

An Introduction to VCR/LTER Information Management Systems

John Porter – 1/24/2012

The VCR/LTER operates a variety of software running on several different computer platforms with the aim of providing information to researchers. The goal is to create systems that, once created, provide automated services with little or no day-to-day oversight, rather than solving the same problem day after day. Put differently, we want to do the job right once!

However, with changes in technology even the best system at time X will not be the best system at time X+5 and may be a totally unreasonable solution by time X+10. For this reason, we balance revamping older parts of our system with the development of new parts. Whenever possible, open source solutions are used and we try to strongly avoid software “black holes” where data goes in, but is difficult to extract later. This document describes the system as of January 2012.

Key Technologies

Web Server – We use the open-source Apache web server running on a Linux virtual machine. A few minor functions are still served by our old UNIX (Solaris) platform, but these are rare and are being transitioned to the Linux system.

SQL Databases – We operate two major SQL database systems. The MiniSQL (MSQL) system dates back to the mid-1990s. It incorporates a web scripting language (EMBER) that is used for many of the on-line forms, linked to MSQL databases. It is small, simple and fast. Most importantly, it was functional years before competing products. However, at this point we are not developing new databases in MSQL and instead are concentrating on the second major database management system: MySQL. MySQL is a larger, more complete implementation of the SQL standard. The web interface for MySQL is primarily through PHP programs. New web forms that we create now use PHP, rather than the proprietary EMBER language. However, we have not re-create all the web forms currently implemented in EMBER since that would require substantial work, with no net increase in system functionality. The MSQL database is mirrored to the MySQL database several times each day and most build-on applications operate only against the MySQL database.

Programming Languages – We use a variety of programming languages to implement different parts of the system. PHP plays an important role in web-database implementations. However, we also use PERL programs (using the standard DBI interface for databases) for document processing (independent of databases) as CGI (Common Gateway Interface) programs. We make limited use of JAVASCRIPT in web forms and for displaying sequences of images. For on-line analysis of data and both “eager” and “lazy” creation of graphics, we use the SAS statistical analysis system, and to a lesser degree SPSS and R statistical packages. We anticipate an increased use of R in the future since it is open source and extensible.

Content Management System – We use the Drupal Content Management System as the front-end for our web site. It uses our MySQL database as a back-end for data storage and provides many valuable built-in functions (e.g., bibliography, password access control). However, because the data model for Drupal is complex, documents that we believe have archival value have “master versions” stored in a traditional file system, with documents grouped into directories.

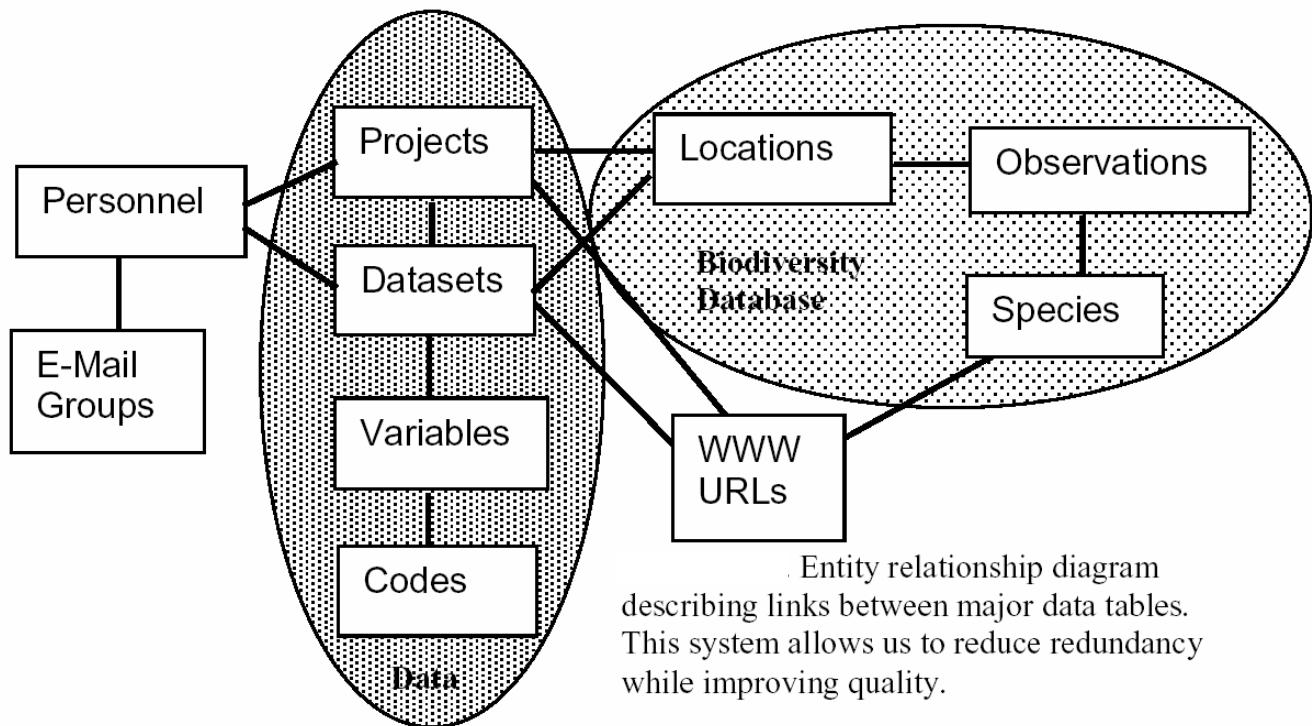
GIS – For online maps we use MAPSERVER, which is an open-source tool. Although not as “flashy”

as some commercial products, it is quite stable and provides the critical functionality we need. ESRI ArcGIS and ERDAS Imagine software are used for off-line processing to create the needed data layers for MAPSERVER.

Data Catalog – The data catalog is EML-based. It uses search services from the LTER Metacat and displays datasets using stylesheets that transform the raw EML into the web pages for display.

Data Model

The data model for VCR/LTER metadata uses normalization and multiply-linked tables to eliminate redundancy, so that researchers need only enter a single piece of information in one location, regardless of the number of products that piece of information is used in (figure 1). For example, the personnel table is linked to both Project and Datasets tables, so that a change in address recorded in the personnel table immediately is updated when projects or datasets are displayed. Similarly, Locations and URL tables are shared across functions.



Locally-written Software Tools

We have a number of locally-written software tools that we use to facilitate system functions:

- **Metadata database to EML** – a PERL program uses standard DBI calls to extract data from the metadatabase to populate an XML document following Ecological Metadata Language specifications. It simultaneously creates an EML Harvester Document used to update the LTER Metacat several times each week.
- **EML to Statistical Package Program** – we have developed XML Stylesheets that convert our EML documents into SAS, SPSS and R statistical programs. With minor editing, the programs created can be run and include input of the data, labeling and rudimentary statistical analyses. A web page is available at: <http://www1.vcrlter.virginia.edu/data/eml2/eml2stat.html> for a tool

using the stylesheets.

- **Web forms to Text Documents** – the process_doc PERL program operates as a CGI program to facilitate creation of text documents from web forms. We use this program extensively for capturing special purpose data that are needed for analysis or display, but not for updating or editing (e.g., annual research reports). A second set of PERL programs automate the production of the templates used by process_doc by analyzing the web form document. These tasks can now also be accomplished using XML-based tools. However, these tools were unavailable when we started using process_doc in 1995-1996.
- **Webcam Image Database** – A MySQL database is used to store information on over two million images from our webcams, including location, time, size of image and the filesystem path to the image. PHP is used for web forms and pages that produce animation and index displays on command.

Milestones

Here is a list of milestones in the continuing evolution of the VCR/LTER information management system:

- 1989- Metadata system created using Dbase III
- 1990- GIS Lab Established
- 1990 - Data Management Policy
- 1992 - Electronic Mail Calendar
- 1992 - Gopher Information Server
- 1993 - WWW Server
- 1994 - Online Research Summaries
- 1995 - Web-based Personnel Directory
- 1996 - Automated System for Research Summaries
- 1996 – ClimDB harvest document created
- 1996 - Biodiversity Database
- 1997 - Web form-based Information Management Tools, Dbase III system ported to MiniSQL
- 1999 - Automated Statistical Programs
- 2000 – EML 1.4 Metadata
- 2001 – ClimDB harvest document revised
- 2002 – Wireless Internet connection to island field site
- 2003 – Mapserv online maps created
- 2004 – Upgrade of computer systems
- 2004 – EML 2.1 Metadata
- 2005 – Web Page revised using PostNuke Content Management System
- 2005 – EML to SAS, SPSS and R software converters
- 2007 – Hog well wireless network installed
- 2008 – Upgrade of computers to Linux Virtual Machines
- 2009 – Web Page revised Drupal Content Management System
- 2010 – EML 2.1 Metadata
- 2011 – Web service for converting EML to R, SAS and SPSS programs
- 2010 – 40% of datasets made accessible through LTER Data Access Server
- 2011 – Keywords converted to LTER-standard keywords, “suggest keyword” tool added to forms
- 2012 – Dataset displays shifted from locally-written programs to XSL transforms of EML