Metadata database models and EML creation at LTER sites

<coverage> <attribute> <dataTable> <contact> <series> <unit> <dataset> <access> <project> <citation> <purpose> <maintenance> <eml> <methods>

– compiled by M.Gastil-Buhl (MCR) from contributions by D.Henshaw & S.Remillard (AND),
J.Laundre (ARC), J.Walsh (BES), P.Tarrant (CAP), K.Baker, M.Kortz & J.Conners (CCE/PAL), D.Bahauddin (CDR), J.Chamblee (CWT), L.Powell (FCE), W.Sheldon (GCE), J.Campbell (HBR), E.Boose (HFR), K.Ramsey (JRN), S.Bohm (KBS), A.Skibbe (KNZ), E.Melendez-Colom (LUQ), S.Welch (MCM), C.Gries (NTL), H.Humphries (NWT), H.Garritt (PIE), M.O'Brien (SBC), K.Vanderbilt (SEV), N.Kaplan (SGS), J.Porter (VCR), I.San Gil (LNO/NBII)

Abstract

The purpose here is to spark discussion. Preparing for data integration, we will each examine our IM System to ask if it will meet potential new metrics. Some LTER sites already generate PASTA-ready EML. Will their design work at my site?

Survey

Please correct or update your site information. Write directly on this participatory poster.

	How is metadata stored?	How is EML created?	Past or Potential collab-oration	Drupal today?	EML populates local data catalog?	EML document→dataTable ratio
AND Don Henshaw, Suzanne Remillard, Theresa Valentine	SQLServer 2005 (input via ASP.NET)	scripted	BNZ, KNZ		no	1:n
ARC Jim Laundre	text now; future Drupal	Excel forms modified from FCE now; future Drupal	FCE past; PIE, LUQ, SEV, NTL current	yes	no	1:1 data table but 1:n entity
BES Jonathan Walsh	SQL Server 2005	Scripted (ASP)	LNO, LUQ, AND, BNZ		no	1:n (estimate)
BNZ Jason Downing	MySQL	Perl script harvesting data from a mySQL database	LNO, ARC		no	f:n (most are 1:1)
CAP Philip Tarrant	EML files are stored in XMLdb (eXist). Metadata for each dataset are also stored in MySQL. (reverse engineered)	PHP script from standard metadata tables plus data table descriptions in MySQL. EML for spatial data is created via ESRI metadata editor and conversion XSLT to EML.			yes	
CCE Karen Baker, Mason Kortz, James Conners	MySQL	scripted: PHP	PAL, LNO		no	1:1
CDR Dan Bahauddin	Oracle					glance appears 1:1
CWT John Chamblee	Adopted GCE Metabase in SQL-Server 2005 (as of 2010)	currently manually. New system will generate metadata thru GCE toolbox, to db, then via stored proc. to EML	GCE now, possibly MCR & SBC			glance appears 1:1
FCE Linda Powell	Metadata stored as Excel EML templates, EML xml files; and some metadata entered into Oracle DB	Excel Metadata Template which is converted into valid EML xml documents using the FCE Excel2EML tool.	Excel form adopted by other sites			glance appears 1:1
GCE Wade Sheldon	SQL Server 2005	scripted				glance appears 1:1
HBR John Campbell	MySQL					
HFR Emery Boose	EML files stored in eXist	Manually using XMLSpy and Morpho	SBC, CAP, GCE	yes		1:n
JRN Ken Ramsey	[text + SQL Server 2005] migrating to SQL Server 2008	xml stylesheets with php				1:n
KBS Sven Bohm	Postgresql	scripted Ruby (on rails)			no	appears 1:0
KNZ Adam Skibbe	SQL Server 2008	an XML editor; this will change	presently in-house; future open	yes	mix; indirectly. Expect change.	1:n
LUQ Eda Melendez-Colom	was fileserver docs; now Drupal MySQL	future Drupal	ARC, PIE, LNO, SEV	yes		1:n
MCM Sue Welch	Oracle for data; flat files for metadata	manual in oXygen				
MCR M.Gastil-Buhl	PostgreSQL	manual in oXygen and Morpho	SBC, GCE, CWT	yes		1:n
NTL Corinna Gries	was: Oracle for database stored data and EML files for spatial data in file system. Now metadata stored in MySQL back end to Drupal. Future (2011) GIS in progress.	was: oracle metadata data are generated by a JavaScrip script. EML, LNO, SEV, for spatial data is created via ESRI metadata editor and conversion XSLT to EML. New: Kepler workflow.	ARC, PIE, LNO, SEV, VCR		no, we use 1:n queries into the Oracle metadata tables	1:n
NWT Hope Humphries	MS SQL Server	a script that reads from text files			no	glance appears 1:n
PAL Karen Baker, Mason Kortz, James Conners	MySQL	scripted: PHP	CCE, LNO		no	1:1
PIE Hap Garritt	text, will be migrating to Drupal	now Excel forms from FCE; future Drupal	FCE past; ARC, LUQ, NTL, SEV; future Drupal	yes	no	1:1
SBC Margaret O'Brien	text + xml (eXist)	some EML is scripted (with bash/perl, and Matlab), some	MCR, CAP, GCE	yes	yes	1:n
SEV Kristin Vanderbilt	was MySQL, now Drupal + MySQL	Was text file to EML via Perl script. Future Drupal	LUQ, ARC, PIE, LNO, NTL	yes		1:n
SGS Nicole Kaplan	Access for web and MySQL	export xml from db, then run perl script to parse xml, and then xalan to map to EML	LNO, CCE/PAL	no. Db query	glance appears 1:1
VCR John Porter	MySQL + legacy msql	scripted Perl, web forms	ARC, LUQ, NTL, PIE, SEV, LNO			glance appears 1:1

Summary

19 sites use a relational database system for metadata (8 MySQL, 7 SQL-Server, 4 Oracle). Of these, 14 sites generate EML from their metadata RDB. 6 sites plan to use Drupal.

5 sites serve a local data catalog from EML. 13 sites have multiple data tables within single EML documents.

Scope

Here we focus only on metadata, not the data per se. Metadata-data congruency can be enhanced when the data are coordinated within the meta-data system. So this is an incomplete picture.

GIS is not covered here.

Commonalities

All LTER sites share common things. Entity-Relationship diagrams show how these things are related. Each thing corresponds to one or more EML elements.

All sites need to present metadata on websites, in EML documents and other uses, such as other metadata standards.

Generic

Longevity and Continuing Design

Some LTER sites' models designed in the 1990s are still in use today, such as at VCR and AND, having migrated to new servers and new applications as technology changed. They remain useful because their schemata inherently model the characteristics of metadata and through continuing design to keep pace with evolving standards.

Mature Models

DataZoo at CCE/PAL, GCE Metabase and AND Metadata Database are three examples of mature models, in production, and part of a larger IM System at these LTER sites. These models continue to undergo improvements. Web page display is just one of their uses. EML is currently generated by scripts from all three of these metadata databases. The AND and GCE metadata model designs pre-dated EML; the extraction of EML was developed after the initial design. EML is just one of several metadata standards these three are designed to serve. All three undergo continuing development.

EML generated from the constrained model of a database is more likely to meet future metrics, especially if the data itself is filtered through a connected system.

Metabase collects data descriptions as part of a data ingest application.

DataZoo uses a data access layer to synchronize data with its metadata.

GCE Metabase

CCE/PAL DataZoo

Future

Web services add options for development and use of data and metadata. The Unit Registry web service will soon be followed by the Controlled Vocabulary of Keywords and then subsequently by the NIS Administrative modules (bibliography and personnel). With this approach, sites may connect to services, replacing or synchronizing those parts of their local database. How will this affect our metadata database architecture?

Several sites are looking to participate in future development of metadata data models.

The GCE Metabase has been adopted by CWT and is planned to be ported to PostgreSQL at MCR and SBC.

Six LTER sites (LUQ, SEV, PIE, ARC, NTL, VCR) are pooling resources to develop a Drupal-based metadata storage, display and EML creation system.

Parts of the metadata data model represented in the MySQL tables back end to the Drupal system as Content Types are color coded to match the figure below.

Legacy EML from LUQ, SEV, and NTL has been uploaded to the Drupal back-end database. This is now in use to serve their web pages. Export to EML is being programmed currently.

Drupal Environmental Information Management System (DEIMS)

EML Specification

Acknowledgement