

From Scientist and Sensor to Synthesis: Metabase Metadata Management System

Wade Sheldon

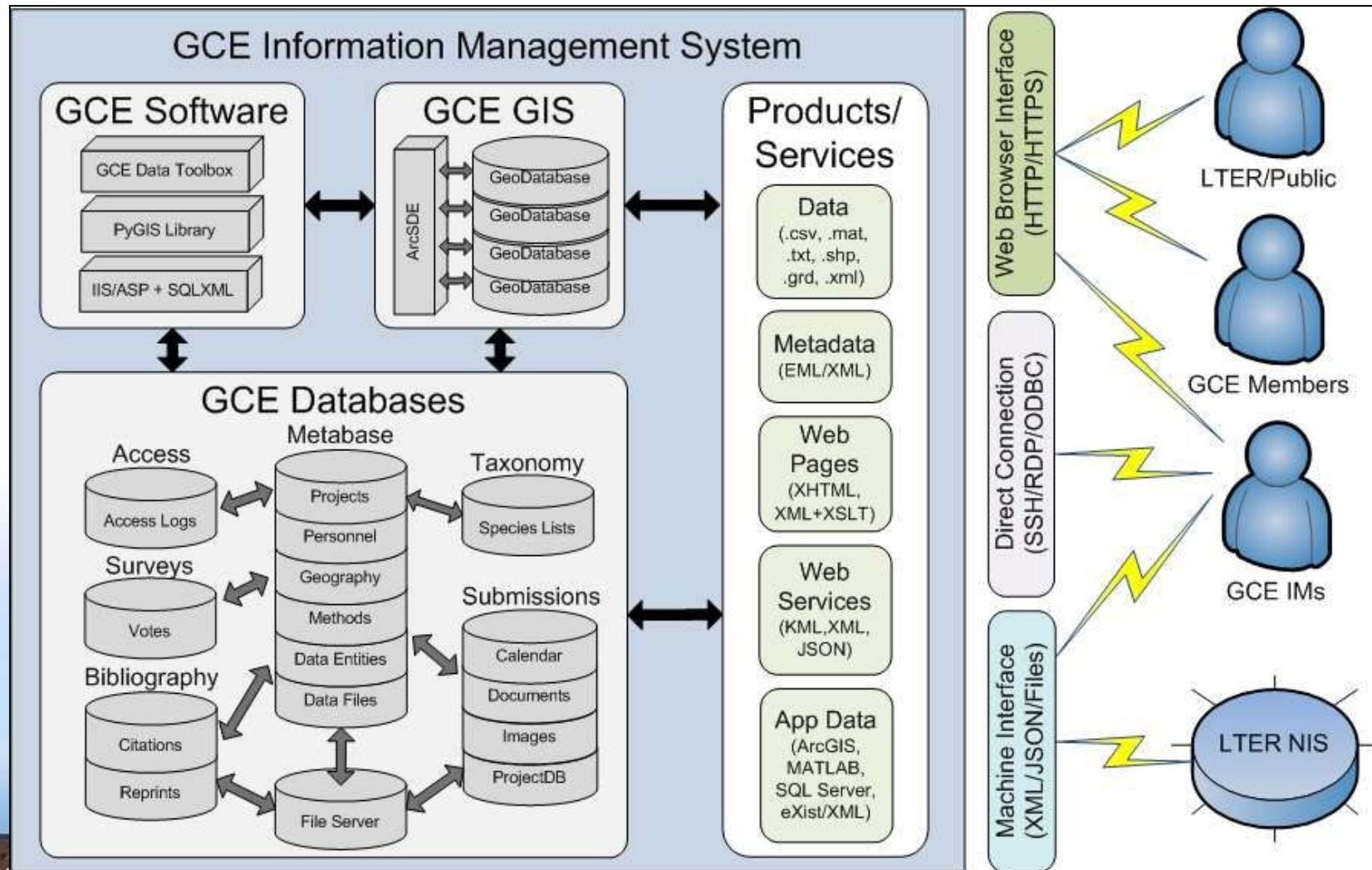
Georgia Coastal Ecosystems LTER

John Chamblee & Richard Cary

Coweeta LTER

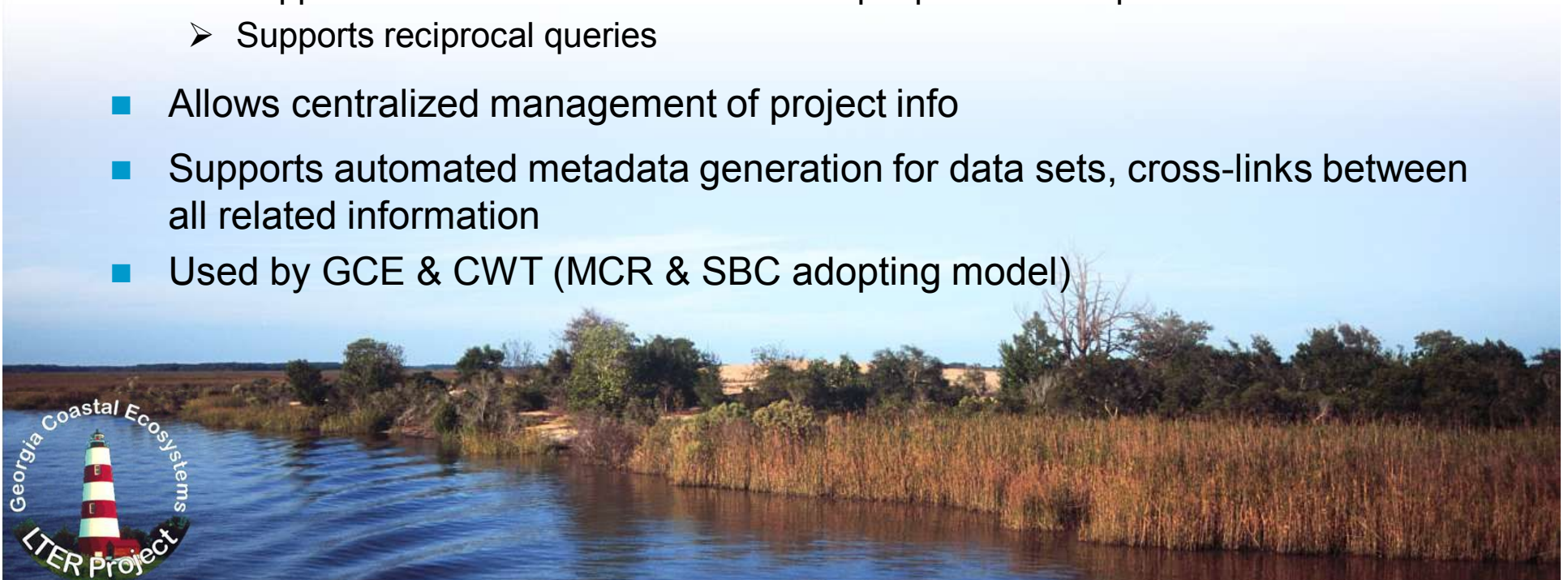
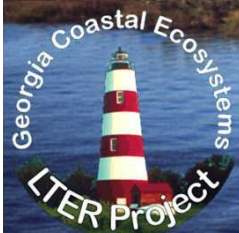


GCE IMS



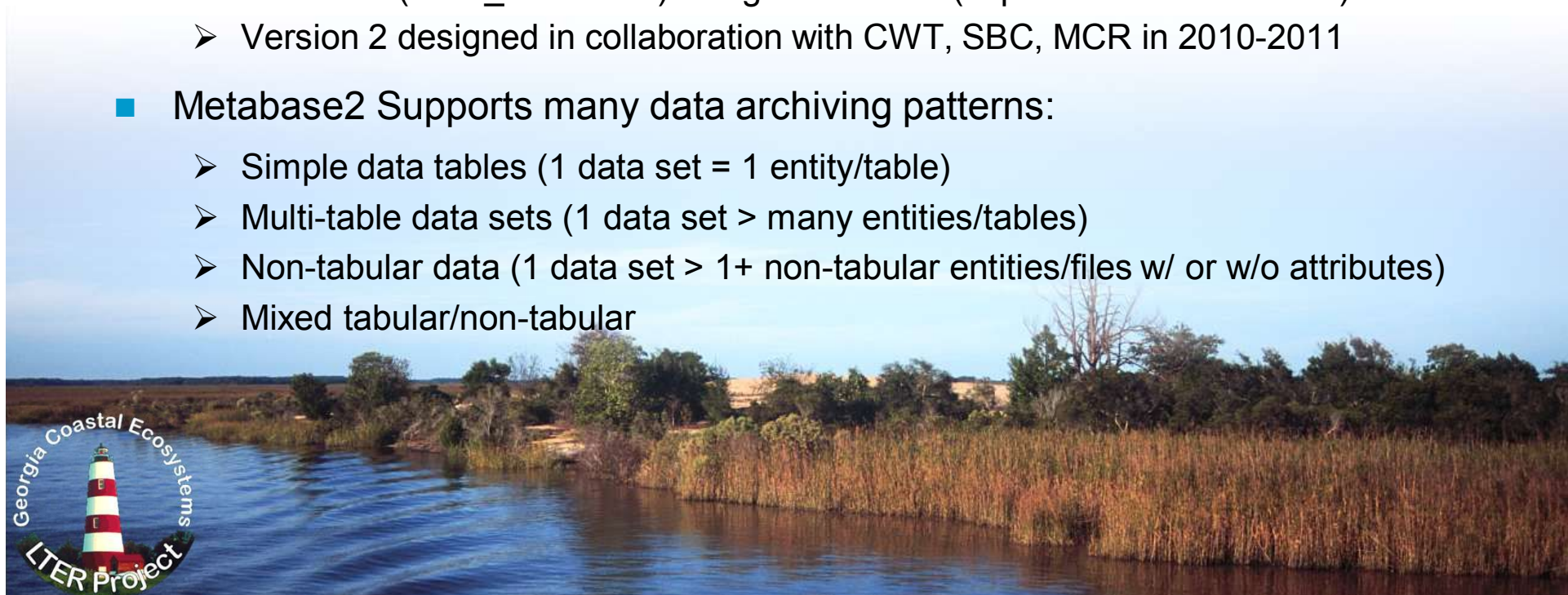
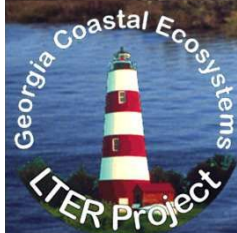
Metabase MMS

- 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



- Georgia C



Metabase Services

■ Web applications

- Data Cataloging (http://gce-lter.marsci.uga.edu/public/app/data_search.asp)
- Personnel (<http://gce-lter.marsci.uga.edu/public/app/personnel.asp>)
- Study Sites (<http://gce-lter.marsci.uga.edu/public/app/sitedetails.asp?site=GCE6>)
- Location Search (http://gce-lter.marsci.uga.edu/public/app/geo_query.asp)
- Research Projects (<http://gce-lter.marsci.uga.edu/public/research/projects.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

Data Set Metadata

General Keywords Studies Methods Supplementary References Dependencies Entities

DataSetID: 430 Accession: INV-GCEM-1208a2 ☒ Display On Web

Contributor: Alber, Merryl Study Type: GCE-II - Monitoring ☒ Georeferenced

Data Set Title: Mollusc population size distribution monitoring: Fall 2011 mid-marsh and creekbank infaunal and epifaunal mollusc size distributions based on collections from GCE marsh monitoring sites 1-10

Abstract: This data set is the Fall 2011 report of infaunal and epifaunal mollusk species size distributions at the GCE-LTER marsh sites used for population monitoring. Infaunal and epifaunal molluscs were hand-collected from within quadrats of known area from mid-marsh and creekbank zones (n = 4 quadrats per zone) at all sites. The molluscs were returned to the lab, preserved in ethanol, measured and counted (count data is reported separately). Length of each measurable individual was determined using calipers or an ocular micrometer mounted in a stereomicroscope. Species abundance and density data for these collections may be found in the GCE-LTER data set INV-GCEM-1208a1. Numbers of individuals of each species in the abundance data file may not correspond exactly to the numbers of individuals in the size data file because some individuals may not have been measureable.

Submit Date: 8/22/2012 Project Release: 8/22/2012 Public Release: 8/22/2013 (2 years after collection, 1 year for monitoring)

Data Set Themes

Theme Code	Primary
Aquatic Invertebrate Ecology	☒
*	☐

Data Set Personnel

Researcher Name	Order	Dataset Role
Alber, Merryl	1	Investigator
Pennings, Steven C.	2	Investigator
Shalack, Jacob	3	research coordinator
Reddy, Caroline M.	4	research technician
Manley, Justin P.	5	research technician
Guo, Hongyu	6	postdoctoral associate
Wieski, Kazimierz	7	postdoctoral associate
Penn, Aaron	8	sample processor
*		

Review Status: New submission

Record: 14 346 of 348 No Filter Search



Content Management

■ Web Forms – some content

GCE-LTER Program Website x

https://gce-lter.marsci.uga.edu/private/app/update_bio.asp

You are logged onto the Private GCE Website as: **Wade Sheldon**

Georgia Coastal Ecosystems LTER
Private Website for GCE-LTER Participants and Guests

Home > Private Site > [Change Password](#) | [Update Bio](#)

Home
GCE News >
Research >
Study Site >
Field Planning >
Bibliography >
Data Products >
GIS Resources >
Maps & Imagery >
Documents >
Outreach & Ed >
Informatics >
Personnel >
Affiliates >
Private Site >

Update Personnel Information *(highlighted fields are required)*

Select a personnel record to update: Sheldon, Wade M.

Name: Honoric: First: Middle: Last: Suffix:
Display Name:

GCE Role: GCE-II - Information Manager *(contact the [GCE IM](#) to revise)*

Organization:

Address: Line 1:
Line 2:
Line 3:
City: State: Country: Zip:

Email:

Web Page:

Phone/FAX: Office: Other: FAX:

Demographics: Gender: ☒ Male ☐ Female
(only for NSF) Race: Ethnicity:
Citizenship: Disability:

Photo: 

Content Management

■ GCE Data Toolbox

➤ Import content

- Geography (all)
- Data set metadata (all)
- Personnel (all)

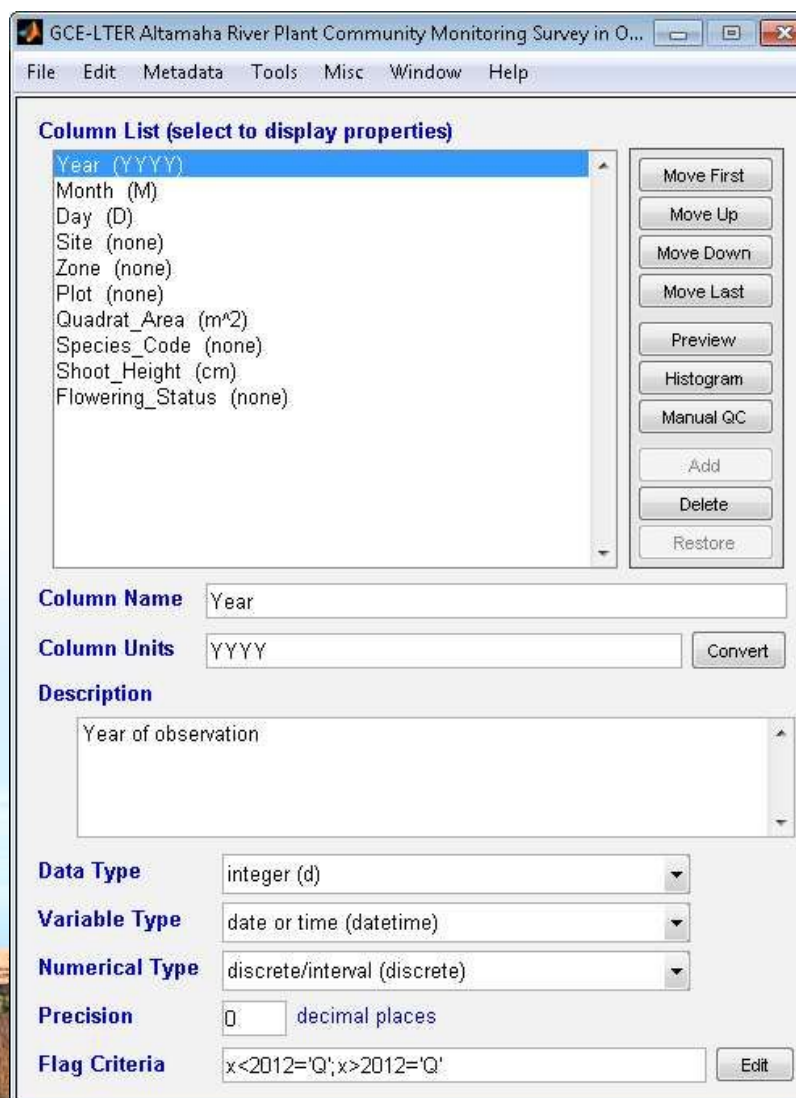
➤ Update content

- Geography (most)
- Data set metadata (most)

■ PyGIS

➤ Import content

- Geography (all)
- Data set metadata (some)



The screenshot shows a software window titled "GCE-LTER Altamaha River Plant Community Monitoring Survey in O...". The window has a menu bar with "File", "Edit", "Metadata", "Tools", "Misc", "Window", and "Help". The main area is divided into two sections. The top section, titled "Column List (select to display properties)", contains a list of columns: "Year (YYYY)", "Month (M)", "Day (D)", "Site (none)", "Zone (none)", "Plot (none)", "Quadrat_Area (m^2)", "Species_Code (none)", "Shoot_Height (cm)", and "Flowering_Status (none)". To the right of this list are buttons: "Move First", "Move Up", "Move Down", "Move Last", "Preview", "Histogram", "Manual QC", "Add", "Delete", and "Restore". The bottom section, titled "Column Name", "Column Units", "Description", "Data Type", "Variable Type", "Numerical Type", "Precision", and "Flag Criteria", contains fields and buttons for configuring the selected column. The "Column Name" field contains "Year", the "Column Units" field contains "YYYY", and the "Convert" button is visible. The "Description" field contains "Year of observation". The "Data Type" dropdown is set to "integer (d)", the "Variable Type" dropdown is set to "date or time (datetime)", the "Numerical Type" dropdown is set to "discrete/interval (discrete)", the "Precision" field is set to "0 decimal places", and the "Flag Criteria" field contains "x<2012='Q';x>2012='Q'" with an "Edit" button.

Column Name	Column Units	Description	Data Type	Variable Type	Numerical Type	Precision	Flag Criteria
Year (YYYY)	YYYY	Year of observation	integer (d)	date or time (datetime)	discrete/interval (discrete)	0 decimal places	x<2012='Q';x>2012='Q'



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?

