

GCE Data Toolbox for MATLAB – An Introduction

Wade Sheldon
Georgia Coastal Ecosystems LTER



Background & Motivation

- Georgia Coastal Ecosystems LTER project started in May 2000
 - Major data collection effort
 - NSF & LTER require data archiving and sharing
 - LTER requires detailed “metadata” for every data set
 - Needed to standardize data processing, quality control, documentation
- No ready-to-use software for LTER data management
 - Lab management software (LIMS) useless for field data, expensive
 - Most LTER sites were using “flat files” – limiting
 - A few sites using relational databases, client/server apps – proprietary, complex, unfamiliar, require constant network access
- Chose to develop custom data management software (MATLAB)
 - Experienced using MATLAB for automating data processing, GUIs
 - Better code-reuse potential than database/web solution
 - Best compromise: file-based but supports fully dynamic operations



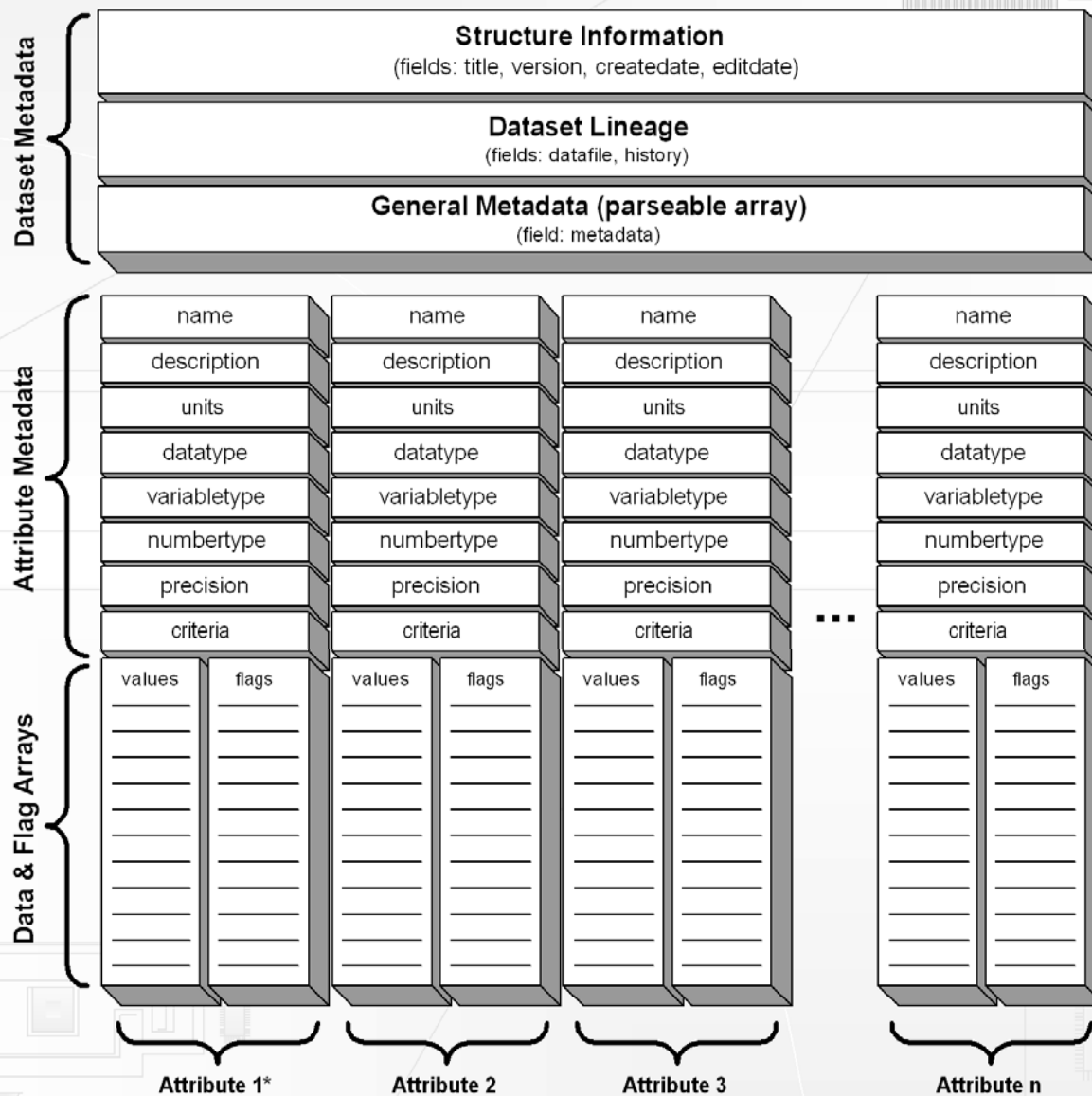
GCE Data Model

- Started by reviewing ESA's "FLED" report
 - Gross, Katherine L. and Catherine E. Pake. 1995. Final report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED). Volume I: Text of the Report. The Ecological Society of America, Washington, D.C.
- Identified information storage requirements
 - Any number of numeric (integer, float, exponential) and text variables
 - Structured attribute metadata for each variable (name, units, desc., type, precision, ...)
 - Structured documentation (dataset metadata) for dynamic updating, formatting
 - Versioning and processing history info (lineage)
 - Added later: quality control rules for every variable, flags for every value
- Designed data model: "GCE Data Structure"
 - MATLAB "struct" array with named fields for each class of information
 - Detailed specifications for allowed content in each field
 - "Virtual table" design based on matched arrays for linking attribute metadata, data, flags
 - Same philosophy as relational database table plus additional descriptors



Conceptual Model of the GCE Data Structure Specification (version 1.1, 29-Mar-2001)

Note: GCE Data Structures are created and managed using the GCE Data Toolbox, a MATLAB software library for metadata-based analysis, visualization, transformation and management of ecological data sets
(url: http://qce-lter.marsci.uqa.edu/lter/research/tools/data_toolbox.htm)

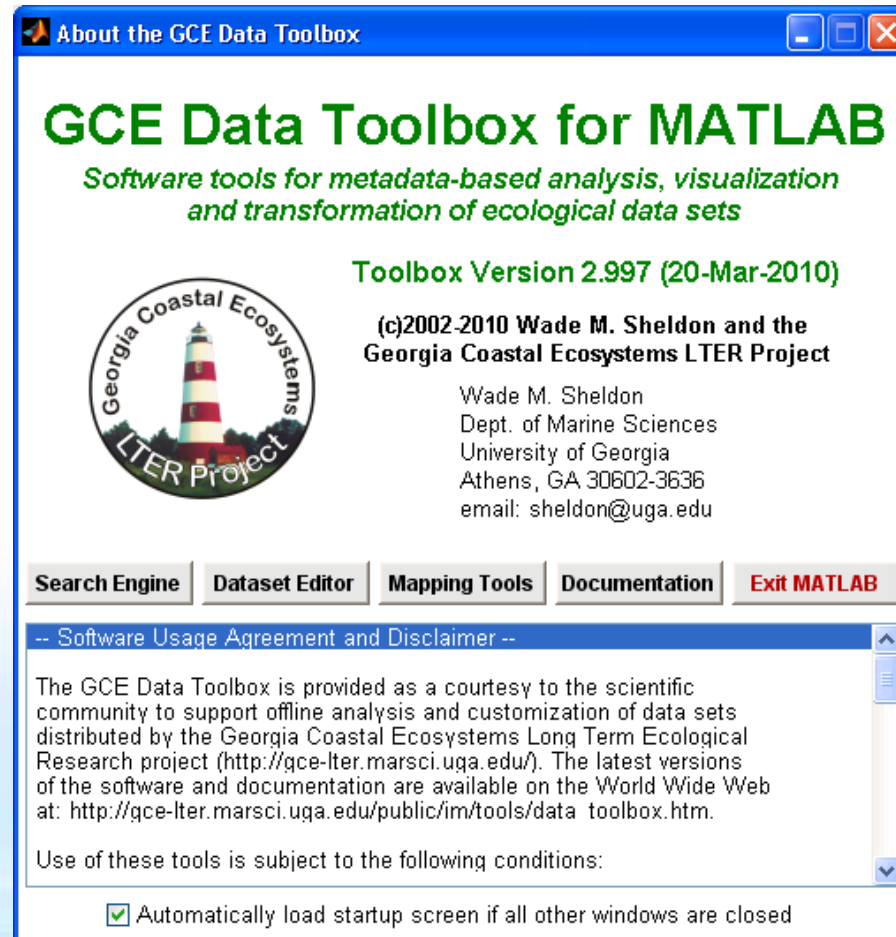


Toolbox Development

- Developed MATLAB software library to work with data structures
 - Utility functions to abstract low-level operations (API)
 - Create structure, add/delete columns, copy/delete rows
 - Extract, sort, query, update data, update flags
 - Analytical functions for high-level operations
 - Statistics, visualizations, geographic & date/time transformations
 - Unit inter-conversions, aggregation/re-sampling, joining data sets
 - GUI interface functions to simplify using the toolbox
 - All functions use metadata, data introspection to auto-parameterize and automate operations (semantic processing)
- Developed indexing and search support (and GUI search engine)



Startup Window (ui_aboutgce.m)



Search Engine (ui_search_data.m)

GCE Data Search Engine (1202 data sets)

File Options Tools Help

Indexed Directories

D:\GCE_Datasets\catalog (316 data sets)
D:\GCE_Datasets\portal (246 data sets)
D:\GCE_Datasets\provisional (76 data sets)
http://gce-iter.marsci.uga.edu/iter/asp/dbi/ (564 data sets)

Search Criteria Subdirectories?

Any Text Contains: Study Site:

Date Range: to Bounding Box: N Lat
W Lon Map E Lon
S Lat Datasets enclosed by boun...
 Datasets overlapping boun...

Author Name:

Keywords: (multiple)

Species Names: (multiple)

Data Columns: (multiple)
(0 selected)
Agency
Alkalinity
Alkalinity_Tot
Aluminum_Conc

Search Options:
 Match all criteria Match any criteria
 Case sensitive text searches
 Save new queries to history list

Clear Bounds
Clear All
SEARCH

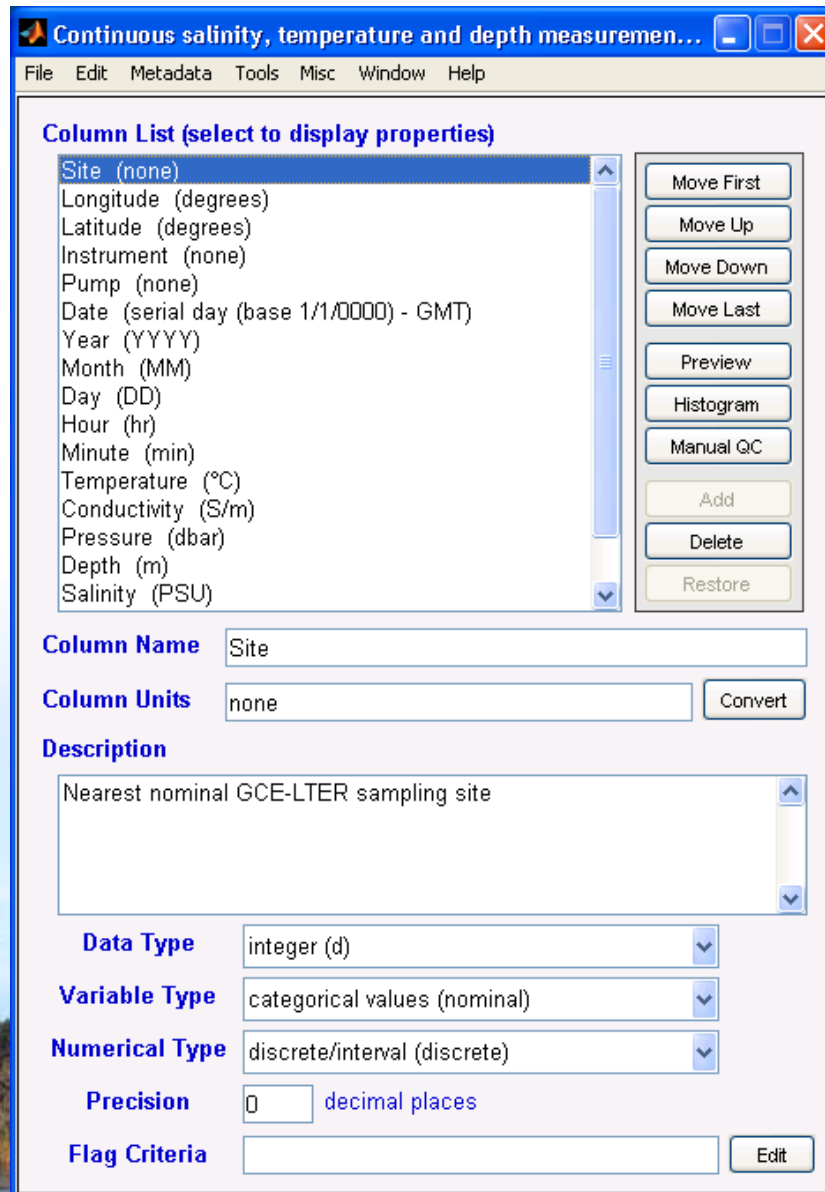
Cumulative Search Results (8 matches)
(double click on title to view metadata)

local/ [no accession] - Continuous salinity, temperature and depth measurements from moored hydrographic data loggers deployed
local/ [no accession] - Daily Summary of Continuous salinity, temperature and depth measurements from moored hydrographic data
local/ [no accession] - Daily Summary of Data from Sea-Bird Electronics 37-SM MicroCAT sonde S/N 1795 deployed at the Hammers
local/ [no accession] - Daily Summary of Data from Sea-Bird Electronics 37-SM MicroCAT sondes deployed at the Hammersmith Cre
local/ [no accession] - Data from Sea-Bird Electronics 37-SM MicroCAT sonde S/N 1795 deployed at the Hammersmith Creek hydrog
local/ [no accession] - Data from Sea-Bird Electronics 37-SM MicroCAT sondes deployed at the Hammersmith Creek hydrographic d
web/ portal - Continuous salinity, temperature and depth measurements from moored hydrographic data loggers deployed at GCE7,
web/ portal - Daily Summary of Continuous salinity, temperature and depth measurements from moored hydrographic data loggers

Sort
Select All
Select None
Remove
Remove All
View/Edit
Plot X/YY
Plot Groups
Map Plot
Summary



Data Editor (ui_editor.m)



Continuous salinity, temperature and depth measurement...

File Edit Metadata Tools Misc Window Help

Column List (select to display properties)

- Site (none) **▲**
- Longitude (degrees)
- Latitude (degrees)
- Instrument (none)
- Pump (none)
- Date (serial day (base 1/1/0000) - GMT)
- Year (YYYY)
- Month (MM)
- Day (DD)
- Hour (hr)
- Minute (min)
- Temperature (°C)
- Conductivity (S/m)
- Pressure (dbar)
- Depth (m)
- Salinity (PSU) **▼**

Move First
Move Up
Move Down
Move Last
Preview
Histogram
Manual QC
Add
Delete
Restore

Column Name Site

Column Units none

Description

Nearest nominal GCE-LTER sampling site

Data Type integer (d) **▼**

Variable Type categorical values (nominal) **▼**

Numerical Type discrete/interval (discrete) **▼**

Precision decimal places

Flag Criteria



Data Viewer (ui_datagrid.m)

All	Site	Longitude	Latitude	Instrument	Pump	Date	Year	Month	Day
None	(none)	(degrees)	(degrees)	(none)	(none)	(serial day (base	(YYYY)	(MM)	(DD)
<input type="checkbox"/> 1	7	-81.475500	31.338383	2398	none	0 733043.000000	2007	1	1
<input type="checkbox"/> 2	7	-81.475500	31.338383	2398	none	0 733043.020833	2007	1	1
<input type="checkbox"/> 3	7	-81.475500	31.338383	2398	none	0 733043.041667	2007	1	1
<input type="checkbox"/> 4	7	-81.475500	31.338383	2398	none	0 733043.062500	2007	1	1
<input type="checkbox"/> 5	7	-81.475500	31.338383	2398	none	0 733043.083333	2007	1	1
<input type="checkbox"/> 6	7	-81.475500	31.338383	2398	none	0 733043.104167	2007	1	1
<input type="checkbox"/> 7	7	-81.475500	31.338383	2398	none	0 733043.125000	2007	1	1
<input type="checkbox"/> 8	7	-81.475500	31.338383	2398	none	0 733043.145833	2007	1	1
<input type="checkbox"/> 9	7	-81.475500	31.338383	2398	none	0 733043.166667	2007	1	1
<input type="checkbox"/> 10	7	-81.475500	31.338383	2398	none	0 733043.187500	2007	1	1
<input type="checkbox"/> 11	7	-81.475500	31.338383	2398	none	0 733043.208333	2007	1	1
<input type="checkbox"/> 12	7	-81.475500	31.338383	2398	none	0 733043.229167	2007	1	1
<input type="checkbox"/> 13	7	-81.475500	31.338383	2398	none	0 733043.250000	2007	1	1
<input type="checkbox"/> 14	7	-81.475500	31.338383	2398	none	0 733043.270833	2007	1	1
<input type="checkbox"/> 15	7	-81.475500	31.338383	2398	none	0 733043.291667	2007	1	1
<input type="checkbox"/> 16	7	-81.475500	31.338383	2398	none	0 733043.312500	2007	1	1
<input type="checkbox"/> 17	7	-81.475500	31.338383	2398	none	0 733043.333333	2007	1	1
<input type="checkbox"/> 18	7	-81.475500	31.338383	2398	none	0 733043.354167	2007	1	1
<input type="checkbox"/> 19	7	-81.475500	31.338383	2398	none	0 733043.375000	2007	1	1
<input type="checkbox"/> 20	7	-81.475500	31.338383	2398	none	0 733043.395833	2007	1	1
<input type="checkbox"/> 21	7	-81.475500	31.338383	2398	none	0 733043.416667	2007	1	1
<input type="checkbox"/> 22	7	-81.475500	31.338383	2398	none	0 733043.437500	2007	1	1
<input type="checkbox"/> 23	7	-81.475500	31.338383	2398	none	0 733043.458333	2007	1	1
<input type="checkbox"/> 24	7	-81.475500	31.338383	2398	none	0 733043.479167	2007	1	1
<input type="checkbox"/> 25	7	-81.475500	31.338383	2398	none	0 733043.500000	2007	1	1

Current Toolbox Use

- Major use is GCE data processing, harvesting, distribution
- Also used for general data mining, instrument data acquisition
 - Internet Import: USGS NWIS, NOAA HADS, LTER ClimDB
 - File Import: NOAA NCDC, NERR CDMO, generic MATLAB, generic text
 - Instruments: Seabird ctd/sonde, Campbell loggers, Hydrolab, Aquatroll, Hobo Tidbit
- Being used by other LTER sites (CWT, SEV, PIE?, VCR?)
- ~3000 downloads by non-GCE web visitors to date
- Potentially very beneficial for GCE data users, but under-used
 - Customize data set layout, format for analyses
 - Handle QA/QC flagged and missing values
 - Re-sample data (date-time aggregation, grouping, binning)
 - Convert units, harmonize column names for comparison with other data



Getting the Toolbox

- Public “compiled” & GCE “source” versions available on GCE web site
 - http://gce-lter.marsci.uga.edu/public/im/tools/toolbox_download.htm
- Source code provided to other LTER sites, collaborators on request
- Code copyrighted to control redistribution
 - Download/usage tracking for NSF
 - Prevent “forking” development, external dependencies
 - Preserve funding opportunities (GCE not funded for software dev)
- New: Software development site/SVN repository
 - https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox
 - Requires login - email sheldon@uga.edu for access

