resulting from long-term ecological research. These pages will be updated as new high-impact findings and publications are identified, and replicated on new web pages being developed for the LTER Network web site.

We also developed a comprehensive web-based application for permitting and tracking research activities on Sapelo Island. Research guidelines and application forms were created in collaboration with the Sapelo Island National Estuarine Research Reserve and UGA Marine Institute (http://gce-lter.marsci.uga.edu/public/site/research_requests.htm [117]), and we designed a relational database and dynamic web forms for submitting, reviewing and approving research applications and generating printable field tags for display at the site. Supporting information can be uploaded and attached to applications, including GPS data, maps, documents, reports and photographs. GIS maps describing all research activities, organized by PI and study characteristics, are being developed to provide site managers from GCE, SINERR, and UGAMI information they need to manage sites, mitigate research conflicts and prevent unnecessary environmental damage. This database will also be a critical resource for investigators as they plan their field research and seek data for synthesis projects in the future, and for project leaders and IM staff as they track research progress and data submissions. Applications are currently limited to GCE-affiliated investigators, but work will begin in fall 2011 to open up the application to SINERR- and UGAMI-sponsored investigators as well. Dynamic search and web-based mapping applications are also under development, using funds from a 2011 LTER IM supplement.

Software Development

We have continued to enhance the GCE Data Toolbox for MATLAB software and offer this toolbox for public download on our web site (http://gce-lter.marsci.uga.edu/public/im/tools/data_toolbox.htm [118]). In 2010 we established a dedicated software development web site for this toolbox (https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox [119]) to provide web-based SVN access, project tracking, support tickets, and a Wiki for producing documentation. Beginning in October 2010 we began releasing the source code for this toolbox to the scientific community under a GPLv3 open source license and added additional documentation and tutorials on the web site. Approximately 3000 web visitors have registered and downloaded this software to date.

The GCE Data Toolbox software can be used to mine data from the NOAA National Climatic Data Center directly over the Internet, in addition to data from the USGS National Water Information System, LTER ClimDB/HydroDB, and NOAA Hydro-meteorological Automated Data System, then re-sample and integrate data to produce derived data products with structured metadata. This software is therefore useful for a broad range of data synthesis tasks beyond primary data processing. The GCE Data Toolbox has been adopted by the CWT-LTER project for processing data from their new synoptic monitoring program. Other LTER sites (e.g. MCR, NTL, NWT, PIE, SBC, SEV and VCR) have also expressed interest in using this software for managing sensor data, and a demo will be provided at the Environmental Information Management Conference in September 2011.

Database Development

This year we collaborated with John Chamblee (CWT), Gastil Buhl (MCR) and Margaret O'Brien (SBC) to revise the GCE metadata database model (GCE_Metabase) to support multi-entry data sets, non-tabular data (e.g. GIS, remote sensing and genomics data), and LTER controlled keyword and attribute name vocabularies that are under development. The new database model (GCE_Metabase2) is now fully implemented at GCE, and NSF supplement funding was used to update the GCE Data Toolbox software and EML implementation to support this new model. These upgrades allowed us to add new geospatial data sets to the GCE Data Catalog this year, and

additional GIS data from a Hammocks project is currently being added to provide access through the LTER data catalog, KNB Metacat and NBII metadata clearinghouse.

New IM/IT Projects

GCE obtained funding in 2009-2010 to deploy an eddy flux tower and horizontal acoustic Doppler current profiler on Sapelo Island, and test deployments of these instruments were completed in September 2011. These new instruments require real-time data telemetry, so ruggedized computers, power systems and 900 MHz communications equipment were also acquired and installed. This is a major sea-change for GCE IM, requiring hands-on work with data logger wiring and programming, maintaining computing and radio communication equipment in the field, and archiving of high volume streaming data. However, improvements to network capacity at the GCE field site on Sapelo Island, funded by an NSF FSML grant, will open up new opportunities for remote monitoring and management of these new installations.

IMC Question: Do you have outside data sets stored in your system?

Yes. We mine data from NWS, USGS and NOAA monitoring sites near to our domain, and synthesize and document these data for inclusion in our IS. We are also planning to manage non-LTER data for project affiliates in support of their NSF data management plans, and are exploring establishment of a broader data management effort at UGA in collaboration with CWT and other entities on campus.

IMC Question: How did you use the 2010 IM supplement?

We used our 2010 IM supplement for the following activities (some work still in progress, since funds were not received until late November 2010):

- Updating the GCE_Metabase relational model to support multi-entity data sets, controlled keyword vocabularies, non-tabular data files, and make corresponding changes to the GCE Data Toolbox software. This task was completed in Summer 2011.

- Adding support for EML metadata publishing from MATLAB-based data processing software: this is about 50% done, and will be completed as part of my preparations for the Workflow Tiger Team workshop in winter 2012. I intend to test our MATLAB tools (also used by CWT, SEV and others) as a platform for executing NIS workflows, and native EML support will be crucial to this (also, MCR and SBC have indicated a keen interest in this activity, since they have many MATLAB data sets and users)

- Developing an EML-compliant analytical sample management database: this work has not started due to personnel turn-over and re-tooling of our cruise sample analysis strategy - it will probably morph into an even more generalized system than originally envisioned. Work is expected to resume in winter 2012 after our renewal proposal is submitted.

- Developing EML metadata and data distribution capabilities for geospatial (GIS) data: this is now completed and fully implemented in our IS. It's rolled into our Python Geospatial Library, portions of which are already available online.

Harvard Forest SiteBytes 2011

Activities over the past year included: (1) All HF servers were migrated from local machines to virtual servers on campus, with generous help from the university. (2) Our local (Windows) domain was also migrated to the Harvard domain, again with generous help from Harvard. (3) Deployment of permanent equipment on our new field wireless network has grown to include about 60

computers, cameras, and dataloggers. (4) HF staff worked closely with NEON staff to facilitate site surveys and develop plans for electrical infrastructure. (5) The Analytic Web project continued this summer with three computer science REU students.

Two new projects are currently underway: (1) redesign of the Harvard Forest website and migration to Drupal, with help from a consulting company, and (2) conversion of our seminar room to a video-teleconferencing facility, with grant funds from the university.

Our practice has been to include in our information management system all datasets resulting from work at the Forest, or conducted by HF researchers elsewhere, regardless of the source of funding. As a general rule, datasets are included if they support a publication or are deemed to have long-term scientific value.

In our 2010 LTER Supplement we requested salary support for a research assistant to help with various improvements to our EML metadata. Accomplishments to date include: (1) all EML files were converted from version 2.0.1 to version 2.1.0, (2) all units were updated to comply with current best practices and submitted to the Unit Registry, and (3) detailed spatial coordinates were obtained for each dataset. With the remaining funds we plan to (1) correct errors uncovered by the EML congruency checker, (2) incorporate more detailed spatial coordinates into each EML file to support online mapping, and (3) implement the new LTER controlled vocabulary for all datasets.

Emery Boose, Julie Pallant, and Liza Nicoll

Hubbard Brook SiteBytes 2011

Do you have outside data sets stored in your system? That is, are you managing data in your system that are produced by a project funded separately from LTER? Is that project considered to be leveraged funding and directly related to LTER?

At Hubbard Brook, we attempt to get all data stored in the HBR system regardless of whether the project that produced the data was directly funded through LTER. Our justification for this is that all studies are reliant on LTER funding in some way, even if individual projects are not specifically funded by the LTER grant. Our major focus is on our long-term core datasets that are collected on a continual basis. We also try to get data from short-term studies, although this is often more challenging, especially when it involves transient researchers or students. We have an approval process for new research studies at Hubbard Brook and make people aware of the expectations for data submission before they start a project.

How did you use the 2010 IM supplement? Please tell us your success story of how you improved data availability at your site with this supplement

AT HBR we used the 2010 supplement to help establish a streaming sensor network, building on years of work in this area. We have finally completed the network and have automated data collection at all our precipitation collectors, stream stage height recorders, and weather stations. Data are transmitted via radio, and we have complete coverage of the entire Hubbard Brook Valley. A series of strategically placed repeater stations enable communication with our headquarters building over mountainous terrain. At headquarters, the data are pushed to a MySQL database where they are displayed on-line in near real time. We are experimenting with data display tools and methods (see prototype at http://hbr1.lternet.edu/data/realtime/ [120]). Once this is complete, we will focus our efforts on developing streaming QA/QC methods that make changes on the fly or flag questionable data. HBR is hosting the SensorNIS workshop, with the intention of moving this and other aspects of the wireless sensor network forward.

Jornada Basin SiteBytes 2011

This year we have been busy preparing for our LTER site renewal proposal and migrating our website content and functionality to a Drupal based website. We plan to have the new website completed prior to submitting the renewal proposal in February 2012.

Using 2010 site supplement funds, JRN was able to increase the storage capacity of our storage area network (SAN), hire a database person for a year, and to increase the capacity of our production virtual server pool, We fully populated our older SAN storage enclosure by adding 18 TB of storage, We hired a person for 1 year to assist in populating all JRN long-term datasets to relational database tables (SQL Server 2008 R2), All long-term datasets should be uploaded to the database by the end of the year. We purchase a new server to add another server to the production server pool to support the increasing number of virtual servers in use at the Jornada, such as the EcoTrends server.

Barbara Nolen has been working with ESRI to upgrade the Jornada's GIS metadata from ArcGIS v9.3.1 to v10. ESRI changed the metadata format in version 10 which required us the either upgrade the older FGDC metadata or to re-enter all GIS metadata. Unfortunately, we were not able to upgrade the existing metadata. Barbara will be re-entering the GIS metadata for all long-term datasets by the end of the year.

Ken Ramsey is implementing the Drupal Information Management System (DEIMS). DEIMS will provide data catalogs for the new Jornada website and support generation of EML for all JRN datasets. DEIMS has been deployed to a development server to minimize impacting the new production website. The data catalog view pages are being constructed to provide the interfaces to the data catalogs for all Jornada research projects, including the Jornada Basin LTER project. The new website will host and provide access to data and information collected by the Jornada and our research partners. The next step will be to populate the content in Drupal from our existing relational databases. We plan to finish populating and implementing DEIMS and the new website prior to submitting the renewal proposal.

Jim Lenz and Ken Ramsey successfully deployed the EcoTrends website to the Jornada. The EcoTrends website is now hosted and maintained by JRN. During the migration, we updated content on the website and corrected some problems that had not been identified previously. We updated the ClimDB metadata reports for all sites and some of the site description documents that were either missing or contained editing comments. Problems were corrected that kept handful of datasets' metadata from being uploaded into EcoTrends. We also updated the EcoTrends bibliography page.

Overall, this has been a busy and productive year for information management at the Jornada. Many of our site efforts coincide with the upcoming cross-site IM workshops. We look forward to completing the the website and to participating in the SensorNIS, DEIMS, and GeoNIS workshops.

Kellogg Biological Station SiteBytes 2011

Do you have outside data sets stored in your system?

Yes, we collect and store data, publications and protocols for the Great Lakes Bioenergy Consortium's Area 4 research group. GLBRC is funded separately.

How did you use the 2010 IM supplement?

We greatly improved test coverage of the data managment application (from around 30% to over 90%), refactored some of the modules especially the eml generation and added eml dataset import capability. We upgraded several data entry applications, including the aglog. We bought some additional disk space.

Konza Prairie SiteBytes 2011

At the site level this past year was largely dedicated to preparing for Konza's midterm site review. The focus up until this was catching up dataset for 2 year compliance and a complete revamp and release of a new website for Konza LTER (http://www.konza.ksu.edu [121]) which is 100% dynamically driven against SQL Server backend. The preparation for and subsequent review have left us in a good place having caught up with many of the things we had been working hard on, and should allow more proactive advancements in as we move forward.

Outside datasets at Konza: We do currently store a few outside datasets but per the NSF requirements this is getting to be such that we've had to write up language to provide researchers in search of a repository for their data. That said, we currently only offer a means of archiving said data and nothing more. Additionally I've been notified of sizable projects coming in who will most likely work against our infrastructure... LTER input vs. guidance levels are yet to be determined.

The bulk of the 2010 supplement was used to hire student hourly programmers who assisted with web development and a few necessary tools to assist in working against our SQL backend. Additional funds went to hiring a part time GPS/GIS worker, and attendance to the fall 2010 GIS2EML meeting at the Andrews.

Luquillo Experimental Forest SiteBytes 2011

2010 IM supplement, LUQ IMS and the NIS
In 2010, we requested resources to participate in several workshops and activities to continue building up LUQ´s new IMS framework. These included participating in the LTERMapS, DEIMS and Vocab workshops and hiring a graduate student for compiling all LUQ´s publications and assigning keywords to those publications which do include them (specially those of the beginning of LUQ LTER in the late 80´s and early 90´s).

The success of these events are a working progress. In regard to LTERMapS, this workshop gave LUQ IM the opportunity to coordinate for the first time a real collaboration with the Remote Sensing staff by joining efforts to get GIS data formally documented using EML. GIS data have been present in LUQ´s website for more than a decade now but without being documented. At present we are looking for resources that will assist us to translate the already documented layers to EML.

The DEIMS workshop was a smashing success. It marked the turning point for LUQ IM to achieving the completion of the new website-IMS that will constitute not only a web site and a information management system but a data management framework featuring a database that holds all the information in the system, the automatic-on-line generation of the data sets´s EML packages performed by the DRUPAL module constructed with the DEIMS group´s resources, forms for entry, web pages, blogs, personnel as well as metadata and data databases and a place to interact with LUQ´s scientific community as well as with the public in general.

The assignment of keywords to LUQ´s data sets started in the early 90´s with the development of LUQ´s metadata standards and the archiving of LUQ metadata in a RDBMS. The real effort lies now not on their assignment but of the re-assignment of keywords that belong to a controlled vocabulary. The ultimate goal of this effort is to make a searching engine out of the vocabulary of keywords.

Two major breakthrough has been achieved in this area: participating in the development of the LTER Network´s controlled vocabulary and getting key members of the LUQ scientific community involved in the related tasks to achieve this goal. Todd Crowl participated in the Vocab group workshop with the scientists and William McDowell and Robert Waide joined Todd and I to form the LUQ Vocab group.

The scientists of the LUQ Vocab work-group have decided to adopt the LTER Controlled Vocabulary and have engaged themselves in the task of re-assigning keywords to all the current 130 LUQ data sets.

In general, the completion of all LUQ´s non-spatial data EML documentation with the DRUPAL module is the major goal achievement facilitated by the 2010 supplement money. The effort and expertise needed for producing the data sets´ EML packages are beyond our resources, hence its automation marks a new era in LUQ´s IMS. To complete this achievement, the spatial metadata must also be translated to EML, a precess that started thanks to the economic resources provided by this supplement. Finally, the assignment of keywords to LUQ´s publications will give the infrastructure LUQ IMS needs to join all the system´s information, thus completing the LUQ IMS´s data management framework.

LUQ hosts data sets external to LTER data collection:
There are two types of collaborations in this area: Publishing data sponsored by other programs outside the LTER and serving as consultant to other programs in the development of their IMS, including metadata standards and web publishing development processes.

The publication of other Sponsors´ data sets are done in two ways: publishing private scientific data sets and publishing data from programs other than LTER. Data sets from R. Myster, L. Woolbright, J. Sharpe and M. and B. Richardsons include the former and data set from the STREAMS Project at La Selva Biological Station, Costa Rica is included in the later.

In the past two years, LUQ´s information manager has been the IM consultant to the San Juan Ultra Program (http://sanjuanultra.org/ [122]). The Information Management system of this program has adopted and customized the LUQ LTER Policies and metadata standard. For LUQ, this has been a great exposure and introduction to the merging of ecological and social science data and knowledge.

Additional achievements:
The area of major achievement besides the development of the new LUQ website-IMS has been the continuing involvement of LUQ IM with LUQ IM has participated in all their scientific symposiums and internships for the last 4 years but this year we have planned a Metadata Hand-On workshop (that will be held on October 12, 2011) where all the vegetation data sets form all the 4 participant schools will be documented and uploaded to the website. This exercise will be the first event to start the synthesis of the LUQ Schoolyard data.

McMurdo Dry Valleys SiteBytes 2011

SiteBytes go here

Moorea Coral Reef SiteBytes 2011

How did you use the 2010 IM supplement? Please tell us your success story of how you improved data availability at your site with this supplement.

MCR LTER, together with SBC, used 2010 IM supplement funds to hire an assistant for one year to help us maintain our dataset, project, and publication updates to allow us time to implement the metadata database developed at GCE for our sites. Our version, now ported to Postgres and filling up with metadata, we call “SBC_Metabase2”. We have kept our version as close as possible to the original to be able to upgrade ours in parallel with refactoring at GCE. Once our site adaptations are complete, the clone for MCR will be “MCR_Metabase2”. Most recently we are writing xslt transforms to reverse-harvest our metadata out of our eml in Metacat and transformed it from eml into csv files for bulk upload of all datasets, a one-time process. Project eml can now be generated but we have not yet begun exporting dataset eml from Metabase.

We have trained our assistant in EML Best Practice, keyword selection, and unit reference. Prior to MCR reverse-harvest, we asked him to subject every MCR eml document in comparison to its data files to a list of manually-performed quality controls. The focus was on data congruence such as verifying record delimiters and missing value codes and collecting file size and row count for all data tables. But we also checked keyword standardization to thesauri, all urls for broken links, address updates, namespace and schema declarations. This took a week, including training and checking, and provides a cleaner initial upload into our new metadata database.

We asked our assistant to teach himself how to use the GCE Data Toolbox and prepare for us a demo. We can see how this may make data ingest and quality control more efficient. In a future phase of our integration of GCE_Metabase2 we hope to feed metadata directly out of Data Toolbox as is done at GCE. To this end we have avoided modifying the tables in Metabase which are coupled to the Toolbox.
MCR and SBC are converging in our metadata practices between our two sites, as well as adopting some GCE practices. As we observe the diversity of eml usage at other sites we have found new practices we want to incorporate.

Do you have outside data sets stored in your system? That is, are you managing data in your system that are produced by a project funded separately from LTER? Is that project considered to be leveraged funding and directly related to LTER?

An NSF Long Term Research in Environmental Biology (LTREB) project with Caribbean coral reef data back through 1987 has similar information management needs to MCR LTER. That project supports one month per year of the MCR site IM. In 2010 we created a data catalog and website for that project using LTER guidelines and a clone of the MCR IM system.

This year a new Ocean Acidification NSF research project began in Moorea, led by two co-PIs from MCR LTER. Similar in size to MCR LTER itself, that project separately funds a technician, but not an information manager. This project will share some infrastructure with the LTER.

We have considered a long-term vision of being able to provide IM services to outside-but-related groups, once our IM system is fully developed and more efficient.

Niwot Ridge SiteBytes 2011

SiteBytes go here

North Temperate Lakes SiteBytes 2011

Do you have outside data sets stored in your system? That is, are you managing data in your system that are produced by a project funded separately from LTER? Is that project considered to be leveraged funding and directly related to LTER?

Yes, NTL has always been a repository for data from associated projects, i.e., projects that are thematically connected to NTL LTER, but are separately funded.

How did you use the 2010 IM supplement? Please tell us your success story of how you improved data availability at your site with this supplement
The supplement was used to migrate data and metadata out of Oracle and a custom Java/jsp website system into MySQL and Drupal. Specifically, into the 'Drupal Environmental Information System (DEIMS). Details are outlined below.

NTL IM activities
All major projects reported to have started the previous year and leading to a modernization of NTL's information management system have been continued this past year and have reached or are nearing production stage. Most of the IM applications have been moved to a virtual Linux server off site and commercially hosted. The new server now hosts MySQL, Postgres/PostGIS, GeoServer, SVN repository, and Drupal. Data have been migrated from an older version of Oracle to the latest version of MySQL, and spatial data from various file storage systems to Postgres/PostGIS. Only the DataTurbine server is still running on an in-house server.

Employing a workflow management system for basic information management has turned out to be an amazing time saver. The Kepler workflow engine has proven to be very well suited for the day to day tasks performed at NTL. Most field data are entered into previously-standardized Excel spreadsheets, which are submitted to the IM office and have to be quality controlled and transformed, calculated, and parsed into the database. During the past year we translated well-documented but labor-intensive procedures that involved Excel, MS-Access, Perl, Fortran, PHP scripts/programs, the Oracle specific SQL language PL/SQL, and Oracle triggers into Kepler workflows. A Kepler workflow provides all functionality that used to be scattered over several applications, and the pertinent documentation for those applications, in one place. The graphical display makes it easy to see how processing steps are linked together, and on-screen annotations can provide what few instructions are needed to run a workflow. Kepler provides basic capabilities in the form of "actors" which can be reused in different workflows, and the graphical user interface allows for easy drag and drop composing of the workflow. Workflows may be designed with limited programming experience, but custom actors may be developed as needed. However, we found Kepler's existing functionality sufficient for our needs because it provides actors designed to wrap existing Perl (or other) scripts. The workflows are then managed in an SVN repository which takes care of versioning and documenting changes to the workflows.

Likewise, using Drupal as a platform to rebuild NTL's website has improved content management of our website dramatically. The original content has now been moved to the new system, expanded, and updated and we are enjoying the ease with which Drupal allows different view of and access to the information. This system is not yet in production because we are in the process of debugging data access functionality. Sophisticated data search and sub-setting functionality that was available in the old website has been reprogrammed as a Drupal module. This module uses metadata for each dataset as its basis for generating a user-friendly custom interface allowing inexperienced users to design fairly complex SQL queries. This system has proven very successful on the original NTL website and was therefore implemented as a Drupal module without major changes to its functionality. In developing the new website in Drupal, NTL partners with six other LTER sites (SEV, LUQ, PIE, ARC, VCR, and recently JRN) which all have adopted Drupal for their websites. The sites collaboratively developed a content type (a Drupal data model) for handling metadata in Drupal. This allows for a very high degree of collaboration between the sites, as page layouts, views, custom modules and search functionality may be shared, while still maintaining a different look and feel for each site. Several participating sites successfully collaborated on an EML generating module. The code was shared in a Google Code repository and several developers were involved, including NTL's programmer. The module now produces limited, but valid EML.
An open source DataTurbine (OSDT) server has replaced custom Java code for managing near real

time sensor data streams. This system is in production and we are working with the OSDT team at Calit2 (SDSC) to adapt DataTurbine's functionality to the field requirements at NTL, i.e. making the server more robust for situations where the whole system may be unreliable as well as where single sensor streams may fail. We were able to provide substantial feedback which has led to a joint proposal submission to the NSF SI2 program. OSDT itself has proven to be very reliable, the problems we are still encountering are mostly related to the off-ramp which is responsible for parsing the data streams into a database. For this system to function as reliably as the previous custom Java code, some work will still have to be done. Another area that we are anticipating to make some progress over the next year is in QA/QC modules for OSDT. Currently the procedures are still encoded in database triggers, i.e. OSDT parses the data into a temporary table from which the database triggers move them to final storage with range checks and appropriate flags applied. We are looking forward to establishing LTER network-wide QA/QC standards and routines during the next year in collaboration with the sensorNIS initiative, and then implementing them in OSDT in collaboration with the OSDT developers. We will also evaluate new developments from the Kepler 'Real-time Environment for Analytical Processing' (REAP, https://kepler-project.org/users/projects-using-kepler-1/reap-project [123]) project and possibly adopt their approaches when the code becomes available.

NTL's spatial data management has been improved in a major way. Our new assistant information manager and geospatial specialist installed a Postgres database with PostGIS extension and a GeoServer for handling spatial data. After locating and organizing a vast supply of spatial data files in various formats, on various media, and in various physical locations, the data are now in PostGIS and available through GeoServer in several common formats (e.g., tiff, KML, ESRI shapefile). They are accessible from NTL's new Drupal website in the format of ESRI shapefiles automatically supplied by GeoServer as zip archives. Next year's emphasis will be on filling out the currently somewhat sparse metadata for these spatial data. In addition, we will seek input from NTL's researchers and perform a use case analysis to determine the need for specific online tools that may ease the discovery and use of our spatial data.

Palmer Station SiteBytes 2011

Palmer LTER's primary data system, DataZoo, stores data from multiple research projects. The data system interface presents dataset catalogs specific to project ownership. Currently, there are no non-LTER datasets contained within the Palmer DataZoo data catalog. The data system as a whole, however, does contain datasets from projects outside of LTER. The type of association with LTER varies across this set of data. Some, like a long term weather timeseries from San Diego's Lindbergh field, are not connected at all with LTER, and are stored in our data system for improved access and to leverage tools built for data exploration and comparison. Other data may come from sampling done by projects within our local institution, but outside of LTER, like pier measurements.

Supplement funds for Palmer IM in 2010 were used in multiple ways for improving data quality and availability. Unit Registry development and coordination efforts were supported as well as opportunities for additional members of the information management team to participate in LTER network events. Funds were also used for work on improving site data quality, which included a re-factoring of the underlying data model for Palmer's primary data system and reviews and improvements to metadata content for site data. Funds also supported work to improve various data system components, such as the personnel and controlled vocabulary databases, in order to maintain system alignment and integration.

Efforts for the next year for information management include a focus on the quality of the workflow from data collection to network submission. The development of resources for use in coordinating efforts and improving communication between researchers and information management is

planned, all with the goal of improving the quality and completeness of data and metadata as well as increasing the amount of available data.

Other news includes the departure of two information management team members, Karen Baker and Mason Kortz. An article co-authored by the pair and containing a fuller description of their transitions and goals is included in this Spring 2011 issue (http://databits.lternet.edu/spring-2011/lter-information-management-cont... [113]).

Plum Island Ecosystem SiteBytes 2011

2010 Activities
During the past year, 2010-2011, PIE has been working with ARC on migrating our Excel based metadata and data management system to a Drupal based content management system. The Drupal Environmental Information Management System (DEIMS) is under development in collaboration with LNO, ARC, SEV, LUQ, NTL and VCR. Supplement money from 2010 was used to hire a programmer to develop modules for converting Drupal content to EML and converting Excel metadata content to Drupal. Unfortunately the programmer left for another job before the modules were complete and debugged. Debugging both modules is still underway. Migrating to Drupal has been more challenging and time consuming than originally thought but should prove to be a quality information system once completed.

PIE installed an eddy flux system in the marsh off Nelson Island Creek this past summer. The system was developed by Campbell Scientific and is still undergoing tests of sensors. Spread spectrum, 900MHz, is used for data telemetry. PIE also migrated to using Campbell's LoggerNet software for our Marshview weather station (previously Campbell PC208W) and IBYC water quality station (previously YSI Ecowatch DCP). PIE also installed an OTT RLS water level sensor at IBYC. Previous within the water column strain gage pressure sensors had to be removed during the winter due to ice flows. The RLS is mounted above the surface and uses impulse-radar technology to determine the water level. Development of a CR Basic program using SDI-12 protocol commands for both the YSI 6600 sonde and OTT RLS was not as straight forward as both vendors suggested. Using 2010 supplement funds, in collaboration with ARC, Vista Data Vision software has been purchased to help us manage sensor streaming data and publish it on the web.

Outside Data
We have a mixture of data republished from USGS discharge sites, NOAA/NOS tidal water level sites, NCDC cooperative weather stations and NADP sites that have been useful for PIE researchers. Links are also provided to the above primary data sites, which allow the user to check for metadata updates provided by the original data source. Several collaborators in coastal research have expressed interest in publishing their data on the PIE web site. Due to NSF's requirement for data management many researchers are looking for options for how to comply but are not familiar with the time and cost commitments that will be required.

Santa Barbara Coastal SiteBytes 2011

2010 Supplement funds: SBC (with MCR) used supplement funds to adopt the GCE-Metabase model for dataset and project descriptions.We chose to use the supplement funds to hire a shared assistant who could maintain our current systems for dataset generation and free up time for the primary site IMs (O'Brien and Gastil-Buhl) to work on the new database. We were able to partition the new work according to our individual strengths: Gastil focused on porting the model from Microsoft SQL Server to Postgres and extracting content from existing content to populate tables. Margaret focused on scripts to build EML from Metabase and web services. As of September 2011,

we have ported the model (Metabase2) and begun EML imports. The first database exports have been SBC “project EML” which are somewhat smaller than datasets but of the same format. Lessons learned from the initial exports will be applied to dataset generation. All code is deliberately generalized for these two sites, and to enable further development as a network resource.

Other data: As one of the larger computing groups at the Marine Science Institute, SBC houses data from long-term collection programs in the area which date to the mid-20^th century. These include a) abundance and distribution of fish, benthos and algae from Santa Cruz Island reefs, b) urchin recruitment, c) kelp biomass from commercial harvesting, d) manually collected sea surface temperature, and e) beach erosion and character. These data have been incorporated into our data catalog. Our file system also houses data for research partners with various non-LTER funding, but whose projects have leveraged SBC’s sampling program. These projects generally do not have information management staff, but have begun to realize the need for IM plans and dataset publication. Consequently, they depend on SBC’s information manager for coordination and occasional advice. The highest priority for SBC’s information manager’s time has always been LTER-funded data, but one of the driving forces for adopting the GCE Metabase model has been to make our dataset production process more agile and better serve a growing number of research partners. We plan to incorporate these additional data products into our catalog as resources permit.

Sevilleta SiteBytes 2011

What did SEV do with the 2010 supplement?

The 2010 IM supplement at the Sevilleta was used to 1) support development of DEIMS and 2) support programmers to streamline processing of core Sevilleta datasets. Not all the supplement has been spent yet; the remainder will help pay a programmer who will assist with implementation of a sensor data management system.

The Sevilleta LTER is participating in the development of the Drupal Ecological Information Management System (DEIMS), along with NTL, PIE, ARC, VCR and LUQ. In 2010, these sites each contributed $9K towards the hire of programmers at NTL and the Marine Biological Lab (MBL) to advance DEIMS. At MBL, the programmer created a PHP program that translated the metadata entered in to the Drupal EML editor (developed by Inigo San Gil and Marshall White or LNL) to EML. This program had lots of errors, which Inigo has been obliged to fix. This program now works for SEV metadata, and metadata is now being harvested from a Drupal-generated harvest list. I am not sure that I will be able to use the product developed at NTL, which generates a query page based on EML.

Two graduate student programmers were hired during summer 2011. One wrote perl scripts to QA/QC and archive data from several core Sevilleta NPP studies. The other attempted to decipher a very old data processing system implemented on a Mac to perl programs that would run on the SEV unix server. She succeeded in getting the programs to again run on the Mac, but they were not transferrable to the UNIX machine.

As a result of the 2010 supplement, SEV is now delivering EML 2.10 to the LNO metacat. Also, the efficiency with which we process NPP data has been improved.

What external data does SEV host on its website?

Sevilleta hosts a water quality dataset for the Bureau or Reclamation on the website. These data were collected by a graduate student whose work was tenuously considered a Sev project. Many hours of programmers’ time were used to develop a search interface to these data under the

guidance of the SEV IM. This project was a huge headache because the student didn't know what was needed and it dragged on for about a year and through three student programmers. The lesson learned was to not agree to host data like this unless it comes with an explicit plan for what is expected and money for development and maintenance.

Short Grass Steppe SiteBytes 2011

SiteBytes go here

Virginia Coastal Reserve SiteBytes 2011

This year major Information Management activities of the VCR/LTER, included implementation of a new server, moving many VCR datasets to using the LTER Data Access Server., editing of VCR metadata keywords to conform to preferred forms from the LTER Controlled Vocabulary, addition of a new "suggest keywords" button in our online metadata editor, semi-automated congruency checking comparing VCR/LTER metadata with the underlying data using the online metadata tool from Taiwan and continuing to improve quality assurance and data workflows.

The VCR LTER was very active in the Controlled Vocabulary Working Group ("Vocab."), which had a banner year, with creation of a polytaxonomy/thesaurus, integration into LTER Metacat browsing and searching and development of autocomplete and keyword suggestion tools. The VCR IM lead two workshops of Vocab. – one in March aimed at development of the polytaxonomy, and one at the Science Council, which included site PI's as well as Information Managers, which focused on vetting the previous work and development of policies for managing the list. The VCR IM created several web services that were used to test and implement searching and keywording using the LTER Controlled Vocabulary. The VCR/LTER IM also co-chaired the LTER Network Information System Advisory Committee and was a member of the Organization of Tropical Studies Science Committee.

The VCR/LTER 2010 supplement was dedicated to development of Drupal-based metadata tools, improvement of metadata quality and development of data workflows. The funds were used to hire a student, Bridget Long, who worked on IM tasks, including editing of keywords, adding new metadata and datasets, congruency checking using the online tool from Taiwan (see her article in the Spring 2011 Databits). Additionally, a server computer was purchased to serve both as a backup for our existing server and to provide additional computational power for running virtual machines, which are widely used by the VCR/LTER to allow us to have a wide array of specialized computers (e.g., web server, database server, online-GIS server, development server and data processing server).

On the topic of hosting of externally generated data, the VCR/LTER does so at a variety of different levels depending on site research priorities and accessibility of the original data. In order of priority:
• Linked Data Sources: for the lowest priority data, or data that is easily accessible via the data collector's system, the VCR provides links to the sites of data providers.
• Filesystem Access: The VCR/LTER web site includes some data from external sources that are stored in simple directory structures with minimal documentation. These data are not listed in the VCR Data Catalog and no EML is produced. Typically, this is data such as remote sensing data where there are either well-developed standards, or metadata provided by the original providers. For the most part this data is available elsewhere on the web, but it is useful to our investigators to have copies in a stable location and some value-added products (e.g., data clipped to site boundaries). Most of these data were originally obtained for specific projects and simply retained in an accessible location.
• Comparative Data: The VCR hosts a few external datasets that contain data that are useful for comparisons with VCR/LTER data. These data are fully documented with metadata, including

Ecological Metadata Language, and are included in the VCR/LTER Data Catalog. Typically, these datasets are “orphan” datasets that are not available in any other repository.

The rationale for this prioritization is that the VCR/LTER IM resources are limited and the highest priority should therefore be placed on datasets that are a unique contribution of the VCR/LTER. If we don’t make those data available to the scientific community, no one will. A limited number of comparative datasets have a relatively low-cost and again, if we didn’t make them available, they would remain unavailable to the scientific community. For the data for which we provide filesytem FTP-style HTTP access, these are low-cost because they are typically something left over from a past project and are consumptive only of (cheap) disk space. Adding web links on the VCR web site to external data sources is restricted to frequently needed data because web links change all the time and require near constant checking and updating. For the most part, researchers are skilled at locating online data so we usually focus on links that are obscure or difficult to browse.

2010

Site Bytes are intended as a general update on what has happened at your site during the last year. Please highlight new developments, ideas, and issues. Site bytes help us to stay informed about what is going on at other sites. Please also be forward thinking and include your thoughts on the following subjects with regards to the developing Network Information System. We will discuss these question further at the annual meeting:

Given limited resources, how do you prioritize data contributions to the NIS?

- Do you give a lot of attention to a few datasets or limited attention to a lot?
- Do you deliver derived or raw data?
- Do you communicate with your PIs regarding prioritization of data?

Andrews Forest SiteBytes 2010

Suzanne Remillard, Don Henshaw, Theresa Valentine
September 3, 2010

(1) Site activities over the past year

Development and implementation of new Administrative Interface
An improved web-based administrative interface has been implemented that allows interactive researcher submission of study metadata, managing of personnel profiles, and managing research projects including an online project application form. This interface is designed to improve the efficiency of IM operations by allowing site members to directly update study metadata and personal information. This application was written in ASP.NET and contracted to Business Solutions Group on the Oregon State University campus. We chose to go this direction because our previous interface, written in 1999-2000 in ASP, had sever vulnerabilities and needed to be updated, our computer support group is heavily invested in Microsoft products, and there was a group (BSG) willing and ready to develop the application for us.

Migration of ClimDB/HydroDB and training of new administrator
ClimDB/HydroDB is now being hosted at the LNO. The production data was moved to the LNO servers and the web page updated to point to the new production server. Yang Xia is the new LNO database manager and is being trained by Suzanne. This training will continue at least through the end of the year or until the management of ClimDB/HydroDB is fully assumed by LNO.

Redesign of long-term sampling databases
EcoTrends and the great work of Christine Laney really exposed some issues with some of our core LTER sampling databases, especially the vegetation sampling. We have known that our metadata badly needed improving, but it became clear that it was impossible for anybody, including us, to use the data properly. Therefore, we have put significant effort into redesigning these databases to be clear about the data, standardize the metadata, provide the derived data people want, and to make the maintenance and management much easier. The implementation of this redesign is still underway. In addition to the vegetation databases, critical updates were made to climate, stream chemistry and streamflow databases including the digitizing of legacy Lookout Creek streamflow data from 1949 to 1986, greatly improving the depth of data we can now provide.

Preparation and timeline for 2011 mid-term review
An Information Management planning committee was established and met several times to consider the efficiency of IM efforts, preparation for the 2011 site review, data access and release issues, webpage development, and prioritization in the preparation of online data. A survey was distributed to key Andrews PIs to prioritize study database development efforts, define Andrews core data sets, and identify new LTER6 data sets for establishment online. A committee will be established to make content and design recommendations for enhancement of the webpage.

LIDAR data
Standard products were developed from LiDAR data such as shaded relief, stream layer, roads, contours, and watershed boundaries. Two GPS forays were conducted to establish precise coordinates for all Andrews study locations and communication towers. Base GIS data was converted to standard Datum Nad83 from Nad27. The data and layout was determined for development of a new publicly accessible Andrews map that will be published Fall 2010.

Development of wireless communication network throughout the Andrews
Two radio towers are under construction to provide line of site with the Andrews Forest Headquarters in order to enable wireless communication throughout the 6400-ha watershed. A combination of solar, methanol fuel cell and state-of-the-art wind turbine should adequately provide uninterrupted power and allow for year round operation even at seasonally inaccessible sites.

Digital orthophotography for historical aerial photography for the Andrews
Working with US Forest Service, we will create digital orthophotography for all historical aerial photography for the Andrews (back to 1949). Scanned photos will be corrected using the LiDAR data.

(2) Personnel activities

Don servers on our local Andrews Executive Committee, assuring communication between the Information Managers and the site PIs. The team also provides a training workshop for LTER graduate students and other researchers on the basic elements of information management and GIS. Concepts of data modeling are taught to allow students to properly prepare and document their study data sets for long-term archival.

Suzanne, Don and Theresa are all actively involved in LTER Network activities. Suzanne is a member of IMEXEC and will continue to assist with the administration of ClimDB/HydroDB during the transition to the LNO. Don participated in the development of a network-wide prospectus for LTER legacy data, organized a Sensor Network Workshop at ASM 2009, and co-chairs the IMC. Theresa chairs the GIS subcommittee, led the development of LTERmapS, chaired a workshop on Visualization at ASM 2009, and conducted a related post-ASM workshop that successfully migrated this LTERmapS interface to the LTER Network Office. She also attended the EML Best Practices Workshop.

Andrews personnel also represent LTER outside of specific network activities. Don attended an ILTER EAP workshop on large dynamic vegetation plot analysis in Malaysia and consulted in an LTAP proposal and an ULTRA exploratory grant.
Theresa represented LTER at ESRI Space/Time Analysis Workshop, presented a poster on 'Andrews Tall Trees' based on LiDAR data at 2010 ESRI International Users Conference, and presented a talk on LTERmapS at the Andrews Annual Symposium.

(3) Issues faced by Andrews IM:

Inefficient QA/QC of streaming data constrains timely access to these data
This is one of our biggest stumbling blocks. Our streaming data is obtained by a mix of automated and manual procedures. Each manual procedure severely constrains our ability to get the data online and available. We have been investigating options and are likely to explore the use of Matlab to process these data. However, we face a huge learning curve.

EML improvements for "pasta-ready" data access
This is more of an issue with person-power. We need to sit down and evaluate our EML with the standard. We have procured funds to hire help with the programming. Our current EML writer is written in ASP.NET.

Accuracy of LiDAR data
The accuracy of LiDAR data requires updates of other GIS layers and study site locations. We have initiated GPS forays to collect more accurate data. However, there are a lot of study locations that need better data, so this will be a time consuming process.

(4) Development of the Network Information System

Given limited resources, how do you prioritize data contributions to the NIS?

Do you give a lot of attention to a few datasets or limited attention to a lot?

The IM team is responsible for several large core datasets (climate, streamflow, vegetation monitoring). The QA/QC for these datasets generally falls onto the team. Additionally, there are datasets that have some sort of regular monitoring (typically annually). Typically the PI takes responsibility for the QA/QC of these data. The team also interacts with researchers and graduate students to archive single monitoring datasets. Again, we do have generic QA/QC reports that we run, but the research is responsible for the QA/QC of these data.

Do you deliver derived or raw data?

Our data is a combination of raw and derived data. We are working on standardizing some of the core datasets (especially vegetation) in order to provide some derived data, which is what data users really want.

Do you communicate with your PIs regarding prioritization of data?

The IM team works closely with the Andrews executive committee to determine prioritization of data.

Arctic LTER SiteBytes 2010

Information management at the Arctic LTER is in the process of updating the web. Currently our web site and data sets are static with updates done periodically. While this has served us well there is a need to implement a more dynamic design. With the varied projects and institutes we are looking for a system that will allow researches more control and input into data and project

management. Therefore we are moving our web site into the Drupal Content Management System. Drupal is a hybrid between a content management system (CMS) and a development framework. This hybrid nature provides content mangers the ability to easily updated web pages and allows programmers to extend the system with custom modules. The developer community surrounding Drupal is large, providing a large suite of general functionalities which we can tap into in addition to the more custom programming for strictly LTER needs. For example modules for group calendars, maps, photo galleries and sharing files already exist extending the functionality of a Drupal site.
A group supplement proposal was funded to help develop ARC and PIE sites and to support a LTER network-wide Drupal collaborative effort. We will be using MBLWHOI Library Drupal programmers to help develop to develop data ingestion tools that automate some aspects of metadata creation and to implement web services clients that will incorporate LTER standards for units, keywords and personnel into metadata.
Network Discussion Questions
• Do you give a lot of attention to a few datasets or limited attention to a lot?
o We have a list of datasets that are put in the NIS with others added or updated are time permits.
• Do you deliver derived or raw data?
o Depending on the dataset it will either be raw or derived or both.
• Do you communicate with your PIs regarding prioritization of data?
o Yes

Baltimore Ecosystem Studies SiteBytes 2010

BES Site Byte 2005-08-03

Jonathan M. Walsh

Highlights

- We now have 10 years of stream chemistry data online.

- We are now hosting a "Moodle" Learning Management System to deliver curricula online.

- We have the curricula and materials for the Math And Science Partnership (Berkowitz, et. al.) online on the BES www site.

- Our data access page is now completely dynamic, generated from a search form that builds a SQL query based on the terms the user enters. The code allows the user to enter as many, or as few words as desired. Instead of using an "AND" logical approach, which would dictate that all words must be present, an "OR" linking is used so that records that match any of the words are displayed. This approach lets the user freely add as many words as desired.

- Our metadata is kept in 100% EML format. The data display ("human readable") page on the website is generated from the EML, as is the information that is harvested by the LTER Metacat system.

- Our Geodatabase continues to grow. We are in the midst of changing the structure to a file-based geodatabase.

- BES Publications online: The BES publications have been imported into a database and are now searchable online. See http://www.beslter.org [124] . There is a button on the top to access the publications database. Additionally, they are available through thge LTER Network Office portal via their harvesting system. We re-standardized to EndNote format to accomplish this.

Network Activities

I have enjoyed membership on the LTERMaps project and ProjectDB and others.

Miscellaneous

- Working on a Drupal system for the BES Intranet site.

- Working on an imporved online tabbed metadata entry/edit form.

Discussion Questions

Do we give a lot of attention to a few datasets or limited attention to a lot?

A limited number of datasets get the most attention. Some datasets receive very little, if any attention. They are valuable to the collection nonetheless!

Do you deliver derived or raw data?

We deliver both. Our "real" public data is not posted until it has been quality assured. In many cases, however, we have provided new, unchecked data to interested parties.

In all cases, we deliver "raw" data, in other words, the actual measurements. Of course some "measurements" are in actuality derived data, perhaps based on the voltage output from a transducer, so the question is tough to answer!

Do you communicate with your PIs regarding prioritization of data?

I communicate with PIs regarding many aspects of data, some more than others. I have not, however had many discussions with them regarding the prioritization of data.

Bonanza Creek SiteBytes 2010

Activities:

This last year at BNZ there has been substantial activity on top of maintaining our core operational systems. Maintaining system integrity and security continues to be our prime task and it requires a considerable amount of resources before we can tackle any larger issues; but there are plenty of additional projects vying for our attention. Preparation and submission of our new research proposal to the National Science Foundation played a leading role for many months. There were two major focus points related to data management addressed in our proposal prepared for NSF. First of all, a great deal of time was spent to update and improve the content and functionality of our website. Secondly, we spent considerable effort to ensure that all core data sets are current and have the most complete metadata.

The website improvements we initiated have significantly improved the quality of our internet presence and outreach capability. While the design of previous web system was somewhat dated and potentially in need of a more complete overhaul we felt limited in the resources we could commit to this effort in light of other ongoing projects. Rather, we chose to upgrade our ColdFusion Web Software and server OS to improve system integrity before enhancing the functionality and content of specific web pages. This provided us with a revitalized platform that could provide more robust services and stability. Content that could be transferred into tables within the database and displayed programmatically was migrated and additional web-based interfaces were created. Additional database content was also displayed on pre-existing and new web pages. Several key researchers contributed additional text and graphics to update and enhance content while new pages were created to highlight recent art and humanities focused projects. While the temptation to

migrate to Drupal for site and data management was significant, in the end we really felt that such a conversion was not feasible given the timeframe and the anticipated learning curve. However, this option is still possible for the future as our IM system evolves. As the network continues to move in this direction and offer more support and modules, it will become a more attractive option.

The data review process included reviewing database outputs and EML for core data sets to ascertain their accuracy and completeness. As issues were identified, the appropriate information was gathered and inserted into the database. Procedures for the generation and harvesting of data by network systems were also reviewed and we are currently in the process of updating the scripts for data submissions to ClimDB. Thanks to some improvements with our EML generation scripts, the Metacat harvests are working to pull in the highest quality metadata on record for BNZ.

This past year, Jason Downing also completed serving as editor for the bi-annual IM newsletter LTER Databits. The opportunity to assist in its production for three full issues was educational, and valueable. Jason would like to encourage any and all LTER IM's to volunteer as editor as vacancies emerge and to submit articles whenever possible.

Also, in January of this year, we saw the departure of our information management consultant Rebecca Koskela. Her previous responsibilities have been migrated to our site IM and operations were able to continue without significant interruption. To support this transition and to help enhance the data managers' competence, Jason has been pursuing additional training in database administration, computer programming and scripting languages through a variety of delivery methods.

Our site GIS systems are maintained by Jamie Hollingsworth who has recently been active in the LTERmaps development team. He has helped with development, organization and presentation for a recent ASM session, numerous VTC's and an upcoming workshop to be held later this fall. In conjunction with these efforts, we have reviewed and improved our sites' GIS and spatial data management system and hope to remain a testing ground for future network systems.

Additionally, we have been working to refine and expand the capabilities of our sensor network database and web visualization package. Additional climate data is going through the final stages of a significant review process before it will be used to repopulate this database and be available through the VistaDataVision interface. Some additional core data not previously in a database format has been migrated into this system as well when it seemed appropriate and the visualization capabilities are desirable. As this system is capable of hosting many more sensor stations the BNZ manages, we have been working with other research teams as well as the Arctic LTER to tap into these services and allow them to take advantage of this BNZ hosted system.

Discussion Questions Response:

Do we give a lot of attention to a few datasets or limited attention to a lot?
There is definitely a subset of data sets that gets a more significant level of attention than others. For these data, we actively reach out to core scientists and technicians to make sure that these select data are as current, complete, and accurate as possible because we have determined them to be of the highest value for their duration and versatility. These are also some of the most requested and longest data sets. The next level of priority most likely goes to new dataset additions and from students and researchers. Getting the most current data sets documented and input into our system before personnel changes or data entropy for the given project becomes excessive. Transient researchers provide a valuable data stream to be captured but the window of opportunity is short; when they offer new submissions it is critical to free up the appropriate level of attention before the window closes. Our final level of attention goes out to system-wide metadata and EML improvements that should benefit the entirety of our data catalog. This mostly deals with database design and EML generation procedures.

Do you deliver derived or raw data?
We do both. Core climate data and the like are generally left in its rawest form possible but certain data must be aggregated or processed to some level before it can be utilized. Individual project and student data is generally of a more aggregated type and we generally leave the decision of data aggregation up to the submitting researcher.

Do you communicate with your PIs regarding prioritization of data?
We have recently embarked on a campaign to personalize our core data sets by assigning a senior researcher to have oversight over each of the core data sets. The goal of this effort is to get an individual person to assume an ownership role over each of the long-term data sets that are not directly part of any individual research project but are more communal in nature. The "Core Data Contact" will help to address quality control and methodology issues that are best lead by an individual who has experience with that particular data type.

Information management issues are also discussed at our almost-monthly BNZ-LTER Staff Meetings. These include all the site and data management personnel, the BNZ PI and co_PI's, as well as any other specific reseachers that may be particularly suited to a particular discussion item. Agenda items cover a variety of topics but data management issues are often the major topic of discussion.

California Current Ecosystem SiteBytes 2010

CCE LTER Information Management Site Byte - September 2010
Karen Baker

Information Management Committee Inquiry

Introduction
Site Bytes are intended as a general update on what has happened at your site during the last year. Please highlight new developments, ideas, and issues. Site bytes help us to stay informed about what is going on at other sites. Please also be forward thinking and include your thoughts on the following subjects with regards to the developing Network Information System.
We will discuss these questions further at the annual meeting:
Given limited resources, how do you prioritize data contributions to the NIS?
1) Do you give a lot of attention to a few datasets or limited attention to a lot?
2) Do you deliver derived or raw data?
3) Do you communicate with your PIs regarding prioritization of data?

CCE Responses
1) Do you give a lot of attention to a few datasets or limited attention to a lot?
Between these two extremes, we strike a balance. Our strategy has been to create the technical capability to rapidly create and update datasets together with a management interface for submitting to community collections. Last year we focused on improving our information system in ways that put us in a position to work with the NIS development team. We also improved our metadata approach while working with several local groups with complex datasets who had not yet submitted data. In addition, site work focused on streamlining of weather data handling so as to have monthly updates for ClimDB as well as summary plots posted in DataZoo. Quality control has been left to the data submitter to date but we plan to develop some capabilities with this in the next year as part of the information management services and hence to begin more routine submissions of more assuredly high quality data to NIS.

2) Do you deliver derived or raw data?
The raw-derived question seems misleading when stated as a binary. We see multiple stages as

part of a processing continuum that varies by dataset depending upon sampling, instrumentation, calibration, and analysis and so we consider the decision as to what dataset stage to deliver as dataset dependent. For many researchers, the questions of what to submit is still under investigation as they may submit abundances from counts obtained via microscope analysis and then submit carbon converted counts as they work with the data. We find that over time and with experience, our understanding of the data and the data audiences as well as the shared projects such as EcoTrends influences our data delivery.

3) Do you communicate with your PIs regarding prioritization of data?
Yes, this is the strength of an LTER site to have IM guided in practice by local data needs and by research scientists who collect, analyze, submit, and use the data. Data inventories are performed regularly on the local system. The significant technical capacity improvement this year coupled with our planned focus on metadata and data quality reviews next year put us in a position to include submission of data to Metacat as part of our routine data inventory next year.

CCE Site Byte

1) Local Development
Focus this year was on technical development of DataZoo, a site-level information system now used by four oceanographic sites (two LTER sites (PAL, CCE) and two CalCOFI sites (SIO, SWFSC)), in an effort to better meet local needs while developing web services as a foundational element for upcoming site-network development. Redesign of the study-project schema for DataZoo added flexibility in response to metadata needs identified in practice and by the IMC working group effort with project collections. Significant updates were made to the DataZoo management interface that allows creation and editing of organizations, studies, datasets, and their related metadata. New approaches to web delivery through development of middleware and standardized, asynchronous access to databases through a web service layer successfully addressed issues relating to the increasing size and diversity of data. Expanding DataZoo to a three-component system accommodated the development of FileFinder, an interface to very-large collections of hierarchically structured datasets. The prototype of FileFinder was populated with oceanographic CTD data.

Additional web-related category work (core and significant datasets) involved delivery of derived or value-added datasets. An online methods manual was designed and is delivered online. Additional work relating to DataZoo included adding the support for taxonomic identifiers to the attribute qualifier system and use of the Integrated Taxonomic Information System (ITIS) as a taxonomic authority that prompted conversion of NODC taxonomic codes. Two other major tasks involved redesign of query selection so users can select studies within a designated time interval and initiation of station matching required for integration of datasets. Plot performance was improved by replacement of the plot library and options added for contour and bubble plots. External dataset handling like for weather was streamlined to facilitate monthly updates for ClimDB as well as summary plots posted in DataZoo. Improvements were made to dataset documentation as datasets were submitted for publication in EcoTrends. An IM media gallery was created in order to preserve posters as a historical record of development. Finally, backups are now done via the Computational Infrastructure Service at IOD in conjunction with the San Diego Supercomputer. In terms of physical infrastructure, a new server with virtual machine capability was purchased (vSurf).

2) Partner Activities and Leadership
Our team (Mason Kortz, Lynn Yarmey, James Conners) is leading a network-level LTER unit registry and unit best practices effort that received support from a LTER post All Scientist Meeting proposal award. Two cross-site visits (KBS, JRN) and a visit to the Network Office were carried out as part of this project. This work has informed redesign of DataZoo architecture to a web-services orientation. The new unit registry together with its management interface replaces a static, isolated unit dictionary thus enacting one type of data comparability as well as demonstrating a new type of site-network model that enables site participation in NIS development through web services. Our team (Mason Kortz, James Conners) is also leading a joint IMC-LNO web services working group

and serving as a NIS tiger team member (James Conners). Karen Baker is co-leading the IMC Governance Working Group (GWG). The GWG was created in order to facilitate the conduct of LTER business in an arena of growing complexity and responsibility and to support the development of Terms of Reference or by-laws for committees and working groups.

3) Professional Development and Publication
Professional development of information managers was provided through exposure to site-level and network-level development activities with the Unit Registry project. In addition, a weekly summer informatics reading group stimulated discussion and joint learning. A sociology graduate student studying the Design Studio stimulated reflection on our work as well as on our work arena. Lynn Yarmey received a fellowship to pursue two-year masters in Library and Information Science with a concentration in Data Curation through the University of Illinois, one of the national iSchools.

At the LTER Information Management Committee (IMC) annual meeting CCE/PAL members contributed as co-chairs of the IM Governance Working Group and Unit Working Groups. Lynn Yarmey reported on a Unit Dictionary Best Practices Guideline. Karen Baker was elected a member of the Network Information System Advisory Committee (NISAC). Mason Kortz and James Conners prototyped the LTER Unit Registry and contributed to formation of a new Web Services Working Group. At the LTER All Scientist Meeting, contributions included a series of posters, participation on a panel about Network level efforts, and a workshop focusing on curriculum for information managers.

Collaborating on two chapters about the EcoTrends Project contributed to description and understanding of network level information management efforts as well as development of a set of recommendations for future work (Laney et al, in press). Another book chapter explores the topic of digital infrastructure growth (GCBowker, KSBaker, FMillerand, and DRibes, 2010. Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In Int. Handbook of Internet Research). Collaboration with Science Studies partners led to publications in peer-reviewed journals. An interdisciplinary investigation into work with data over time spanning decades provides insight into ‘Big Data’ efforts and the LTER network model (Aronova, Baker, and Oreskes, 2010. From the International Geophysical Year to the International Biological Program: Big Science and Big Data in Biology, 1957-present. Historical Studies in the Natural Sciences 40(2):183-224). A study of design provides insight into the complexity of the development processes at an LTER site (Millerand and Baker, 2010. Who are the users? From a single users to a web of users in the design of a working standard. Information Systems Journal 20:137-161).

4) In Summary
We plan to focus next year on metadata population, quality control, and data delivery as well as on documentation of the DataZoo system. Technical work with the unit registry and web services will continue at a network level. Technical issues that remain to be addressed include integration of the personnel module with DataZoo, on geolocation based on the earlier gazetteer effort, and on management of dataset use logging. Our website content will be updated but migration to drupal/wordpress/etc is not yet scheduled. The application module MediaZoo was redesigned as an API. The move from a photo gallery to a plot gallery and eventually to the media gallery module has reached an appropriate level of abstraction so that it is time to make uniform the five instances of installation currently in use.

Despite significant experience with information management practices and issues at sites, innovation at sites is limited by lack of resources and incentives. The 2010 information management supplement stabilizes our work on DataZoo and collaboration on the LTER NIS this year but highlights the need either to continue such support or to collaborate locally with additional projects in order to maintain a team with technical expertise at the site-level over the long-term.

Cedar Creek SiteBytes2010

The main goals for advancing Cedar Creek's information management in 2010 centered around improving the quality and completeness of our historic datasets and associated metadata. Particular emphasis was placed on reducing the number of datasets measuring the same parameters at the same locations but during different time periods. This effort afforded us the oppertunity to impliment the recommendations of the Units Best Practices created by the Units Working Group. The end result was a reduction in the number of datasets in our data catalog, but an increase in the usability and quality of our data and metadata.

There were two major changes with regards to Cedar Creek's IT infrastructure in 2010. The first was the installation of a site wide radio system. The primary purpose of this system is for communication during our prescribed burning season, but it also has the potential for handling some data collection processes. Second, we transferred our website to a new server. During this process, we examined and fixed potential security problems.

Given limited resources, how do you prioritize data contributions to the NIS?

* Do you give a lot of attention to a few datasets or limited attention to a lot?*

We do both. Data that is considered high priority is given as much attention as is needed to create and make available high quality datasets and metadata. For continuing data for established datasets with established protocols for QC, this may mean a small time commitment. New datasets may require significantly more time. Likewise, lower priority datasets, which are generally historic, may require more or less time depending on the quality of the documentation available.

* Do you deliver derived or raw data?*

Because data exist on a continuum of "rawness", it is difficult to lable data as "raw" or "derived". The general approach at Cedar Creek is to discuss with the researcher responsible for the data to consider how such data is normally presented by other researchers, and the usefulness of various levels of derivation.

* Do you communicate with your PIs regarding prioritization of data?*

Yes, directly and indirectly. We do not have a formal procedure for determining priority, however.

Central Arizona Phoenix SiteBytes 2010

Site Bytes go here

Coweeta SiteBytes 2010

Overview
In 2009, John Chamblee took over from Barrie Collins as Information Manager. The focus of this transitional year was an evaluation of existing resources, an effort to reach out scientists and to community partners connected with the Coweeta LTER, and the transformation of Coweeta's Information Management Office into a production-oriented mentoring lab. The audit covered everything from hardware and physical infrastructure, through to server architecture, databases and applications, and user interface designs. New ties were established with both Coweeta PIs and citizen scientists via Information Management's participation in the design of two large databases

for storing and sharing regional data on water quality in the Upper Little Tennessee Watershed. CWT IM is now a place where undergraduate students gain experience in Geographic Information Systems, database development, and IT support services while helping the Information Manager fulfill the CWT IM Office mission.

In 2010, efforts have been directed at the completion of the two large regional databases that were started in 2009 and an overhaul of Coweeta IT and IM infrastructure. IT upgrades have been undertaken at both the Coweeta IM Office at the University of Georgia and at the Coweeta Hydrologic Laboratory campus in Otto, North Carolina. A "new" Information Management Architecture is also now in development. This upgraded architecture is new only to Coweeta and is based on proven technology. Coweeta's upgrade has been made possible by a collaboration between John Chamblee, Coweeta, Wade Sheldon and the Georgia Coastal Ecosystems LTER. Since Wade and GCE are also located at UGA, this partnership is a wise use of LTER resources. Coweeta and Chamblee have both benefited greatly from Sheldon's long experience and the generous contribution of his time and expertise. Coweeta is also benefiting from the use of the GCE technology Sheldon developed. The collaboration promises to be mutually beneficial over the long term. The remainder of this article goes into detail concerning IT and IM upgrades at Coweeta and also briefly introduces the aforementioned regional database projects.

IT

January of 2009 marked the third month of Coweeta's most recent renewal. As a result, many of the hardware components both in the Coweeta IM Office and the CWT field site were near their end of life. After completing an audit of all systems, we began the process of replacing network appliances, servers and workstations. On the Coweeta Hydrologic Laboratory campus, we have replaced 80% of the routers, switches, security appliances, and media converters. We will have a new integrated, enterprise-scale wireless access cloud installed by late 2010 or early 2011. We have also upgraded the backbone serving the IM Office at the University of Georgia, adding 24 gigabit ethernet lines to our existing set of 18 10/100 ports. In terms of workstation management, we are engaged in an ongoing process of replacing a patchwork of workstation models with two standard model configurations. We are taking advantage of University of Georgia license agreements to provide a standard suite of software. This new approach allows us to cut our OS-related troubleshooting time through the use of disk images. At the UGA CWT IM office, all work-related files are now stored on a central file server. If a workstation fails, we simply follow a typical campus computer laboratory model by erasing the drive and installing a clean drive image. We expect to have a similar model implemented on the Coweeta campus after all of the more recently purchased computers reach their end of life.

Coweeta IM is also in the process of overhauling our server platforms and backup systems. We have already established interim production platforms for MS SQL Server, mySQL server, ArcGIS license server, as well as file and print services. We have also built a new web hosting platform that is being used to develop the new Coweeta website and a new server for handling off-site backups. File and print services, MS SQL Server, and the ArcGIS license server are being hosted on a Windows Server Platform, while all other services depend on an enterprise-scale release of Linux. The physical devices hosting these systems are stand-alone towers featuring RAID 1 storage to ensure redundancy. Our older, manually operated off-site backup system relied on external hard-drives, bootable CDs, and manual transportation offsite. This labor-intensive model has been replaced with a new, automated, network driven disk-to-disk backup model. By employing bacula, a free backup client/server system for Linux, we now maintain daily incremental backups of our file and web servers. Database backups are handled with cron jobs and scheduled tasks, depending on the OS, and daily database backup files are handled with normal bacula routines. The bacula server is a 3 TB RAID 5 rack mounted system with redundant power supplies. At present, this server possesses the capacity to handle two months of full backups for our production and development systems, augmented by daily incremental backups for the same period. Through a cooperative agreement with University of Georgia Laboratory of Archaeology, the backup server is hosted in one

of their curation facilities. This facility is more than a mile from the IM Office, is climatically stable, and is one of the most secure locations on the UGA campus. By the end of the fiscal year, the new Coweeta production web server, mySQL server, file server, MS SQL server, and ArcGIS license server will be hosted as virtual machines on a single, 6 TB RAID 5 rack-mounted system running VMWARE ESX. This system will reside in the IM lab. The disk capacity for the offsite backup system will also be expanded to handle this upgrade.

Coweeta's final IT update involved the planning, funding, design, and implementation of a teleconferencing room for Coweeta. This project, dubbed the Sustainable Human Ecosystems Teleconferencing Room (or SHELTR) was funded by a grant obtained by Coweeta for the purpose of increasing interaction between meetings in spite of the fact that Coweeta PIs are scattered across ten different institutions. SHELTR provides seating for eight, a Power Mac Desktop computer with bluetooth keyboard and mouse, and two 40" LCD monitors. The machine runs any OS X application, and, thanks to a VMWare Fusion installation, supports many Microsoft Windows programs. Facilities also include a high quality, echo-canceling dedicated speakerphone and a webcam. Voice Over IP (VOIP) capabilities are available through a noise canceling portable USB microphone and wall mounted speakers. Teleconferencing services are provided through a Coweeta LTER subscription to GoToMeeting [125]. This program provides screen-sharing, VOIP, and conference call bridging for participants from up to 16 separate locations. SHELTR users may participate in bridged teleconferences hosted elsewhere via X-meeting and point-to-point meetings are possible via Skype and iChat. SHELTR's webcam provides a clear, continuous picture of SHELTR participants. When used together with the conference phone, the webcam provides a simple, easy one-way video-conferencing solution.

IM

As was the case with Coweeta hardware, many of the applications powering Coweeta's website and database tools were also in need of refreshing. Our specific needs revolved around the need to integrate a content management system into the website development process and the need for more robust relational database models for handling our publication catalog, personnel databases, and metadata. Since Chamblee already has experience with PHP mySQL, and some object-oriented applications, Coweeta IM settled on Drupal as a content management solution. In terms of relational databases, the goal at Coweeta IM was to implement relational models that would adequately allow us to document the linkages between personnel, projects, project data, and publications. Fortunately for us, such a model existed in the form of the GCE Biblio and GCE Metabase ER-models. Even more fortunate was the fact that GCE is another University of Georgia based LTER site and the fact that GCE staff -- and Wade Sheldon, in particular -- were willing to participate in a cross-site technology transfer project whereby GCE provided working models of their systems. Coweeta IM staff then adapted these models and are currently populating them programmatically from existing Coweeta databases. This project has also provided the opportunity to build and test (in a production environment) the use of a joint hosting solution in which CWT-adapted versions of GCE's ASP code stack are hosted on GCE's IIS-based resources, but served through Coweeta's Apache-based systems via reverse proxy-passes.

As of this writing, GCE has provided the technology transfer and the IIS services needed for the proxy system. Coweeta has successfully migrated the entire Coweeta bibliographic database to the GCE Biblio model and code stack. These resources are available at http://coweeta.uga.edu/pubcatalog [126]. In addition, Coweeta has nearly completed the process of migrating our Personnel database to the Personnel module within the GCE Metabase ER model. We are still awaiting updates from many faculty and graduate students to flesh out the content, but the system is otherwise fully functional and is available here [127]. We are also in the early stages of migrating all of the metadata for our core data to the Metabase. This process will be complete by the end of the 2010 calendar year.

Coweeta's collaboration with GCE is part of a larger collaboration with SBC and MCR Information Management to develop common data models for handling metadata storage and production in relational database frameworks. At MCR and SBC, the focus is on adapting GCE models to operate on the postgreSQL platform, while we all make contributions to the overall ER model and help to expand already robust modularity of the GCE Metabase. In addition to collaborating with GCE, SBC, and MCR on changes to the ER model, Coweeta will be developing x-query based tools to import our own existing legacy EML -- originally coded individually for each dataset using Morpho -- into the CWT implementation of Metabase. In late 2010 or early 2011 CWT implementations of Biblio and Metabase will be integrated into a new Drupal-based Coweeta website. This site is well into the development phase and undergraduates are currently populating and updating our less dynamic content. The SQL Server based modules will be integrated using iFrames that utilize the recently established reverse proxy passes.

Coweeta's other major Information Management projects have involved the development of large-scale regional databases concerning water quality in the Upper Little Tennessee Watershed. The first collaboration is a citizen-science initiative that involves Coweeta Information Management, the Highlands Biological Station [128], and, principally, the Little Tennessee Watershed Association [129]. Since 1988, Dr. William O. "Bill" McLarney has been conducting fish species inventories at stream sites across the Upper Little Tennessee River Basin. He has conducted surveys sufficient to calculate an adapted version of the Tennessee Valley Authority's Index of Biotic Integrity at over 168 sites on over 500 different occasions -- including at 16 sites annually for over 15 years. This database is one of the largest of its kind and includes over 8,000 individual observations on hundreds of thousands of specimens from dozens of species categories.

Before this collaborative effort began, the data were stored in a semi-structured excel spreadsheet that was not amenable to machine driven analysis of any kind. Coweeta's role was to develop an entity relationship model for the data, work with LTWA staff to restructure the data for import into the new relational model, and then oversee undergraduate student interns from the Highlands Biological Station's fall internship program, the LTWA, and the CWT IM lab itself to perform quality control. All data were checked against the original and, at this point, the database is current through 2009. Due to the fact that internet connectivity at LTWA facilities is so poor, updates are handled via a standalone Microsoft Access application that is then periodically mailed to CWT, placed into simple version control folder, and updated, on annual basis to a mySQL database hosted on Coweeta's production mySQL server. The mySQL database will serve as the basis for a simple php, mysql, and Google maps-based application that will be completed by the end of calendar year 2010.

The 20 year (1988-2008) dataset has already yielded dividends in terms of student-driven citizen-science, conservation and evidence-based decision-making. The students from Highlands Biological Station produced a report on a preliminary comparison of the data against Coweeta's forthcoming 2006 land cover classification. That report is available here [130]. More importantly, however, has been the collaboration's contribution to citizen-science-based decisions by the North Carolina government. By providing a quality assured copy of the data to the North Carolina Natural Heritage Program, Coweeta and the LTWA were able to jointly provide the NCNHP with the data they needed to significantly expand the spatial footprint of the aquatic species diversity heritage zone for the Upper Little Tennessee Watershed by adding several previously unlisted tributaries.

In addition to collaboration with our external partners, Coweeta IM has been instrumental in a large scale regional water quality analysis effort involving nearly all Coweeta's PIs. The Synoptic Sampling Program was initiated in 2009. In this inaugural year of a multi-year effort, Coweeta scientists took samples and made observations relevant to the fields of stream chemistry, stream morphology, hydrology, species diversity, human land use, and socio-economic diversity. Coweeta IM designed the GIS that allowed for the integration of the field program, built the database for the stream chemistry and geomorphic measures, and assembled the public records on socio-economic diversity. Given that 58 sites across a gradient of exurban land use were sampled, and given the

long-term nature of the project, these data are likely to be an important addition to Coweeta's Core Data catalog. Metadata are currently being written and the release of the data in accordance with NSF guidelines is anticipated soon.

Coweeta IM's final project has been a re-organization of the work model at the IM office. Coweeta IM has always consisted of a single, full time Information Manager, augmented at times with temporary or undergraduate assistance. This model has now undergone a slight but important shift in which the focus on temporary employment is not just on data production, but also on student mentoring. The Coweeta IM office now employs between 2 and 5 undergraduate students each semester. We also provide occasional graduate assistantships. These students learn GIS production, metadata authorship, and the value of collaboration. Six out of the seven students who have graduated from UGA and participated in this program have gone on to advanced studies or careers in conservation, informatics, or GIS. The changes that Coweeta IM has undertaken over the last two years will provide a firm foundation for handling the growing size and complexity of the data sets in our care and will provide IM staff and students with plenty of exciting challenges.

Florida Coastal Everglade SiteBytes 2010

Overview
The FCE Information Management team consists of the FCE Information Manager, Linda Powell, and the FCE Project Manager, Mike Rugge. The FCE IMS continues to facilitate the site's scientific work and to ensure the integrity of the information and databases resulting from the site's coastal Everglades ecosystem research. The major focus of its Information Management effort during this past year has been on FCE web site (http://fcelter.fiu.edu) enhancement in preparation for its 2009 National Science Foundation (NSF) mid-term review. In October 2009, the National Science Foundation's LTER Review team recognized our IMS as 'one of the finest in the LTER Network' and wanted to see the FCE IMS team play a leading role in the continued development of 'best practices' across the LTER network.

Website Development
Although the FCE web site (http://fcelter.fiu.edu) went through a major redesign in 2007, this year's focus has been on FCE web site enhancement where we incorporated several LTER working group initiatives to improve standardization of data search and access across LTER sites through adoption of controlled vocabularies and common interface features. The website serves as the primary portal for dissemination of information about the FCE LTER program, for distribution of datasets, to coordinate our Education and Outreach activities, and to aid FCE scientists and students in their research so it is important that the FCE information management team continue to improve existing web pages and expand the web site capabilities. The small changes made to the FCE 'Data' web page structure allow users to more easily choose between FCE data products like our signature research datasets, LTER core research data and the FCE physical/chemical online database. Links to important LTER Network and outside agency data resources remain an important part of the FCE website 'Data' section. This year, the FCE IMS team worked with members of FIU's Periphyton Group to migrated their diatom image database to the FCE LTER data and web servers. The FCE IMS team created a new web interface to the database (http://fcelter.fiu.edu/data/database/diatom/) and linked the database to the FCE Data section of the FCE website.

Support for EML Metadata
The FCE IMS has fully adopted the LTER network metadata standard Ecological Metadata Language (EML) and one hundred percent of the FCE tabular data are accompanied by a Level 5 (Data Identification, Discovery, Evaluation, Access and Integration) EML (XML) metadata documents. FCE EML documents are harvested daily to the LTER network metacat XML database. The FCE Excel2EML metadata converter tool and template have been made available to the LTER network and broader ecological community via the LTER CVS repository and as download link on

the FCE web site (http://fcelter.fiu.edu/research/information_management/tools/). The FCE IM team lends its expertise to site and network researchers when necessary by providing application support for the Excel2EML tool and assisting with metadata entry.

FCE Data
All of the FCE LTER core data and metadata files from individual research studies are stored in a hierarchical flat file directory system. FCE project information and minimal research data metadata are stored in an Oracle10g database that drives the FCE Web site. This hybrid system (flat file and database) gives FCE researchers, network scientists and the general public an option to download complete original data files submitted by individual FCE scientists in addition to downloading queried data from the Oracle10g database. Core data are made available to the public within two years of data collection and are accessible on-line in accordance with the FCE Data Management Policy.

Because we feel that it is extremely important that published online data be accessible at all times, the FCE IMS has implemented a versioning system where all previously published data are unchanged. Changes in data values or newly appended data will result in the creation of a “new” version of the dataset as described under the “Data Organization” section of the data management policy. Currently, the FCE archive contains 429 FCE datasets, of which a total of 397 are publically available.

Contributions to the NIS
Given limited resources, how do you prioritize data contributions to the NIS?

• Do you give a lot of attention to a few datasets or limited attention to a lot?

1. I give a lot of attention to a lot of our data sets as each data set submitted gets equal processing time. Our long-term monitoring data gets versioned with each new submission and the majority of the FCE has been tabular data.

* Do you deliver derived or raw data?
2. We deliver both derived and raw data but have asked the scientists to submit their data in the ‘raw’ format when possible.

• Do you communicate with your PIs regarding prioritization of data?
3. No, the prioritization of data is determined on the order of data submission.

Georgia Coastal Ecosystem SiteBytes 2010

Major Happenings
Fall 2009 IM activities were dominated by preparations for our second mid-term site review, which included an all-day IM review session prior to the main review. It was an exhausting experience (on both sides of the table), but it was satisfying to have an opportunity to discuss our IM approach and information system design in depth. Just prior to the review we finished implementing a comprehensive research projects database linked to dynamic research question web pages (below), which helped acquaint reviewers with our project goals and productivity. This database proved both helpful and "hurtful", though, as it highlighted the number of concluded projects that have not yet submitted publications or data. The review went quite well, though, and we received excellent feedback on our IM program and site research.

After the review we focused on deploying some new technology to extend our information system, including an eXist native XML database server, Apache reverse proxy server (for server aggregation and url abstraction), and Trac software development CMS for several software projects including

the GCE Data Toolbox (below). After installing and hardening eXist, we deployed a production version of the XQuery-based research project search form [131] developed during the LTER ProjectDB working group meeting at LNO in April 2009.

We also continued our efforts to use the Google Maps API to provide dynamic map visualizations on the GCE web site. For example, we deployed a production version of the interactive site map [132] John Carpenter developed in collaboration with the LTERMaps working group, and we designed SQLXML query templates to produce dynamic KML files directly from geography tables in our metadata database. We have now implemented several mapping web services using SQLXML to provide Google Map displays and Google Earth KML downloads for study sites, point locations, and data set coverages. Recently we extended this capability to our EML implementation, so we can automatically add KML entitites (referenced in methods/sampling/spatialSamplingUnits elements) for every data set in our catalog.

Unfortunately John Carpenter announced his departure in May 2010 for a wildlife job in NC, so our GIS work went on hold and the remainder of the Spring was spent developing transition documents and preparing to re-post his position. We hired Travis Douce in July to take over for John, and he's been working hard to get up to speed. We hope to resume our productive collaboration with the LTERMaps effort at the 2010 IMC meeting, which Travis will attend.

Projects Database

As mentioned above, we developed and populated a research projects database in 2009 that is fully compliant with the ProjectDB schema. We implemented the back-end database using SQL Server rather than eXist, though, because the majority of content referenced in GCE project descriptions is already managed in our existing SQL databases. However, the database schema was designed to produce lter-project XML from the start, and we were able to leverage our existing EML-generation infrastructure to produce Project EML from this database with minimal effort. We also developed code to periodically retrieve and cache Project EML documents in our eXist database in order to leverage the XQuery search form and web services developed for the ProjectDB effort. We also leveraged and extended the XSLT stylesheets developed for ProjectDB for styling both documents streamed from the SQL database (via ASP scripts) and pulled from eXist to provide uniform functionality and avoid duplicating effort.

We additionally developed dynamic research question [133] web pages that draw content from the research projects database. Each question page lists all projects addressing that question (nested by research component), along with all associated publications, data sets and personnel. We also provide dynamic cross-links to questions and research projects on data set summary pages in our data catalog [115], and plan to add links to researcher personnel pages in the near future. The projects database therefore serves as a powerful link between research questions, personnel and products, which was a key design goal of the ProjectDB working group.

LTER IM Collaborations

In addition to our successful IM collaborations with the ProjectDB and LTERMaps working groups, we have also begun to collaborate with other LTER sites to develop and share technology and insights. For example, CWT underwent a significant transition in both staff and research focus in their latest funding cycle, and they approached us about leveraging significant portions of our information system for data processing and information management. We have worked closely with John Chamblee this year to transition GCE software (GCE Data Toolbox) and databases (GCE_Biblio and GCE_Metabase, along with ASP and SQLXML middleware) for use by CWT, and they are now starting to put those systems into production for their site. Feedback from John and lessons learned from this process have been very valuable for improving our own system, and using common technology will provide future collaboration and cost-sharing opportunities for both our sites. Both John and I have also been collaborating with Gastil Buhl (MCR) and Margaret O'Brien (SBC) on developing a next-generation metadata database, and porting elements of the GCE_Metabase to Postgres for their system. Explaining design and operation of a 9-year-old

production database that's already been through several paradigm shifts (particularly EML mapping) can be a cringe-inducing experience, but it's a great opportunity to revisit and strengthen our systems for the next generation of challenges. I am also working with Kristin Vanderbilt (SEV) to explore use of the GCE Data Toolbox software for processing SEV sensor network data. After many years of working in relative isolation it has been rewarding to work closely with other IMs on these collaborations.

Software Development
We have been developing and sharing software code for several years at GCE, but in early 2009 we established a more robust infrastructure for software development, including an SVN server, better code documentation practices, and automated software release builds. This year we extended this infrastructure by setting up a Trac software development CMS [119] for the GCE Data Toolbox software to provide milestones, bug-tracking, code distribution and web-based SVN access. In September 2010 we also began releasing source code for the GCE Data Toolbox under a GPL license to foster broader use in the Ecoinformatics community. We have also been using our SVN for managing all web content and script code, SQL database code, SQLXML templates and other software resources. This has already proved indispensable for tracking down and fixing problems after updates to various code-bases.

Database Redesigns
We are preparing to enter our 11th year as an LTER site, and many of our initial information system components are starting to show their age. In addition to the metadata model redesigns described above, we will be reviewing and updating other database models (e.g. taxonomic database, cruise database) and continuing to develop web service and AJAX application layers to improve portability and lessen dependence on web application middleware (e.g. ASP). We also set aside funds from our 2010 IM supplement to develop a lab sample management database that can connect to long-term monitoring databases for moorings, cruises, and weather stations.

Flux Tower
Last, but not least, we received supplemental funding to purchase a flux tower package to install in the marsh. We are collaborating with PIE, FCE and VCR to develop common approaches for characterizing CO2, heat and other gas fluxes from marsh-dominated environments. This is our first foray into wireless networking at a remote location, and we are currently conducting site surveys for configuring 900MHz communication to and from the candidate tower locations. We hope to learn from experiences at VCR, AND and elsewhere as we establish our field data network. Another issue is that off-island bandwidth is currently very limiting (several shared asymmetric DSL lines over microwave), but a recent field station improvement grant to the UGA Marine Institute coupled with a block grant to our local Telco for beefing up our microwave system should dramatically improve that issue over the coming two years.

Given limited resources, how do you prioritize data contributions to the NIS?

* Do you give a lot of attention to a few datasets or limited attention to a lot?

Effort varies, but given the growing number of long-term datasets that GCE IM staff are primarily responsible for acquiring, processing, QA/QC'ing and documenting, "limited attention to a lot" is becoming our primary operating mode.

* Do you deliver derived or raw data?

Both, but raw data continue to be our priority. We generally view derived data as value-added products to be generated on request, but not archived in our IS (or contributed to the NIS) unless called for by a cross-site synthesis effort (e.g. EcoTrends). There are so many potential derived data products that could be produced, and so many decisions to be made when producing them, that I

believe specific synthesis requirements or network-wide standards should be established to guide generation of derived data on a large scale if that becomes a stated goal of LTER.

* Do you communicate with your PIs regarding prioritization of data?

Occasionally, but we set broad goals and priorities in the GCE Executive Committee and that guides most decisions.

Harvard Forest SiteBytes 2010

1. Activities over the past year

After years of planning, the Harvard Forest Field Wireless Network (HFFW) became operational in the spring of 2010 with funds from NSF (LTER and RAPID), DOE, and Harvard University. The HFFW provides high-speed Internet access to field sites with line power, enabling scientists to monitor and control their equipment remotely and to collect and process data in real time. The HFFW is implemented as an extension of the Harvard University network and is jointly administered by the Forest and Harvard Network Operations. In its current form the HFFW connects to major research sites in the Prospect Hill Tract, including three eddy flux towers and two global warming experiments. Users in the vicinity of these sites have direct access to the Internet via wired or wireless connections. Authenticated users who are offsite can access their equipment remotely using a dedicated VPN tunnel. Current examples of (near) real-time datasets include the Fisher Meteorological Station and four phenology webcams. The HFFW is designed to support future expansion as the use of sensor and wireless technologies grows and as additional funding becomes available.

A snow pillow was installed at the Harvard Forest last fall (with LTER supplement funding) to provide continuous measurements of the water content of snow pack, a critical factor in the water cycle of central New England.

The Analytic Web project, a collaboration between ecologists at Harvard Forest and computer scientists at UMass Amherst and Mt. Holyoke College, continued this summer as part of the summer REU program at the Forest. Two computer science students focused on the problem of creating precise audit trails for the processing of sensor data from hydrological gages.

In recent years Harvard University has increased its support for information technology at the department level. We are currently in the process of migrating some of our production servers from the Forest to virtual machines on campus. If this move is successful it will help to relieve us of some of the cost and effort required to maintain production servers onsite.

2. Contributions to the NIS

The Forest has a long tradition of historical studies and document archiving, extending back more than 100 years. Probably because of this, our approach to electronic data archiving has been to include as much research data as possible without regard to the source of funding. As a result the IM staff can spend less time on each dataset. Researchers who wish to work here must submit an online research project application each year and must be up to date in meeting their data obligations. For the most part we post raw data (with a few exceptions, such as eddy flux data, where the level-0 data are not of general use). In general all data are expected to be posted, but we do confer with our site leadership if there are important projects that seem to be lagging behind.

Hubbard Brook SiteBytes 2010

Site Bytes go here

Jornada Basin SiteBytes 2010

2010 Jornada Basin LTER Site Byte

The last year has been very busy. The Jornada Basin LTER (JRN) had its mid-term site review last year. We also virtualized our server farm to prepare for making them highly available using free enterprise class software. We have more than doubled our storage capacity in the last year within our storage area network (SAN) to meet the needs of Jornada research efforts and collaborations. During the process, we migrated and upgraded server operating systems, server platforms (32-bit to 64-bit), server applications, and data stores for all production servers. We also used 2 older servers to deploy virtual servers to support site and research project development efforts.

Server Purchases
The Jornada Experimental Range (JER, USDA ARS) purchased 2 new Dell PowerEdge R710 servers in the last year or so. The Jornada Basin LTER (JRN) is currently purchasing another R710 server using 2010 LTER Site Supplement funds. The R710 servers are used to support all Jornada production virtual servers. 2 older PowerEdge 2950 servers are being used to support virtual servers used for development.

We are using Citrix XenServer software for server virtualization. XenServer is a free enterprise-class virtualization hypervisor. Virtualization reduces administration and maintenance costs of our server farm and allows our servers to be highly available and load balanced to improve performance.

In addition, JRN and JER are cost sharing the purchase of Citrix XenServer Essentials Enterprise for 5 servers to allow automated load balancing and high availability of virtual servers running within 2 XenServer resource pools; production (3 - R710 servers) and development (2 - PE2950 servers).

Storage Expansion
JRN purchased a new IceWeb TE444 SAN storage enclosure half-populated with 24 TB of storage last year. In addition, JRN is currently purchasing hard drives to completely populate the older TE444 storage enclosure with site supplement funds to provide an additional 18 TB of storage. The SAN provides storage for our virtual servers and will provide redundant storage once the new hard drives arrive.

Server Virtualization, Migrations, and Upgrades
All Jornada servers were virtualized within a production resource pool using Citrix XenServer this year, with the exception of 1 physical Active Directory server. The following upgrades (including system and application patches) and migrations were performed during the virtualization process.
• Services, imagery, data, network shares, and permissions were migrated from Novell OES2 to Windows Server 2008 R2
• Operating systems were migrated from 32-bit to 64-bit along with appropriate software applications
• Databases were upgraded from Microsoft SQL Server 2005 to 2008 R2
• All Windows servers were upgraded from Windows Server 2003 to 2008 R2 during virtualization
• 2 Active directory servers (physical, virtual) were updated from Windows Server 2003 to 2008 R2
• The mail server was upgraded from Novell GroupWise 7 to 8 and the operating system was migrated from NetWare 6.5 to Windows Server 2003 and mail accounts were migrated

Novell cluster services (NCS) were leveraged to allow the websites and FTP sites to failover and be hosted by the file server while the web server was added to the production resource pool and the

web and FTP sites were virtualized.

The GIS server was configured as virtual host server to support a new development resource pool. The development resource pool contains virtual machines running Windows Server 2008 R2. The virtual servers that were created within the development resource pool support active directory, GIS, database, file, image, and web services needed to support software integration and development and prototyping within a separate environment from our production environment. The Jornada can now quickly clone and export virtual servers from the development resource pool to the production resource pool without impacting services running in either pool. As we complete migration of data from the file server, it will be added to the development resource pool. The development resource pool is being licensed for high availability. Older servers will be used to support development efforts and science projects until ready for deployment in the production resource pool. The file server will be added to the development pool to provide failover and automated load balancing capabilities for the development pool.

LTER Unit Registry Implementation
The Unit Registry Working Group funded Ken Ramsey and Justin Jensen (student programmer) to travel to San Diego in August to implement the LTER Unit Registry into the Jornada Information Management System (JIMS). The units table in our SQL Server database is now populated and synchronized from the LTER Unit Registry with units, metadata, and XML snippets used to create EML custom units for Jornada metadata. Our existing documentation forms (PHP pages) were also modified to allow use of the unit registry. The last step will be to update our XSLT transformations in the Spring Semester to use the custom unit STMML snippets in our units table to create level 5 EML for all Jornada datasets.

Discussion questions:

Given limited resources, how do you prioritize data contributions to the NIS?
This has not been determined as the new NIS has not been developed. We are planning towards integrating JIMS into the new NIS as it is developed. To that end, we are focused on populating all research data into the JIMS relational databases and to generate level 5 EML for all datasets stored in the relational databases. We are also implementing current and evolving best practices from the LTER IMC into JIMS. I answered the following questions in relation to JIMS, not the LTER NIS.

Do you give a lot of attention to a few datasets or limited attention to a lot? We do both. Mostly give a lot of attention to a few datasets. More attention is typically given to new datasets and to updating ongoing datasets. However, we have and continue to update all metadata in our relational databases. In the coming year, we plan on focusing on populating data within the relational database and deriving EML level 5 for all Jornada datasets. This will require verifying all metadata and performing quality assurance data.

Do you deliver derived or raw data?
Both.

Do you communicate with your PIs regarding prioritization of data?
Yes, but the level of communications will ramp up in the next year as data is being documented and populated into the databases.

Kellogg Biological Station SiteBytes 2010

We have not made a lot of changes this past year in our information management. The big effort was writing tools to add the GLBRC data to the LTER metadata system. GLBRC wanted to restrict

access to data so we had to implement access controls on the web for the first time. As a side benefit it allows us to serve PDF's from the citation database to people who are signed in.

Do you give a lot of attention to a few datasets or limited attention to a lot?
We generally lump datasets together. I like to think in terms of data streams, and not discrete units. I think this results in fewer datasets and so we can spend a little more time writing tools to monitor the quality of the datasets.

Do you deliver derived or raw data?
We concentrate on delivering good quality raw data. Derived data is added if we can get it but we don't generally seek it out. I would like to start collecting workflows from some of the greenhouse gas balance papers so that the calculations can be rerun as new data becomes available.

Do you communicate with your PIs regarding prioritization of data?
Not regularly. Our base datasets are the highest priority, followed by researcher submitted datasets, followed by student datasets.

Konza Prairie SiteBytes 2010

In the past year information management at Konza has been largely focused on re-inventing our web site (currently nearing finish) to provide a much more user friendly environment as well as developing a back end "CMS" to allow for easier updates to the system. These products should allow for a site which is much easier to maintain as well as better, easier access to our data other resources. Additional focus has been on bringing our GIS and burn history data into a more usable format.

Luquillo Experimental Forest Site Byte 2010

LUQUILLO LTER ANNUAL Site Bytes
From: Eda C. Melendez-Colom, LUQ LTER IM

Outreach and Schoolyard activities:

Presented information management workshop in the third LUQ LTER Internship held at El Verde Field Station from November 22 to 23, 2009. Published data, metadata and presentations given at the activity in the following web page:
http://luq.lternet.edu/outreach/schoolyard/Activities/2009Internship/200... [134]

Organized an activity with the NASHUA SOUTH AND CAMPBELL high schools from New Hampshire to give students the opportunity to learn about LUQ research at the EL VERDE FIELD STATION, including an activity involving data gathering at the field.

Publications resulting from this activity:

- LUQ LTER Information Management Education Exchange, Communication & Collaboration with Scientists and other Community Members. 2010. In LTER Databits Spring Issue. Commentary.(http://databits.lternet.edu/issues/115 [135])
- Developed a web page at: http://luq.lternet.edu/outreach/SCHOOLS/NashuaMain.html [136]

Information Management Development Activities:

The LUQ Information Management Committee (LUQ-IMC), comprised by three LUQ LTER scientists and information manager, developed a set of scientific keywords that have been primarily applied to the LUQ data sets and will be used as part of the new IMS data and metadata model.

Three weeks mini sabbatical at the LTER Network Office (LNO) to advance the knowledge in the use of the content management system and further develop the new LUQ Information Management System (IMS) and web site. LNO staff members Inigo San Gil and Marsh White mentored the information manager in this development progress. Progress was done in the development of the data model as the LUQ data set vocabulary was incorporated into the system and the scripts to automatically produce EML packages for the data sets incorporated into the system using forms was incorporated.

Publications resulting from this visit:

- Developing a Drupal "website-IMS" for Luquillo LTER while learning Drupal . 2010. In LTER Databits Spring Issue. Commentary. (http://databits.lternet.edu/issues/115 [135])
- Use cases of ecological integrative information systems: The Luquillo and Sevilleta Information Management Systems. Inigo San Gil ,1, Marshall White1, Eda Melendez and Kristin Vanderbilt. Submitted to Metadata/semantics International Conference (Madrid, October 2010).

LTER Network Activities:
Member of the Drupal working group of the LTER Information Management Committee: Developed diagram of LUQ Metadata schema to share with group and that will serve as part of the documentation for the new IMS of LUQ. Obtained an LTER Supplement grant that will be used in part for developing modules in DRUPAL by this group.

Member of the LTER IMC Governance Work group. Helped developed the concept and template for documenting IM activities, Terms of Reference (ToR).

Other Network Activities:

CTFS Workshop to learn SQL and add LUQ Plot data to the LUQ CTFS database in Panama.

Participated in the SECOND INTERNATIONAL ANALYTICAL WORKSHOP ON DYNAMIC PLOT ANALYSIS TOOLS DEVELOPMENT to be held on the 19 - 23 July 2010 at our Forest Research Institute Malaysia campus. The main theme of the workshop is “Promotion of Dynamic Plot Database Application and Tools Design”. During the period, discussion of dynamic plots data sharing and application internationally is planned.

- Contribution: Presentation on Metadata posted at: http://www.facebook.com/video/?oid=114899015228942 [137]
- Post workshop contribution: On stimulating data sharing by scientists: http://www.facebook.com/topic.php?uid=113263762058952&topic=33#!/topic.php?uid=113263762058952&topic=41&post=48#post48

Participated in the selection of an Information Manager for the San Juan ULTRA Project and started his training in the metadata standards and information management protocols of the LTER

Specific data management issues:

- Do you give a lot of attention to a few datasets or limited attention to a lot? A lot of attention is given to the data that LUQ IM manages. These are about 14 long term data sets. In terms of the metadata, continuous attention is given to a bigger subset of the LUQ Data Catalog.

- Do you deliver derived or raw data? Mostly raw data, but some derived (especially meteorological data)
  Do you communicate with your PIs regarding prioritization of data? Yes, constantly. Prioriy is established based on the field activities for data gathering of the projects.

McMurdo Dry Valleys SiteBytes 2010

Site Bytes go here

Moorea Coral Reef SiteBytes 2010

Flying the plane while building it
This year, 2009-2010, effort has been divided between developing the IM system and maintaining the datasets. The MCR site leadership has recognized the need for an improved IM system and has supported Gastil in these efforts. The science coordinator and the technicians at MCR have contributed to metadata and data maintenance.

Most of the MCR IM System was redesigned in 2010, retaining some of the existing architecture while adding more dynamic web content. At the same time, MCR and SBC migrated to new web, file, login and database servers. Throughout this process the system continued to be used. The IM at SBC describes this as “flying the plane while rebuilding it.” The MCR and SBC Information Managers actively collaborate.

An Information Management Plan [138] was formed to guide development of new features.

A recipe for PASTA
We recognized the need to create EML dynamically in congruent data packages, as systems at some other sites do. Rather than re-invent such a system, we plan to port the GCE system (based on Microsoft SQL-Server and ASP.NET) to our architecture (PostgreSQL and Perl/PHP). Wade Sheldon is advising this effort.

However, realizing our old website was insufficient for short-term needs, a stub database, just sufficient as a back-end for PHP web pages [139], was designed and deployed in February and March of 2010. These pages cross reference MCR publications, people, datasets, and research themes.

Having your EML and eating it too
The local data catalog displays data packages as HMTL pages from the EML which is queried directly from the LNO Metacat, and transformed with a Perl script and XSLT provided by the SBC-LTER site IM. The same code is used, differing only in CSS presentation and menu header.

This readable, tabbed display of EML leverages the effort invested to add metadata content, which in turn motivates the further enrichment of EML documents. More complete taxonomic coverage was added, using the taxonomy database. Keywords were revised to fit the NBII Thesaurus or MCR controlled theme vocabulary. Methods were attached at more granular levels. MCR investigators responded enthusiastically to the new tabbed display and were responsive to requests to provide further metadata.

Frequent news from the LNO about NIS development inspired Gastil to scrutinize MCR data packages to see how well we might meet new metrics expected soon. Gastil ran 12 MCR data packages through the EML congruency checking application at the Taiwan Forestry Research Institute http://metacat.tfri.gov.tw/modules/ [140] and corrected the EML or data tables to comply. The

remaining datasets, with too many columns or Type II data, could not be checked, so we look forward to a similar LTER tool.

Stretching EML till it breaks
A different kind of dataset, that has not been fully described in EML, is the time series photo record of coral reef quadrats. This photo data is viewed [141] on a web page. In 2010 this was enhanced to view a time series rather than one year at a time.

Streaming real-time data does not fit static XML documents. MCR collects oceanographic bouy (CTD) and weather station data in near real time. In collaboration with a group at CalIT2, these data are streamed through DataTurbine and can be viewed though web [142] and the RDV desktop interfaces. But for the data catalog, static snapshots are appended once per month. Daily aggregates are queried from the database and reformatted for ClimDB harvest once per month.

Do you give a lot of attention to a few datasets or limited attention to a lot? MCR only has 25 data packages. Of these, the 19 with Type I data get more attention since data is appended and has to be re-validated. The two new packages received the most attention, as their data differ from any existing template. The physical oceanography datasets require the least effort since the data appended comes out of Matlab, and is structurally valid. All EML docs were significantly enhanced in the past year, even the Type II.

Do you deliver derived or raw data? Both. Meteorological data are posted before QC (and clearly labeled as such) since timeliness is of value. Biologic survey data are cataloged without aggregation, to allow flexibility in how researchers apply statistical models. But these data *are* subjected to rigorous QC. Some datasets are cataloged as the result of aggregation, with their corresponding raw data being Type II. Physical oceanographic data are only useful after extensive filtering in Matlab; those data are only posted as a derived product.

Do you communicate with your PIs regarding prioritization of data?
MCR entered a new phase this year, with our sixth year of data collection now allowing some synthesis of time-series data. The PIs advised which datasets were eagerly awaited by investigators to complete their manuscripts. The PIs requested that Quality Control of the data be given a high priority, which lead to a significant effort to re-process data based on the results of QC scripts. They also requested web-based data search and cross-reference be given priority. This is the first year non-core and thesis datasets were added to our catalog. The PIs prioritized datasets as (1) core time series, (2) non-core time series, (3) thesis data, (4) non-core, one-time studies.

Niwot Ridge SiteBytes 2010

This is the first full year that Hope Humphries has been the information manager for Niwot. In addition, she is responsible for collecting yearly plant survey data at Niwot (such involvement of an information manager with data collection and use may be of interest to IMC). Finding time for IM activities other than data processing, such development of the IM system, has proved to be challenging this year, especially given the time required for conducting fieldwork. At this point, it is unknown how Niwot's system will need to be adapted to accommodate network efforts.

Hardware. We replaced the old tape-based backup system for our UNIX server with a ReadyNAS NV+ desktop server, which is housed offsite and is used for automated incremental and full backups. An additional ReadyNAS NV+ server now acts as our FTP server, delivering all spatial data and associated metadata.

Website. Our online climate data display system now includes data from 3 additional stations: Albion townsite, Green Lake 4, and Arikaree, in addition to the D1, Saddle, and C1 stations. Niwot

now hosts two outreach websites. A website has been developed to accompany the “My Water Comes from the Mountains” children’s book to display information about the book and, importantly, artwork contributed by each individual student from various schools in the Rocky Mountain region (http://culter.colorado.edu/MyWater [143]). A website has been initiated for the Rocky Mountain Lake Algae outreach effort that contains images, taxonomic information, and data for hundreds of algae taxa, mostly from lakes in the NWT LTER site, but also from nearby Rocky Mountain National Park (http://culter.colorado.edu/lake-algae/ [144]). In addition, the Alpine Microbial Observatory (AMO) website complements NWT LTER microbial research (http://amo.colorado.edu [145]) and when the AMO grant (DEB 0455606) expires in 2010, NWT will host and integrate the AMO website and database. We plan in the coming months to overhaul our website to incorporate better presentation of mapped data including clickable maps.

Communications. Currently 10 meteorological sites are serviced by frequency-hopping radios. A wireless connection from MRS to Boulder, replacing the previous T1 connectivity, has been completed. This system consists of a fiber-optic line from MRS to the TundraLab, and a wireless connection from the TundraLab to the world with a bandwidth of 3mbs.

Spatial Data. Spatial data sets are searchable and can be downloaded from an FTP site accessed via our web page (http://culter.colorado.edu/exec/Database/gis_layer_query.cgi [146]). A major accomplishment was the release, in the fall of 2009, of a set of high-resolution orthophoto mosaics and accompanying DEMs and accessory map layers for NWT, including twelve “timeslices” encompassing the past seven decades. These map layers are available from our FTP site and are expected to be very useful for change detection. A new fine-scale raster land cover map is currently under development by Hope, and will be used to determine the current spatial distribution of vegetation, document phenological changes over time for the National Phenology Network, examine treeline and shrub movement, and conduct biogeochemical modeling. Hope, in collaboration with Patrick Bourgeron, is also in the process of developing classified land cover maps for time slices spanning 1938 to the 2000s as part of LTER’s Maps and Locals (MALS) project. New LiDar covering Niwot was flown this year for the Boulder Critical Zone Observatory project, but is not yet available to Niwot.

Do you deliver derived or raw data?
Both. For example, raw climate data are displayed graphically on our website. Downloadable climate data sets on our website are subjected to varying levels of QAQC, including, for selected data sets, procedures to infill missing values.

Do you give a lot of attention to a few data sets or limited attention to a lot?
Core ongoing time-series data tend to get highest priority, with less attention given to one-time data sets, unless there are specific requests. Climate and snow data sets tend to command a lot of IM attention, because demand for them is high and they tend to be time-consuming to process.

Do you communicate with your PIs regarding prioritization of data?
Yes, but we don’t have a formalized process. PIs frequently request raising the priority level of particular data sets of interest to them.

North Temperate Lakes SiteBytes 2010

The information management team at NTL-LTER has seen major changes during this past year. Barbara Benson retired from her position and was replaced by Corinna Gries, formerly information manager at CAP-LTER. David Balsiger will be retiring at the end of July and his replacement, Aaron Stephenson, was hired recently. Stephenson brings extensive GIS and general IT project development expertise to the team. In addition a full time programmer, Preston Alexander was recently hired for one year.

Some time during this past year was devoted to the hand-over of the information management system, a system that has been accumulating data for over 30 years. Gries was introduced to and has assumed responsibility for the NTL sample collection at the Zoology museum in addition to the day to day data management.

After many years of successful additions, changes and improvements it was decided that the NTL-LTER website needs a major overhaul in technology as well as look and feel. Particularly better responsiveness to news and exciting findings from NTL and LTER in general is desired. Therefore, NTL is joining a growing group of LTER sites that are collaborating in using the content management system DRUPAL for website development. Content models have been developed in DRUPAL by LNO staff for the most common aspects of LTER site data (personnel, EML metadata, research projects, research sites etc.) and data have been transferred from the original data model used at NTL and stored in Oracle to the MySQL backend for DRUPAL. Standard DRUPAL modules will be used to handle publications and images, both are already deployed in other applications locally. A coordinated supplement proposal among six LTER sites to develop DRUPAL modules which can be used by all sites involved and will provide LTER specific functionality was funded. Development will take place at NTL and the Marine Biology Lab at Woods Hole where experienced DRUPAL developers have been involved with the Encyclopedia of Life DRUPAL implementation.

The first application built in DRUPAL captures information on field research activities at the Trout Lake field station, allowing the management to better coordinate research activities and subsequent cleanup. This application will be implemented for NTL LTER project as well as the Microbial Observatory project for their project management needs and is currently being field tested.
The second major project started this year is to update workflows for processing long term data sets. Most of these data are submitted to the NTL IM lab periodically as standardized spreadsheets and need to undergo Q/C routines, some calculations and rearrangements before uploading into the central database. Currently this is accomplished via Perl, Fortran, PHP sripts/programs, the Oracle specific SQL language PL/SQL, and Oracle triggers. The Kepler workflow system has been chosen to accomplish this task and several new workflow have been developed already.

In collaboration with Calit2 (formerly San Diego Super Computing) at UCSD we are currently implementing DataTurbine for harvesting streaming sensor data from buoys in NTL research lakes. Kepler is being employed to process the data streams directly from DataTurbine, apply basic Q/C procedures and upload data into final storage. To fully take advantage of DataTurbine several upgrades in the buoy system are necessary, which will be implemented over the next year. For final storage we are currently experimenting with CUAHSI's Observation Data Model. However, this is a longer term project because the basic NTL data query application is laid out for matrix type tables, not the attribute – value approach taken by the ODM, and new local data access application will have to be developed. The advantage of using the standard ODM lies in instant participation of the data in the large CUAHSI Hydrologic Information System (HIS), the use of their applications, the possibility of developing more stable applications locally, and extensive collaborations.

All steps described above are taken with the goal in mind to outsource NTL's web and database servers to a commercial hosting company. For being able to rent server space rather than maintain a server locally it is necessary to move the data from Oracle to an open source database like MySQL. MySQL has been chosen for the table data because of the large code base available for DRUPAL and MySQL. However, GIS data storage will be accomplished in the upcoming year in PostgeSQL/PostGIS with display applications built on GeoServer.

Palmer Station SiteBytes 2010

Palmer Station LTER Information Management Site Byte - September 2010
Karen Baker

Information Management Committee Inquiry

Introduction
Site Bytes are intended as a general update on what has happened at your site during the last year. Please highlight new developments, ideas, and issues. Site bytes help us to stay informed about what is going on at other sites. Please also be forward thinking and include your thoughts on the following subjects with regards to the developing Network Information System.
We will discuss these questions further at the annual meeting:
Given limited resources, how do you prioritize data contributions to the NIS?
1) Do you give a lot of attention to a few datasets or limited attention to a lot?
2) Do you deliver derived or raw data?
3) Do you communicate with your PIs regarding prioritization of data?

PAL Responses
1) Do you give a lot of attention to a few datasets or limited attention to a lot?
Between these two extremes, we strike a balance. Our strategy has been to create the technical capability to rapidly create and update datasets together with a management interface for submitting to community collections. Last year we focused on improving our information system in ways that put us in a position to work with the NIS development team. We also improved our metadata approach while working with several local groups with complex datasets who had not yet submitted data. In addition, site work focused on streamlining of weather data handling so as to have monthly updates for ClimDB as well as summary plots posted in DataZoo. Quality control has been left to the data submitter to date but we plan to develop some capabilities with this in the next year as part of the information management services and hence to begin more routine submissions of more assuredly high quality data to NIS.

2) Do you deliver derived or raw data?
The raw-derived question seems misleading when stated as a binary. We see multiple stages as part of a processing continuum that varies by dataset depending upon sampling, instrumentation, calibration, and analysis and so we consider the decision as to what dataset stage to deliver as dataset dependent. For many researchers, the questions of what to submit is still under investigation as they may submit abundances from counts obtained via microscope analysis and then submit carbon converted counts as they work with the data. We find that over time and with experience, our understanding of the data and the data audiences as well as the shared projects such as EcoTrends influences our data delivery.

3) Do you communicate with your PIs regarding prioritization of data?
Yes, this is the strength of an LTER site to have IM guided in practice by local data needs and by research scientists who collect, analyze, submit, and use the data. Data inventories are created and discussed periodically. The significant technical capacity improvement this year coupled with our planned focus on metadata and data quality reviews next year put us in a position to include submission of data to Metacat as part of a more routine data process next year.

PAL Site Byte

1) Local Development
Focus this year was on technical development of DataZoo, a site-level information system now used by four oceanographic sites (two LTER sites (PAL, CCE) and two CalCOFI sites (SIO, SWFSC)), in an effort to better meet local needs while developing web services as a foundational element for upcoming site-network development. Redesign of the study-project schema for DataZoo added flexibility in response to metadata needs identified in practice and by the IMC working group effort with project collections. Significant updates were made to the DataZoo management interface that allows creation and editing of organizations, studies, datasets, and their related metadata. New approaches to web delivery through development of middleware and standardized, asynchronous access to databases through a web service layer successfully addressed issues relating to the

increasing size and diversity of data. Expanding DataZoo to a three-component system accommodated the development of FileFinder, an interface to very-large collections of hierarchically structured datasets. The prototype of FileFinder was populated with oceanographic CTD data.

Additional work relating to DataZoo included adding the support for taxonomic identifiers to the attribute qualifier system, use of the Integrated Taxonomic Information System (ITIS) as a taxonomic authority that prompted conversion of NODC taxonomic codes, and redesign of query selection so users can select studies within a designated time interval. Plot performance was improved by replacement of the plot library and options added for contour and bubble plots. Handling of external datasets like weather was streamlined to facilitate monthly updates for ClimDB as well as summary plots posted in DataZoo. Improvements were made to dataset documentation as datasets were submitted for publication in EcoTrends. An Information Management media gallery was created in order to preserve posters as a historical record of development. Finally, backups are now done via the Computational Infrastructure Service at IOD in conjunction with the San Diego Supercomputer. In terms of physical infrastructure, a new server with virtual machine capability was purchased (vSurf).

2) Partner Activities and Leadership
Our team (Mason Kortz, Lynn Yarmey, James Conners) is leading a network-level LTER unit registry and unit best practices effort that received support from a LTER post All Scientist Meeting proposal award. Two cross-site visits (KBS, JRN) and a visit to the Network Office were carried out as part of this project. This work has informed redesign of DataZoo architecture to a web-services orientation. The new unit registry together with its management interface replaces a static, isolated unit dictionary thus enacting one type of data comparability as well as demonstrating a new type of site-network model that enables site participation in NIS development through web services. Our team (Mason Kortz, James Conners) is also leading a joint IMC-LNO web services working group and serving as a NIS tiger team member (James Conners). Karen Baker is co-leading the IMC Governance Working Group (GWG). The GWG was created in order to facilitate the conduct of LTER business in an arena of growing complexity and responsibility and to support the development of Terms of Reference or by-laws for committees and working groups.

3) Professional Development and Publication
Professional development of information managers was provided through exposure to site-level and network-level development activities with the Unit Registry project. In addition, a weekly summer informatics reading group stimulated discussion and joint learning. A sociology graduate student studying the Design Studio stimulated reflection on our work as well as on our work arena. Lynn Yarmey received a fellowship to pursue two-year masters in Library and Information Science with a concentration in Data Curation through the University of Illinois, one of the national iSchools.

At the LTER Information Management Committee (IMC) annual meeting CCE/PAL members contributed as co-chairs of the IM Governance Working Group and Unit Working Groups. Lynn Yarmey reported on a Unit Dictionary Best Practices Guideline. Karen Baker was elected a member of the Network Information System Advisory Committee (NISAC). Mason Kortz and James Conners prototyped the LTER Unit Registry and contributed to formation of a new Web Services Working Group. At the LTER All Scientist Meeting, contributions included a series of posters, participation on a panel about Network level efforts, and a workshop focusing on curriculum for information managers.

Collaborating on two chapters about the EcoTrends Project contributed to description and understanding of network level information management efforts as well as development of a set of recommendations for future work (Laney et al, in press). Another book chapter explores the topic of digital infrastructure growth (GCBowker, KSBaker, FMillerand, and DRibes, 2010. Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In Int. Handbook of Internet Research). Collaboration with Science Studies partners led to publications in peer-reviewed journals. An interdisciplinary investigation into work with data over time spanning decades

provides insight into ‘Big Data’ efforts and the LTER network model (Aronova, Baker, and Oreskes, 2010. From the International Geophysical Year to the International Biological Program: Big Science and Big Data in Biology, 1957-present. Historical Studies in the Natural Sciences 40(2):183-224). A study of design provides insight into the complexity of the development processes at an LTER site (Millerand and Baker, 2010. Who are the users? From a single users to a web of users in the design of a working standard. Information Systems Journal 20:137-161).

4) In Summary
We plan to focus next year on metadata population, quality control, and data delivery as well as on documentation of the DataZoo system. Technical work with the unit registry and web services will continue at a network level. Technical issues that remain to be addressed include integration of the personnel module with DataZoo, on geolocation based on the earlier gazetteer effort, and on management of dataset use logging. Our website content will be updated but migration to drupal/wordpress/etc is not yet scheduled. The application module MediaZoo was redesigned as an API. The move from a photo gallery to a plot gallery and eventually to the media gallery module has reached an appropriate level of abstraction that it is time to make uniform the five instances of installation currently in use.

Despite significant experience with information management practices and issues at sites, innovation at sites is limited by lack of resources and incentives. The 2010 information management supplement stabilizes our work on DataZoo and collaboration on the LTER NIS this year but highlights the need either to continue such support or to collaborate locally with additional projects in order to maintain a team with technical expertise at the site-level over the long-term.

Plum Island Ecosystem SiteBytes 2010

Hap Garritt (PIE IM) attended the LTER Post ASM 2009 MIRADA (Microbial Inventory Research Across Diverse Aquatic) LTERs workshop in Woods Hole, MA on March 9-10, 2010. After the workshop Hap helped the MIRADA investigators with the various cross site (13 LTER sites) environmental data to create a standardized nomenclature for variable descriptions of variable names and then applied the LTER Unit Dictionary best practices for standardizing the corresponding variable units.

PIE hosted the May 2010 LTER Science Council Meeting at the Marriott in Peabody, MA including an afternoon field trip to various PIE field sites throughout PIE watersheds and estuaries.

PIE in collaboration with ARC is in the process of developing a content management system utilizing the Drupal platform. Utilizing Inigo San Gil’s experience/expertise with SEV and LUQ, we will migrate our existing datasets into Drupal, which will be hosted through the Marine Biological Laboratory Woods Hole Oceanographic Institution (MBLWHOI) Library. Data ingestion tools will check the quality of metadata for conformance to LTER network best practices. Once in Drupal a common EML document generator will insure the quality of the metadata content and improve metadata to data linkage. Drupal as a content management system will allow research scientists the additional ability to edit web content as well as enter new metadata through a user-friendly web interface.

PIE is collaborating with five other LTER sites (ARC, NTL, SEV, LUQ, VCR) toward further development of LTER content in the Drupal platform. The intent is to leverage development work across sites to provide rich relational information content. The collaboration is utilizing programming expertise from the MBL/WHOI Library to convert Drupal content to EML and to begin incorporating web services (for example Unit Dictionary, Integrated Taxonomic Information Service).

PIE is in the process (Fall 2010) of hiring personnel to manage the installation, maintenance and data management of an eddy flux tower for deployment in the marsh areas of Plum Island Sound.

Given limited resources, how does PIE prioritize data contributions to the NIS?
* Do you give a lot of attention to a few datasets or limited attention to a lot?
Core data sets get most of the attention as most of them are under the direct control of the PIE IM. Datasets are primarily updated on an annual basis once the main field season is complete, analyses have been conducted and data QA/QCed by the respective researcher.

* Do you deliver derived or raw data?
Both

* Do you communicate with your PIs regarding prioritization of data?
Yes, as stated above annual updates are the norm as that time frame coincides with field season and laboratory analyses schedules.

Santa Barbara Coastal SiteBytes 2010

Margaret O'Brien, September 2010

ACTIVITIES OVER THE PAST YEAR
SBC LTER's IM activities have continued to focus on metadata components for multiple uses (i.e., datasets, bibliography, website). Most work in the past year has been concentrated in three areas: sampling location descriptions, scripted EML creation from centralized data sources, and development of a measurement dictionary using the OBOE ontology. Additionally, SBC is in the midst of migrating its IT system from a local lab-supported system (the PISCO project) to a more centralized model at the UCSB Marine Science Institute (MSI).

IT system migration: As part of our migration from the PISCO IT system to MSI, we began using the LTER Network Metacat catalog as our primary repository. We have developed our own interfaces for displaying and querying EML datasets. We present datasets in a modular "tabbed" view using Perl CGI, Javascript and XSL transformation stylesheets. The stylesheets have been deliberately generalized and decoupled from scripting to enable their further development as a community resource. See the result at http://sbc.lternet.edu/data/dataCollectionsPortal.html [147]. The system was also easily adapted for displaying draft datasets (e.g., from a fileserver) to illustrate dataset views and content to scientists during the dataset creation process.

Sampling locations DB: SBC LTER has collected data at over 300 geographic locations. We now have approximately 75% of these entered into standardized forms which will allow us to use the information in multiple ways, e.g., reports for field teams, inclusion into EML datasets for publication, map display on the web site and contributions to the network databases and applications (e.g. siteDB and LTERMapS). Our schema is compatible with EML and GML.

Measurement Ontology: We have begun development of a dictionary of measurements as part of the OBOE ontology and a related NSF project, Semantic Tools for Data Management (Semtools, DBI-0743429). This project is providing software for increased data access, discovery, and integration using semantically annotated metadata. By providing datasets for the development process, SBC takes advantage of knowledge-modeling experts to build a detailed dictionary of its own ecological and environmental measurements. In addition to their use in the ontology, these measurement descriptions and terms can be exported for use in SBC EML metadata, and also will contribute to the broader measurement standardization efforts in the LTER network.

Collaborative development of a metadata model: We are taking advantage of the opportunity to collaborate on development of a common relational data model with three other sites (MCR, CWT, GCE). We have continued to develop our scripts for EML creation, and their further development will make use of the relational metadata database as it develops, using the concept of a "switchboard". Along with MCR, we are beginning the search for a shared IM assistant.

DISCUSSION QUESTIONS
Given limited resources, how do you prioritize data contributions to the NIS? We group data generally into priority types. Highest priority for publication are data from long-term time-series observations supported by the NSF core funding. There are 20-30 datasets which have been ongoing since SBC's start in 2000, with 2 or 3 added each year. Second priority are data from discrete core-funded experiments, such as the body of cruises carried out in 2001-2006. Third comes student data, although if supported by core funding these would be included in priority-2. Lastly comes data from ancillary funding, inherited ("legacy") data, or other data of interest. Occasionally these will be boosted up on the priority list if they are of timely interest or highly desirable. Given the volume and diversity of data, and the paucity of good tools or resources, very little from the last 2 categories (student data, ancillary or legacy data) has been published.

Do you give a lot of attention to a few datasets or limited attention to a lot? Usually, when a dataset is first identified, it gets a lot of attention -- at least a week or two per data table as data and metadata are finalized and/or cleaned up. Ideally, datasets which are ongoing (about 40% of our total) take less and less time to update in later years. The IM tries to engage the scientific staff as much as possible in the preparation of data for publication.

Do you deliver derived or raw data? Generally, we publish the data product that the scientist would have chosen to share with his/her colleagues. Usually this means derived or aggregated in some manner applicable to general SBC research.

Do you communicate with your PIs regarding prioritization of data? Our executive committee decided on the publishing priorities, but the implementation is left up to the IM. The project's lead PI and project manager stay in frequent contact with the site IM, and are aware of progress and general effort required to get data published.

Sevilleta SiteBytes 2010

Activities:
Sevilleta is adopting the Drupal system developed by Inigo San Gil (USGS) and Marshall White (LNO) for managing EML. Content from the old SEV web page (implemented in PostNuke CMS) has been migrated to Drupal, and Inigo developed a script that imported all SEV EML documents into the MySQL database backend of Drupal. Drupal allows tagging of content by keywords, and IM Kristin Vanderbilt extracted a subset of the LTER Controlled Vocabulary to use on our website. Extensive QA of all imported metadata will now be done and all web content will be tagged to provide the user with a “no dead-ends” web experience.

Sevilleta continues to be challenged by management of high frequency, high volume sensor data. We are planning to adopt Vista Data Vision software for rapid visualization of the data, but a system for QA/QC of the data needs to be found.

Mike Friggens is investigating options for updating the SEV’s spatial data management system.

Kristin has been actively involved in the network information management scene. She is a member of NISAC and also chaired the ILTER IM Committee until September 2010. She is co-chair of the

US LTER International Committee. She also co-organized and attended the “Second International Analytical Workshop on Dynamic Plot Analysis” in Malaysia in August 2010.

Discussion Questions:
Do we give a lot of attention to a few datasets or limited attention to a lot? A few core data sets get most of the IM attention. There aren’t typically lots of data being submitted beyond our core datasets collected by the field crew, so the core gets priority. Sensor data are becoming core data sets, and those are getting the lion’s share of attention.

Do you deliver derived or raw data? Both. For instance, NPP data are calculated from volume measurements made in the field and then processed through regressions to calculate biomass. We offer both the volume measurements and the final calculated NPP.

Do you communicate with your PIs regarding prioritization of data? PIs do not have time to discuss information management. I find out from other SEV staff what data is being collected, and then do the best I can with that info.

Short Grass Steppe SiteBytes 2010

Updates to SGS-LTER Website and Back-end Database
Data Enhancements: Over the past year we continued to integrate our years of data, standardize attribute codes, keywords and units with SGS and LTER Network standards where available, and obtained and documented more detailed metadata. A few of the core datasets were integrated with historical records dating back to the International Biome Project from the early 1970s. In all we migrated fifty-six more datasets to our database and were able to harvest most of them to the metadata with no errors. In addition, data aggregation and summarization steps were taken by IM staff to facilitate data analysis. Working on these value-added datasets afforded opportunities for staff and scientists to work together, as well as scientists to collaborate on reporting trends and testing ecosystem responses to different drivers.
Web Site Enhancements and GIS Data Access: On the front end of the website, we added an improved search capability and new mapping tools. The maps page now has static images of various physical features and treatments on our study area. A capability to download all GIS layers was added as was an interactive mapping tool that allows viewing all these layers via an internet browser. More detailed coordinates of the field study designs are currently being collected by marking fence lines, as well as locations of transects and plots with GPS units in the field. This will be available for researchers and for website data downloads and serve as another means to integrate tabular and spatial data within the SGS information management system.

Virginia Coast Reserve SiteBytes 2010

Below are some responses to questions posed for this site byte, followed by a brief precis on activities in the past year.

Given limited resources, how do you prioritize data contributions to the NIS?

At the Virginia Coast Reserve LTER (VCR/LTER) we use several criteria to determine the priority and resources applied to a given data set:
1. Importance of the data to achieving research goals
2. Generality – is it something many users will benefit from?
3. Uniqueness – is this data available from multiple sources, or only from the VCR/LTER?
4. Degree of difficulty – what resources are required?

The priority assigned to a given dataset balances all these factors. An example of some of our most important data, are measurements from Surface Elevation Tables that assess accretion and erosion of sediments on the marsh soil surface. These data are critical for assessing whether salt marshes will survive in an environment of rapid sea level rise, and are widely used by that subset of our researchers who focus on salt-marshes. The data are unique in the sense that no other group is making similar measurements in our system (although similar measurements are made in other coastal locations). Finally, the data is relatively easy to manage – with updates every 6 months. This gives this data a huge benefit:cost ratio and makes this a high-priority dataset. In contrast, satellite images from LandSat are an important tool for examining the VCR/LTER landscape and they are used by several of the VCR/LTER investigators. However, they are provided by NASA on its own web sites and would be expensive and difficult to manage due to the high data volumes. So we give them only a modest priority – maintaining copies of a few critical images, but encouraging and aiding investigators in getting the data from existing government web sites.

• Do you give a lot of attention to a few datasets or limited attention to a lot?

The answer to this question is “yes.” We deal with many datasets, covering a wide array of activities from GIS to detailed process studies. For each we attempt to have the data available with high quality metadata. However, there are some datasets that, because of their importance and wide use, we put additional resources into quality assurance and control and making the data easy-to-use and add features such as online graphs etc.

• Do you deliver derived or raw data?

Primarily raw data. However, for some high priority datasets there are some derived products.

• Do you communicate with your PIs regarding prioritization of data?

Yes, we periodically review the status of priority datasets. That review includes discussion of both existing and new datasets.

The past year: At the Virginia Coast Reserve LTER we’ve been working on some housekeeping and learning more about how to use advanced workflow tools, such as Kepler and R. The “housekeeping” activities consist of continuing to port applications and capabilities developed on our old Sun/Unix-based web server to our new Linux/virtual machine-based web server. The new Linux server now supports the primary VCR/LTER web site using the Drupal Content Management System, along with the needed MySQL support, Web Map Server, photo gallery and bibliography, but some functionality, such as the mini-SQL-based personnel database, biodiversity and metadatabase applications, along with EML generation and a host of special-purpose applications (e.g., data file uploads, biodiversity database) still are being provided by the Sun server. On a priority basis these remaining functions are either being ported, or replaced by more modern and general approaches. We are trying to balance the desire to center applications around a few core technologies (e.g., LAMP), with the need to continue to support some legacy applications (e.g. mini-SQL) because they still do the job well and would be expensive or difficult to replace. For some of the legacy applications, rather than recreate them using more modern tools, we’d prefer to concentrate on developing solutions that could be applied across multiple LTER sites.

During the summer the VCR/LTER IM spent an extended period of time in the East Asia Pacific region working with partners in Taiwan, Nanchang China and Malaysia. A significant part of this time was spent working on developing KEPLER workflows for data analysis, including extensive use of the “R” statistical package. The R software is gaining increasing popularity with ecologists because it is easily extended to provide custom-made functions for a wide array of applications. Moreover, because it is open source, there are no barriers to sharing it with partners, worldwide (something that can’t be said of Matlab, SAS and SPSS). Learning R still remains a challenge because its antecedents do not overlap broadly with older statistical packages (other than S) nor with many of

the Unix/Linux-based programming environments, so there is little transference of knowledge, but it is clear that it has tremendous capabilities for analysis and visualization. Although the KEPLER workflow system can be used as a “wrapper” for “R” programs, it is at its most useful when it brings together a number of different applications into a single workflow. For example, using a web service to feed data to an R application that then displays data using ImageJ. R programs are now used for creating near real-time graphs of groundwater levels from a network of wells on Hog Island, and we anticipate increased use in the future.

In the coming year, we look forward with collaborating with the Web Services Working Group and on improving the nascent Controlled Vocabulary for use across all LTER sites, as well as continuing to explore how content management systems, such as Drupal, might be applied to a wider array of functions than simply providing an attractive and functional web site for the VCR/LTER.

2008

Site Byte content can include any information about the site and typically reflects activity related to the information system at your site during the past year.

Please also include in your site byte:

1. Identify at least one major Cyber-Infrastructure (CI) need at your site (e.g. software, hardware, communications, staffing) and describe:
a) What is the the science issue that drives this CI/IM need
b) How does the CI need constrain site activities
c) How does the CI need impact your site's participation in network activities

2) If applicable, list any CI development projects at your site that could potentially be used by other sites or shared and collaboratively developed into network-level CI implementations as part of a general Service Oriented Architecture (SOA) framework (i.e. projects with a wide range of use cases that could support decadal plan CI initiatives and cross-site science projects).

Please submit site bytes to the IM web site, http://intranet.lternet.edu/im/siteprofiles/SiteBytes/2008 [148]. Select your LTER site from the list, select Edit (next to the View button), then paste your site byte in the Body portion of the page. You will need to sign in under your LTER User name and password to make any edits.

Andrews Forest SiteBytes 2008

LTER Site Byte: Andrews Experimental Forest
Suzanne Remillard, Don Henshaw, Theresa Valentine, Fred Bierlmaier
September 4, 2008

(1) Site activities over the past year:

The primary activity of the Andrews LTER Information Management (IM) Team for the first half-year revolved around IM development in association with the LTER renewal proposal process. The IM Team participated in the proposal writing process and prepared the IM section for the proposal that included a complete list of online databases and a history of data downloads over the past decade. Key publications and data sets were also listed in conjunction with long-term core measurements. This LTER6 proposal was successful.

Andrews personnel conducted a cyberinfrastructure (CI) planning workshop with funding from an NSF FSML planning grant. The workshop focused on development of the “cyber forest” concept including bandwidth requirements from the site to OSU as well as throughout the forest, plans for deployment of sensor networks, and consideration of streaming data needs including QA/QC requirements. The workshop featured participation from invited guests Susan Stafford and Judy Cushing and John Porter via videoconference. Planning has resulted in a successful FSML proposal which will fund the initial phase implementation of wireless connectivity throughout the forest. Additionally, a new QA/QC process for our climate data is designed and under development that will allow more efficient data processing and assignment of preliminary qualifiers in near real-time.

Site activities include development of an interactive web form for investigators to request approval of new projects and a database to track ongoing research projects. The Andrews will participate in the LNO-funded project proposal, which is highly related to this work. Another web application under development is a web-based administrative interface to allow PIs to enter and edit database metadata.

The migration of ClimDB/HydroDB to the LTER Network Office is planned and collaborations with James Brunt and Mark Servilla (LNO) have been initiated to accomplish this. The goal is to replicate the current system at the LNO within 6 months.

The Andrews Experimental Forest will be obtaining LiDAR data. The projected flight was to be this past winter during leaf-off season, but due to greater than average snow accumulation, the flight didn’t occur until August. The data should be available within the next couple of months.

The Andrews LTER web pages were victims of the recent SQL injection attacks hitting sites running Cold Fusion. Vulnerabilities in the Cold Fusion code allowed malicious attackers access to our database through SQL “where clauses”. Revising and tightening our code has given us some resistance to these attacks, and we are adding filtering and error trapping capabilities to our web site. Recently, however, our asp code has also been targeted and so this fight is ongoing.

(2) Identify at least one major Cyber-Infrastructure (CI) need at your site (e.g. software, hardware, communications, staffing) and describe:

Staffing issues and development of a wireless communications network within the Andrews are predominant needs and currently command the most discussion and attention at the site. Another key CI issue is quality control (QA/QC) of streaming sensor data and improvement in the QA/QC process to more quickly screen and post this data.

a) What is the science issue that drives this CI/IM need?
Near real-time access to hydrometeorological and ecophysiological data streams has become essential for many atmospheric studies, and snow and subsurface hydrology studies, biodiversity studies, and modeling activities. While some provisional data is continually posted on the web, this data is not checked for potential problems, and there is a significant time lag before final corrected data sets with complete metadata are posted online.

b) How does the CI need constrain site activities?
Site activities are constrained by lack of timely access to some data streams. This is partly a staffing issue, but more efficient QA/QC processing and metadata development would alleviate much of this problem.

c) How does the CI need impact your site's participation in network activities?
Participation in network-level databases such as ClimDB is delayed, and the hope is to post data to ClimDB/HydroDB in near real-time.

(3) If applicable, list any CI development projects at your site that could potentially be used by other sites or shared and collaboratively developed into network-level CI implementations as part of a general Service Oriented Architecture (SOA) framework (i.e. projects with a wide range of use cases that could support decadal plan CI initiatives and cross-site science projects).

Certainly, we would be interested in any project related to QA/QC of streaming data and assignment of qualifying flags.

Another potential SOA-like project (we are not currently working on this) would be the use of EML to do some generic QA/QC checks on data sets. Now that the EML-parser is in place, it would be nice to build on this to allow QA checking such as metadata completeness reports or checking data against prescribed min-max values, or other checks (entity integrity, domain integrity relational integrity, etc.). As every site has issues with data quality, this would be a tool that could be easily shared across sites, would enforce best practices for EML, and would provide a key incentive for developing EML. This would help establish a more meaningful relationship between the metadata and actual data.

Arctic LTER SiteBytes 2008

2008

Baltimore Ecosystem SiteBytes 2008

2008

Bonanza Creek SiteBytes 2008

Bonanza Creek LTER

Site Update (Year in Review):

The last year has seen a lot of activity as we worked to integrate a new data manager into our management team. Jason Downing was hired in September of 2007 as our site information manager and has been working hard to learn the current system and protocols as well as working to plan and implement improvement projects for the data management system.

Over the past months we have deployed a new system for data submission into the BNZ database. The old system relied upon personal or e-mail communications with the data manager to receive information about metadata requirements and progress through the submissions process. It was easy for submissions to become misplaced or abandoned and fell apart with personnel changes. This process first involved creating a new metadata submission spreadsheet with documentation and examples. Next we developed an on-line submission interface that links with our database to create relational records of each upload action for future processing and tracking. Training was also developed and provided to researchers, staff, and students on the basics of metadata, the BNZ metadata form, and how to use our data submission system. Feedback from this training was positive and it will be an annual workshop offered to our personnel.

In an attempt to become more familiar with each of the senior investigators, their research, and their data files in our system, the data manager and a supervising Co-PI for the site met individually with each senior researcher. These meetings were a forum to discuss status of their data in the

database, provide information and resources to one another, and to foster improved working relationships between the scientists and the data management staff. In follow-up to these meetings, the data manager has begun to have field/lab visits with each of the scientists in their home 'element'. These are proving to be very beneficial in increased development of the working relationships and interactions among staff.

The big hardware activity and upgrade was to replace our outdated web application server with a new and multifunctional server providing virtual platforms for various operations. This new server currently provides a virtual web site server (Linux), and an ArcGIS server (Windows). The virtual platform allows for easy development and upgrading for services. The next addition will be a dedicated SAS interface for quality control operations for implementation with our streaming climate data input.

Currently our database holds data, metadata, and generated EML (Levels 3- 5) for 245 distinct data files. The EML is generated with a PERL script developed by Inigo SanGil from the LNO to efficiently produce versioned metadata for harvest. Through the increased IM outreach activities we are working to boost the availability of electronic versions of published materials and establish links between the listed publications and actual data files currently in our system.

Site Plans (Year to Come):

Our current website is what was salvaged from the previous server but has many legacy issues that make it ripe for an upgrade and we would like to have this be part of the server upgrade process. The virtual platform will allow for simultaneous development and production on the same server with limited changeover complications. Our big step will be to end our use of Coldfusion as the web-database linkage in favor of a more simple and universal format like PHP. The new web interface will provide a visual and structure facelift to our site as well as hasten the development of more useful interactive interfaces for generation of files and graphics from the database.

The ArcGIS server is only in the beginning stages of development but will replace our current ArcIMS web service that will no longer be supported. The spatial data server will continually become a more useful and critical component in our site and data management activities.

There are also additional sites that now have radio communications and can be added into the streaming weather data system. With this we want to implement a system for filtering all the streaming data through a SAS system that will run basic filtering and will produce graphics for the technician staff to view and inspect for sensor equipment issues.

CI Need at BNZ:

The major CI needs for BNZ are software and staffing to implement real-time QA/QC filtering and flagging of streaming data as it enters our database. As we have transitioned to more loggers sending data straight into our database it has become possible to loose connection with the quality of the incoming data or be unaware of serious sensor issues. Also, the volume of data and wide range of possible issues make playing catch-up after the fact extremely difficult. These core datasets are supposed to be the foundation of data that supports our other scientific research but as we try to use the data we are finding significant quality issues that require significant amounts of attention before the data can be utilized. These issues cause an undue burden on field and data staff as they try to make this data available and in a usable state. This deficiency limits our ability to contribute some basic data to our site scientists as well as to network level collaborations.

California Current Ecosystem SiteBytes 2008

LTER Site Byte 2008: California Current Ecosystem
Karen Baker

In addition to advancements in design and population of our cross-site information management system DataZoo, there were three major accomplishments this year: 1) the development of CCE partnerships in general and links with CalCOFI in particular, 2) the publication of an information management paper in the oceanography journal Deep-Sea Research (Baker and Chandler, 2008, 'Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics'), and 3) initiation of a division-level computer infrastructure recharge facility as an organizationally situated support for systems administration.

DataZoo has reached 'flagship' status as evidenced by a change in inquiry from 'Where is dataset X' to 'Why isn't dataset X in DataZoo?'. Local design efforts focused in particular on improving the user interface to DataZoo, creating an online help system, and shifting code practices to use of libraries and an object oriented approach. Design and flow of the system as a coherent whole was revisited in order to clarify the design as well as to add consistency to users' interactions with the data and metadata models. The management interface was improved in terms of useability. We continue to address design decisions such as whether code sets are associated at the attribute or the column level (decision: move to column level), how to deal with metadata that changes over the course of a dataset (decision: handle at section level), and what keyword classification approaches to take.

The PeopleZoo personnel application module has been redesigned as an API and is being used to track project and cruise participation. Three additional information system elements stabilized: 1) definition of relationships between units, attributes, and qualifiers as an interdependent set of dictionaries, 2) development of web services in a locally developed hybrid approach using both SOAP and REST, and 3) launch of a multi-element cooperative dataspace composed of a set of related applications that capture and make accessible multi-faceted datasets.

In support of field sampling, we deployed the eventlogger and created event glossaries. Support for the outreach component included development of an education web portal and investigation of communication options resulting in redevelopment of the picture-of-the-day activity as a blog to support the new teacher-at-sea activity. A partnership with researchers conducting bird and marine mammal surveys resulted in a series of datasets available in DataZoo and contributed to initial discussions regarding a West Coast bird consortium and data policy development. LTER network collaboration included returning to design of a community-developed, network-hosted, service-oriented unit dictionary database. The Ocean Informatics team worked closely with the LTER Information Management Committee Dictionary Task Force to define and coordinate continuing development of the unit dictionary. In addition, we worked to capture site datasets submitted to EcoTrends for submission to DataZoo. After a third unsuccessful proposal to obtain NSF funding for the Ocean Informatics conceptual approach to local infrastructure building and to distributed networking dictionary efforts as alternative models for development, we obtained auxiliary support for our team approach from NOAA for a new database project and from Cal Fish and Game for work on fisheries program data management. This has supported migration and ingestion of data collected within the CCE LTER sampling region.

Collaborative science studies research continued with a best paper award for a Digital Curation Conference contribution that will be published in the International Journal of Digital Curation (Karasti and Baker, 'Digital Data Practices and Long Term Ecological Research Program Growing Global'). In addition, the concept of 'Community Design' was investigated and presented at the Participatory Design Conference (Karasti and Baker, 2008). As part of our focus on articulation we continued to write for Databits: 3 articles in Fall 2007 ('Tools: Web-based data visualization with JPGraph, 'YUI: An Open-source JavaScript Library', 'Professional Learning Opportunities: Conferences, Meetings, and Mindsets'; 4 articles in Spring 2008 ('Big Science and Local Meetings', 'Developing and Using APIs in System Design', 'Preservation Metadata: Another Chapter in the Metadata Story' and 'Data

Quality: Yet Another Chapter in the Metadata Story'). Four Ocean Informatics participants attended the 2008 IMC meeting using support from this year's supplement grant. Three posters were presented at the EIMC Environmental Informatics Conference: 'Abstracting Functionality and Access: Facilitating Data System Manageability and Site Coordination'; 'LTER IMC Community of Practice: A Learning Environment'; and 'Local Information Management and Information Infrastructure: Roles, Responsibilities and Practices'.

1. Major Cyber-Infrastructure (CI) Challenge
Our major need is for increased staff support in order to be able to make long-term plans as well as to establish and maintain a minimum local infrastructure (MLI) capacity that would allow us to participate actively in development of a wider variety of tasks in a more timely manner, e.g. tasks related to networking, enactment of existing standards, community prototyping, standards-making, web services development, design work for new data types, and pre-federation activities.

1a) Science Issues Driving the Need for a Minimum Local Infrastructure (MLI) Capacity
Traditionally within the LTER, local science is supported by local information management. In order to expand this arrangement to include support for synthetic science as well as cross-project, cross-institution, and cross-network informatics, a new way of doing information management is required. We have grown from preserving well-defined local datasets for immediate use locally to needing to plan for new sets of expectations, new types of organizational arrangements, and new kinds of learning environments. The multiple levels of work involved in developing long-term data stewardship and networked data repositories involving diverse data types have only begun to be recognized and incorporated into the information landscape. Maintaining usefulness of local endeavors means being prepared to address data interoperability. Data interoperability, a requirement for a growing number of contemporary large-scale scientific undertakings, will require development of capabilities that enable LTER site participants to contribute to community standards-making, local infrastructure-building, and sustainable innovation.

Without support to local infrastructures, a divide grows between top-down and bottom-up approaches to information management. Cyberinfrastructure efforts often focus on high end computing, massive storage, and grid capabilities while local infrastructure efforts focus on data capture at the source, organization, and description. Local information systems typically evolve through development of modular functionality, informal prototyping and dataset-driven design. Local information management tasks frequently involve rapid responses to unplanned opportunities that occur during data capture, analysis, or preservation. The work involves both data-handling and articulation work; both are critical to contemporary cyberinfrastructure-building. Development of local infrastructure will insure that sites are prepared to manage, innovate, and engage locally as well as to design and participate in community collaborative environments. Work involves
-implementation of improved information systems able to address contemporary data and metadata exchange requirements in support of a web-of-repositories vision
- data organization that includes real-time and new types of datasets as well as their integration into our information system as well as for basic system administration
- data integration that would be supported by implementation of the proposed living dictionary in partnership with the LTER IMC
- availability of high quality data through QA/QC services addressed in an extensible manner at the information system level in partnership with the LTER IMC
-mechanisms for working with data through development of local toolkits for matrix manipulation as a resource for local contributors and users of the information management system
-new types of data access that involve design and development of web services as well as support for a half dozen enactment scenarios involving diverse users, datatypes, and interface needs
-new types of data discovery and query through implementation of a geographic module and framework datasets compliant with existing standards and thus able to interface to existing GIS or KML applications

-taking leadership role in articulating, learning and teaching of informatics so local researchers and students benefit from and engage with the local information environment

1b) Impact of Insufficient Minimum Local Infrastructure (MLI) Capacity
Without support of local infrastructure approaches, network design is conceptualized and funded top-down.

Lack of staff precludes a site from addressing the topics listed above (1a) in a timely or a practice-based rather than theory-based manner. Due to the increased volume of data and maintenance of existing systems, site participants are increasingly limited in their ability to contribute time for communication, design innovation, or practice-based experience that would enrich theoretical understandings of data organization and information management. There is also no time or support to participate in opportunities to articulate and/or teach what has been learned in practice. Site information management work would benefit from a more formal approach that would lead to site infrastructure readiness as well as further advancements in local information management and innovative contributions to cross-site coordination.

1c) Impact of Lack of Minimum Local Infrastructure (MLI) Capacity on Network Activities
Without site-based MLI support, a top-down approach to information infrastructure-building influences design of contemporary networks in terms of centralization versus federation.

We have had to become highly selective in identifying a few tasks that synergize with site tasks to work on at a community level. We have focused on the unit dictionary since 2004. Several years ago design was initiated collaboratively; this effort was developed, prototyped, reported verbally at a meeting and in writing via Databits. The effort was discussed in conjunction with other IMC work on controlled vocabularies and ontologies. Having completed the first prototype, it proceeds at an unfunded pace rather than at the speed of a critical NIS component. Progress was made at the 2008 IMC meeting on the dictionary although no mechanisms to support this community work are currently available. Discussions at the annual IMC meeting resulted in agreement to try a new site-network design approach. The plan is to carry out unit dictionary development in a site development arena and deploy it into an LNO production area using SVN as a coordinating mechanism.

2) Site CI development projects pertinent to a general Service Oriented Architecture (SOA) framework

The unit dictionary described above is being developed in a SOA framework. We will investigate development using the site-network design approach mentioned in 1c. With appropriate site support several Ocean Informatics (PAL & CCE) modules could be used community wide, e.g. unit, attribute, and qualifier sets in addition to personnel, bibliographic, and media gallery modules. These modules, designed with cross-project sharing in mind, have been developed as APIs with management interfaces.

Cedar Creek SiteBytes 2008

2008

Central Arizona Phoenix SiteBytes 2008

This year started with our NSF site review, which of course had been the main focus for several months of preparation. It has also been a year of gaining experiences with using ASU’s high performance storage system and virtual servers. Although overall successful and recommendable it

was not smooth sailing and I am sure our systems administrator, Wayne Porter, would be happy to share details with anyone interested in taking that route.

Mostly routine work has gone into the CAP website. Several datasets via EML files and php data entry applications were added. The most exciting of which is for our new storm water auto-sampler installed at Indian Bend wash. This application accesses a CUAHSI webservice to retrieve height and discharge for the closest USGS gage whenever a new date and time for a set of water quality data is entered. On the backend we have been experimenting extensively with CUAHSI's Observation Data Model (ODM) and taken it to the test of many different datasets. As expected, this has been very successful for simple time series and is serving us well for the streaming data from our two climate towers. However, we found ourselves extending the schema to hold additional metadata describing sites or experimental treatments of the site when trying to go beyond the simplest time series and it was quickly clear that these types of datasets will need a different approach.

Our INTRANET, an application that lets researchers enter and update project descriptions, proposals, annual reports, publications, and manage vehicle and equipment reservations has seen the end of its useful life with the latest upgrade of Java, Struts and Tomcat. This has led me to look into the new Visual Java Server Faces in combination with AJAX as implemented in NetBeans. This technology looks promising, although it still seems to change with every upgrade of NetBeans.

The projects described in the 2007 site bytes funded by the Arizona Water Institute are currently being pulled together into a coherent application that will allow the search and retrieval of water related data. In this context especially the online metadata editor is of interest, where Raul Aguilar has reached a milestone and a first beta release can be found on this website under projects.
The year ended with the award of a NSF BDI grant to add infrastructure to SEINet and improve its online character and character state data model and keying algorithms. Ed Gilbert, who started programming SEINet will join us again for the three year funding cycle.

The metadata editor is one of the project at our site that we hope other sites will find useful and will join into the development / feedback process set up here. In addition we are currently collaborating with the ASU Mars lab to start build a GIS data management system based on PostgeSQL, GEOServer, GoogleEarth and related open source technologies.

The CI needs at our site are centered around gearing up to managing real time sensor data streams. We are intending to implement data Turbine, everything we can learn on QA/QC proceedures, data models, and data delivery through webservices.

Coweeta SiteBytes 2008

2008

Florida Coastal Everglades SiteBytes 2008

With last year's primary focus being the FCE web site (http://fcelter.fiu.edu [149]) redesign, this year's focus has been on FCE web site enhancement. The second phase of FCE research (FCE II) entered its second year and several new FCE II working and cross-cutting theme groups have started to work with the FCE information management team on ideas about how to present their research and data on the FCE web site. Two groups in particular, Modeling & Synthesis and Human Dimensions, have plans for dynamic web interfaces that may allow the web user to enter their own set of parameters into one of the FCE models and see the real-time model results or possibly allow

users to see and listen to human interest interviews about living in the South Florida area from the early developmental stages of the area to the present day.

A web-based query interface tool, which is linked to FCE physical and chemical research results stored in the FCE Oracle10g database, is nearly finished and is expected to facilitate data discovery and data access for FCE and LTER network scientists. Over 7 million physical and chemical data values from outside agencies, such as Everglades National Park, South Florida Water Management District and the USGS, have also been added to the FCE Oracle10g database to enhance data discovery and access of important ‘support’ data from outside agencies.

There are at least three major cyber-infrastructure (CI) needs at the FCE LTER at this point in time:

1. The FCE needs a large Geodatabase (software, hardware, and staffing all related to the Geodatabase).
a. The ability to share GIS data within and among working groups, especially the Human Dimensions and Climate groups, would facilitate data analysis and synthesis. GIS data stored in a database would help us to distribute GIS datasets and allow us to use the data for interactive maps.

b. Researchers have to transfer GIS data via email, FTP server, optical media, or thumb drives and large GIS files transfers using any of the methods described can be very problematic.

c. GIS data from the Everglades and FCE researchers would be available for cross-site comparisons.

2. Additional storage for the Oracle database.

a. FCE researchers have requested that outside Agency physical and chemical Everglades data be added to the FCE research physical and chemical database. Currently, there are over 7 million rows of outside agency data included in our database and I expect to add at least 5 million more rows in the next few weeks. These data, coupled with the previously mentioned GIS data, will quickly fill hard drives!

b. FCE researchers want to have access to their data AND any support data via one user friendly portal. Many have complain that it is too 'difficult' to learn database interfaces for several outside agencies like Everglades National Park, South Florida Water Management District and the USGS. If we can't collect and house their support data, research would be more difficult.

c. It is evident that LTER network scientists also need these support data and having the information in the FCE database would facilitate discovery and access for the community. Especially for those in the LTER network who are not familiar with what agency data may be available.

3. A new FTP server

a. Transfer of large data files between researchers is necessary for collaborations.

b. The computer currently used for the FTP server has software RAID instead of hardware RAID. Additionally, it is a small, older computer with performance issues.

c. A more reliable FTP server would facilitate the sharing of data and files among researchers.

Georgia Coastal Ecosystems SiteBytes 2008

GCE Site Byte for 2008

This past year we finished the update of our IT infrastructure we began last year. We acquired a new web server and database server, both equipped with quad-core cpus, hardware RAID arrays and other fault-tolerant features. These servers compliment the high capacity GIS/Geodatabase server we acquired last year. We also re-deployed our older database server as an application testing platform and SVN repository server for the project. All GCE source code, including MATLAB data processing software, web sites, and database designs (i.e. SQL scripts for all objects) are now under version control with appropriate permissions.

We also deployed a completely redesigned GCE website this year. The new site brings together the content from the original public GCE web site, private project web site and GCE data portal site within a single framework and navigation system. Legacy static and dynamic pages were converted to valid XHTML and redesigned using Dreamweaver templates to provide consistent scaffolding and linked library items to manage navigation elements. Newer pages and MATLAB-generated content from real-time data harvesters are now implemented as XML data files, using client-side XSLT and AJAX to generate all web scaffolding and navigation menus. Having the capability to produce dynamic web content from presentation-free XML data is an exciting alternative to server-side scripting. We have also begun using SQLXML 3 to create SQL query templates (which provide secure access to data storeded in SQL Server directly as XML with support for input parameters), making it possible to generate dynamic web pages directly from the database with no web server code to maintain at all.

As part of the web site redesign we also implemented a more rebust authentication system that is linked to role and status information in the GCE personnel database. Now that all GCE members authenticate individually, and we can provide appropriate access control, we have started to provide more web tools for maintaining information in GCE databases (e.g. end-user updates of personnel bio pages, login password recovery and changes, adding/editing news items, direct updates to the bibliography, etc.). Other recent additions are a searchable/browsable file and imagery archive with web-based management (http://gce-lter.marsci.uga.edu/public/app/resource_search.asp [150]), and a new application for managing citation information and reprints in the GCE bibliographic database.

Our new GIS program suffered a significant set-back last Fall, when Kris Meehan (our first GIS specialist) left to take a job in Iceland. We hired John Carpenter to replace Kris, though, and he has been able to pick up where Kris left off. We're now back on track and providing analytical support to GCE researchers and general GIS resources for the LTER community. John will be attending this year's IM meeting and he's looking forward to participating in discussions on improving GIS capabilities for the network.

Our long series of server migrations and web site redesigns put us behind on basic data archiving, so a major goal for the coming year is to get all the 2007/2008 data online as we prepare for our mid-term review next summer. We are also developing a strategy to use the Google Maps API to provide basic spatial display and query capabilities to the GCE data catalog, following the excellent work that FCE and VCR are doing in this area. We are also expanding the real-time data displays and tide information available on the public web site to make it easier to assess field conditions at our site.

Cyber-Infrastructure Needs:

The major CI need at GCE is additional IM staffing, particularly in area of system administration and sensor deployment and management. Our IT infrastructure has grown to include 4 production servers, 5 workstations and several laptops, and securing and maintaining these systems is significantly limiting effort that can be devoted to processing core GCE data, acquiring ancillary data to support GCE research, and contributing to network-level projects like Eco-Trends. We have also proposed to add an Eddy Covariance tower at Sapelo, and funding appears likely; this sensor installation will present major new challenges for instrument installation and maintenance, and

provide a major new data stream to be managed, which will put even more strain on current IM staff.

CI Development Projects:

The GCE Data Toolbox software, described in numerous DataBits articles and IMC presentations, is a very mature product that could be useful to both LTER IMs and site scientists alike. For example, these tools can be used to acquire, standardize and synthesis USGS NWIS data, NOAA climate data, and LTER ClimDB data on the fly, and also includes a powerful Q/C system that could be very useful for automating processing and Q/C of high volumn sensor network data (described in a paper that will be presented by Wade Sheldon at EIMC 2008). With additional development effort these tools could be updated to support EML metadata (instead of or in addition to text-based FLED metadata). This would provide IMs with a feature-rich tool for both data processing and EML generation, and scientists with a flexible client tool for retrieving and analyzing EML-described data produced by LTER sites.

Harvard Forest SiteBytes 2008

2008

Hubbard Brook SiteBytes 2008

One of the major advances at Hubbard Brook this past year was an overhaul of the physical sample archive. While this effort is not yet complete, a strategy has been developed to ensure that this valuable resource will continue to stimulate new research. The physical sample archive now contains over 40,000 samples (e.g., water, soil, leaves, tree rings) that have been collected at the site since the early 1950's. A team was developed to assess the state of the archive and make recommendations for improvement. The sample archive database is central to the archive and allows researchers locate samples and associated metadata. Major improvements have been made to the structure and content of the database this past year. An interface is in development and it will soon be possible to query the database on-line. As part of the renewed interest in the archive, the Hubbard Brook Research Foundation will host a workshop that will bring in collections experts to make recommendations and establish procedures for establishing archives.

We are continuing to enhance the wireless sensor network at Hubbard Brook. We now have 9 wireless radios that have been deployed at each of the nine experimental watersheds. The data are posted and graphed on our website in near-real time. The next step is to develop quality control procedures in real-time. In addition to installing wireless hydrometeorological sensors, we have also been experimenting with a wireless camera and light sensors that will be used to measure phenology at the site as part of the the larger Northeast Regional Phenology Network (www.nerpn.org [151]).

Hubbard Brook recently received funding from NSF to establish a site REU program. This past summer was the first of 3 years of funding, with the hope that the tradition will continue in the future. Ten students participated in the program and both the students and mentors thought it was highly successful. While the first year went smoothly, we identified several improvements that would make the program run more efficiently. One item pertaining to IM is the implementation of on-line application forms. We have begun developing these forms and have been modeling them after the forms Harvard Forest uses.

The coming year will be a busy one for Hubbard Brook. Our LTER renewal proposal is due in 2010 and much of the focus of our quarterly meetings will be dedicated to shaping the new proposal.

Information management will be an integral part of the proposal and we will continue to make improvements by incorporating suggestions we received during our mid-term review. One of these suggestions was to develop a better method of tracking research projects at the site. This suggestion was fairly simple to initiate because there is a proposal form for conducting research at Hubbard Brook, which provides much of the necessary project level information. This information is now written directly to a database so we are better able to track which projects were approved and can keep tabs on progress. This recent effort is timely because it will be useful as the larger IM group begins developing the recently funded IM project database. We are looking forward to contributing to this effort and may make adjustments to the information we collect at Hubbard Brook based on the larger effort.

Major Cyberinfrastructure need: We need additional personnel to help move projects along. We currently rely heavily on NSF supplements to hire students, but this is not a long-term solution.

CI development project: Integration of physical sample archive data across sites that would enable researchers to see what historical samples are available.

Jornada Basin SiteBytes 2008

Ongoing Activities
We continue to perform ongoing tasks of data documentation, collection, archive, and backup for all research data archived within the Jornada Information Management System (JIMS). We continue planning how to effectively integrate geographic information system (GIS) data layers and research site locations with research data and associated metadata to enhance the quality and availability of all JRN data and to generate more detailed and precise EML documentation. We also continue to provide GPS of research site locations and production of new GIS layers, to provide GIS and GPS support to researchers and students including training and map production. We continue to administer NMSU site licenses for GIS and remote sensing software (Nolen). We continue to populate the JIMS database with research project and associated dataset metadata to support EML generation and the data catalog for all Jornada Basin LTER (JRN) datasets.

EML
We are currently modifying our XML style sheets to generate EML level 5 using PHP for all JRN data holdings. The GIS metadata stored in XML (ESRI ArcCatalog) will be used to generate EML documentation for all JRN GIS layers stored within JIMS, which will include geographic bounding coordinates within EML documents describing research datasets.

Storage Solution
The Jornada Experimental Range (JER) purchased a new storage area network (SAN), a fiber channel switch, network attached storage (NAS) server, and fiber channel tape library. This storage solution gives JRN nearly 24 TB of storage capacity with the ability to add 28 TB by adding 1 TB hard drives.

File Server
JRN is purchasing a new file server using the LTER site supplement this year. The operating system for the new server will be Novell Open Enterprise Server on SUSE Linux Enterprise Server. Upgrading the operating system from Netware 6.5 will give us the capability to organize our data easier as our volume size limit will be increased from 2 TB to 16 TB per volume; e.g., our image archive is currently spread out over 6 volumes on the old file server.

GIS/Image Archives, Services, and Web Applications
We are working closely with JRN users of imagery to improve organization of and access to JRN GIS and remote sensing (RS) data. With the growing quantity of imagery at the Jornada, consistent

organization, use of naming conventions, processes for adding and acquiring GIS and RS data, and search interfaces and map services for gaining access to the image archives is critical. We are also exploring new methods for backup of imagery to minimize the backup window and reduce impacts on the local area network (LAN). We are preparing to connect 2 imagery users to the SAN to demonstrate the speed and performance of the SAN versus traditional local disk storage to our imagery users. Backup of the SAN is much faster than LAN-based backups of desktop computers and does not impact LAN performance. We deployed ESRI ArcGIS Server for Java and ArcGIS Image Server on the new GIS server and have begun to develop and deliver map services and web mapping applications for research projects. We attended the Annual ESRI International User Conference in San Diego, CA again this year to gather more detailed technical information necessary to develop search interfaces for the aerial photographs (> 5,000) and other image archives and to stay current on advances in ESRI software.

Jornada Website
The new web server has been deployed. The new data cart will be deployed soon and will enforce the JRN data access policy by requiring user registration and authentication prior to download of JRN data. We will be phasing in the data cart in a prioritized manner; ongoing, long-term datasets first, followed by climate and all remaining unrestricted datasets. Prior to deployment of the new server, the website was updated. The new website has the familiar look and feel of the old website, but under the hood, site authentication now uses the JRN LDAP directory and XML configuration files for dynamic web page creation. The LDAP integration also supports the new user registration and data cart systems.

Cyber-Infrastructure Challenges and Projects

1. Major Cyber-Infrastructure (CI) Challenge
Adding full-time, permanent personnel to the JRN site to support site and cross-site research projects and other LTER Network activities such as K-12 education is a major CI challenge for JRN. This section of the site byte will describe the CI staffing needs to support site and cross-site research and education activities and a planned CI project and how the position impacts current and planned activities. JRN needs a full-time computer programmer/systems administrator position in the site office to reduce current workloads within the site office and maintain the computer network systems of the JRN, including the wireless networks in the field. This position would also develop computer programs for collection, transmission, quality assurance, and analysis of data collected from sensor networks and supporting wireless systems. There will also be a need for a full-time sensor/wireless network technician position to install and maintain the planned wireless/sensor networks and related equipment, and when those responsibilities allow, to assist in data collection and maintenance of other field research equipment. Jointly, the 2 positions would maintain the wireless and sensor networks described below; one from a software perspective and the other from hardware.

Planned Jornada CI Project Goal
Extend wireless coverage and infrastructure across research sites and increase bandwidth to the field station to support research projects, wireless sensors and sensor networks, and research and educational activities; making data collection, quality assurance and control, and site management and maintenance more efficient and cost effective.

1a. Science issue driving the need for increased wireless instrumentation and staffing

Since the mid 1990s, we have extended our LTER research to include all of the Jornada Experimental Range in addition to our original site, the NMSU CDRRC. The total area being addressed by our research activities is > 100,000 ha. This area is diverse in topography, vegetation, soils, and management history. In addition, summer thunderstorms are isolated events that often occur only in a small part of the study site. Because the focus of our LTER research since LTER IV (2000) is on patch to landscape-scale variation in pattern and processes, we need to be able to

measure and monitor both the abiotic and biotic characteristics of the entire area as well as its spatial and temporal context within the county of Dona Ana and the northern Chihuahuan Desert region. We have been expanding our wireless network capabilities as funds are available, but our staffing needs have not kept up with our instrumentation.

One limitation to conducting site and cross-site research is the limited coverage of the wireless networks across the research area (104,166 hectares). Remote, semi-automated collection and processing (QA/QC, storage, backup) of data collected over the wireless network from sensor networks connected to scientific data loggers and storage modules offers the promise of increasing the frequency, quality, and amount of data collected across the Jornada. By leveraging technology we can minimize field technician time needed to collect data and perform maintenance while minimizing data errors by automating the process of data collection, quality assurance, and archival storage and subsequent backup. The extended coverage will also support near-realtime event based analysis. Currently, several research sites and the JRN weather station are connected to the wireless network for data collection. The need to add further instrumentation at varying locations across the research area requires extending the wireless network to include most of the research area. Given the topology and lack of high canopies, the Jornada Basin research area is ideal for extending the WiFi and spread spectrum wireless clouds across most of the research area. This system will support a more automated approach to data collection and processing and maintenance.

1b. Impact of insufficient staffing and wireless instrumentation on site activities

An extended wireless network, and staff to install and maintain it, will support research, education, and public outreach activities across the research area. Some of the activities that the extended wireless cloud will support include interactive distance learning from the field to classrooms (K-12), data collection and use (upload/download data) in the field, near real-time streaming data collection, and efficient use of field technician time when responding to event driven data collection efforts. The wireless coverage will support the automation of historic rain gage networks and weather stations by supporting wireless data collection as the instrumentation is upgraded. The USDA ARS Jornada Experimental Range is currently updating its recording rain gage network in a phased approach to support wireless data collection.

The extended wireless coverage will also help reduce trips to/from the field to retrieve information from the JRN servers and reduce technician and travel time to remote sites for event driven data collection by allowing the technician to monitor events remotely prior to going to the field, thus minimizing trips to remote sites that were not impacted by the event (e.g. localized thunderstorm). This should help reduce fuel costs and make more efficient use of technician time. The system will also support a more automated sensor data collection and processing approach as more sensors and data loggers are established, reducing field technician time collecting data and providing more time for technician maintenance and will reduce the turnaround time from data collection to data availability and use.

Increased bandwidth is needed to support increased use of the connection between the field station and NMSU campus to support video and data streams. The current T-1 connection is not sufficient for this purpose as the T-1 bandwidth (1.54 MB) is shared with the phone system and the Intranet connection at the field station. The most economical solution would be to connect the campus and field station using point-to-point high speed wireless from the field station to Tortugas Mountain, near the NMSU campus. NMSU currently has the fiber cable to connect Tortugas Mountain and the campus and is in the process of obtaining easements with the electric company and the city. Some down time due to atmospheric conditions is acceptable and future sensor networks and data loggers will be designed to accommodate for this limitation.

1c. Impact of insufficient staffing on site’s participation in network activities
The ability of JRN scientists to conduct cross-site research with other sites is severely limited by

incomplete wireless coverage at our site. We can participate in location-based research where data are collected manually or from our current wireless setup, but that includes < 5% of our large research site.

Our site is also limited in the ability of our IM to fully participate in network level activities because of time constraints. Currently, two of the responsibilities of the JRN Information Manager (IM) are system administration and computer programming. Given the many duties of the Information Manager to support the site, it is extremely difficult for the IM to participate fully in LTER Network initiatives. The proposed positions would augment site staff and reduce the workload of the IM, Research Site Manager, and field crew. The new positions would allow the IM to focus on information management and maintenance of the Jornada Information Management System while allowing for more active participation in the LTER Network including cross-site research projects and information management working groups. The assistance of a sensor/wireless technician would give the site flexibility to assist in research site setup, data collection, and maintenance for cross-site research projects. Adding the 2 staff positions to the site is the JRN's top priority as funding is made available. If funding is sufficient for only one position, the computer programmer/system administrator position is the higher priority.

The JRN has and will continue to pursue funding for the equipment and software to extend the wireless coverage and bandwidth from grant proposals and collaborations. The JRN plans to extend the wireless coverage within the next year or 2. Increasing the bandwidth to the field station will be delayed until final easements are established from Tortugas Mountain to the NMSU campus. The extended wireless system will also facilitate research and education activities within the site and across the LTER Network. It will enable cross-site sensor network deployment at the Jornada to support synthetic research by providing the infrastructure for sensor networks and remote K-12 ecological education within Southern New Mexico and West Texas and potentially across the LTER Network. The extension of the wireless coverage across the research area is critical to emerging site and cross-site research efforts and approaches. The greatest challenge now is to fund the additional staff positions.

CI Project needs (prioritized):
a. Additional Staffing:
i. Full time position: Computer Programmer/System Administrator. This position would maintain the wireless systems and network infrastructure and develop programs for data loggers and sensor networks to support collection, download, and quality control and assurance for near-realtime streaming data and quality assurance routines and algorithms to monitor the wireless system and sensor networks for performance issues, such as detecting sensor drift.
ii. Full-time position: Sensor/Wireless Network Technician. This position would install and maintain the wireless system and sensor networks, collect and process data from scientific instrumentation in the field, and perform other duties of a research assistant as required.
b. Support Equipment: Additional equipment and software (rugged laptop, computer, etc.) to support new position
c. Vehicle: 4WD vehicle for the Sensor/Wireless Network Technician position.
d. System: Wireless network and related equipment to extend wireless coverage across research area and sites (infrastructure only, not sensors and related equipment) and to establish a low recurring cost point-to-point high speed wireless connection to/from the remote field station and the servers and site offices on the NMSU campus (25+ miles). This would increase bandwidth from existing < 773 KB partial T-1 connection to > 1 GB bandwidth to the Jornada servers with minimal recurring costs beyond initial acquisition.

Kellogg Biological Station SiteBytes 2008

No major new developments occurred during the last year. We spent some time experimenting with using Google Earth and map overlays to deliver GIS data, improving the metadata especially the methods and protocols, and took advantage of virtualization to experiment with caching, replication and backup strategies.

This past year the KBS internet connection was upgraded from 1.5 to 10Mb/s allowing for faster offsite transfer and better use of video conferencing. The next CI need at KBS will be establishing internet connections to outlying sites to provide communications for micro flux towers.

Konza Prairie SiteBytes 2008

2008

Luquillo Experimental Forest SiteBytes 2008

Outreach and collaboration in synthesis activities:
This past year the LUQ LTER IM staff dedicated a great deal of time to develop a closer collaboration with the Institute's professors, specifically with the director of the Institute Collaboration consisted in developing presentations, managing data to be used in cross-site efforts of synthesis and presenting Information Managements methods to educate the Institute students in the development of metadata. The following URL contain the links related to those events and/or presentations:

Generation of the database files that Dr. Elvia Melendez-Ackerman, Institute for Tropical Ecosystem Studies (ITES) director, had to prepare to bring to the Ecophylogenetics Working Group held at Harvard Forest from Oct. 11 to 14, 2007
(http://www.cbs.umn.edu/cavender/ecophylogenetics.shtml)

Talk to the ITES staff and faculty on the importance of data documentation and the methods used by the LUQ LTER Information Management staff. Presentation entitled "INFORMATION MANAGEMENT: a path less taken. A PRESENTATION TO THE ITES PERSONNEL" can be reached and downloaded from: http://luq.lternet.edu/datamng/index.html

Collaboration with Dr. Elvia Melendez-Ackerman, Institute for Tropical Ecosystem Studies (ITES) director, in the preparation of the 50th Anniversary: "ITES Historical Photo Album.:

Presented at the Symposium (http://www.ites.upr.edu/HistoricalPhotos/ITES50thAnniversary.html), and

Web page with photo, press conference clips and other materials
(http://www.ites.upr.edu/HistoricalPhotos/ITES50thAnniversary.html)

LUQ Information Management CyberInfrastructure development related activities:
The generation of Level 5 with High Quality Control standards of LUQ LTER Metadata in collaboration with Inigo San Gil of the LTER Network Office (LNO). Approximately one third of the online metadata has been revised and updated to achieve this high quality standard although around 84% of our online metadata has unrevised level 4 to 5 EML packages generated for them.

Meetings, and specific plans have been developed, in conjunction with LNO technical staff to build a community-based services and a service-oriented architecture (SOA Main target has been to re-structure and update the design the LUQ LTER Web site in order to provide a scalable, community-based, service-oriented architecture that will provide data services to ensure secure and efficient

access to data stored in site data repositories. This is an LTER communities collaborative effort itself since it involves the collaboration of the LUQ LTER scientific community (represented by the local IM Committee composed by 4 LUQ Investigators, several LNO's technical staff, and the LUQ Information Management staff), where the information manager is the liaison that coordinates the activities of the 3 community members.

McMurdo Dry Valleys SiteBytes 2008

2008

Moorea Coral Reef SiteBytes 2008

2008

Niwot Ridge SiteBytes 2008

2008

North Temperate Lakes SiteBytes 2008

The information management team at the North Temperate Lakes (NTL) LTER has been developing enhanced functionality for the NTL information management system in several areas: data acquisition, data access including expanded access to spatial data, and the management of sensor network data.

NTL has a wealth of online data available through the NTL website. The existing data catalog interface is innovative in that its web pages are generated dynamically from metadata in the NTL database. The current catalog is organized by categories, which works well in "browse mode" for scientists who are familiar with the NTL data. Recently we have developed a new search interface to enhance data discovery, especially for other user groups such as K-12 educators and the public. This interface allows the user to select a Project Type, Theme, Location, and Period of Interest as well as to add search criteria for text strings in various metadata fields such as dataset title, investigator or species name. We are currently beta-testing this new search interface and anticipate public release this fall.

NTL-LTER completed a data migration and upgrade of the spatial data server in conjunction with a transition to a new spatial data manager (Jeff Maxted). This process involved the purchase of a new Dell PowerEdge SC440 Server and the installation of Red Hat Linux, Oracle, and ArcSDE. We upgraded our server-based GIS architecture by installing ArcGIS Server. Migrating to this application is enabling us to provide spatial data through web mapping services. These new web mapping services enhance our capacity to serve vector and raster data from our extensive spatial data catalog. This upgrade has also led to new plans for developing applications that will enhance a user's ability to visualize and explore our data catalog.

Z3, a program originally designed for zooplankton counting and measuring, is being reworked to include extra features and to extend its applicability to other counting and measuring projects, such as those with benthic invertebrates or fish scales. A new "overlay" mode will allow Z3 to work with any Windows-compatible camera, including any microscope cameras that provide a live feed window.

The management of data from high-throughput field sensors deployed on lake buoys has challenged us to investigate new data models and collaborate to develop new tools. We are currently evaluating and testing these new approaches for use in the production information system. The Vega data model is a flexible database architecture designed to accommodate additions and changes in sensor deployments without database structure changes. Tools based on the Vega database handle quality assurance/quality control tasks, exceptions in streaming data, and insertion of data into a repository. In addition, a web-based application has been developed for discovery and retrieval of sensor data stored in Vega.

Response to questions:
1. Identify at least one major Cyber-Infrastructure (CI) need at your site (e.g. software, hardware, communications, staffing) and describe:
a) What is the science issue that drives this CI/IM need
b) How does the CI need constrain site activities
c) How does the CI need impact your site's participation in network activities

The major NTL CI need is increased staffing. An expanding set of activities for information management both at the site and network levels is difficult to handle at the current staffing level. Particularly challenging is processing the data volumes from the NTL sensor network. These data are important for understanding multi temporal and spatial scales of control of ecological processes. New initiatives, while important, such as the Trends database have taken considerable time. While NTL has participated at a significant level in the past in network activities, the work load on staff does constrain our participation at the network level.

2) If applicable, list any CI development projects at your site that could potentially be used by other sites or shared and collaboratively developed into network-level CI implementations as part of a general Service Oriented Architecture (SOA) framework (i.e. projects with a wide range of use cases that could support decadal plan CI initiatives and cross-site science projects).

Many of our databases could be made accessible as part of an SOA solution, including our Lake Information Database, which provides variable and sensor information about sites participating in the GLEON network (http://www.gleon.org/lakes/ [152]), our raw sensor data via a web-accessible tool (dbBadger; see http://www.gleonrcn.org/index.php?pr=Products [153]), and a wide range of ecological data via our dynamic data catalog (http://lterquery.limnology.wisc.edu/ [154]). These applications could also be extended to process data from other groups that expose their databases via web services.

Palmer LTER SiteBytes 2008

LTER Site Byte: Palmer Station
Karen Baker

In addition to advancements in design and population of our cross-site information management system DataZoo, there were three major accomplishments this year: 1) the writing and acceptance of the site’s fourth 6-year proposal in which information management is recognized as one of ten components where both science and information management summarized their objectives in an itemized list, 2) the publication of an information management paper in the oceanography journal Deep-Sea Research (Baker and Chandler, 2008, ‘Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics’), and 3) initiation of a division-level computer infrastructure recharge facility as an organizationally situated support for systems administration.

DataZoo has reached ‘flagship’ status as evidenced by a change in inquiry from ‘Where is dataset X’ to ‘Why isn’t dataset X in DataZoo?’. Local design efforts focused in particular on improving the

user interface to DataZoo, creating an online help system, and shifting code practices to use of libraries and an object oriented approach. Design and flow of the system as a coherent whole was revisited in order to clarify the design as well as to add consistency to users' interactions with the data and metadata models. The management interface was improved in terms of useability. We continue to address design decisions such as whether code sets are associated at the attribute or the column level (decision: move to column level), how to deal with metadata that changes over the course of a dataset (decision: handle at section level), and what keyword classification approaches to take.

The PeopleZoo personnel application module has been redesigned as an API and is being used to track project and cruise participation. Three additional information system elements stabilized: 1) definition of relationships between units, attributes, and qualifiers as an interdependent set of dictionaries, 2) development of web services in a locally developed hybrid approach using both SOAP and REST, and 3) launch of a multi-element cooperative dataspace composed of a set of related applications that capture and make accessible multi-faceted datasets.

Support for the outreach component included development of an education web portal and investigation of communication options resulting in redevelopment of the picture-of-the-day activity as a blog to support the new teacher-at-sea activity. LTER network collaboration included returning to design of a community-developed, network-hosted, service-oriented unit dictionary database. The Ocean Informatics team worked closely with the LTER Information Management Committee Dictionary Task Force to define and coordinate continuing development of the unit dictionary. In addition, we worked to capture site datasets submitted to EcoTrends for submission to DataZoo. After a third unsuccessful proposal to obtain NSF funding for the Ocean Informatics conceptual approach to local infrastructure building and to distributed networking dictionary efforts as alternative models for development, we obtained auxiliary support for our team approach from NOAA for a new database project and from Cal Fish and Game for work on fisheries program data management that supported development of web services and a new generation targeted application.

Collaborative science studies research continued with a best paper award for a Digital Curation Conference contribution that will be published in the International Journal of Digital Curation (Karasti and Baker, 'Digital Data Practices and Long Term Ecological Research Program Growing Global'). In addition, the concept of 'Community Design' as an approach to infrastructure building was investigated and presented at the Participatory Design Conference (Karasti and Baker, 2008). As part of our focus on articulation we continued to write for Databits: 3 articles in Fall 2007 ('Tools: Web-based data visualization with JPGraph, 'YUI: An Open-source JavaScript Library', 'Professional Learning Opportunities: Conferences, Meetings, and Mindsets'; 4 articles in Spring 2008 ('Big Science and Local Meetings', 'Developing and Using APIs in System Design', 'Preservation Metadata: Another Chapter in the Metadata Story' and 'Data Quality: Yet Another Chapter in the Metadata Story'). Four Ocean Informatics participants attended the 2008 IMC meeting using support from this year's supplement grant. Three posters were presented at the EIMC Environmental Informatics Conference: 'Abstracting Functionality and Access: Facilitating Data System Manageability and Site Coordination'; 'LTER IMC Community of Practice: A Learning Environment'; and 'Local Information Management and Information Infrastructure: Roles, Responsibilities and Practices'.

1. Major Cyber-Infrastructure (CI) Challenge
Our major need is for increased staff support in order to be able to make long-term plans as well as to establish and maintain a minimum local infrastructure (MLI) capacity that would allow us to participate actively in development of a wider variety of tasks in a more timely manner, e.g. tasks related to networking, enactment of existing standards, community prototyping, standards-making, web services development, design work for new data types, and pre-federation activities.

1a) Science Issues Driving the Need for a Minimum Local Infrastructure (MLI) Capacity
Traditionally within the LTER, local science is supported by local information management. In order to expand this arrangement to include support for synthetic science as well as cross-project, cross-institution, and cross-network informatics, a new way of doing information management is required. We have grown from preserving well-defined local datasets for immediate use locally to needing to plan for new sets of expectations, new types of organizational arrangements, and new kinds of learning environments. The multiple levels of work involved in developing long-term data stewardship and networked data repositories involving diverse data types have only begun to be recognized and incorporated into the information landscape. Maintaining usefulness of local endeavors means being prepared to address data interoperability. Data interoperability, a requirement for a growing number of contemporary large-scale scientific undertakings, will require development of capabilities that enable LTER site participants to contribute to community standards-making, local infrastructure-building, and sustainable innovation.

Without support for MLI, a divide grows between top-down and bottom-up approaches to information management. Cyberinfrastructure efforts often focus on high end computing, massive storage, and grid capabilities while local infrastructure efforts focus on data organization, description, analysis, and capture at the source. Local information systems typically evolve through development of modular functionality, informal prototyping and dataset-driven design. Local information management tasks frequently involve rapid responses to unplanned opportunities that occur during data capture, analysis, or preservation. The work involves both data-handling and articulation work; both are critical to contemporary cyberinfrastructure-building. Development of local infrastructure will insure that sites are prepared to manage, innovate, and engage locally as well as to design and participate in community collaborative environments. Work involves

-implementation of improved information systems able to address contemporary data and metadata exchange requirements in support of a web-of-repositories vision
- data organization that includes real-time and new types of datasets as well as their integration into our information system as well as for basic system administration
- data integration that would be supported by implementation of the proposed living dictionary in partnership with the LTER IMC
- availability of high quality data through QA/QC services addressed in an extensible manner at the information system level in partnership with the LTER IMC
-mechanisms for working with data through development of local toolkits for matrix manipulation as a resource for local contributors and users of the information management system
-new types of data access that involve design and development of web services as well as support for a half dozen enactment scenarios involving diverse users, datatypes, and interface needs
-new types of data discovery and query through implementation of a geographic module and framework datasets compliant with existing standards and thus able to interface to existing GIS or KML applications
-taking leadership role in articulating, learning and teaching of informatics so local researchers and students benefit from and engage with the local information environment

1b) Impact of Insufficient Minimum Local Infrastructure (MLI) Capacity
Without MLI support, network design is conceptualized and funded top-down.

Lack of staff precludes a site from addressing the topics listed above (1a) in a timely or a practice-based rather than theory-based manner. Due to the increased volume of data and maintenance of existing systems, site participants are increasingly limited in their ability to contribute time for communication, design innovation, or practice-based experience that would enrich theoretical understandings of data organization and information management. There is also no time or support to participate in opportunities to articulate and/or teach what has been learned in practice. Site information management work would benefit from a more formal approach that would lead to further

advancements in local information management and innovative contributions to cross-site coordination.

1c) Impact of Lack of Minimum Local Infrastructure (MLI) Capacity on Network Activities
Without site-based MLI support, a top-down approach to information infrastructure-building influences design of contemporary networks in terms of centralization versus federation.

We have had to become highly selective in identifying a few tasks that synergize with site tasks to work on at a community level. We have focused on the unit dictionary since 2004. Several years ago design was initiated collaboratively; this effort was developed, prototyped, reported verbally at a meeting and in writing via Databits. The effort was discussed in conjunction with other IMC work on controlled vocabularies and ontologies. Having completed the first prototype, it proceeds at an unfunded pace rather than at the speed of a critical NIS component. Progress was made at the 2008 IMC meeting on the dictionary although no mechanisms to support this community work are currently available. Discussions at the annual IMC meeting resulted in agreement to try a new site-network design approach. The plan is to carry out unit dictionary development in a site development arena and deploy it into an LNO production area using SVN as a coordinating mechanism.

2) Site CI development projects pertinent to a general Service Oriented Architecture (SOA) framework

The unit dictionary described above is being developed in a SOA framework. We will investigate development using the site-network design approach mentioned in 1c. With appropriate site support several Ocean Informatics (PAL & CCE) modules could be used community wide, e.g. unit, attribute, and qualifier sets in addition to personnel, bibliographic, and media gallery modules. These modules, designed with cross-project sharing in mind, have been developed as APIs with management interfaces.

Plum Island Ecosystems SiteBytes 2008

We had a successful Fall 2007 site review and have made great progress in generating attribute level EML metadata. However, it’s amazing how many new datasets are becoming available for inclusion in the PIE LTER database, primarily legacy and other non PIE LTER directly funded research grants in the Plum Island Ecosystems area. Goals for the future include: documenting our raster and vector based GIS data in EML; archiving unprocessed/raw non QA/QC data (from MBL researchers and other PIs around the country); completing the redesign of PIE web site (at this point repopulating content); and developing system for more easily sharing GIS data.

Major cyber-infrastructure (CI) needs at PIE LTER:

1. PIE LTER needs more resources (time, personnel and expertise)
a) Additional resources related to staffing are necessary to continue productive scientific research and is necessary to fulfill unfunded mandates for IM activities at the site and network level.
b) Under funded research program staffing constrains both the scientific and IM site activities as solo staffers are expected to become experts in everything and multitasking across a variety of disciplines is the norm.
c) IM activities at the network level are not performed at as high a level as some LTER sites with leveraged IM resources. IM Network activities seem to be judged in relationship to the highest standard of other LTER sites with abundant IM resources.

2. Centralized GIS web site
a) Centralized GIS web site will allow PIE related GIS (geographic, social and ecological) data to be more easily viewed and subsequently shared.

b) It is presently difficult to view /share GIS data across different GIS platforms (ArcGIS, IDRISI).
c) A centralized GIS web site would allow GIS data to be more easily shared at larger scales (regional) than the PIE LTER domain.

Santa Barbara Coastal SiteBytes 2008

2008

Sevilleta SiteBytes 2008

2008
The Sevilleta has three studies that collect data using soil moisture, temperature and carbon dioxide probes. These data are ingested into the Sevilleta MySQL database on a daily basis. In Fall 2008, a student programmer was employed to write scripts to use JPGraph to create plots of the data at daily, weekly, and monthly intervals. The scripts were found to be CPU intensive and would cause services on the Sevilleta UNIX server to fail. Two new Sun servers have been ordered to replace the Sun E450 server that has been the heart of the Sevilleta IMS for ten years.

Managing all the Sevilleta data collected by the sensors is a challenge. Range checks are done on the data as it is entered into the database, but more sophisticated error checking needs to be done. Sevilleta lacks the personnel at this time to write scripts to do more extensive QA/QC. The carbon dioxide data will eventually need to be processed to yield carbon fluxes, and a programmer would also expedite the implementation of scripts to process the data.

The Sevilleta still lacks a mechanism for generating EML from both legacy and new metadata. We continue to work with Inigo San Gil at LNO to resolve this situation, and are close to having a Perl program that parses metadata from a text file into EML.

Short Grass Steppe SiteBytes 2008

The SGS-LTER Information Managers, Nicole Kaplan and Bob Flynn, have focused on four main areas during the past year. First, we are nearly finished with development of our new website, supported by improvements in our underlying information architecture. Recommendations developed by the LTER Information Management committee were taken into consideration during the development process. The new SGS-LTER website was designed by LTER staff, researchers, and web development professionals at the Colorado State University (CSU) Creative Services. The website beta version is implemented ((http://ccsbeta.colostate.edu/ [155]) and will be launched into production November 2008. We used a revised data model for our relational database management system (RDBMS) to drive the new website content, which includes dynamic delivery of data and EML level 5 metadata, improved relationships between different types of information within the RDBMS, and more efficient navigation and search capacities for our ends users.
Second, we've continued to work closely with SGS researchers and graduate students on data discovery, integration and analysis. We've also worked closely with other national and international LTER sites to provide data, expertise, and tools to facilitate integration and analysis for cross-site and network-wide synthetic research projects. Information managers and researchers from the SGS, SEV, KNZ, and JRN LTER sites, Kruger National Park, in South Africa, and EcoTrends have worked together with computer scientists from The Evergreen State College on the Grasslands Data Integration Project (GDI). The GDI brings together ecologists, information managers and computer scientists to address the challenges of integrating long-term annual aboveground net primary productivity data sets, which were collected at different temporal and spatial scales, with

site-specific species nomenclature and codes, and applied various methodologies and experimental designs.
Third, we've consulted with telecommunications and information technology professionals on cyberinfrastructure improvements for our SGS campus facilities and the new SGS Research and Interpretation Center (SGS-RIC). SGS-RIC includes a classroom building, with an attractive seminar room lined with windows for views of the prairie and mountains, a computer lab, and conference room, as well as two, five bedroom houses for lodging and working on-site.
Lastly, we continue to participate in planning future activities for the LTER IM committee. Nicole Kaplan serves as co-chair of the IM Committee, which establishes priorities and strategies to facilitate network-level and synthetic research that requires greater integration of information and interoperability of systems.
Site Byte content can include any information about the site and typically reflects activity related to the information system at your site during the past year.

Please also include in your site byte:

1. Identify at least one major Cyber-Infrastructure (CI) need at your site (e.g. software, hardware, communications, staffing) and describe:
At this time SGS needs more CI support, including staffing and collaboration tools. Information Managers will be required to design, implement and maintain CI in support of new directions in research as a new proposal is being developed, network-wide projects and interoperability initiatives, and CI capabilities at SGS-RIC, a new research and education facility. We recognize such efforts will require investment and maintenance from staff. There is a greater interest in collaboration tools to facilitate communication and exchange of ideas between researchers in different geographic locations as opportunities continue to grow for our site.

a) What is the the science issue that drives this CI/IM need
SGS researchers are reconsidering how they integrate research projects, and apply knowledge gained about the shortgrass steppe ecosystem into future research plans. This knowledge relies heavily on our long-term data sets and therefore increases our PIs' appreciation of a robust CI.

b) How does the CI need constrain site activities
Data discovery and sharing, as well as collaboration and communication are limited at present. More effective tools for data discovery have been developed and will soon be in production on our website. Researchers need training on available data discovery functions with the SGS and Network-wide CI, as well as collaboration tools, such as Video Teleconferencing.

c) How does the CI need impact your site's participation in network activities
Responsibilities of our full time IM staff, leaves limited time for engaging deeply with many Network-wide CI projects. However, we have participated in efforts that don't require long-term time commitments. Nicole participates in leadership and governance rolls and has contributed to a cross-site database, which contains tools and data of interest to the greater ecoinformatics community (please see the grasslands data integration (GDI) paper at EIMC). In addition, Bob participated in XML training at the LNO training lab.

2) If applicable, list any CI development projects at your site that could potentially be used by other sites or shared and collaboratively developed into network-level CI implementations as part of a general Service Oriented Architecture (SOA) framework (i.e. projects with a wide range of use cases that could support decadal plan CI initiatives and cross-site science projects).

We currently have non-automated SOA that transfers data between our meteorological download and parsing service (PC208W) and our Matrix QA/QC and formatting system for input to into

ClimDB.

Virginia Coast Reserve LTER SiteBytes 2008

2008
This has been a busy year for the installation of new hardware and monitoring capabilities at the Virginia Coast Reserve Long-Term Ecological Research (VCR/LTER) Project. In August 2007 we completed installation of a network of ten water-level monitoring stations on Hog Island, Virginia. The new stations use 900 MHz serial wireless communications to connect to the Wi-Fi network installed on the island in previous years. This upgrade replaces a mix of mechanical well monitors (using paper chart recorders) and electronic recorders (that required manual monthly dumps). The upgrade has dramatically cut data outages, because problems are identified at UVA, using graphs that are produced three times per day, and fixed before they become serious. Additionally, we have added an additional tide station and a flux tower to our wireless network.

We also have been doing a lot of work on using VMWARE to implement virtual machines on a variety of low and high-end hardware. On the high end, in spring 2008 we purchased an eight processor Linux server that will, ultimately, replace the VCR/LTER web server (currently running on a Sun workstation). Although that transition is only starting, we have used the new server to develop new analysis and graphical tools for specific datasets. On the low-end, we have used PCs that were discarded due to low speed or lack of memory required to run new versions of Windows to install Linux variants which are much less consumptive of resources. These machines have been used to replace unreliable DHCP network address servers in our laboratory at the Anheuser Busch Coastal Research Center (ABCRC), and to take over primary electronic data collection tasks using the wireless network at the ABCRC. By using VNC (Virtual Network Console) all the machines can be administered using graphical tools from anyplace in the world.

We have continued our work with Taiwan Ecological Research Network (TERN) and participated in an East-Asia Pacific ILTER Urban Forestry and Information Management Workshop in Seoul Korea in the fall of 2007.

In terms of site cyberinfrastructure, there are two major areas where we would like to move forward. The most critical is staffing. With only a single Information Manager, who needs to do a large number of functions: from system administration, to system development, to data management, to managing GIS data, to participation in network initiatives and authoring scientific publications, there are few opportunities for extensive or in-depth development activities. This limitation does not affect any specific scientific area more than the others, but has modest impacts on most parts of the VCR/LTER by making data harder to find and access and reduces the overall time available for network participation. The second area is computational power and storage. The addition of the new server and a 1 TB network appliance have helped to keep our head above water for now, but it would be good to have a backup server in case of hardware failures. Expanded storage and computational power are needed because new monitoring equipment (such as the recently installed flux tower in Fowling Point Marsh) generate data at much higher rates. For example, 18-months of data from our flux tower requires over 30% of the **total** disk capacity of our old Sun server.

These CI limitations constrain site and network activities, primarily because with overextended personnel and computational infrastructure, it becomes difficult or impossible to devote protracted periods of time, or computer power, to particular projects. Only the highest priority projects (e.g., EcoTrends) or those that can be accomplished in relatively brief periods of time (e.g., translator of EML documents into Google Earth KMZ files), can be undertaken.

Despite these limitations, we believe we can substantially contribute to LTER Network-Wide activities. Wherever possible, we design site software to be open ended, so that sites other than

VCR/LTER could be served. For example, in the mid-1990s, we developed the first web-accessible personnel catalog that was used by the VCR /LTER and some other LTER sites, until the LTER Network Office (LNO) came out with a new and improved version. In the late 1990s we prototyped the Data-Table-of-Contents Data Catalog, which served as the primary LTER data catalog until the implementation of an LTER Metacat in 2003. We see exciting new capabilities that are facilitated by Ecological Metadata Language (EML) and have developed EML-based tools for facilitating the reading of data into statistical packages and plotting geographical locations. We also have significant expertise with managing image data.

2007

Site Byte content can include any information about the site and typically reflects activity related to the information system at your site during the past year.

Please also include in your site byte:

A description of IM projects planned for the upcoming year.

Andrews Forest SiteBytes 2007

LTER Site: Andrews LTER (AND)

Contributor: Suzanne Remillard, Don Henshaw, Theresa Valentine (Jul 26, 2007)

Site Byte:

1. Annual Site Byte

After spending 3 days backpacking in Rocky Mountain National Park, following last year’s meetings, the IM team was rejuvenated and ready to move ahead with the next year of IM activities. The main focus has been a planning process in conjunction with our LTER6 proposal submission (due Feb 2008). Planning activities have included significant interaction between the IM Team and site PIs with the objective of developing a strategy for keeping pace with the large IM workload. The strategy includes streamlining certain IM processes, better defining the responsibilities of PIs and IMs with respect to providing and posting data online, better defining the bases for assigning priorities to individual data sets, and actually setting major priorities for this proposal writing year. One of our streamlining strategies is the revamping of our QA/QC process for our climate data with the goal of being more efficient and reducing the hands-on need for data processing and graphical checking. A strategy for improving the data submission process includes improving guidelines for data submission and improving our web administrative interface to allow PIs to enter and edit their upper level metadata (abstract, coverage, methods, etc.). Additionally, the Andrews Team received a $25K planning grant and will conduct a fall workshop to consider planning for increased streaming sensor data and associated Q/C, larger bandwidth and WiFi or WiMax capability, and considering long-term personnel needs.

---

2. Planned IM Projects

a. Improve the efficiency of data processing and Q/C for our climate, streamflow, air shed and other planned sensor networks
b. Develop a proposal for improving telemetry to accommodate new, streaming data sets,

establishing WiFi or WiMax throughout the Andrews, and increasing our bandwidth for communications
c. Review all data and metadata and set priorities for updating, establishing online, improving the metadata content, and improving EML
d. Improve the data submission process through 1) better defining the responsibilities of the PI and IM, and 2) improving the web interface for managing and editing metadata to better accommodate site researchers
e. Create GIS web services using ESRI ArcGIS Server technology to provide remote web access to Andrews GIS data by researchers/staff at the Andrews site

Arctic LTER SiteBytes 2007

LTER Site: Arctic LTER (ARC)

Contributor: Jim Laundre (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

This past year Arctic LTER updated its web pages and continued work on implementing our new Excel metadata entry form (http://ecosystems.mbl.edu/arc/metadata_forms/MetadataBlank.xls [156]). Since most ARC-LTER investigators use Excel this form allows the metadata and data to reside in the same workbook. VBA macros are then used to output EML, rich text, ASCII and html files for the web.

Our site review occurred this past June. I was not able to be at the field site during the review but reports back indicate that the site review team had a good time and that many useful discussions were held. The report will be out soon.

Several new large projects will provide opportunities for enhancing the data information available for the Toolik Region:

- International Polar Year (IPY): Collaborative Research on Carbon, Water, and Energy Balance of the Arctic Landscape at Flagship Observatories and in a PanArctic Network (AON). This project will provide landscape-level data and analyses of terrestrial carbon, water, and surface energy. An integrated terrestrial carbon/water/energy database for use in modeling and synthesis activities will be developed. Proposed is the development or extension of three databases: (1) the Toolik/Imnavait database (already a part of the Arctic LTER database, (2) a Cherskii database (to be developed and archived with the Arctic LTER) and (3) the PanArctic database (to be developed in partnership with our collaborators, archived with the Arctic LTER).

- Toolik Field Station is the tundra candidate for The National Ecological Observatory Network (NEON).

These projects will expanded the environmental baseline datasets and provide exciting opportunities for science at Toolik. The task of integrating the databases will both be challenging and rewarding.

Unfortunately because of family commitments I will not be able to attend this year’s meeting. The IM team member who was planning on attending also had to cancel.

_____

2. Planned IM Projects

1) Add new and update existing datasets.

2) EML implementation:
Continue work on the Excel metadata worksheet entry form. More documentation is need and taxonomic coverage and protocols sections need to be added in the VBA macro that output EML files.
Our legacy metadata has been converted to Ecological Metadata Language (EML) but only at EML Best Practices level 2/3 (no attribute EML). In bringing the files up to level 4, the files will be reviewed and where appropriate consolidated into multi-year files. Differences in methods and personnel will require that some years remain separate.

3) Train of new ARC IM team members in metadata and data methods.

4) Integration of non-LTER data such as AON (see above in section 1).

5) Evaluate relational databases for use with some of our long-term datasets.

Baltimore Ecosystem SiteBytes 2007

LTER Site: Baltimore Ecosystem Study (BES)

Contributor: Jonathan Walsh (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

We underwent our mid-term review. Our report was for the most part very good. With regard to information management the report indicated we should be investing more but other than that the system was met with praise.

We are “adopting” our metadata system. The Open Research System (ORS) which we have utilized for years as our metadata repository and clearing house has run into funding shortfalls. Up until now it has been hosted by Dr Charlie Schweick’s public policy GIS lab at the University of Massachussetts Amherst. In order to continue using it, we have installed the requisite hardware – a server – and software – ColdFusion, Server 2005, SQL server 2005 – at the Institute of Ecosystem Studies and we are currently porting the existing code and data onto it.

We have continued to work with the synthesis of demographic and social data and physical data.

We now have close to seven years of stream chemistry data online and available to the public. These data represent several sites along the urban rural gradient of the city of Baltimore.

_____

2. Planned IM Projects
We plan to facilitate a means to combine geodatabases. This will be accomplished by means of the ORS metadata system. The metadata generated by these geodatabases – in Arccatalog – will be synthesized so the geodatabases can be polled in larger searches.

We plan to increase bandwidth of our sensing networks and move toward more wireless solutions.

We plan to collect data as part of educational outreach and involve high school students in the design and planning of the sampling and collection efforts.

We plan to improve our online databases used in collection (rather than dissemination) of data. We plan to increase the number of fields collected, adopt wireless connectivity to the data, and facilitate more streaming data collection.

Jonathan M Walsh, August, 2007

Bonanza Creek SiteBytes 2007

LTER Site: Bonanza Creek LTER (BNZ)

Contributor: Brian Riordan (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

The BNZ IM group has been very busy the last year as we prepped for our site review that occurred in June. Some of our major accomplishments are:
· Level 4-5 EML for all of our data files (Thank You Inigo!!!)
· Versioning for our EML that adjusts as metadata changes
· 10 hourly streaming weather stations

We are in the process of implementing field data loggers to help stream line our core datasets into our database. This will help to improve our QA/QC as well as allow us to make ecological observations as they occur.

Our site review went well and we are feeling rejuvenated by many of the responses from the review committee. Thank you Kristin for taking the time to come up to AK.

We are in the process of a Data Manager change as Brian has left for the Minnesota Dept. of Natural Resources to pursue GIS/Remote Sensing. We hope to have a new IM hired by the end of August.

_____

2. Planned IM Projects

a) Rebuild our Intranet to allow our Pis the ability to adjust their datasets as well as add new ones.

b) Rebuild the way we maintain site information (super sites, sub-sites, ect) This could possible entail a large GIS project

c) Continue to develop web applications for data that is stored in the database for on the fly analysis

d) Explore historical data files and metadata to ensure that the information is correct and robust.

Brian Riordan

California Current Ecosystem SiteBytes 2007

LTER Site: California Current Ecosystem (CCE)

Contributor: Karen Baker (Sep 13, 2007)

Site Byte:

1. Annual Site Byte

There were two major foci this year: growth of the information infrastructure and launch of a queriable site information system. In terms of physical infrastructure, a second server (iSurf) was deployed providing increased storage together with isolation of web and collaboration services and also providing an LDAP server for single sign-on authentication. In terms of organizational infrastructure, department arrangements for a computational recharge facility were supported as a growth strategy central to cross-project infrastructure and resource sharing. Our physical work area was moved from a long-time laboratory to a more compact Design Studio, centered on a refurbished design table and posting spaces that serve as memory traces. In terms of conceptual infrastructure, we continue to develop the tie between theory and practice through the framework of Ocean Informatics. In terms of social infrastructure, an informatics team approach coalesced to include KBaker, MKortz, JConners, LYarmey, JWanetick together with part-time associates and stronger ties with other local data initiative efforts (CalCOFI and SWFSC).

Throughout the year, Datazoo, an online information system, was designed and developed, combining a MySQL backend with an object-oriented, PHP-based web frontend. A data link and a site data policy were added to the CCE website data page and connected to the information system. Data practices that included the information system were discussed with site researchers and students during demonstrations of the system held throughout the summer. The system architecture addresses critical data integration issues so that multi-dataset browsing, comparison, and joins are possible. Though the core functions had all been addressed by June, subsequent recoding improved the interface and created new functionality. We expect this process to continue according to the iterative design principle that “Data-using informs Information-Managing informs Data-using”.

Population of the database began over the summer, during which the data ingestion procedure and template were streamlined. Datazoo is now a cross-site information system using an augmented EML specification where metadata describes datasets to the column level through use of attribute, unit, and qualifier dictionaries. Work on attribute naming conventions and re-naming conventions was carried out. Capacity to publish dataset metadata to the community catalogue was demonstrated by submitting our first dataset to the Metacat server this summer.

Tools were added as a feature of the information system. A grid converter, which calculates line and station given latitude and longitude (and vice versa) and uses Google maps for visual display, was given a second interface for batch mode input. Code was refactored to separate input from display functionality. A time formatting tool permits construction of time in a format required for dataset ingestion. Finally, a joining tool permits matching a dataset file against an existing eventlog dataset so that designated columns can be added to the original dataset file.

Support was provided for the second CCE process cruise in May 2007. The eventlogger was installed onboard, providing centralized activity coordination. In addition, procedures were developed for Picture-of-the-Day in support of Education/Outreach. The development of cruise web sites continued with the cruise glossary and eventlog among the first files posted. A cruise map was generated this year using the previously developed dynamic mapping tool, and the participant list was ingested using the existing standard template.

Three posters were presented at the annual IMC: Data Integration in the Decade of Synthesis, (MKortz, LYarmey, JConners, and KBaker); Environmental Data Management: Infrastructure Studies Insights (FMillerand and KBaker); and Long Term Informatics (KBaker, CChandler, AGold,

FMillerand, and JWanetick). More than a dozen Databits contributions over the year included news articles, reports, and good reads as well as an editorial and an FAQ. Work with the NSF Comparative Interoperability Project continued bringing forward the concepts of ‘enactment’, ‘articulation work’ and of ‘knowledge provinces’. A joint keynote with FMillerand at the IMC07 presented to the community the notions of roles, informatics, and sociotechnical systems. A history associate worked at SIO early summer focusing on historical data models.

---

2. Planned IM Projects

Planned projects include continuing to migrate data into the new Datazoo system and populating our augmented EML metadata schema by working through the metadata web forms with participants. We recognize this as an opportunity to elicit tacit data collection information and to facilitate category building and vocabulary development with local participants. Having added a user-friendly view of the metadata this year in addition to the xml view, we will be considering methods for download of the metadata with the data. With Datazoo, we will reconsider both project-study and cross-project relations as well as dataset creation and ingestion. Further, there will be a more extensive online help system developed in coordination with user input. Finally, as an outgrowth of working groups on Matlab and on plotting reviews held in the last year, we are planning to develop visualization in general and a plot gallery that builds from the redesign of existing media galleries in particular.

Ongoing partnerships within LTER and with Ocean Informatics Woods Hole and MIT will continue as will collaboration with Science Studies, NOAA, and international partners. Within LTER, we will continue to contribute to efforts to articulate and/or demonstrate strategies that enable site contributions to network efforts. Specifically we are working locally on units, dictionaries, quality control, semantic bridges, and learning environments so are engaged with the LTER Information Management Unit Registry Task Force and Dictionary Process, Quality Control, and Training Working Groups.

Cedar Creek SiteBytes 2007

LTER Site: Cedar Creek LTER (CDR)

Contributor: Dan Bahauddin (Jul 27, 2007)

Site Byte:

1. Annual Site Byte

IM activities at Cedar Creek over the last year have focused on standardizing datasets and migrating them to an online database. Metadata has been systematically examined in an effort to bring our content to EML level 4.

Concurrent with these activities, we are completely redesigning out website to better Cedar Creek’s outreach efforts while making information more accessible to researchers. We have also upgraded many proprietary software packages Cedar Creek had come to rely on to more accessible, efficient systems.

---

2. Planned IM Projects

1. Complete web redesign (September 20007)
2. Create procedures for EML file generation from database.
3. Collect and convert legacy GIS data and make GIS database web accessible.
4. Improve and expand metadata.
5. Make database schema EML compliant.

Central Arizona Phoenix SiteBytes 2007

LTER Site: Central Arizona - Phoenix (CAP)

Contributor: Corinna Gries (Jul 31, 2007)

Site Byte:

Site Bytes 2007

Many important changes have taken place for CAP data management. We moved all our data out of SQL server and into MySQL, which is now our main database server. In addition we have moved all data that were stored on CAP servers to storage space at ASU's High Performance Computing Cluster. The web applications hosted on our servers were moved to a virtual server in the same cluster. Currently, we are maintaining only our development server on site. The Global Institute of Sustainability has provided personal storage space as well as collaboratively accessible space for researchers and administrators. Therefore, this move affected the entire personnel at GIOS which implies major efforts in user support.

The CAP website saw many additions this year. Several password protected areas were added for data entry, and the regular management activities for data provided to the LTER Network databases. The data entry applications are written in php using and improving the earlier developed template. Entry applications for all long term data collection efforts at CAP are in place. These applications replace the formerly used ACCESS front ends and have already proved to be much easier to access and use by the data entry personnel. In the process the data models were streamlined, metadata tables added and attribute level metadata stored in comment fields directly in MySQL table descriptions. The earlier developed reverse engineering tool now produces valid level 5 EML automatically with minimal editing of the resulting document. The applications to manage contributions to the LTER Network databases was added in an effort to document these activities in one central location.

Public data access was changed on the CAP website. Download applications for the large long term monitoring databases were developed and table level data download was enabled via EML files. EML files now allow direct data access which may be used by other applications as well as the CAP website. Currently, a user registration and logging of downloads are developed.

Ed Gilbert, who originally developed our biodiversity portal within the Southwest Environmental Information Network (SEINet http://swbiodiversity.org/seinet/index.php [157]) as a grad student has been working for GIOS this past year as web developer. In this capacity he developed new online keying application within SEINet. The applications provide major improvements over traditional matrix based online keys like DELTA intkey. The underlying data model combines hierarchical taxonomic information obtained from ITIS with the traditional character and character state information. This allows for strict normalization of information and character inheritance within taxonomic groups which greatly reduces data entry and management efforts. In addition, this approach allows for easy integration of new species if the key is to be expanded geographically. It also provides the means to dynamically build keys based on a subset of species and makes the key

very user friendly as characters may be added and dropped dynamically depending on where in the keying process a user is.

Last year the Arizona Water Institute received funding from the Governor’s office for water related research and the development of an Arizona Hydrologic Information System. This is a tri University effort (University of Arizona, Northern Arizona University and Arizona State University) in which we are developing infrastructure for data storage, curation and access. A federated data storage system is accessed via web services which provide the data in standard format modeled according to CUAHSI’s standards. One full time programmer and two computer science grad students are now working at GIOS to develop a metadata editor and management system, web services accessing groundwater data from the Arizona Department of Environmental Quality and data in the SRP flood warning system, and a search engine for data sets.

The School of Sustainability is growing in leaps and bounds. Five new faculty positions have recently been added through an endowment to the School. Charles Redman, has given up the directorship of GIOS to fully concentrate on running the school and ASU’s former Vice President of Research, Jonathan Fink, has taken over as director of GIOS.

---

2. Planned IM Projects

online EML editor
data set search engine
onine specimen keying application

Coweeta SiteBytes 2007

LTER Site: Coweeta LTER (CWT)

Contributor: Barrie Collins (Jul 19, 2007)

Site Byte:

1. Annual Site Byte

I think that the comments below serve as a pretty nice site byte...Thanks, Barrie.

---

2. Planned IM Projects

2007 has largely been about preparation for Coweeta LTER's 2008 review. Thus, our focus has been oriented towards filling gaps in our data, publications, and metadata.

a. Improvement of Data metadata. (Thanks to John Porter for helping get PI's motivated).
b. Working with the redoubtable Inigo San Gil to improve metadata in EML format to a higher plane.
c. Classification of 2006 imagery from Ga to Virgnia to fit with our current coverage on five year increments back to 1986 or so.
d. Support of EcoTrends Demographics.

Point a especially is grueling work, as it's one thing to write: Improve metadata. It's another thing to be the combination of Sherlock Holmes and wolverine required to get the job done...but the payoff

has been a more professional site and I say this with full sincerity: I'm glad we're biting the bullet and doing it right.

Florida Coastal Everglades SiteBytes 2007

LTER Site: Florida Coastal Everglades (FCE)

Contributor: Linda Powell (Jul 27, 2007)

Site Byte:

1. Annual Site Byte

As the Florida Coastal Everglades LTER program heads into its second round of funding (FCE II), the IMS team (Mike Rugge and Linda Powell) is in the process of completely redesigning the FCE website to reflect its new research phase and expects to launch the new website sometime this fall. The new website was designed to be more graphical and user-friendly and will hopefully capture the interest of the 'general public' in addition to facilitating LTER research. FCE program information is now organized with a series of tabs that give a 'file folder' look to the web page and once a tab is selected, the user will find several related subcategories. Many new features have been added to enhance a user's visit to our website such as 'What we do', 'About the Everglades', and 'Featured Movies'. If you're interested in taking a look at the new website, please see my FCE website demo at the upcoming IM meeting in San Jose.

Although the IMS team has made many changes to the FCE website, two important sections have been preserved and enhanced: 1) FCE Projects and 2) FCE Sampling. Following the new organizational template, the 'Projects' section is organized specific projects and their related information using a series of tabs that give a 'file folder' look on the web page (http://fcelter.fiu.edu/research/projects/projects.htm?pid=14 [158]). For any given project, a user can easily find the abstract, research sites, personnel, sampling attributes, datasets and publications related to that specific project all on one web page. Additionally, projects can be searched by keyword, researcher or funding organization.

The 'Sampling' application (http://fcelter.fiu.edu/research/projects/sampling.htm [159]) is extremely useful in information and project management. Users are able to search for sampling attributes by entering keywords or manually selecting an attribute of interest from a list. Once the selection has been made, a map with a series of tabs give the web user access to research site information, dataset listings and project information related to their sampling attribute selection. This feature greatly facilitates project management and site science as users are able to easily access related information from one portal. For example, researchers interested in porewater salinity can enter the sampling attribute in the keyword field and the results will return a map of the FCE research sites where porewater salinity is collected as well as tabular links to all datasets containing porewater salinity and projects responsible for the collection of that particular attribute.

This coming year, the IMS team will focus on building a web-based query interface for FCE physical and chemical data values. In conjunction with the query capabilities, a graphing application will be added so that FCE researchers can graph their query results in real-time. Allowing users to query, download, and graph data for specific sampling attributes across all FCE data files will be a welcomed addition to the FCE website.

---

2. Planned IM Projects

A. Finish FCE website redesign and launch by late Fall, 2007

B. Create synthetic physical and chemical data tables (and appropriate related tables) in the FCE Oracle DB.

C. Build a web-based query interface for FCE physical and chemical data values and add a web-based graphing application.

D. Create FCE ArcGIS data.

Georgia Coastal Ecosystems SiteBytes 2007

LTER Site: Georgia Coastal Ecosystems (GCE)

Contributor: Wade Sheldon (Jul 25, 2007)

Site Byte:

1. Annual Site Byte

Our major focus this past year has been transitioning from GCE-I to GCE-II, which began in January 2007. Our GCE-II study plan emphasizes marsh processes and landscape-scale ecology, so we added a major GIS component to our project to support this research. We began by hiring a full time assistant IM / spatial data manager (Kris Meehan) to develop GIS resources for the project and assist investigators and students with GIS analyses, as well as provide backup and support to our lead IM (Wade Sheldon). We also greatly expanded our IT infrastructure to accommodate large volumes of geospatial data. For example, we added a high speed GIS workstation, 2 TB spatial data server (running ArcGIS SDE 9.2 with MS SQL 2000) and LTO-3 tape backup system, and we have just ordered an updated web server with 1 TB RAID to support hosting of GIS data and imagery on the web. We also maintain a GIS workstation and file server at our field site, as well as 50-seat ArcGIS license servers both at Sapelo and UGA Marine Sciences.

We also acquired a high precision field GPS unit (Trimble GeoXH, with subfoot accuracy) and post-processing software this spring, and are finally starting to acquire good geo-location information for our land-based sampling sites and plots. Hydrographic sites are already geo-referenced using high-precision GPS on research vessels, so now all our research locations can be properly registered in the GIS. Kris has made major progress in acquiring maps, imagery, and GPS data relevant to our study area and organizing them into documented geodatabases; unfortunately, she will be leaving the project this fall for a job in Iceland so we are about to begin the hiring and training process all over again.

These major additions to our IM program (as well as moving into a new office at UGA) have dominated work this year, but we did make progress in a few other areas as well. Don Henshaw, Ken Ramsey and I received post-ASM funding to conduct a workshop on quality control for derived data products, which was held at JRN in conjunction with a Trends editorial meeting in February 2007. Several other LTER IMs attended, as well as representatives from LNO, CUAHSI, SEEK and the Canopy Databank Project. Follow-up work is planned during this IM meeting in San Jose, and I will also be demonstrating new Q/C capabilities in the GCE Data Toolbox software inspired by discussion in the workshop. I also collaborated with Barrie Collins to provide access to near-real-time USGS streamflow data on the CWT web site, as well as automatic data harvesting for HydroDB. Similar capability could also be added for other sites very quickly, with results "skinned" using html templates to provide appropriate look and feel, so let me know if anyone is interested in a similar collaboration.

The other major project we undertook this spring was to redesign the GCE web site. We needed a more flexible menu system as well as better support for dynamic content and user contributions. We also wanted to simplify maintenance, update the code for better W3C standards compliance, and incorporate more of the LTER web design group recommendations. We finished developing a prototype framework early this summer, and we are currently gathering feedback and moving our existing content to the new site. We will launch the new site in conjunction with deployment of our new web server this fall. Planned additions to this site are listed below.

_____

2. Planned IM Projects

a. finish migration of legacy content to new web site (designed in summer 2007)
b. develop web forms to allow GCE participants to directly upload bibliographic citations, reprints, project announcements and documents to the database
c. implement a "current conditions" page on the GCE web site for integrated display of tide predictions and near-real-time weather and hydrographic plots (with access to documented data sets)
d. expand GCE data catalog coverage to include ancillary data holdings (to provide EML support and integrated searching)
e. add support for multiple data set selection and zip archive downloads to data catalog (i.e. shopping cart metaphor)
f. add search and browse interfaces to data catalog
g. extend GCE data catalog to include new geospatial data sets (with EML metadata)
h. implement map query interface (Google maps) to data catalog

Harvard Forest SiteBytes 2007

LTER Site: Harvard Forest (HFR)

Contributor: Emery R. Boose (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

Activities over the past year included:

1. Design and initial field testing of a wireless network to reach experimental sites in the forest.
2. Installation of optical fiber and wiring for a new building.
3. Upgrades and reconfiguration of Harvard Forest servers.
4. Creation of attribute-level EML for most datasets using Morpho.
5. Continued collaboration with computer scientists at UMass Amherst on analytic web project.

_____

2. Planned IM Projects

1. Installation of 100 Mbps optical fiber connection to university through Verizon (to replace T1 line).
2. Completion of wireless network design and submission of proposals for funding.
3. Completion of level 5 EML for all online datasets.
4. Implementation of new policy requiring researchers to update data & metadata annually before granting permission to conduct or continue field work.
5. Continued collaboration with computer scientists at UMass Amherst on analytic web project.

Hubbard Brook SiteBytes 2007

LTER Site: Hubbard Brook LTER (HBR)

Contributor: John Campbell (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

In March, 2007 we released a new redesigned Hubbard Brook webpage (http://www.hubbardbrook.org/ [160]). The new webpage is a vast improvement over the old one and includes features such as better search capabilities, image archive with an upload feature, curricula vitae that can be edited by investigators, password protected intranet site, near real-time data graphing, and more.

Last summer we received an NSF supplement to install a wireless sensor network at Hubbard Brook. A repeater station was erected which made the entire Hubbard Brook Valley accessible with 900 Mhz radios. We are currently collecting real-time data from 5 locations and are planning to expand the network over the coming year. An example of the online graphical interface is posted at http://www.hubbardbrook.org/data/Realtime_Data/index.php [161].

We recently received another NSF supplement this year to improve the physical sample archive database. In the coming months we will post the database online and make a number of other enhancements.

In June, we had a NSF midterm site review. This was a positive experience and the reviewers offered a number of constructive ideas for improving information management at the site that we have begun to implement.

_____

2. Planned IM Projects

Develop QA/QC for near realtime data

Develop XML database for EML

Put database for physical sample archive online

Improve method for tracking research projects

Jornada Basin SiteBytes 2007

LTER Site: Jornada Basin (JRN)

Contributor: Ken Ramsey (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

Information Management

Ken Ramsey and Barbara Nolen have been planning on how to effectively integrate GIS data layers and research site locations with research data and associated metadata to enhance the quality and availability of all JRN data and to generate more detailed and precise EML documentation.

Barbara continues to provide GPS of research site locations and production of new GIS layers, to provide GIS and GPS support to researchers and students including training and map production, and to administer NMSU site licenses for GIS/RS software.

Ken is currently populating the JIMS database with research project and associated dataset metadata to support EML generation for all Jornada Basin LTER (JRN) datasets as well as the new dynamic data catalog and data cart that will enforce the JRN data policies for download notification using a user registration system that is in beta testing. This process includes validating data structure consistency between data and metadata.

Database Server

The USDA ARS Jornada Experimental Range (JER) purchased a new database server and the JIMS' databases and Trends administrative databases have been migrated to the new server, which required upgrading from Microsoft SQL Server 2000 to 2005. During this process we discovered that Xanthoria (CAP LTER product) would not work on SQL Server 2005. This required that we develop a customized EML creation solution from metadata stored with JIMS. New style sheets have been developed that generate EML level 3 from the JIMS database.

Integrating GIS Data, Metadata, and Services

The GIS metadata in xml will be used to generate EML documentation for all JRN GIS layers stored within JIMS, which includes geographic bounding coordinates within EML documents describing research datasets.

JRN purchased a new server to support ESRI ArcGIS Server and ArcGIS Image Server software to allow JRN researchers to query and access the JRN data, GIS layers, and imagery. Ken and Barbara attended the Annual ESRI International User Conference in San Diego, CA to gather technical information necessary to develop and serve the GIS services on the new GIS Server. The new GIS Server will support the following GIS Services:

- query and access to aerial photograph archive
- access research site locations (Intranet)
- online shapefile production to support research site location selection and approval
- Jornada Interactive Map (2D) on website
- Jornada Interactive Map (3D) using ArcGIS Explorer

Deployment of ESRI ArcGIS Explorer is being evaluated to support interaction with the globe services and research site locations and their associated data and metadata as well as GIS thematic layers such as vegetation, soils, and base maps.

Jornada Website

The new web server (purchased by JRN) has been installed and configured to support the new Jornada website currently in beta testing. The new website will include a new XML-based data catalog and data cart that will enforce the JRN data access policy by requiring user registration and authentication prior to download of JRN data. Users that download data will be required to supply an intended use statement upon download. Users will provide contact information, affiliation, and acknowledgement of the JRN data policies when they register with JIMS. When the beta testing of the new website is completed, the new server will be renamed and replace the web server that currently serves the Jornada website.

---

2. Planned IM Projects

1. Data:
a. Production of Chihuahuan Desert Landforms map
b. Update Desert Project map data
c. Review and update metadata for legacy research datasets
d. Upload legacy dataset files into databases (JIMS and ArcSDE) and correct any problems discovered while moving from ASCII text files to RDBMS
2. EML:
a. Generate EML level 3 for all Jornada datasets including GIS layers (soils, vegetation, etc.)
b. Enhance style sheets used to create EML to generate EML level 5 (attribute information).
c. Implement and enhance the esri2eml style sheet to create EML for GIS layers and incorporating research site location bounding coordinates within research dataset EML documentation
3. JIMS:
a. GIS Services:
i. Incorporate vector drawing functionality to website to give prospective researchers at the Jornada the ability to draw proposed research site locations on the Jornada base map to aid in selection and evaluation of impact of proposed research on nearby ongoing/historic research efforts for approval/disapproval.
ii. Spatial and tabular query interface to aerial photograph archive
iii. Information management notification system
b. Develop research notification administrative and user interfaces
c. Develop automated dataset tracking system
d. Implement QA rules for dataset attributes within metadata and enforce when uploading research data to JIMS database
e. Consistently document and implement QA rules and flags within JRN research dataset documentation and JIMS databases and interfaces
4. Systems and networking:
a. Upgrade to ArcSDE 9.2 on Database Server
b. Upgrade to ArcIMS 9.2 on the Map Server.
c. Deploy development server as new web server to add new data catalog and cart functionality
d. Extend wireless coverage in the field
e. Increase storage capacity for imagery using SAN/NAS solution

Kellogg Biological Station SiteBytes 2007

LTER Site: Kellogg Biological Station (KBS)

Contributor: Sven Bohm (Aug 02, 2007)

Site Byte:

1. Annual Site Byte

This year was a mid term review for us and we redesigned our website which was an opportunity to redesign the dynamic part of the site. I implemented the data catalog, directory and publication parts of the site using the Ruby on Rails web framework, which provided me with a testing framework so that it will be easier to evolve both the database(s) and the website.

---

2. Planned IM Projects

Write data upload applets to help researchers do a first pass quality control of their data.

Konza Prairie SiteBytes 2007

LTER Site: Konza Prairie LTER (KNZ)

Contributor: Jincheng Gao (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

The main focus of Konza IM activities this past year have been the QA/QC checking of our datasets and metadata, updating of GIS coverages and spatial metadata, and web site design. We generated window-based interfaces to convert text file data into a format for uploading into our SQL Server database and to perform initial QA/QC checking based on data ranges within specified minimum and maximum values. We also updated the historical Konza Prairie climate data and reharvested into CLIMDB to effectively recover missing data. We also performed a quality check for our metadata database in the SQL Server, which is dynamically connected with our web site. The geospatial datasets were updated and checked for accuracy with GPS unit. Spatial metadata were updated and were converted into EML. The Konza geospatial data were stored in shapefiles and arcinfo files, and as well as in a spatial feature format which is managed with ArcSDE and SQL Server. A spatial database of Konza LTER research activities (locations of plots, sampling transects, experimental infrastructure) was generated and added to interactive web sites managed with ArcIMS v9.2. The new Konza LTER web site was redesigned, and more functions which are dynamically connected to the metadata database were added.

In the past year, we also updated our IM hardware systems to support better data service. We expanded the capabilities of our database server and added one more SQL Server instance to accommodate the large volume of GIS data and images associated with the development of several new databases. We also replaced our old web server, which served us more than 9 years, with a new server with faster speed and more storage space. The new server has an Intel CPU of 3.0 GHZ and a 3.0 GHz and 2.0 GB RAM. Our new server will support our expanding datasets and the increased demands of our new GIS server.

_____

2. Planned IM Projects

a. Continue with QA/QC checking of our LTER datasets and metadata, and add spatial attributes and methodology in EML metadata based on EML Best Practices.
b. Improve our dynamic web sites with integration of our metadata database in order to make our web site more flexible and more convenient for users.
c. Create GIS services to manage Konza geospatial data and the geospatial data of several related projects, and harvest spatial EML in Metacat.
d. Work with Konza LTER PIs as we prepare for our next LTER renewal proposal.

Luquillo Experimental Forest SiteBytes 2007

LTER Site: Luquillo LTER (LUQ)

Contributor: Eda C. Melendez-Colom (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

In this last 12 months the LUQ Information Management (IM) was dedicated to three types of projects and tasks: the EML further development of LUQ metadata, improving the services given to the LUQ scientists and outreach type projects.

• Once the EML level 3 – 4 for all the LUQ databases were achieved, we started the path to achieve the attribute level (5). We received a second visit from Iñigo whose assistance has been crucial to this task. We realized that further QA/QC was needed on the attribute-related metadata items such that the EML metadata could be suitable for web services processes. We decided that although the process to achieve level 5 will be longer, we will assure the quality of these metadata before producing new level 5 EML packages. Nineteen percent of the metadata files has gone the necessary QA/QC to be EML level-5 ready and the transitory files to produce their EML packages have been delivered and EML-proved by Iñigo's perl scripts for LUQ's metadata.

This year, four new databases were published on our web site (http://luq.lternet.edu/data [162]) and four additional databases' metadata whose data dissemination are still restricted, were published on the web site. All of the new complete published databases have gone the QA/QC processes.

• Special attention was given to assist the LUQ investigators in their synthesis-related tasks. A private Plone site was installed and sectios were made accessible to different science projects communities within the site; the Canopy Trimming Experiment, the Synthesis Book, and the Schoolyard projects' groups were given special groups' access and rights to share documents and information and to make collaboration easier. A section was established to give access to all the LUQ PIs to long-term databases and publications. The interesting thing about these last two sites was that it was an idea created and promoted by two of the LUQ PI's.

• We further developed a perl script that handle online input forms. The script was modified to send the data to a readily-database-importable ASCII-file, to produce a screen that displays the just entered information to the user's screen on the fly, and to add the URL of the updated ASSCII file to the email that the original script sends to the person that is defined as a recipient. With few modifications done to the script it can process any kind of input in the same way.

The first version of the script handles 4 different forms designed to get group, personal, and project information from groups visiting the El Verde Field Station (http://ites.upr.edu/EVFS/reservations.htm [163]). The second modification of the script was designed to get LUQ LTER graduate students' personal and thesis project information (http://luq.lternet.edu/people/StudentdPers/PersonalDataEntryForm.html [164]).

_____

2. Planned IM Projects

• Further develop LUQ metadata database by:
o Finish exporting LUQ metadata to an EML level 5 with attribute data that will make the LUQ's metadata ready to be processed by present and future web services.
o Re-structuring the metadata database to accommodate the new uses given to the metadata such that the process of producing EML packages or any other kind of metadata products can be automatized (that is, re-structure LUQ metadata database to a more modular structure).
• Change the LUQ LTER site to a more dynamic, databased web site that will facilitate the installation of effective searching mechanisms for the users.
• Create scripts that will automate the input of metadata information online. Two kinds are to be produced:
o Script(s) that process one set of forms that get the direct metadata input of the user and stores it in an online database (preferably MYSQL)

o Other scripts that accept the LUQ LTER MS Word metadata standards (http://luq.lternet.edu/datamng/imdocs/dsetallfrm.doc [165]) process it and get the metadata information into the online metadata database (preferably MYSQL).

McMurdo Dry Valleys SiteBytes 2007

LTER Site: McMurdo Dry Valleys (MCM)

Contributor: Chris Gardner (Aug 10, 2007)

Site Byte:

First of all, I apologize for the lateness - I thought I had submitted this several weeks ago, but I guess I wrote it and forgot!

1. Annual Site Byte

Information Management at MCM over the past year has focused on improving our GIS capability, increasing data redundancy, better tracking analytical samples in the field, and preparing for our site review in January, 2008 in Antarctica.

Many GIS layers in our database were cleaned up and properly attributed, and protocols were developed for handling spatial layers. Several new spatial layers were created, including a DEM and bathymetric lake layers. Finally, an interactive online map using ESRI's ArcGIS Server technology was implemented on www.mcmlter.org [166]. This map is JSP based, and pulls layers directly from our Oracle database. The map also serves as a data portal - clicking on certain features returns actual data and metadata.

An older Sun workstation was also acquired for free. This machine is now running Oracle 10g and automatically replicates any changes in the main database every 7 days using materialized views. Rsync open source replication software was also installed on this machine, and it replicates key directories from the main server.

Chain of Custody (COC) forms for analytical chemical samples were redesigned for the 2006-2007 Antarctic field season. These forms are now inserted into the database and viewable both by users of Analytical Services in the field, and by team members running the samples. A page in the restricted section of our website gives users new flexibility for tracking samples and assuring data continuity from the field to the database. The COC database can be used to compare actual samples and sample names received to data submitted after the field season. This will solve many problems with missing data and incorrect sample names

_____

2. Planned IM Projects

Our main task is now to prepare for our site review in January, 2008. Will will be updating data and protocols and ensuring that we meet the review criteria for LTER Information Management Systems. Another goal is to continue to make the database more useful by offering complex web-based queries and joins on data. Finally, the online map will be expanded and GIS capabilities will be increased. We will also be creating EML for our spatial layers.

Moorea Coral Reef SiteBytes 2007

LTER Site: Moorea Coral Reef (MCR)

Contributor: Sabine Grabner (Jul 25, 2007)

Site Byte:

1. Annual Site Byte

The past year's IM activities focused on setting up an environment to post the site's long term datasets online. Datasets were described in EML varying from level 3 to 5. Posting EML metadata online through XSLT transformation was adapted from SBC. Most datasets can be downloaded automatically and downloads are registered in the database.

Our back end is a combination of disk storage and database. For insertion of manually collected data we developed database interfaces using web forms and Java applications. Maintaining the database schema is a continuous task and will become a major task in the coming year, when we will make the schema fully compliant with the used EML tags.

_____

2. Planned IM Projects

1. develop generic instrument data harvester for harvesting TCP connected instruments/loggers/fieldpcs (Java, XML, RBNB)
2. make database compliant with used EML tags
3. develop web interface to edit metadata
4. generate EML from DB

Niwot Ridge SiteBytes 2007

LTER Site: Niwot Ridge LTER (NWT)

Contributor: T. M. Ackerman (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

This year was our mid-term site review. The review process proved once again a trying but very important process. The comments from the review team were extremely helpful and will result in an improvement in our IM system.

This year also saw Niwot Ridge being selected as a Core Site for the NEON.

Our Relational Database server was upgraded from MSSQL 2000 to MSSQL 2005. The production database server (Dell PowerEdge 500SC) has been upgraded to a more powerful server (Dell PowerEdge 2600) capable of handling more requests. We migrated the existing databases designed in Microsoft SQL Server 2000 to redesigned SQL Server 2005. Currently this server is handling the generation of the EML metadata, and once version upgrade issues are resolved, will be able to allow query of the climate data. All of the meteorological data that collected via the wireless radios are imported into the current schema.

Three sites have recently been enhanced with High Throughput radios which support Ethernet as well as serial communications. The Ethernet communications are in use at the Soddie lab in order

to remotely control and manage the computer which interfaces with the instrumentation. These radios allow for troubleshooting problems quickly, previously such problems may have gone weeks without notice. The installation of these radios has saved time and energy for researchers and MRS personnel adding the ability to perform minor tasks such as changing data logger programs or rebooting upon system failure.

A wireless connection from the MRS to Boulder replacing the current T1 connectivity is in development. The current system consists of a fiber optic line from the Tundralab to the MRS and a T1 from the MRS to the world. The new system will consist of the existing fiber optic line from the MRS to the Tundralab and a wireless connection from the Tundralab to the world with a bandwidth of 3mbs. We have installed two relay stations, ground-truthed the sites, and are currently awaiting the installation of the receiver on the CU campus.

The precipitation and temperature datasets for C-1 (1952-2006), Saddle (1981-2006), and D-1 (1952-2006) have been analyzed in order to fill in missing values and correct erroneous values. Adjacent sites were used to statistically insert values. These new datasets provide more a robust ability to look at some of the longest continuous high elevation climate data sets.

_____

2. Planned IM Projects

1) Improve Site Research Application to include formal project definitions, for better data acquisition tracking.
2) Web metadata form.
3) Attempt to get data into eml file.
4) Climate data query system.

North Temperate Lakes SiteBytes 2007

LTER Site: North Temperate Lakes (NTL)

Contributor: Dave Balsiger, Barbara Benson, Jonathan Chipman (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

We have continued our development in the area of sensor networks and our leadership in the Global Lake Ecological Observatory Network (GLEON). We are collaborating with computer scientists from the University of California San Diego Supercomputer Center and SUNY-Binghamton on automating the configuration and quality assurance components of the NTL information system for near-real time data streaming from instrumented buoys. We are developing a database of lake characteristics information on lakes associated with GLEON including web forms for users to enter and edit information from their sites.

Two major upgrades to NTL-LTER data collection software were accomplished by our student programmer during the past year. Software used in counting and measuring zooplankton was made more user friendly and efficient. Software used by the fish crew for field data collection was migrated to the latest IPAQs with substantial program enhancements based on feedback from the fish crew and information management staff.

During the past year we developed complete EML metadata for the NTL spatial data sets. Inigo San Gil was very helpful with the development of a set of methods for converting ESRI-format metadata

into EML. At NTL, most spatial metadata are originally developed using ESRI's ArcCatalog utility. We then save the metadata as XML, use a stylesheet to reformat them into EML, and then use the oXygen XML editor to correct various errors and make the metadata fully EML-compliant. EML from the spatial datasets are included in the Metacat harvest.

We currently provide access to the NTL spatial data through several interfaces -- a "static" data catalog for downloading individual files, a set of interactive ArcIMS map applications, and an Oracle/ArcSDE geodatabase for direct interactive access to all spatial data. We are exploring a series of improvements to these interfaces, including investigating methods for efficient storage of and access to very high-volume (multi-gigabyte) raster images.

_____

2. Planned IM Projects

1. Design and implement an alternate search interface for the NTL data catalog to allow users to query by investigator, subject, keywords, lake name, geographic location, and taxonomic criteria.

2. Continue our progress in automating data processing of near-real time data from instrumented buoys and the NTL meteorological station including testing a new data model.

3. Develop a generic version of our custom zooplankton software that would enable microscope counting and measuring of other organisms.

4. A new seminar in information management and technology for graduate students will be offered this fall at the Center for Limnology. Organized by Barbara Benson and Paul Hanson, this course will cover topics such as database design, software options, and metadata requirements.

Palmer LTER SiteBytes 2007

LTER Site: Palmer Station (PAL)

Contributor: Karen Baker (Sep 13, 2007)

Site Byte:

1. Annual Site Byte

There were two major foci this year: growth of the information infrastructure and launch of a queriable site information system. In terms of physical infrastructure, a second server (iSurf) was deployed providing increased storage together with isolation of web and collaboration services and also providing an LDAP server for single sign-on authentication. In terms of organizational infrastructure, department arrangements for a computational recharge facility were supported as a growth strategy central to cross-project infrastructure and resource sharing. Our physical work area was moved from a long-time laboratory to a more compact Design Studio, centered on a refurbished design table and posting spaces that serve as memory traces. In terms of conceptual infrastructure, we continue to develop the tie between theory and practice through the framework of Ocean Informatics. In terms of social infrastructure, an informatics team approach coalesced to include KBaker, MKortz, JConners, LYarmey, JWanetick together with part-time associates and stronger ties with other local data initiative efforts (CalCOFI and SWFSC).

Throughout the year, Datazoo, an online information system, was designed and developed, combining a MySQL backend with an object-oriented, PHP-based web frontend. Data practices that included the information system were discussed with site researchers and students during

demonstrations of the system conducted in the summer. The system architecture addresses critical data integration issues so that multi-dataset browsing, comparison, and joins are possible. Though the core functions had all been addressed by June, subsequent recoding improved the interface and created new functionality. We expect this process to continue according to the iterative design principle that “Data-using informs Information-Managing informs Data-using”.

Population of the database began over the summer, during which the data ingestion procedure and template were streamlined. Datazoo is now a cross-site information system using an augmented EML specification where metadata describes datasets to the column level through use of attribute, unit, and qualifier dictionaries. Work on attribute naming conventions and re-naming conventions was carried out. Capacity to publish dataset metadata to the community catalogue was demonstrated by submitting our first dataset to the Metacat server this summer.

Tools were added as a feature of the information system. A grid converter, which calculates line and station given latitude and longitude (and vice versa) and uses Google maps for visual display, was given a second interface for batch mode input. Code was refactored to separate input from display functionality. A time formatting tool permits construction of time in a format required for dataset ingestion. Finally, a joining tool permits matching a dataset file against an existing eventlog dataset so that designated columns can be added to the original dataset file.

Three posters were presented at the annual IMC: Data Integration in the Decade of Synthesis, (MKortz, LYarmey, JConners, and KBaker); Environmental Data Management: Infrastructure Studies Insights (FMillerand and KBaker); and Long Term Informatics (KBaker, CChandler, AGold, FMillerand, and JWanetick). More than a dozen Databits contributions over the year included news articles, reports, and good reads as well as an editorial and an FAQ. Work with the NSF Comparative Interoperability Project continued bringing forward the concepts of ‘enactment’, ‘articulation work’ and of ‘knowledge provinces’. A joint keynote with FMillerand at the IMC07 presented to the community the notions of roles, informatics, and sociotechnical systems. A history associate worked at SIO early summer focusing on historical data models.

---

2. Planned IM Projects

Planned projects include continuing to migrate data from the original PAL data system to the new Datazoo system and populating our augmented metadata schema by working through the metadata web forms with participants. We recognize this as an opportunity to elicit tacit data collection information and to facilitate category building and vocabulary development with local participants. Having added a user-friendly view of the metadata in addition to the xml view, we will be considering methods for download of the metadata with the data. With Datazoo, we will reconsider both project-study and cross-project relations as well as dataset creation and ingestion. Further, there will be a more extensive online help system developed in coordination with user input. Finally, working groups on Matlab and on a plotting review held in the last year. We are planning to develop visualization in general and a plot gallery that builds from the redesign of existing media galleries in particular.

Ongoing partnerships within LTER and with Ocean Informatics Woods Hole and MIT will continue as will collaboration with Science Studies, NOAA, and international partners. Within LTER, we will continue to contribute to efforts to articulate and/or demonstrate strategies that enable site contributions to network efforts. Specifically we are working locally on units, dictionaries, quality control, semantic bridges, and learning environments so are engaged with the LTER Information Management Unit Registry Task Force, Dictionary Process, Quality Control, and Training Working Groups.

Plum Island Ecosystems SiteBytes 2007

LTER Site: Plum Island Ecosystem (PIE)

Contributor: Hap Garritt (Jul 23, 2007)

Site Byte:

1. Annual Site Byte

The Plum Island Ecosystems (PIE) LTER has been working diligently the last year at upgrading our EML metadata from Level 3 to Level 4+. We feel complete attribute level information is a critical component of metadata as most researchers want the actual research data sets. We are currently redoing existing Level 3 EML by reentering our MS Word metadata into a modified version of the ARC LTER Excel based metadata form developed by Jim Laundre. The Excel form is relatively straightforward with metadata in Sheet 1 of Excel and data in Sheet 2 thus allowing the metadata and research data to always “travel” together. With the help of a site visit by Inigo San Gil in February 2007, we are also making progress on generating EML compliant metadata for our ArcGIS data. Generation of EML compliant metadata dominates most IM activity; hopefully this effort will lessen as our legacy metadata reaches Level 4+ status.

_____

2. Planned IM Projects

1) Generate EML level 4+ metadata for all data sets.
a) all non spatial data
b) ArcGIS spatial data
c) IDRISI (Clark University) GIS spatial data
d) RIVERGIS (University of New Hampshire) GIS spatial data
2) Redesign PIE web site
a) Redesign skin template
b) Develop usable GIS web page
c) Develop usable Photo/image web page
3) Continue generation of discrete TRENDS datasets

Santa Barbara Coastal SiteBytes 2007

LTER Site: Santa Barbara Coastal (SBC)

Contributor: Margaret O'Brien (Jul 20, 2007)

Site Byte:

1. Annual Site Byte
Most of the past year's IM activities at SBC have been laying the groundwork for the tasks listed in Part 2.
Metadata: work on a design for a metadata database which can be used by both SBC and MCR; work with PIs to more completely and correctly describe sampling locations and parameters. Start work on a prototype data portal and gather feedback from user groups.
Integrate data processing with IM: assist scientists in a redesign of hydrologic data processing using matlab to lessen the dependence on excel spreadsheets. We plan to create EML packages as a

data processing step.
EML data table query tool: Create an generic application for querying tables which are described in EML (see Spring 07 issue of Data Bits).
In addition, the SBC website was redesigned, adding dynamic menus and incorporating several features recommended by the network. The new design uses templates and libraries in a flexible framework that allows existing pages to be edited and new pages to be added easily.

_____

2. Planned IM Projects
A. Create an interface to browse SBC's data catalog by sampling site and parameter
B. Finish and install the EML data table query application
C. Populate metadata database
D. Write matlab m-files to output EML as a processing step, with metadata drawn from db

Sevilleta SiteBytes 2007

2007

Short Grass Steppe SiteBytes 2007

LTER Site: Shortgrass Steppe (SGS)

Contributor: Nicole Kaplan and Bob Flynn (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

The SGS-LTER Information Management System archives extensive amounts of data, metadata, and other information. We have made significant revisions to our relational database management system to more efficiently organize, relate and deliver information, data and metadata. Our new schema also contains metadata content to generate level 5 EML, which defines each attribute within each dataset and the associated measurement units or code definitions. Implementation of EML level 5 will begin with long term studies consisting of 30 datasets, (15% of our entire database) which supports both legacy and current studies. Perl scripts have been developed to generate EML from our database and store it on our web server for harvest by the LTER Network metacat catalog. This level of metadata and our participation in the communitywide metacat will facilitate data discovery, data integration and synthetic research.

We are contracting with a web development team at our home institution, Colorado State University, to implement our newly designed website that queries, joins, and publishes information from the revised database. The benefits of using the database to drive the website content include delivery of information that is more dynamic, connections between different types of information within the relational database management system, and increased ease of use for our end users with the latest web features. An example of related information that can be delivered within just a few clicks of the mouse includes a publication citation with the supporting data and the authors’ e-mails. The SGS-LTER website is large in scope, presenting an impressive amount of information that requires unique tools compared to other projects on the CSU campus. The tools we will build can be maintained and expanded by SGS-LTER Staff and may be passed along to other LTER sites. Implementation is planned for Fall of 2007.

We are also providing data, metadata and expertise to integrate datasets with other LTER sites. For example, researchers working at the SGS, SEV, and JRN, as well as the TRENDS project are interested in determining drivers of Net Primary Productivity (NPP). The Grasslands Data Integration Project (GDI) brings together ecologists, computer scientists, and information managers to address the challenges of integrating diverse datasets and produce analyses from the integration product. Unfortunately, integrating NPP data over time and over regions is not a straightforward problem. Issues of ecological synthesis and data integration arise from questions of methodology (e.g., sampling, plot sizes, etc.), data semantics (e.g., changing taxonomic information), and straightforward but messy data transformations. These datasets represent a core area of research within the LTER community, and an integrated databank would create a powerful resource for ecologists from several sites to perform cross-site analyses.

Our staff provides support for SGS-LTER researchers and students in various aspects of GIS including gathering data with GPS equipment and imagery, assisting with GIS model development for their particular research, and providing GIS data and maps for field work and modeling. We have also extended existing GIS programs for analysis of SGS-LTER data.

We also continue to participate in planning future activities for the LTER IM committee. Nicole serves as co-chair of the IM Committee, which establishes priorities and strategies to facilitate network-level and synthetic research that requires greater integration of information and interoperability of systems.

Virginia Coast SiteBytes 2007

LTER Site: Virginia Coast Reserve (VCR)

Contributor: John Porter (Aug 01, 2007)

Site Byte:

1. Annual Site Byte

Another busy year at the Virginia Coast Reserve LTER project! We're greatly enjoying our new home at the Anheuser-Busch Coastal Research Center. Its modern, climate-controlled laboratory and dormitory facilities are a giant improvement over the renovated farm house it replaced. However, based on how busy it's been this summer, we may need to expand it soon!

On the IM-front this has been a year for consolidating some activities and improving user interfaces for data access. In particular, we've improved our local data catalog by changing a long list into a matrix format that can be sorted by a variety of criteria, including core area and popularity. We also did some major work on our database of locations (for datasets, species observations etc.), dealing with a large number of locations that were named and described, but had no coordinates. To avoid such problems in the future a new interface for identifying locations was developed. It uses the Google Maps API v.2 (following testing to make sure that the georeferencing was sufficiently accurate in our area) so that a user can use the standard Google Maps controls to zoom and pan the map, then click on a location to record its coordinates into our location data table. Coupled with that new facility is an interactive map (again using the Google API) that plots all locations for VCR/LTER datasets. Clicking on a location pulls up a list of all the datasets associated with that location. Currently the system only supports points and bounding boxes, but we look forward to adding polygon capabilities in the future.

We have also worked on enhancing data systems for important datasets. Traditionally these have been handled using user-designed spreadsheets. However, in some cases these were inconsistent

in format and highly variable in the degree of quality control and assurance. For our water quality dataset we have developed an operational system that incorporates spreadsheet forms (including both data and “flags”), a PERL ingestion program that resolves relational issues that the spreadsheet handles poorly, a MySQL database, R-scripts that conduct standard QA analyses, and data editing capabilities using phpMyAdmin and an ODBC-linked Access database. A web page provides a central location for uploading spreadsheets, running QA analyses and for data editing. We plan to use this system as a prototype for additional types of data in the near future.

We also have experimented with the Drupal content-management system. Although, PostNuke systems are used for all of our major web pages, Drupal is being used for new graduate student and project-specific web pages that depend heavily on relatively unstructured user input, as the Drupal tools are somewhat simpler and easy to use for the naive user, albeit perhaps less powerful than the PostNuke tools for the expert user.

We are also looking at options for the future. Currently our main systems are hosted by a powerful, but aging Sun Blade 2000 server. However, as more software has become available in the Linux environment, we see that as an option for our future. We have therefore set up a “developmental” Linux server on a functional, but obsolete, PC, and are using it to test software etc.

Our wireless network on the shore continues to expand. We had to do a major rebuild on one of the nodes when it took a direct lighting strike so intense that most connections were welded together. However, that gave us a chance to experiment with improved power systems that now permit 24/7 operation. We are adding an extensive network of ground-water wells to the network on Hog Island and developing plans for extending the network to other islands. We have also collaborated with the University of Virginia's Engineering School to deploy a mote-based network of light sensors (LUSTER – Light Under Shrub Thickets Ecological Research).

We have continued our strong interactions with the Taiwan Ecological Research Network (TERN). TERN researchers Chi-Wen Hsaio and Chau-Chin Lin visited during the winter of 2006 and spring of 2007, respectively. We worked with them on developing web-based systems that ingest EML documents and the associated data and produce quality assurance analyses. We also participated, along with Don Henshaw from AND and Kristin Vanderbilt from SEV in a workshop in Shanping Taiwan that helped train East Asia Pacific (EAP) ILTER researchers in the use of Kepler and other tools that exploit EML.

Along with Chau-Chin Lin, John Porter made an invited presentation to the Coastal Environmental Sensing Networks workshop at the University of Massachusetts, Boston in April 2007. We also took the opportunity to visit Emory Boose at HFR during the trip. John Porter also remains a member of the Oak-Ridge National Laboratory Distributed Active Archive Center User Working Group.

---

2. Planned IM Projects

1) Expand use of relational databases in day-to-day data input and management and improve automated QA/QC checking
2) Explore how to best capture polygon-based location data
3) Continue development of EML-based tools, ideally as web services
4) Link VCR Personnel Database with LNO Personnel database, as API becomes available

2006

Site Bytes 2006

Andrews Forest SiteBytes 2006

LTER Site: Andrews LTER

Contributor: Suzanne Remillard and Don Henshaw (Sep 14, 2006)

Site Byte:

The Andrews Information Management (IM) Team includes Don Henshaw (USFS), Suzanne Remillard (Oregon State University), Theresa Valentine (USFS), and Fred Bierlmaier (OSU). Zhiqiang Yang, a remote sensing post-doc at OSU and a former Information Manager for KNZ, is helping us review and edit our EML with respect to best practices and provides expertise for dynamic EML generation and other technical issues.

This past year has been a year of transition for the Andrews Forest. Gody Spycher, the long-time LTER-funded Information Manager for the Andrews, retired in April and Suzanne has officially taken the reins. Gody has been terrifically productive working behind-the-scenes since 1985 in building information systems for the Andrews and will be greatly missed. Fortunately, Gody will still be available to assist our team on a temporary appointment. Additionally, Mark Harmon has stepped down as the lead site PI and Barbara Bond is now the lead.

In light of this transition and the loss of an IM position (Suzanne's previous soft-funded position is eliminated), the Andrews has initiated a planning effort to discuss the scope of Information Management (IM) activities, priorities with respect to database and metadata development, and the roles of investigators with respect to IM. The Andrews also received an NSF-funded field station planning grant. This funding will be used in part to improve efficiency in data processing and quality assurance checking of data from our sensor networks and to streamline the flow of data from the field to the lab.

Study database and metadata development for online access continues to be the highest priority with continual updates to long-term data sets, the migration of legacy data sets, and the development of newly submitted data sets from Andrews researchers. In the last year, we have added or updated 40 data sets. Currently, 140 data sets are now online including 24 spatial (ArcGIS) data sets. A web demo and poster at this year's ASM will feature our metadata-driven information system and interactive data retrieval programs that deliver value-added climate and streamflow data. Gody completed development of an interactive management system before retiring, which will facilitate entry, editing, and quality assurance checks of data.

Suzanne continues as the database administrator for the combined ClimDB/HydroDB effort. The ClimDB/HydroDB warehouse now includes over 7 million daily measurement values from 39 sites (280 stations) and hosts 30 visitor sessions per day. Don and Suzanne are collaborating with Wade Sheldon (GCE) on a ClimDB/HydroDB paper which Don presented at the International Conference on Hydrological Science and Engineering in Philadelphia on September 12, 2006.

Don chairs the LTER Network Information System Advisory Committee (NISAC) and serves on the LTER Executive Board representing the IM Committee. As NISAC chair Don also participates with the Cyberinfrastructure (CI) Team and as a technical representative on the Trends Editorial Board. Don attended an East Asian Pacific (EAP) ILTER Information Management Workshop hosted by the Taiwan Ecological Research Network (TERN) in February. John Porter (VCR) and Peter McCartney (NSF) also participated as instructors and in attendance were representatives from 7 other nations. This visit has enhanced collaboration with TERN and Taiwan sites are now participating in ClimDB/HydroDB. ILTER workshop participants from Taiwan, Japan, and Malaysia will be attending this year's ASM. Don also participated as a panelist in the Central America Workshop for

Information Technology and Biodiversity in Panama to better link information technology efforts with the community of biologists and to consider cyberinfrastructure improvements in Central America.

Theresa remains involved with the LTER GIS working group and will lead a GIS workshop at the ASM 2006. Theresa has led a pilot project to develop WatershedDB, an interactive web application and database intended to provide spatial data layers in conjunction with ClimDB/HydroDB sites. (Look for the demo and poster at this year's ASM.) The project is complete and we are looking for additional funding or collaboration to continue this effort. A website is available: http://wwwgis.forestry.oregonstate.edu/website/wsheddb1/wsheddb.htm. Theresa also co-leads the annual GIS Day in coordination with OSU departments and hosted over 400 Corvallis middle school students this past year. Students interacted in the use of GPS and the Andrews Interactive Mapping Site, and volunteer lectures highlighted the use of GIS in natural resources.

Fred Bierlmaier (OSU) is the on-site Andrews system administrator and maintains the site Local Area Network (LAN), wireless LAN, and digital radio and spread spectrum telemetry networks. The newly installed wireless access has been incredibly useful for researchers and conference attendees at the Andrews. The site has also received OSU matching funds from student fees that will buy two new computers and a Polycom video-conferencing system for the Andrews site.

Arctic LTER SiteBytes 2006

LTER Site: Arctic LTER

Contributor: Jim Laundre (Oct 27, 2006)

Site Byte:

This past year EML implementation and new web design has been a major focus of the Arctic LTER IM. We have successfully moved our legacy data to EML level 2 using a Perl/java script developed by the LNO. For new metadata we have developed a simple one sheet Excel form (modeled after FCE’s form) for new metadata entry (see http://ecosystems.mbl.edu/arc/metadata_forms/MetadataBlank.xls [156]). The one worksheet Excel form will help researchers easily include metadata with their Excel datasheets. A rich text Metadata form is available for researchers who do not use Excel. EML is then generated using a VBA macro. Currently the EML produced is level 3-4. There is still some coding to be completed in order to meet the “Best Practices” guidelines. As soon as this coding is done and documentation completed a version will be placed on the LNO’s CVS site.

In order to update legacy metadata a script was also developed to populate an Excel workbook with a metadata sheet and a data sheet. The code made best guests at the attribute table based on our previous text metadata. We have begun check and update these files. At the same time where possible we will combine yearly files into multi-year files.

We started to redesign our web site more then a year ago but have been waiting to deploy it until after the Web Design Recommendations were completed. In designing our new web site we are moving away from a frame base web site to using templates. We also plan to use style sheets to dynamically display metadata and have the Excel metadata and data workbook file available as a download.

From the field: This winter will be the start of Toolik Field Station having year round power. At least two staff will be on station all winter. For the Toolik Field Station webcam see http://www.uaf.edu/toolik/webcam [167]. With power the fiber optic internet connection will be available

year round and will provide continual downloads of weather data (http://ecosystems.mbl.edu/ARC [168]).

Bonanza Creek SiteBytes 2006

LTER Site: Bonanza Creek LTER

Contributor: Brian Riordan (Sep 18, 2006)

Site Byte:

Bonanza Creek LTER data management has made huge strides this past year. Not only in EML but overall data structure and our website flow. We have managed to be taken off of probation and are looking forward to our site review this coming summer. We have recoded a number of the Coldfusion webpage's to reflect the new database design. We have taken into account as many of the Website guidelines into account during our changes.

Data/website focus redirection: We have changed our philosophies on data structure and focus. Previously we were focused on a "story line" structure. This is were the concepts behind that data files were more important than the data itself. Now we are focused on individual files and letting researchers decide how they want to use it.

Hardware - The purchase of a dedicated database Penguin server has allowed us to generate large climate search engines, robust data file searches, and 3 million + row tables. We now have the Coldfusion html server on a separate piece of hardware.

Bibliography - Half of the references are linked to full pdfs. The bibliography is very dynamic and can be searched on a number of different levels. We hope to merge the data files with bibliographies over the next year.

Personnel - We have modeled our new format after the Andrews site. It makes it look more professional and informative with keywords, publications, data files, and key personnel information.

Data - We now have over 191 diverse data sets. The PIs responded very well to this years data file call and together they produced 40+ new data files this year. We rebuilt the data file search engine. It is formatted after the GCE site. All data files can now be searched for by personnel, keywords, sites, titles, and dates.

EML - We are currently producing 131 EML documents around level 2 - 3. We have made a number of different changes to the database to allow us to support attributes now. We are developing our site section to further round out or EML information. We hope to generate a dynamic EML system for the BNZ DB by Spring 07.

Climate Search engine - We have built a number of different search engines that are modeled after a "tax wizard" format. The engine dynamically builds queries that are executed upon finishing. The results are then displayed in a downloadable formats as well as graphs.

Future:
We are now beginning to look at the site section of the database and generate a new look and feel to it. This will be a large task since we have over 1000 sites in our database and over 500 site alternative names.
With the acquisition of the radios we are now working on fully automating out climate data. This will provide "real time" hourly data that will be available online immediately.

California Current Ecosystem SiteBytes 2006

LTER Site: California Current Ecosystem

Contributor: Karen Baker (Oct 22, 2006)

Site Byte:

The CCE information management effort is in its second year extending information management strategies to support the CCE community of scientists and interfaces with the LTER Network. A new disk share capability augmented last year's technical infrastructure build-out. Two disk mount technologies have been established for immediate use and comparative purposes (see MKortz, File Sharing Options: Elements of a Collaborative Infrastructure, http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/06sp... [169]). In addition, plans include a new server that will isolate web and storage services. Our informatics design studio capacity was updated and used for joint design sessions with California Cooperative Oceanographic Fisheries Investigations (CalCOFI) program participants. Organizational infrastructure was also built-out by initiating collaboration with CCE related projects.

Features added to the CCE web site (http://ccelter.sio.ucsd.edu [170]) include a site dynamic mapper consisting of a web interface to the GMT plotting tool. In addition, a grid station converter was developed in order to make available the standard sampling grid line and station given any latitude and longitude. The eventlogger system designed last year was further developed and deployed on a series of CalCOFI cruises as well as on the first CCE Process Cruise in May 2006. During this cruise, support included insuring daily web access to satellite imagery. After the cruise, a cruise web page is being developed and populated in order to collect cruise information and data. This serves as a data focal point while best practices, data procedures, and the local database are designed. In addition, identifying and gathering of long-term data for the Trends project took place.

Local cross-project work focused on cruise glossaries, eventlogs, and dataset dictionaries. As Ocean Informatics participants, a cross project Matlab User Group was initiated drawing together a variety of participants including programmers, technicians, and graduate students. An Infrastructure Studies presentation was made to the Integrative Oceanography Division in collaboration with a social science project and posters (Ocean Informatics; SCCOOS Data System: A Real-Time Data Acquisition, Storage and Access System) presented at the Marine Monitoring Conference 24-25 April hosted by Aquarium of the Pacific. KBaker became a UCSD Science Studies Affiliate (http://sciencestudies.ucsd.edu/ [171]) and collaboration with science studies projects (http://interoperability.ucsd.edu [172]) and participants continued with a focus on development of the notion of Infrastructure Studies. A student from sociology worked at SIO over the summer with a focus on the development of the eventlogger system. As LTER Network participants, we continued communications via feature article and good read submissions to the Databits Newsletter. A team of five Ocean Informatics participants for PAL & CCE submitted posters (CCE LTER: Information Management (2004-2006); Research in Infrastructure Studies: Social and Organizational Perspectives on Ecological Data Management) and attended the 2006 All Scientists Meeting.

Central Arizona Phoenix SiteBytes 2006

LTER Site: Central Arizona - Phoenix

Contributor: Corinna Gries (Sep 12, 2006)

Site Byte:

During the last year we improved publication search and download functionality on the CAP website. Results from a simple or advanced search can be marked and exported as plain text, EML or EndNote export format. The EndNote export format is also used for bulk upload to the LNO publications database. About 40 protocols used by CAP have been converted to EML format and added to the EML database in eXist an open source xml database system.

As a major update to data management at CAP we have experimented with and implemented online data entry applications written in php and accessing data in MySQL. This represents a move away from SQLServer and ACCESS front end applications. A basic template was developed in php to rapidly add new entry applications. Currently three data entry applications are being used successfully and a fourth is ready.

Thanks to the addition of a student programmer we have made major progress with our intranet. Bugs have been worked out and this year's annual report submission went smoothly. Online equipment and vehicle reservations have been added and the very successful concept of an informal working group was expanded. Working groups allow people to post and download documents, have a threaded discussion and collaboratively edit text documents. After several less successful experiments with commercial and open source software packages we have integrated this functionality into the intranet and at this point the document posting and downloading has been very well accepted by the researchers.

A former graduate student, Ed Gilbert, has returned to work with us full time as web programmer. A plant taxonomist by training, he is mainly interested in biodiversity informatics and during his former tenure has been instrumental in establishing the biodiversity section of the Southwest Environmental Information Network (http://swbiodiversity.org/seinet/index.php [157]). Arizona's large biodiversity collections can all be searched simultaneously, species distributions mapped and dynamic checklist created for any region based on collection records. In his spare time he is working on an application that allows interested people to contribute remotely to a central character database and online key for Arizona. His first duty will be the improvement of our data downloading functionality.

The Global Institute of Sustainability (GIOS) is participating in the state funded Arizona Water Institute, a consortium of Arizona's three universities focused on water sustainability through research, technical assistance, education and technology. The GIOS datalab staff is involved in building the Arizona Hydrologic Information system (http://chubasco.hwr.arizona.edu/ahis-drupal/ [173]) in collaboration with the other universities, local stakeholders and CUAHSI (http://www.cuahsi.org/ [174]).

GIOS has recently added the School of Sustainability (http://sos.asu.edu [175]), offering undergraduate and graduate degree programs in the area of sustainability research.

Florida Coastal Everglades SiteBytes 2006

LTER Site: Florida Coastal Everglades

Contributor: Linda Powell (Sep 05, 2006)

Site Byte:

In preparation for our 2006 funding renewal this past February, we made several enhancements to the Florida Coastal Everglades LTER program's information management system (IMS). With the addition of three new servers to the IMS, Linda Powell (information manager) and Mike Rugge (project manager) have spent numerous hours incorporating the new equipment into the FCE

network. Their focus over the past year has been on several important components of the IMS: 1) upgrading and expanding the FCE Oracle10g database, 2) upgrading and migrating information to the new FCE web server, 3) adding new web applications to the FCE website to facilitate information and project management and 4) converting all 270 FCE metadata files into tier 4 and 5 EML.

We have redesigned the 'Research' section of the FCE website by adding several new categories: 1) Information Management, 2) Projects and 3) Sampling. The Information Management System (IMS) (http://fcelter.fiu.edu/research/information_management/ [176]) is described in detail as it includes an IMS overview, the FCE data management and data submission policies, IMS statistics and a link to the EXCEL2EML tool. This information is also cross-referenced under the 'Data Resources' section.

Our 'Projects' section has greatly improved as we have organized specific projects and their related information using a series of tabs that give a 'file folder' look on the web page (http://fcelter.fiu.edu/research/projects/projects.htm?pid=1 [177]). For any given project, a user can easily find the abstract, research sites, personnel, sampling attributes, datasets and publications related to that specific project all on one web page. Additionally, projects can be searched by keyword, researcher or funding organization.

The 'Sampling' application (http://fcelter.fiu.edu/research/projects/sampling.htm [159]) is extremely useful in information and project management. Users are able to search for sampling attributes by entering keywords or manually selecting an attribute of interest from a list. Once the selection has been made, a map with a series of tabs give users access to research site information, dataset listings and project information related to their sampling attribute selection. This feature greatly facilitates project management and site science. Users are able to easily access related information from one portal. For example, researchers interested in porewater salinity can enter the sampling attribute in the keyword field and the results will return a map of the FCE research sites where porewater salinity is collected as well as tabular links to all datasets containing porewater salinity and projects responsible for the collection of that attribute. The FCE LTER has a written policy whereby researchers beginning new projects must submit their project information, including proposed sampling attributes, to the information no later than six months after the beginning of such project. With a list of sampling characteristics, both real and proposed, this application allows the information manager to track which sampling attribute values have not been submitted to the FCE IMS in a timely fashion as there will be no dataset affiliated with the missing attributes.

Additionally, the FCE student organization now has a section on our home page that delivers information on ongoing events and activities and links to important student features such as graduate student tools, links to other student organizations, poster logo information and student presentations, dissertations, and theses. There is also a 'Featured FCE Student' section that highlights the student's background and interests.

This coming year our focus will be on building a web-based query interface for FCE data values and adding a graphing application so that FCE researchers can graph their query results in real-time. To date, our data are available in ASCII files only. Many sampling attributes, like Total Phosphorus, can be found in several different flat files. Users must download each file to compile one Total Phosphorus data table for the FCE research sites. Allowing users to query download or graph the database for specific sampling attributes across all data files will be a welcomed addition to our website.

The FCE IMS has fully adopted the LTER network metadata standard Ecological Metadata Language (EML) and one hundred percent (100%) of the FCE tabular data are accompanied by a Level 4/5 (Data Identification, Discovery, Evaluation, Access and Integration) EML (XML) metadata documents. FCE EML documents are harvested daily to the LTER network metacat XML database.

Georgia Coastal Ecosystems SiteBytes 2006

LTER Site: Georgia Coastal Ecosystems

Contributor: Wade Sheldon (Nov 14, 2006)

Site Byte:

We submitted our GCE-II renewal proposal to NSF in February 2006, so most IM effort this past year was dedicated to helping researchers acquire, integrate and analyze data for publication and incorporation in the proposal. We also acquired some new ancillary long-term data sets from NWS and NOAA stations along the way, which we documented and added to our online data portal (http://gce-lter.marsci.uga.edu/portal/ [178]). We decided not to significantly overhaul our web site prior to the renewal, so we focused instead on tightening integration between web-accessible databases and web applications. For example, we developed a dynamic personnel list and bio pages that automatically display recent data submissions and complete lists of publications and presentations, with follow-up query links to our data catalog and bibliographic database query page. We also replaced all contact information links displayed by other web applications (e.g. data catalog, web calendar, species lists) and static web pages with query links to these dynamic bio pages to improve cross-linking of information and centralize content control for contact information to one database.

The GCE-II renewal proposal was very well received by reviewers, and both science and IM reviewers were very supportive of our IM program, web site, software development, and network participation. The only suggested changes were to provide a web-accessible version of our MATLAB data search client (already under development) and to "take some risks in thinking about value-added products or interfaces with other digital library efforts beyond the LTER". The latter suggestion will be difficult to address, both logistically and philosophically, so we're not yet sure how we will respond.

Work this summer has largely focused on wrapping up projects left over from GCE-I and preparing for new challenges in GCE-II, particularly a new emphasis on spatial data and GIS to support studies on marsh hammocks. I took over hosting of ESRI licenses in the UGA Marine Sciences Dept. and began supporting several GIS workstations, due in part to the relocation of our Marine Extension Service GIS lab across campus. I also set up a small GIS lab on Sapelo Island as part of a collaboration with the Sapelo Island National Estuarine Research Reserve. We are currently in the process of hiring a full time assistant IM / GIS specialist to assist researchers with collecting GPS data in the field and develop geospatial databases for our project, as well as assist me in acquiring and documenting new tabular data sources and maintaining our IS. This represents a huge and exciting operational change at GCE, more than doubling the FTE dedicated to informatics.

At the network level, I was elected to serve on NISAC to fill out Peter McCarney's term. This committee provides a great opportunity to interact with domain scientists on issues relevant to IM, and I expect to learn a lot this year. I have also participated in the Trends editorial meetings as a NISAC representative, and look forward to more interaction with this group. I have also continued to collaborate with Don Henshaw and Suzanne Remillard at AND on ClimDB/HydroDB development. We recently contributed a paper on ClimDB/HydroDB to the International Conference on HydroScience and Engineering, which is now submitted for review and publication. I also developed a ClimDB/HydroDB data access client for the GCE Data Toolbox for MATLAB (http://gce-lter.marsci.uga.edu/lter/research/tools/data_toolbox.htm [179]), allowing users to retrieve data from any registered station directly into MATLAB for analysis and transformation. Suzanne and I presented a poster and demo on these tools at the 2006 LTER ASM. I am currently collaborating with Barrie Collins at CWT to provide near-real-time USGS data and plots on the web for the CWT

study site, and we are exploring the possiblity of sharing resources on GIS development in the future.

Hubbard Brook SiteBytes 2006

LTER Site: Hubbard Brook LTER

Contributor: John Campbell (Sep 21, 2006)

Site Byte:

The information management team at Hubbard Brook is in the process of installing a wireless sensor network for automated collection of continuous, 15-minute meteorological and stream water data from one remote watershed. The wireless network will be based around 900 MHz spread spectrum Freewave radios. The network will be complemented by a web-based data delivery system through which researchers can view graphic displays of real-time environmental data, or download archived data. In September, a 15 m repeater station was installed on a ridge between the experimental watersheds and the Forest Service headquarters to establish line-of-site radio transmission. Once the data are transmitted to the headquarters building they will be uploaded via satellite internet connection to a MySQL database.

This past summer we hired Pavel Dorovskoy, an undergraduate Computer Science major at the University of New Hampshire. Pavel was the first REU student at Hubbard Brook specializing in information management. His research project was centered on the wireless sensor network and he gave a presentation at the annual Hubbard Brook Cooperators' meeting and will be submitting an article to the Databits newsletter.

Considerable progress has been made in redesigning the Hubbard Brook web page. In January, the Hubbard Brook's Information Oversight Committee approved the new design of the web page. Since that time, the Hubbard Brook information management team has been converting the old web page to the new format. The new webpage will adhere to the new LTER webpage design recommendations. Some of the new features of the web page include: an image archive with an image upload feature; curriculum vitas with forms for making on-line updates; and improved methods for searching and downloading data. A draft of the new website will be completed by October 1st for review by the Information Oversight Committee. The final website will be made public before Hubbard Brook's mid-cycle NSF review in 2007.

The Hubbard Brook sample archive was recently dedicated to Cindy Veen. Cindy Veen served as Hubbard Brook database manager from 1988 until her untimely passing in 1996. She was well-respected at the USDA Forest Service and among her peers in the LTER network. Among her many accomplishments, she was instrumental in the establishment of the Hubbard Brook sample archive and worked diligently to make it operational. In recognition of her achievements, the Hubbard Brook sample archive building was dedicated to her in July at the Annual Hubbard Brook Cooperators' meeting. The exterior of the building now has a sign bearing Cindy's name and there is a plaque on the inside with her picture and a description of her role in the Hubbard Brook Ecosystem Study.

Jornada Basin SiteBytes 2006

LTER Site: Jornada Basin

Contributor: Ken Ramsey (Nov 02, 2006)

Site Byte:

In the last year, the information management team at the Jornada Basin LTER has redesigned the Jornada website and prepared for our site renewal (which was approved) in addition to normal information management tasks and responding to requests from researchers for map products. The Jornada website was redesigned using a draft of the LTER web design guidelines and by visiting all LTER websites. The website was initially developed using the PHP programming language and a SQL Server backend database. We have migrated most of our dynamic web pages from using a relational database to using XML files. This has made our website available when our database server is down for maintenance and updates.

We are populating the metadata for all research projects and datasets in the relational database to allow dynamic generation of EML. We are also editing the style sheets used to create EML documents from the database to correct some initial problems caused by generating blank EML elements and to enable the creation of level 5 EML from our database once the metadata population has been completed. Inigo San Gil is scheduled to come to the Jornada to assist with completing the style sheets in early November.

Kellogg Biological Station SiteBytes 2006

LTER Site: Kellogg Biological Station

Contributor: Sven Bohm (Sep 17, 2006)

Site Byte:

This past year we on consolidated and normalized our databases. Reorganizing tables and trying to eliminate redundancies. We made one more switch in database systems from MSSQL to DB2, to gain spatial SQL capability. We also added a server with 1 terra-byte of disk space to store our annual air photo campaigns. We're currently replacing our old web/db server with a new virtual server. Using a virtual server, we can easily create a copy to test changes, and troubleshoot without disrupting the service.

I wrote a couple of small quality control web applications, to provide an incentive for the lab and other researchers to submit data earlier in the process. Two of the QC apps have been successfully used so far, and they have generated request for apps to process other datasets.

Konza Prairie SiteBytes 2006

LTER Site: Konza Prairie LTER

Contributor: Jincheng Gao (Sep 16, 2006)

Site Byte:

With 18 months work experience as an Information Manager at the Konza site, I have learned a lot and have benefited from the efforts of our former Information Managers, including John Briggs, Brent Brock and Zhiqiang Yang. In particular, I appreciate the efforts of Zhiqiang Yang, who did an excellent job in building the KNZ metadata database and establishing the dynamic web site. In the past year, I have continued to work on the transfer of data from ascii text files to a SQL Server

database and QA/QC checking, as well as GIS data generation and regular updating of the database and website.

So far, we have finished the transfer of historical data from text files into the SQL Server database. The data can now be searched and downloaded based on specific watershed, treatment, or date of data collection criteria. Data input interfaces for each dataset were created with VC# to convert text data files into the SQL Server database. Data quality checking is performed based on data range and logical consistency before data transfer. The data ranges are defined according to expected minimum and maximum values for each dataset, and logical consistency is checked based on general knowledge for each data type. For example, daily mean temperature must be less than the daily maximum and greater than the daily minimum values, or the data are flagged for further checking and correction. We are continuing to update the EML for each dataset to comply with the Best Practice for the Metacat harvest. Currently, over 50 datasets in the KNZ database (more than 80% of our core datasets) are harvested weekly by the Metacat database. After QA/QC checking, four stream flow datasets from the KNZ are being regularly harvested by HydroDB.

Based on the suggestion of the KNZ site review last year, I have also collected and edited the GPS data for historical and long-term sites of data collection, such as the climate and stream gauging stations, and long-term transects for collection of plant species data. The accuracy of other spatial data in the KNZ database has been checked, and updated where necessary. The GIS and remote sensing data in KNZ have been also reorganized and their metadata have been updated with FGDC and EML format. The dataset of annual burning history of various watersheds at KNZ is now stored in two formats. One format is as layers for the various watersheds or fire treatments, and stored as a polygon feature class in our database. The other format is as non-spatial data. The non-spatial data are linked with KNZ watershed spatial data through ArcSDE. The burning history data can be interactively queried by watershed name, burning type, and burn time (year or date) in the KNZ interactive web site.

A new bison dataset was created last year. A GPS collar unit was installed on an adult bison as an initial trial to assess the utility of this approach for tracking bison movement patterns. The geographic coordinates of the bison are automatically collected every two hours. The GPS data of the bison are in the interactive web site. As more animals are fitted with tracking devices, the GPS data for bison movement will be integrated with the vegetation data in KNZ to study bison behavior and their effects on heterogeneity in the tallgrass prairie landscape.

In the coming year, I will focus on our web design and the QC/QA checking of our EML metadata. In addition, the methodology associated with each dataset will be converted from the current text format to EML metadata.

Luquillo Experimental Forest SiteBytes 2006

LTER Site: Luquillo LTER

Contributor: Eda C. Melendez-Colom (Sep 14, 2006)

Site Byte:

LUQ LTER main achievement this year was the unconditional approval of the LUQ LTER 4 proposal. For the last year LUQ IM dedicated a great time in the everlasting task of improving the databases' metadata and updating data on the web site.

LUQ IM is always looking for ways to improve the communication among the different communities of the LUQ LTER. Web sites could be a mean to improve this communication. A new Plone web site

was created for the LUQ LTER Canopy Trimming Experiment (CTE) group which is now being used by its members to share and edit their documents, e.g., the LUQ LTER CTE Poster for the 2006 ASM.

To further enhance the communication and present the community with information management concepts, we created periodical email messages, copies of which that can be accessed on the web site called “Gotitas del Saber” or “Drops of Knowledge” (http://luq.lternet.edu/datamng/GotitasdelSaber/index.html [180]). The explanation of simple concepts like what is a URL was very well welcomed by the community.

Realizing the fact that the LUQ LTER Schoolyard Program is going through a re-birth process, we realized that there was an obvious lack of an IM component in this program. We had collaborated in year 2000 by giving a workshop to the teachers on data management and creating a static directory on the web. Since fall 2005, LUQ IM has been closely collaborating with the investigator in charge by visiting the schools to learn more about the research sites and their data sets, creating a mailing group alias (lterschool@ites.upr.edu [181]), designing a Web Site (http://luq.lternet.edu/outreach/schoolyard/index.html [182]), entering their data, and planning workshops for managing data at the different schools. In these workshops, the student will be exposed to the principles that define an LTER site: sharing and preserving information. A subset of LUQ metadata standards will be used to document their data uniformly across the schools. IM has suggested modifications to the structures of similar data sets in an attempt to standardize the way data is being managed in the schools. The web sites where each school has a page with Photos, a data set catalog, main schools results, publications and contact information. Finally, IM will give a hands-on workshop in an Internship that will be held next November for selected students of all three schools.

LUQ IM continues with the effort of translating its metadata into EML and it will focus in this activity for the next year.

LUQ IM System needs to be enhanced. We plan to devote ourselves in studying ways to create a framework that helps the user (internally and externally) to access information and perform synthesis-related tasks. The best way to do this is still not clear. The LUQ IM intends to learn more about the best infrastructure to do this. We have hopes that EML will be the answer to this.

McMurdo Dry Valleys SiteBytes 2006

LTER Site: McMurdo Dry Valleys

Contributor: Chris Gardner (Oct 31, 2006)

Site Byte:

This year has been one of major IM-related advances at MCM. I went live with our new website at www.mcmlter.org [166] in January 2006. Working with Inigo at the LNO, we were able to take MCM from 0% EML compliance to 100% EML level 5 in the matter of a few short months.

The new web presence represents a total redesign with new database-driven features. Some new features include an updated searchable bibliography and personnel list, a database of freely downloadable high quality Antarctic photos, and several tools to query our Oracle database for different dynamically derived datasets. A main goal of mine is to take full advantage of the relational database structure through the use of derived datasets that are useful to MCM researchers and the greater science community. An example of these derived dataset query tools include the ability to query stream chemistry and have Oracle dynamically convert mass units to moles, calculate TDS

and charge balance, and then round the time the chemistry sample was taken to the nearest quarter hour and join those chemical data to stream discharge measurements. Other dynamic pages calculate the number of days of flow of these ephemeral streams at MCM and the total volume during the summer flow season. Other accomplishments include a site management section in the restricted access portion of the website where ancillary research projects can be documented, news stories can be added to the database, chain of custody forms can be downloaded, etc.

Current projects include documenting our spatial data layers in ESRI's ArcSDE and linking those data dynamically to our tabular database in Oracle. Additionally, this field season I will be implementing improved sample tracking protocols where the chain of custody forms will be entered into the database, which will eventually be used as table joins and to verify the existence of data. Using a new standardized chain of custody form that requires specific sample name formats will also greatly reduce the time spent manually verifying and combining chemical data. These and other new protocols will greatly increase the efficiency and accuracy of MCM's system of dealing with data starting from the point of collection all the way to making the QA/QC'd data publicly available on the web through dynamic query tools.

Niwot Ridge SiteBytes 2006

LTER Site: Niwot Ridge LTER

Contributor: Todd Ackerman (Sep 18, 2006)

Site Byte:

The personnel of Niwot Ridge LTER Information Management over the past year has consisted of Information Manager Todd Ackerman and Student Worker Anobha Gurung. It has been another busy year for the IM staff. We have been working on several new projects on top of the centralized data entry/process/dissemination.

One major dataset undertaking that has occurred this year was to re-visit our 50+ year climate record for the D1 and C1 Station and the 20+ year record at the Saddle Station. We realized a shortcoming in the data when it comes to modeling. We needed to have a solid/standard record of daily values for precipitation and temperature for these main stations, and there were many missing days throughout such a long record causing false trends when the long-term record was analyzed. Using adjacent climate stations and statistical methods we filled and flagged these missing days so that modeling could be better performed when needed. The datasets are nearly complete, as we are revisiting some of the methodology used.

We have also been developing a query-able system for the chemistry data produced by our in-house Kiowa Environmental Chemistry Laboratory run by Chris Seibold. In the past these files existed as MS Excel files and only some of the data posted as ancillary data to the core datasets. This new query system at the moment is available internally only until use permissions and proper metadata can be developed. It has vastly improved the compilation of data for researchers when requested.

We are also now installing the latest FreeWave HT-plus radio system for two new sites (new high frequency data collecting eddy-flux towers run by Peter Blanken) and adding one to the existing Soddie Site to allow for Deltev Helmig's group to better access and control their sampling tower remotely in the winter. Our field personnel (Mark Losleben and Kurt Chowansky) have been key in getting this project going.

This year we have also been participating in the DayCent-Chem biogeochemical cross-site inter-comparison led by Jill Baron. Along with several other LTER sites (NWT, CWT, HBR, and HJA), as well as experimental forest sites and National Parks, we attended workshops and submitted data to Baron's group to look at C and N dynamics.

We have recently upgraded the core LTER file/web server CULTER from a Sun Ultra 60 to a Sunfire V250. This has vastly increased our file storage capabilities allowing us to make larger datasets such as imagery more easily accessible. It also allowed us to consolidate one other LTER server SNOBEAR with CULTER to reduce the number of machines to maintain. We are also upgrading our Windows server from a development server (Dell PowerEdge 500SC) to a much more powerful server (Dell PowerEdge 2600) which will allow us to serve data publicly from the MSSQL RDBMS which the previous machine could not handle from lack of system resources.

North Temperate Lakes SiteBytes 2006

LTER Site: North Temperate Lakes

Contributor: Barbara Benson, Dave Balsiger, Jonathan Chipman (Sep 15, 2006)

Site Byte:

We have continued our development in the area of sensor networks as NTL increases the number of lakes on which instrumented buoys are deployed. Currently data are flowing automatically, in near real-time, into our Oracle database from 8 buoys. We are collaborating with computer scientists from the University of California San Diego Supercomputer Center, SUNY-Binghamton, and Indiana University on automating the configuration and quality assurance components of the NTL information system for near-real time data streaming from instrumented buoys. We have created a metadata model for instrument management. Calibration previously had been done for the dissolved oxygen data in a manual process; we designed and implemented a web-interface for computer-assisted calculation of correction factors for the calibration and updating values in the database based on the correction factors. We designed and implemented a Cyberdashboard based on the GridSphere portal framework to help manage the network of instrumented lake buoys. This web-based application enables the buoy technician or data manager to quickly determine the status of the multiple remote sensors. Information regarding missing data and the automated QA/QC processing of the data is also available via the Cyberdashboard.

NTL is providing leadership in the Global Lake Ecological Observatory Network. The Global Lake Ecological Observatory Network (http://gleon.org [183]) is a grassroots network of limnologists, information technology experts, and engineers who have a common goal of building a scalable, persistent network of lake ecology observatories. Data from these observatories will allow us to better understand key processes such as the effects of climate and landuse change on lake function, the role of episodic events such as typhoons in resetting lake dynamics, and carbon cycling within lakes. In March 2006, we met in Townsville, Australia for a GLEON /CREON workshop. Another workshop will be held in Hsinchu,Taiwan in October 2006. We maintain the GLEON website and are developing a database of information on lakes associated with this global network.

More social science data sets including a manure management survey and a data set on shoreline property sales were added to the NTL Oracle database and are available through the website. We have implemented a second “tier” for the information management system in order to capture data sets as text files along with their metadata when it has been decided that these data are not in high enough demand to warrant incorporation into the Oracle database but need to be permanently archived.

Ongoing spatial data management activities include development of web-based mapping systems and the addition of new data sets to the catalog. The NTL website now has three interactive map servers that provide users with the functionality to create maps using spatial data layers for our two study areas (the Madison region and the Trout Lake region in Vilas County) and for Wisconsin statewide.

Palmer LTER SiteBytes 2006

LTER Site: Palmer Station

Contributor: Karen Baker (Oct 22, 2006)

Site Byte:

The PAL information management effort is developing within a broader Ocean Informatics environment that includes the CCE LTER marine site co located with the PAL information management component at UCSD/Scripps Institution of Oceanography. Technical, organizational, social, and individual infrastructure needs are addressed together in order to support the PAL community and interfaces with the LTER Network.

A new remote mount technology was established for manuscript, presentation, and file sharing. Our informatics design studio capacity was updated with a system able to support multi-user chat sessions and with processor speeds more effective for CPU intensive applications such as google earth. Local efforts are ongoing in preparation for Open Directory LDAP services, an approach intended to enable secure and versatile information exchange. In the interim, local authentication/authorization conventions have been developed as collaborative needs arise for accessibility and security. Community tools developed include JPGraph for web plotting, Matlab for data analysis, and WordPress for blog and web page editing.

The local database design was expanded into a dual component model influenced by identifying, discussing, and gathering of long-term data for the Trends project. Attribute naming conventions are in development providing a framework for dictionaries that ultimately will support interoperability. The PAL web site (http://ccelter.sio.ucsd.edu [170]) was updated for meetings of the site review committee in November 2005 where ice precluded the ship carrying site reviewers from reaching Palmer station and in May 2006. The site bibliography content was updated and electronic copies of manuscripts made available either publicly or to the site. In addition, the bibliography module was refactored to include script delivery for an LNO harvester and to conform with the recently updated LTER Network requirements. An Ocean Informatics web site (http://oceaninformatics.ucsd.edu [184]) is being developed into a cross-project web portal. Design issues have been explored (Design Interfaces; Design Patterns; http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/06sp... [185]) and techniques used to develop a hybrid model incorporating elements of global navigation as well asll hub-spoke design. Cross-project work (CCE, CalCOFI, Ocean Observing System, and NOAA Fisheries) is prompting development of meta level context and stimulates development of best practices.

Collaborative local activities included joint design sessions focusing on eventlogs, and dictionaries. This work was part of a proposal written to frame and support cross-project infrastructure work. Work with the LTER Network included participation on the LTER Governance Committee. Where new text was drafted that develops more fully the role of information management. Participation on the network Web Recommendations Committee provided an opportunity to foreground the SiteDB network module.

KBaker became a UCSD Science Studies affiliate (http://sciencestudies.ucsd.edu/ [171]) with synergies developing along side previous LTER ethnographic studies focusing on articulation work and data stewardship. A team of five Ocean Informatics participants for PAL & CCE submitted posters (Palmer LTER: Designing a Queriable Community Data System; Research in Infrastructure Studies: Social and Organizational Perspectives on Ecological Data Management) and attended the 2006 All Scientists Meeting.

Sevilleta SiteBytes 2006

LTER Site: Sevilleta LTER

Contributor: Kristin Vanderbilt (Oct 30, 2006)

Site Byte:

The Sevilleta IM Team continues to support site and network level science by seeking ways to improve data discovery and availability.

Sevilleta scientists have two studies in which wireless sensor networks play a significant role. Data is being collected in large volumes, and the challenge is to import the data from the Campbell datalogger into our MySQL database, write QA/QC scripts, and graph the data in near real-time. We are presently exploring off-the-shelf software and custom software solutions using SAS for these purposes. Sevilleta LTER graduate student Etsuko Nonaka is working with two scientists at Los Alamos National Labs to develop QA/QC methods for the Sevilleta sensor data.

Sevilleta continues to translate legacy semi-structured metadata into EML using a variety of methods including Morpho and a custom script developed by Inigo San Gil of LNO. We have Level 5 EML for about 50% of Sevilleta legacy data.

Wikis have become valuable parts of the Sevilleta IMS. Sevilleta submitted a successful renewal proposal in 2006, and the preparation of the proposal made use of a wiki for sharing document revisions. A wiki is also being maintained by the permanent four person field crew and the IM, describing all steps needed to collect, manage, and archive the long-term data sets collected by the field crew. A password-protected System Administration wiki has been implemented to document mission-critical sys admin tasks.

IM Kristin Vanderbilt has been active on the US ILTER Committee. She gave two presentations at the EAP-ILTER meeting in Kyoto, Japan in February about information management training. At the ILTER Coordinating Committee meeting in Namibia in August she became Chair of the ILTER Information Management Committee.

Short Grass Steppe SiteBytes 2006

LTER Site: Shortgrass Steppe

Contributor: Nicole Kaplan (Sep 18, 2006)

Site Byte:

Over several weeks, the IM Team and SGS-LTER Staff have focused on setting up new administrative and staff office at Colorado State University. Staff and the IM Team have been responsible for inventorying and organizing legacy data and reports dating back to the 1930s,

located across six offices and over a dozen filing cabinets. We have used this opportunity to create a library of historic International Biome Project and LTER data and reports, and prepare older, richer metadata for entry into our database. These efforts will help preserve the history of project goals, and document the evolution of ecological studies and data that spans over forty years at the SGS field research site. We hope this will serve as a useful resource for LTER Researchers in the future.

The IM Team now consists of Nicole Kaplan, full-time IM Team Leader, Bob Fynn, half-time IT/GIS Manager, a half-time student web developer and two quarter time data entry students. We continue to work on improving data discovery, delivery and interoperability tools within the SGS Information Management system. Our efforts have been guided by both internal evaluation of our strengths and weaknesses, and an external mid-term review, which occurred last summer by National Science Foundation. We have continued to work with Inigo San Gil at the LTER Network office to implement a newly designed relational database management system (RDBMS) and PERL and XSLT (Extensible Stylesheet Language Transformation) scripts that now contain and generate level 5 EML (Ecological Metadata Content) content. The IM Team is working with our core Staff and Principal Investigators to enhance metadata for our current long-term studies to Level 5, which describes tables and attributes in addition to project objectives, methods, locations, and principal investigators.

We are also constructing new information delivery tools within a redesigned SGS-LTER website that serves metadata and other information from the new RDBMS. Nicole has been involved with developing and adopting Recommendations for Website Design within the LTER Network. We are taking into account these recommendations, applying a Java script for menu navigation, installing website search tools and using Macromedia (Adobe) Fireworks for displaying images from the field site.

Bob provides support for SGS-LTER researchers and students in various aspects of GIS including gathering data with GPS equipment and imagery, assisting with GIS model development for their particular research, and providing GIS data and maps for field work and modeling. He has also extended existing programs for analysis of SGS-LTER data.

Virginia Coast SiteBytes 2006

LTER Site: Virginia Coast Reserve

Contributor: John Porter (Sep 17, 2006)

Site Byte:

This has been a busy year for information management at the VCR/LTER in a number of areas. The first has been work on developing tools that use EML to produce documents, such as statistical programs, for use by researchers. Converters were developed that convert suitable EML documents to SAS, SPSS and R statistical languages. In addition, a web interface to the tools was developed and means to link the programs to static web pages were developed.

The VCR/LTER also hosted the second-to-last Coordinating Commitee meeting in the fall of 2005 (the CC was replaced by the "Science Council" in the spring of 2006). We used a system we have developed for easy-to-use online forms to manage the complex logistics of getting everyone to and from the airport.

During 2005 and January 2006, we also wrote our new proposal. During that process we made extensive use of videoconferencing to bring together (virtually) investigators taking the lead on specific proposal sections. The system we used consisted of PC's running the Polycom PVX

software ($120) using an inexpensive ($50) USB-camera/microphone combination. These then linked to a higher-end Polycom FX unit that allowed us to have 4 participants in the conversations. We also continued revamping our web page using the PostNuke Content Management System.

This has also been an important year for international collaboration with the Taiwan Ecological Research Network (TERN). It started in January when John Porter (VCR), Don Henshaw (AND) and Peter McCartney (NSF) participated in a series of workshops with Taiwanese researchers and with other LTER Information Managers from the East Asia Pacific (EAP) region. The trip also included visits to several of their research sites. As a result of the trip, several sites now participate in CLIMDB. Closer to home, we have had a series of visits by members of the TERN "IM Team." Chien Wen "Kevin" Chen visited the VCR/LTER during February-May 2006. The focus of his visit was collaboration on the development of wireless sensor networks for field research. He was followed by a return visit (June-Sept) from his colleague Meei-ru Jeng. She focused on collaborations involving EML, specifically the creation of EML documents, XML editors, stylesheets and "R" statistical language programming. In Sept. 2006 she was joined by her colleagues Yunyin Yeh and Fu-Ching "Tanya" Yang. During their visit they focused on the examination of several US LTER web sites and how they provide services to researchers. This was followed up by visits to the Baltimore Ecosystem Study, Harvard Forest and Hubbard Brook LTER sites, where they met with the information managers from each site. This fall, their colleague Chi-Wen Hsao will be coming to work on GIS-related issues.

We also saw the openning in August of the long-awaited "Anheuser-Busch Coastal Research Center of the University of Virginia" (ABCRC). The new ABCRC provides much improved laboratory and housing facilities, and replaces our old, rented farmhouse as the field headquarters for VCR/LTER activities. IM activities at the new Center involve moving our T-1 connection, developing wireless and wired networks in the new buildings and connecting those networks to wireless field equipment.

We have also been active in several ongoing network-wide activities. We have been involved in the Cyberinfrastructure working group of the LTER Planning Grant activity, and hosted a modeling workshop at the University of Virginia in January 2006. We have also been active with the controlled-vocabulary working group and ILTER training in the East Asia Pacific region.

Plans for the coming year include continued improvement of the EML we produce, specifically in the areas of units and geographical information, development and improvement of tools for using EML , continued collaborations with TERN and EAP researchers, and development of systems to support management of the expanded LTER facilities at the ABCRC.

This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Please contact us with questions, comments, or for technical assistance regarding this web site.

Source URL: http://im.lternet.edu/siteprofiles

Links:
[1] http://im.lternet.edu/profile
[2] http://im.lternet.edu/siteprofiles/AND
[3] http://im.lternet.edu/siteprofiles/ARC
[4] http://im.lternet.edu/siteprofiles/BES
[5] http://im.lternet.edu/siteprofiles/BNZ
[6] http://im.lternet.edu/siteprofiles/CCE
[7] http://im.lternet.edu/siteprofiles/CDR
[8] http://im.lternet.edu/siteprofiles/CAP
[9] http://im.lternet.edu/siteprofiles/CWT

[10] http://im.lternet.edu/siteprofiles/FCE
[11] http://im.lternet.edu/siteprofiles/GCE
[12] http://im.lternet.edu/siteprofiles/HFR
[13] http://im.lternet.edu/siteprofiles/HBE
[14] http://im.lternet.edu/siteprofiles/JRN
[15] http://im.lternet.edu/siteprofiles/KBS
[16] http://im.lternet.edu/siteprofiles/KNZ
[17] http://im.lternet.edu/siteprofiles/LUQ
[18] http://im.lternet.edu/siteprofiles/MCM
[19] http://im.lternet.edu/siteprofiles/MCR
[20] http://im.lternet.edu/siteprofiles/NWT
[21] http://im.lternet.edu/siteprofiles/NTL
[22] http://im.lternet.edu/siteprofiles/PAL
[23] http://im.lternet.edu/siteprofiles/PIE
[24] http://im.lternet.edu/siteprofiles/SBC
[25] http://im.lternet.edu/siteprofiles/SEV
[26] http://im.lternet.edu/siteprofiles/SGS
[27] http://im.lternet.edu/siteprofiles/VCR
[28] http://www.mysql.com/
[29] http://www.microsoft.com/sql/default.mspx
[30] http://www.postgresql.org/
[31] http://www.oracle.com/database/index.html
[32] http://exist.sourceforge.net/
[33] http://www.ibm.com/software/data/db2/
[34] http://www.adobe.com/products/coldfusion/
[35] http://php.net/
[36] http://www.activestate.com/Products/activeperl/
[37] http://en.wikipedia.org/wiki/Common_Gateway_Interface
[38] http://www.w3schools.com/html/default.asp
[39] http://www.w3schools.com/xsl/
[40] http://www.mathworks.com/products/matlab/
[41] http://java.sun.com
[42] http://msdn.microsoft.com/vbasic/
[43] http://www.sas.com/
[44] http://office.microsoft.com/en-us/excel/default.aspx
[45] http://en.wikipedia.org/wiki/Ajax_(programming)
[46] http://en.wikipedia.org/wiki/JavaScript
[47] http://knb.ecoinformatics.org/software/eml/
[48] http://www.lternet.edu/news/Article118.html
[49] http://kepler-project.org/
[50] http://im.lternet.edu/links/goto/129/70/links_related
[51] http://im.lternet.edu/links/goto/130/22/links_related
[52] http://im.lternet.edu/links/goto/131/24/links_related
[53] http://im.lternet.edu/links/goto/131/23/links_related
[54] http://im.lternet.edu/links/goto/131/25/links_related
[55] http://im.lternet.edu/links/goto/132/26/links_related
[56] http://im.lternet.edu/links/goto/144/42/links_related
[57] http://im.lternet.edu/links/goto/134/28/links_related
[58] http://im.lternet.edu/links/goto/133/27/links_related
[59] http://im.lternet.edu/links/goto/135/29/links_related
[60] http://im.lternet.edu/links/goto/136/30/links_related
[61] http://im.lternet.edu/links/goto/137/117/links_related
[62] http://im.lternet.edu/links/goto/137/120/links_related
[63] http://im.lternet.edu/links/goto/137/121/links_related
[64] http://im.lternet.edu/links/goto/139/35/links_related
[65] http://im.lternet.edu/links/goto/138/34/links_related
[66] http://jornada-www.nmsu.edu/datacat.php
[67] http://im.lternet.edu/links/goto/145/43/links_related
[68] http://im.lternet.edu/links/goto/145/46/links_related
[69] http://im.lternet.edu/links/goto/145/45/links_related
[70] http://im.lternet.edu/links/goto/145/48/links_related
[71] http://im.lternet.edu/links/goto/145/44/links_related
[72] http://im.lternet.edu/links/goto/145/47/links_related
[73] http://im.lternet.edu/links/goto/140/36/links_related
[74] http://im.lternet.edu/links/goto/146/49/links_related

[75] http://im.lternet.edu/links/goto/141/37/links_related
[76] http://im.lternet.edu/links/goto/142/38/links_related
[77] http://im.lternet.edu/links/goto/143/39/links_related
[78] http://im.lternet.edu/links/goto/143/41/links_related
[79] http://im.lternet.edu/links/goto/143/40/links_related
[80] http://im.lternet.edu/links/goto/148/53/links_related
[81] http://im.lternet.edu/links/goto/148/52/links_related
[82] http://im.lternet.edu/links/goto/147/51/links_related
[83] http://im.lternet.edu/links/goto/147/50/links_related
[84] http://im.lternet.edu/links/goto/151/56/links_related
[85] http://im.lternet.edu/links/goto/150/55/links_related
[86] http://sbc.lternet.edu/external/InformationManagement/documents/SBC/SBC_IMplan.pdf
[87] http://im.lternet.edu/links/goto/149/98/links_related
[88] http://im.lternet.edu/links/goto/152/57/links_related
[89] http://im.lternet.edu/links/goto/153/58/links_related
[90] http://im.lternet.edu/links/goto/153/59/links_related
[91] http://im.lternet.edu/links/goto/154/61/links_related
[92] http://im.lternet.edu/links/goto/154/60/links_related
[93] http://jornada.nmsu.edu/data-catalogs/jornada
[94] http://jornada.nmsu.edu/data-catalogs/long-term
[95] http://jornada.nmsu.edu
[96] http://lter.kbs.msu.edu/score_cards
[97] https://github.com/kf8a/noodle
[98] http://www.vcrlter.virginia.edu/webservice/PASTAprog
[99] mailto:jporter@lternet.edu
[100] http://luq.lternet.edu/downloads/precipitationandtempdataentryand-conversiontemplatexlsx
[101] http://luq.lternet.edu/IM/IMSProtocols
[102] http://rambutan.ites.upr.edu:8080/itesplone/
[103] http://luq.lternet.edu/
[104] http://lter.limnology.wisc.edu/ltearts
[105] http://lter.limnology.wisc.edu/ltearts/exhibition/panel1
[106] https://portal.lternet.edu/
[107] http://databits.lternet.edu/spring-2007/generating-eml-relational-database-management-system-rdbms
[108] http://andrewsforest.oregonstate.edu
[109] http://andrewsforest.oregonstate.edu/lter/about/site/map.cfm?topnav=219
[110] http://besdirector.blogspot.com/
[111] http://bes-news.blogspot.com/
[112] http://beslter.org/facebook
[113] http://databits.lternet.edu/spring-2011/lter-information-management-continuing-education-and-site-change
[114] http://scoria.lternet.edu:8080/lter-hive-prototypes/emlTagger.jsp
[115] http://gce-lter.marsci.uga.edu/public/app/data_catalog.asp
[116] http://gce-lter.marsci.uga.edu/public/research/research_bullets.htm
[117] http://gce-lter.marsci.uga.edu/public/site/research_requests.htm
[118] http://gce-lter.marsci.uga.edu/public/im/tools/data_toolbox.htm
[119] https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox
[120] http://hbr1.lternet.edu/data/realtime/
[121] http://www.konza.ksu.edu
[122] http://sanjuanultra.org/
[123] https://kepler-project.org/users/projects-using-kepler-1/reap-project
[124] http://www.beslter.org
[125] http://www.gotomeeting.com
[126] http://coweeta.uga.edu/pubcatalog
[127] http://cwt-dev.anthro.uga.edu/dbpublic/cwtv_personnel.asp
[128] http://www.wcu.edu/hbs/
[129] http://www.ltwa.org
[130] http://www.wcu.edu/hbs/CEP%20capstones/2009IEreports.pdf
[131] http://gce-aux.marsci.uga.edu/exist/rest/db/projects/util/xquery/getProjectsQueryForm.xql?xslUrl=http://gce-aux.marsci.uga.edu/exist/rest/db/projects/util/xslt/gceQueryForm.xsl
[132] http://gce-lter.marsci.uga.edu/public/gis/gcewebmap.html
[133] http://gce-lter.marsci.uga.edu/public/research/research.htm
[134] http://luq.lternet.edu/outreach/schoolyard/Activities/2009Internship/2009Internship2.html
[135] http://databits.lternet.edu/issues/115
[136] http://luq.lternet.edu/outreach/SCHOOLS/NashuaMain.html
[137] http://www.facebook.com/video/?oid=114899015228942
[138] http://mcr.lternet.edu/research/information_management

[139] http://mcr.lternet.edu
[140] http://metacat.tfri.gov.tw/modules/
[141] http://mcr.lternet.edu/data/db/census/coralImage.php
[142] http://mcr.lternet.edu/data/realtime
[143] http://culter.colorado.edu/MyWater
[144] http://culter.colorado.edu/lake-algae/
[145] http://amo.colorado.edu
[146] http://culter.colorado.edu/exec/Database/gis_layer_query.cgi
[147] http://sbc.lternet.edu/data/dataCollectionsPortal.html
[148] http://intranet.lternet.edu/im/siteprofiles/SiteBytes/2008
[149] http://fcelter.fiu.edu
[150] http://gce-lter.marsci.uga.edu/public/app/resource_search.asp
[151] http://www.nerpn.org
[152] http://www.gleon.org/lakes/
[153] http://www.gleonrcn.org/index.php?pr=Products
[154] http://lterquery.limnology.wisc.edu/
[155] http://ccsbeta.colostate.edu/
[156] http://ecosystems.mbl.edu/arc/metadata_forms/MetadataBlank.xls
[157] http://swbiodiversity.org/seinet/index.php
[158] http://fcelter.fiu.edu/research/projects/projects.htm?pid=14
[159] http://fcelter.fiu.edu/research/projects/sampling.htm
[160] http://www.hubbardbrook.org/
[161] http://www.hubbardbrook.org/data/Realtime_Data/index.php
[162] http://luq.lternet.edu/data
[163] http://ites.upr.edu/EVFS/reservations.htm
[164] http://luq.lternet.edu/people/StudentdPers/PersonalDataEntryForm.html
[165] http://luq.lternet.edu/datamng/imdocs/dsetallfrm.doc
[166] http://www.mcmlter.org
[167] http://www.uaf.edu/toolik/webcam
[168] http://ecosystems.mbl.edu/ARC
[169] http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/06spring/#7fa
[170] http://ccelter.sio.ucsd.edu
[171] http://sciencestudies.ucsd.edu/
[172] http://interoperability.ucsd.edu
[173] http://chubasco.hwr.arizona.edu/ahis-drupal/
[174] http://www.cuahsi.org/
[175] http://sos.asu.edu
[176] http://fcelter.fiu.edu/research/information_management/
[177] http://fcelter.fiu.edu/research/projects/projects.htm?pid=1
[178] http://gce-lter.marsci.uga.edu/portal/
[179] http://gce-lter.marsci.uga.edu/lter/research/tools/data_toolbox.htm
[180] http://luq.lternet.edu/datamng/GotitasdelSaber/index.html
[181] mailto:lterschool@ites.upr.edu
[182] http://luq.lternet.edu/outreach/schoolyard/index.html
[183] http://gleon.org
[184] http://oceaninformatics.ucsd.edu
[185] http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/06spring/#2gr