From Scientist and Sensor to Synthesis: Metabase Metadata Management System

Wade Sheldon
Georgia Coastal Ecosystems LTER

John Chamblee & Richard Cary
Coweeta LTER
**Metabase MMS**

- Metabase is a generalized RDBMS for managing LTER metadata:
  - Personnel
  - Site geography (study area polygons, point locations)
  - Instrumentation
  - Research Projects
  - Data sets (studies, methods, entities, attributes, files)

- Linked to Biblio and Taxonomic DBs
  - Supports automatic cross-links between people/research/pubs and data
  - Supports reciprocal queries

- Allows centralized management of project info

- Supports automated metadata generation for data sets, cross-links between all related information

- Used by GCE & CWT (MCR & SBC adopting model)
Metabase Model

- Dataset metadata based on ESA-FLED, parallels GCE Data Toolbox
- “Strongly Typed” model
  - Pluses (easy to understand & program against)
  - Minuses (rigid)
  - Works well for LTER since metadata content model well established
- Long history
  - Version 1 (GCE_Metabase) designed in 2001 (in production 2001-2010)
  - Version 2 designed in collaboration with CWT, SBC, MCR in 2010-2011
- Metabase2 Supports many data archiving patterns:
  - Simple data tables (1 data set = 1 entity/table)
  - Multi-table data sets (1 data set > many entities/tables)
  - Non-tabular data (1 data set > 1+ non-tabular entities/files w/ or w/o attributes)
  - Mixed tabular/non-tabular
Metabase Model

ERD online: http://gce-lter.marsci.uga.edu/public/app/resource_details.asp?id=434
Metabase Services

- **Web applications**
  - Data Cataloging ([http://gce-lter.marsci.uga.edu/public/app/data_search.asp](http://gce-lter.marsci.uga.edu/public/app/data_search.asp))
  - Personnel ([http://gce-lter.marsci.uga.edu/public/app/personnel.asp](http://gce-lter.marsci.uga.edu/public/app/personnel.asp))

- **Web services**
  - EML generation (multiple implementations)
  - EML harvest list
  - Personnel (various schemas, XML/JSON)
  - Datasets (lists, details, search)
  - Research Projects (eXist/xquery)
  - KML web services (sites, data sets, search)
  - ...
Content Management

- MS Access forms (ADP) – all content
Content Management

- Web Forms – some content
Content Management

- **GCE Data Toolbox**
  - Import content
    - Geography (all)
    - Data set metadata (all)
    - Personnel (all)
  - Update content
    - Geography (most)
    - Data set metadata (most)

- **PyGIS**
  - Import content
    - Geography (all)
    - Data set metadata (some)
Metabase – GCE Data Toolbox Integration

- Supports complete data lifecycle management
  - Metabase houses research context metadata (projects, people, sites, ...)
  - Toolbox used to create/manage/QC tabular data
  - Push/Pull content sync provides both RDBMS/web and MATLAB access
  - Automated revision management and publishing to Metacat/NIS
  - Automated access control and tracking

- Expanding support for EML provides more options in future
  - GCE Toolbox can be used as PASTA workflow engine
  - Automated ingestion of EML-described data to populate Metabase
  - Generation of PASTA-ready EML from sensors
Future Plans

- Ongoing work
  - GCE Toolbox metadata enhancements to support bidirectional sync
  - Completion of Metabase web services for data catalog
  - Web forms for data submission, tracking

- Future opportunities
  - MS SQL and Postgres implementations of Metabase available
  - MS SQL & middleware hosting at UGA (or LNO)?
  - “Community” Metabase instance/middleware populated from EML?
  - Other ideas?