

The GCE Data Toolbox for MATLAB

Wade Sheldon

Georgia Coastal Ecosystems LTER

University of Georgia

email: sheldon@uga.edu



CUAHSI Virtual Workshop –
Field Data Management Solutions

01-Oct-2014

Background & Motivation

- ▶ Georgia Coastal Ecosystems LTER started in 2000
 - Large and varied data collection effort
 - NSF & LTER require data and metadata archiving, sharing
 - Needed to standardize data processing, quality control, documentation
- ▶ No ready-to-use software for LTER data management
 - Lots of great papers and reports, no tools to download
 - Most LTER sites were using “flat files” – limiting
 - Relational databases emerging – proprietary, complex, require constant network access
- ▶ Developed custom data management framework using 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

What is MATLAB?

- ▶ From Mathworks (www.mathworks.com):
"MATLAB is a programming environment for algorithm development, data analysis, visualization, and numerical computation."
- ▶ Benefits
 - Ubiquitous in engineering and many science branches
 - Rapid development, pre-built functionality, Java integration
 - Code, GUIs and data formats multi-platform (Windows, Unix, Mac)
 - Stable: good support and backward compatibility (~30 year history)
 - Scalable (netbook to HPC) – good performance with huge data sets
 - Broad I/O support (serial ports to web services)
- ▶ Costs
 - Commercial ("licensed source") – limits flexibility, costs \$-\$\$\$
 - Some programming required for maximum use

Toolbox Development

- ▶ Used ESA's "FLED" report to guide approach

- 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.

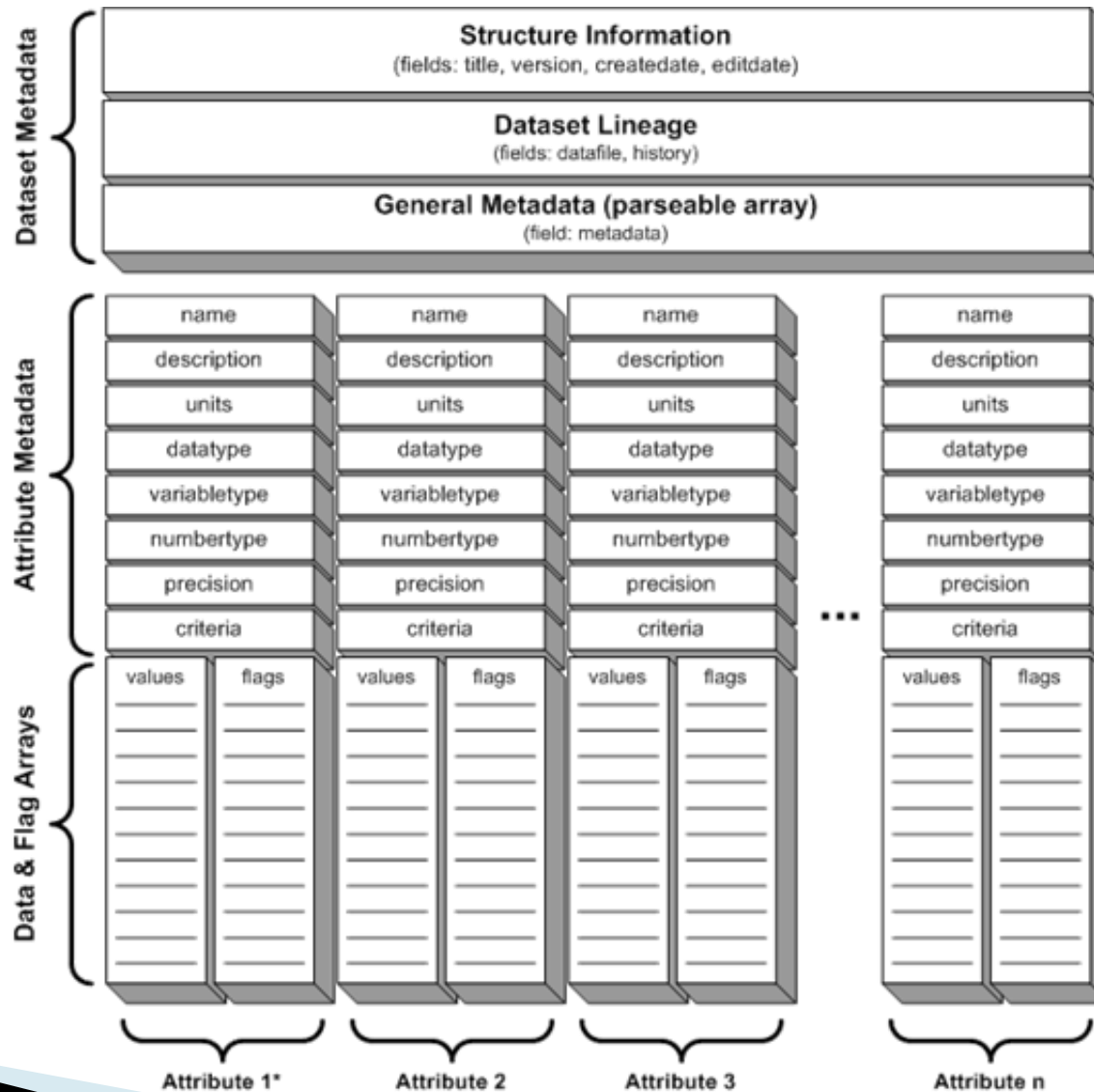
- ▶ Identified information storage requirements

- Any number of numeric and text variables
- Attribute metadata for each variable (name, units, description, type, precision, ...)
- Structured documentation metadata for dynamic updating, formatting
- Versioning and processing history info (lineage)
- Quality control rules for every variable, qualifier flags for every value

- ▶ Designed tabular data model in MATLAB

- MATLAB object ("struct") 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

Data Model (GCE Data Structure)



Toolbox Development

- ▶ MATLAB software library (“toolbox”)
 - Utility functions to abstract low-level operations (API)
 - Create structure, add/delete columns, copy/insert/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)
- ▶ Indexing and search support (and GUI search engine)
- ▶ Data harvest management tools

Command Line Interface (API)

The screenshot shows the MATLAB R2009b environment. The Command Window displays the following code and its execution results:

```
>> [s,msg] = fetch_usgs('0226000','realtime',60,'USGS_Doctortown');  
>> s  
  
s =  
  
    version: 'GCE Data Structure 1.1 (29-Mar-2001)'  
      title: 'Data from USGS Station 0226000 (ALTAMAHA RIVER AT DOCTORTOWN, GA) for 05-Feb-2010 through 06-Apr-2010'  
 metadata: {87x3 cell}  
   datafile: {'usgs_0226000_realtime_20100406_1130_mod.txt' [5797]}  
 createdate: '06-Apr-2010 11:30:48'  
   editdate: '06-Apr-2010 11:30:50'  
     history: {16x2 cell}  
        name: {1x12 cell}  
       units: {'none' 'none' 'none' 'serial day (base 1/1/0000) - GMT' 'YYYY' 'MM' 'DD' 'hr' 'min' 'm' 'm^3/sec' 'mm'}  
description: {1x12 cell}  
   datatype: {'s' 's' 'd' 'f' 'd' 'd' 'd' 'd' 'd' 'f' 'f' 'f'}  
variabletype: {1x12 cell}  
numbertype: {1x12 cell}  
precision: [0 0 0 8 0 0 0 0 0 2 1 2]  
values: {1x12 cell}  
criteria: {1x12 cell}  
flags: {' ' ' ' ' ' ' ' ' ' ' ' ' ' ' ' ' '  
  
>> listcols(s)  
  
ans =  
  
1: Agency -- string  
2: StationID -- string  
3: Provisional -- integer  
4: Date (serial day (base 1/1/0000) - GMT) -- floating-point  
5: Year (YYYY) -- integer  
6: Month (MM) -- integer  
7: Day (DD) -- integer  
8: Hour (hr) -- integer  
9: Minute (min) -- integer  
10: GageHeight (m) -- floating-point  
11: Discharge (m^3/sec) -- floating-point  
12: Precipitation (mm) -- floating-point  
  
>> dt = extract(s,'Date'); discharge = extract(s,'Discharge');  
>> whos
```

Name	Size	Bytes	Class	Attributes
ans	12x63	1512	char	
discharge	5797x1	46376	double	
dt	5797x1	46376	double	
msg	0x0	0	char	
s	1x1	1346932	struct	

The Workspace pane at the bottom left lists variables: ans, msg, s, dt, and discharge.

GUI Application Interface

Real-time climate data from the SINERR/GCE/UGAMI weather station...

File Edit Metadata Tools Misc Window Help

Column List (select to display properties)

- Date (serial day (base 1/1/0000) - GMT)
- Year (YYYY)
- Month (MM)
- Day (DD)
- Hour (hr)
- Minute (min)
- Record (number)
- Temp_Air (°C)
- Max_Temp_Air (°C)
- Time_Max_Temp_Air (hhmm)
- Min_Temp_Air (°C)
- Time_Min_Temp_Air (hhmm)
- Humidity (%)
- Baro_Press (mbar)
- Wind_Speed (m/s)
- Wind_Dir (degrees)

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

Column Name Temp_Air

Column Units °C Convert

Description

Air temperature, 15 minute mean

Data Type floating-point number (f)

Variable Type measurements (data)

Numerical Type continuous/ratio (continuous)

Precision 3 decimal places

Flag Criteria $x \leq -40 = 'I'; x > 60 = 'Y'; x < 10 = 'Q'; x > 40 = 'Q'; \text{flag_value} = \text{flag_value} \cup \text{flag_value}$ Edit

About the GCE Data Toolbox

GCE Data Toolbox for MATLAB

Software for metadata-based processing, analysis, visualization and transformation of environmental data

Toolbox Version 3.9.0 (08-Sep-2014)

(c)2002-2014 Wade M. Sheldon and the Georgia Coastal Ecosystems LTER Project

Wade M. Sheldon
Dept. of Marine Sciences
University of Georgia
Athens, GA 30602-3636
email: sheldon@uga.edu

Search Engine Dataset Editor Mapping Tools Documentation Exit MATLAB

-- Software Usage Agreement and Disclaimer --

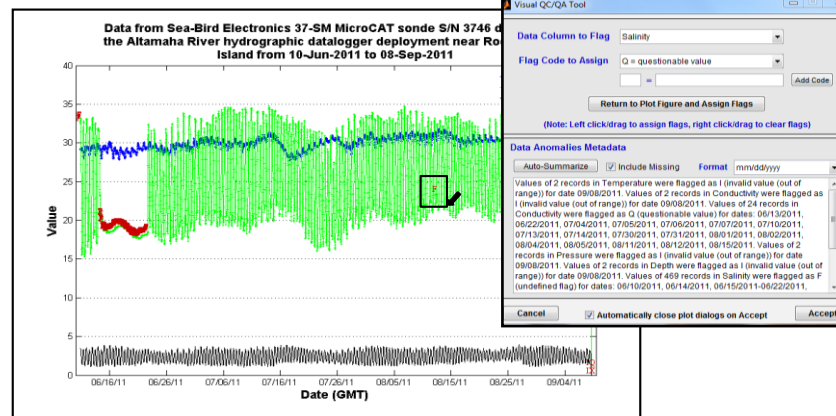
The GCE Data Toolbox for MATLAB is provided as a courtesy to the scientific community by the Georgia Coastal Ecosystems LTER Program (<http://gce-lter.marsci.uga.edu/>), Coweeta LTER Program (<http://coweeta.uga.edu/>), and other collaborators in LTER. The latest versions of the software and documentation, as well as user support resources, are available on the World Wide Web at https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox.

The GCE Data Toolbox is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

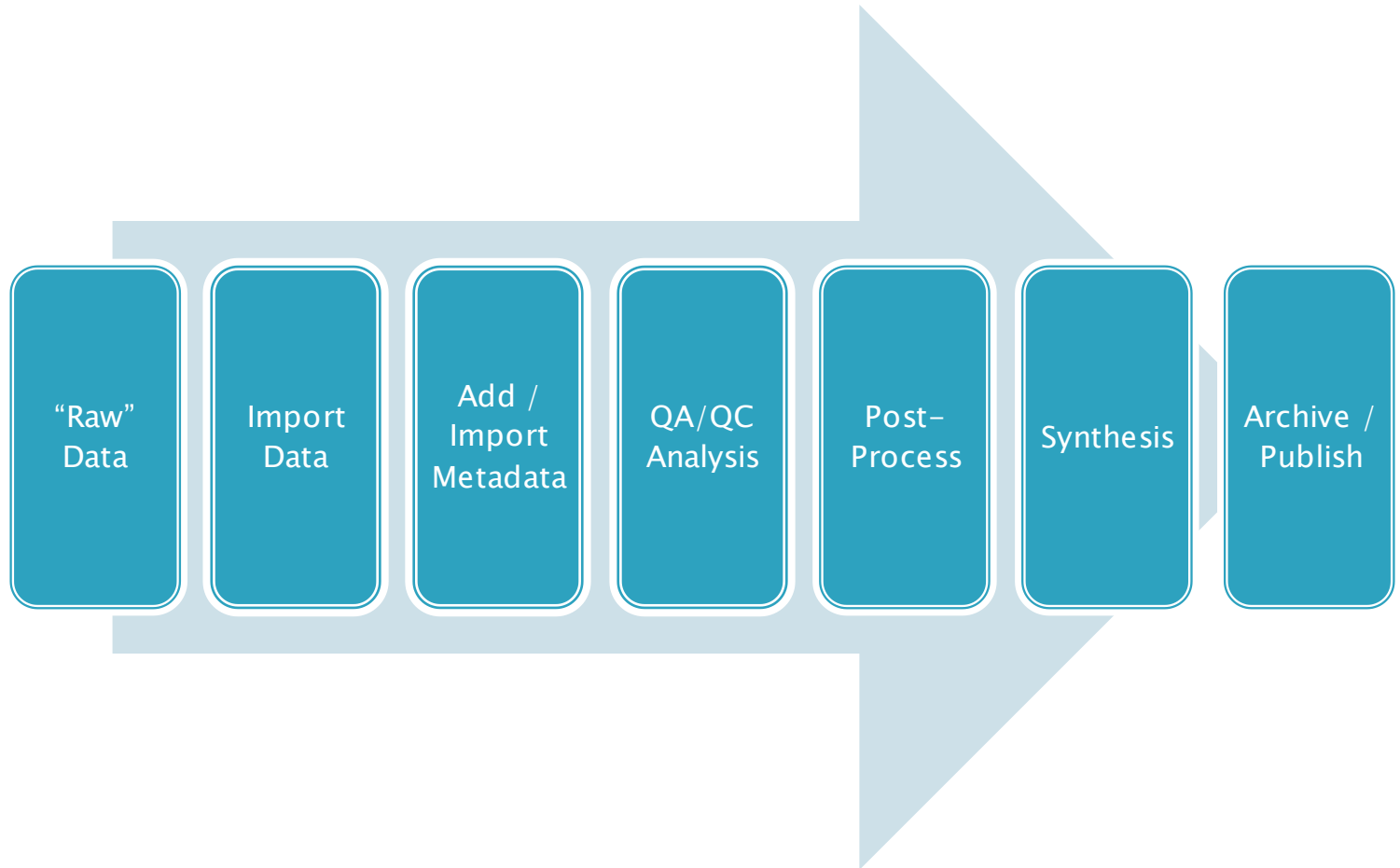
☒ Automatically load startup screen if all other windows are closed

Data Editor

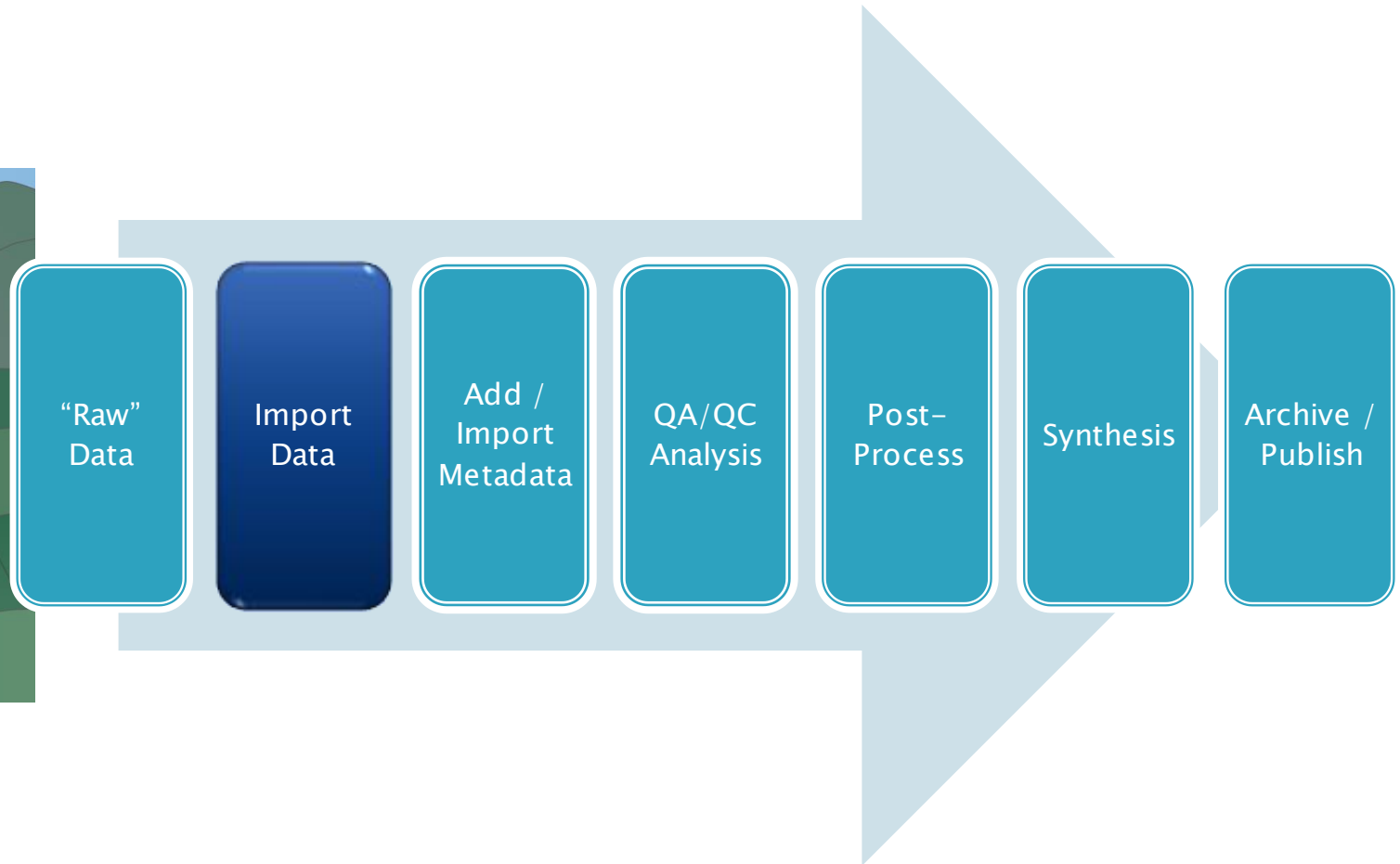
All	Date	Year	Month	Day	Hour	Minute	Record	Temp_Air	Max_Temp_Air
None	(serial day (base 1/1/0000) - GMT)	(YYYY)	(MM)	(DD)	(hr)	(min)	(number)	(°C)	(°C)
1	733627 7063333	2008	6	7	17	15	0	34.730	35.509
2	733627 7167500	2008	6	7	17	15	1	35.157	35.570
3	733627 7291667	2008	6	7	17	30	2	35.576	35.909
4	733627 7396533	2008	6	7	17	45	3	35.369	35.840
5	733627 7500000	2008	6	7	18	0	4	36.094	36.006
6	733627 7604167	2008	6	7	18	15	5	36.412	36.908
7	733627 7708333	2008	6	7	18	30	6	36.591	37.009
8	733627 7812500	2008	6	7	18	45	7	36.849	37.346
9	733627 7916667	2008	6	7	19	0	8	36.632	37.312
10	733627 8020833	2008	6	7	19	15	9	34.669	35.695
11	733627 8125000	2008	6	7	19	30	10	31.681	33.676
12	733627 8229167	2008	6	7	19	45	11	27.943	28.626
13	733627 8333333	2008	6	7	20	0	12	25.417	27.002
14	733627 8437500	2008	6	7	20	15	13	26.084	26.316
15	733627 8541667	2008	6	7	20	30	14	25.098	26.004
16	733627 8645833	2008	6	7	20	45	15	25.690	25.902
17	733627 8750000	2008	6	7	21	0	16	25.964	26.138
18	733627 8854167	2008	6	7	21	15	17	26.564	27.420
19	733627 8958333	2008	6	7	21	30	18	26.662	27.492
20	733627 9062500	2008	6	7	21	45	19	24.648	24.992
21	733627 9166667	2008	6	7	22	0	20	24.795	25.237
22	733627 9270833	2008	6	7	22	15	21	24.714	25.205
23	733627 9375000	2008	6	7	22	30	22	24.470	24.871
24	733627 9479167	2008	6	7	22	45	23	24.970	25.317
25	733627 9583333	2008	6	7	23	0	24	25.116	25.283



Data Management Cycle



Data Management Cycle



Importing Data

- ▶ Generic parsers
 - Delimited text (CSV, space, tab)
 - MATLAB variables (arrays, matrices, structs)
- ▶ Specialized parsers
 - Vendor-specific logger files
 - Campbell Scientific Instruments (tables, arrays)
 - Sea-Bird CTD, sondes
 - Others (Hobo, Schlumberger, OSIL, ...)
 - Network data sources
 - SQL database queries (JDBC)
 - Federal databases (USGS NWIS, NOAA NCDC, NOAA HADS)
 - LTER ClimDB/HydroDB
 - EML repositories (LTER NIS/PASTA, KNB)
 - Data Turbine servers
- ▶ Custom parsers can be added

Importing Data

The image shows two overlapping software windows. The background window is the 'Data Structure Editor' with a menu bar (File, Edit, Metadata, Tools, Misc, Window, Help) and a 'Column List (select to display properties)' section. The foreground window is the 'Retrieve USGS Data' dialog box.

Data Structure Editor

Column List (select to display properties)

Column Name:

Column Units:

Description:

Data Type:

Variable Type:

Numerical Type:

Precision:

Flag Criteria:

Retrieve USGS Data

State/Territory:

Available Stations:

- 313118083325401 Atmosphere RAINGAGE AT NADP SITE GA99 NEAR CHULA, GA
- 02228295 Coastal CUMBERLAND SOUND AT SEA CAMP DOCK, NR ST MARYS, GA
- 02338003 Laboratory or sample-preparation area GWSC Bacteria Lab QA/QC Monitoring
- 02393500 Lake ALLATOONA LAKE NEAR CARTERSVILLE, GA
- 02334995 Lake BERKELEY LAKE NEAR BERKELEY LAKE, GA
- 02196838 Lake BUTLER CREEK RESERVOIR AT FORT GORDON, GA
- 02381400 Lake CARTERS LAKE NEAR CARTERS, GA
- 02382400 Lake CARTERS RE-REGULATION LAKE NEAR CARTERS, GA
- 02334400 Lake LAKE SIDNEY LANIER NEAR BUFORD, GA
- 02400000 Lake CALHOUN LAKE NEAR ANDERSON, GA

Station ID: Data Type: Date Range: to

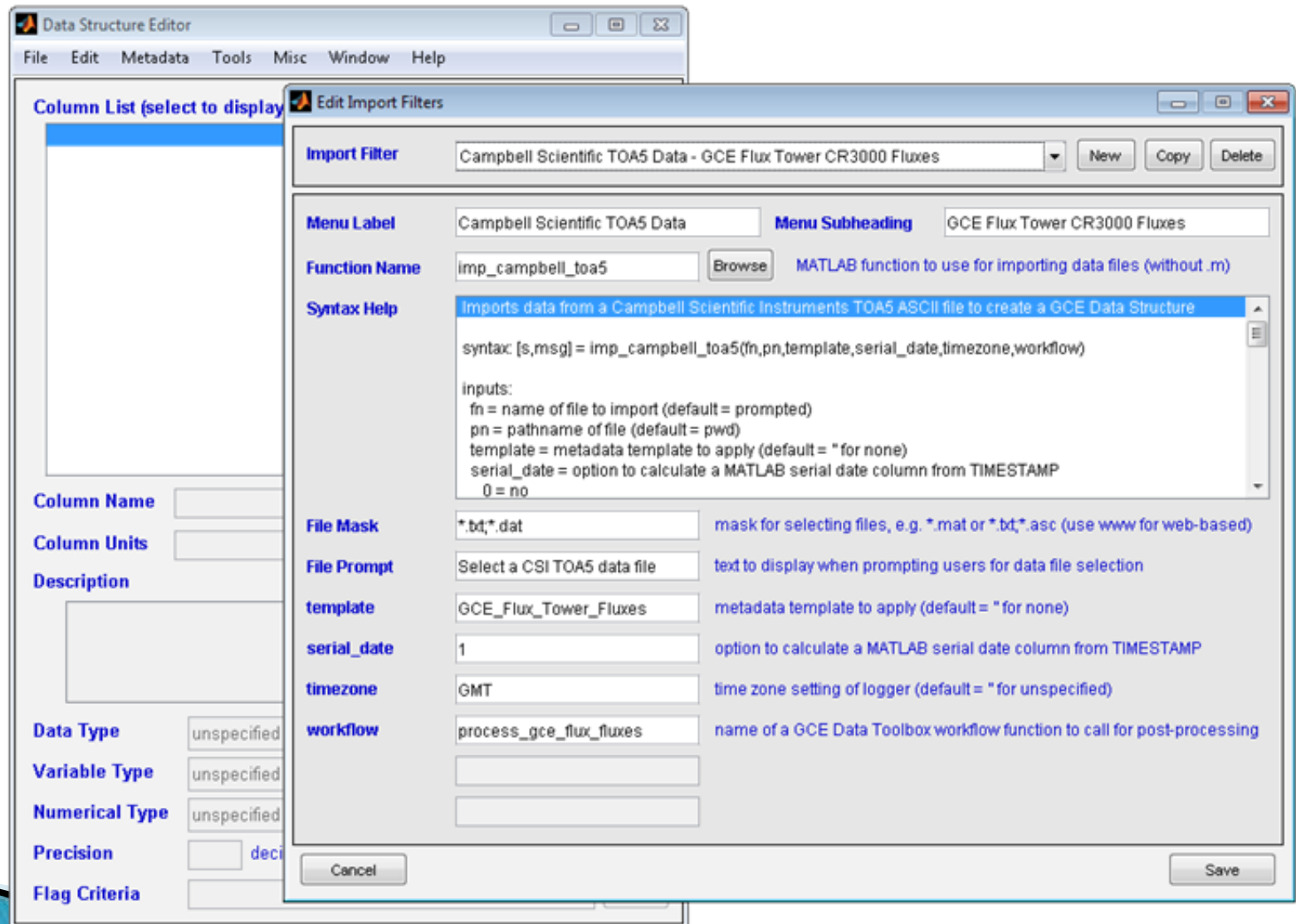
USGS Website URL:

Metadata Template:

☐ Save Temp File

☒ Close dialog after retrieving data ☐ Open in new window

Importing Data



Importing Data

Real-time climate data from the SINERR/GCE/UGAMI weather station...

File Edit Metadata Tools Misc Window Help

Column List (select to display properties)

Date (serial day (base 1/1/0000) - GMT)
Year (YYYY)
Month (MM)
Day (DD)
Hour (hr)
Minute (min)
Record (number)
Temp_Air (°C)
Max_Temp_Air (°C)
Time_Max_Temp_Air (hhmm)
Min_Temp_Air (°C)
Time_Min_Temp_Air (hhmm)
Humidity (%)
Baro_Press (mbar)
Wind_Speed (m/s)
Wind_Dir (degrees)

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

Column Name Temp_Air

Column Units °C

Description
Air temperature, 15 minute mean

Data Type floating-point number (f)

Variable Type measurements (data)

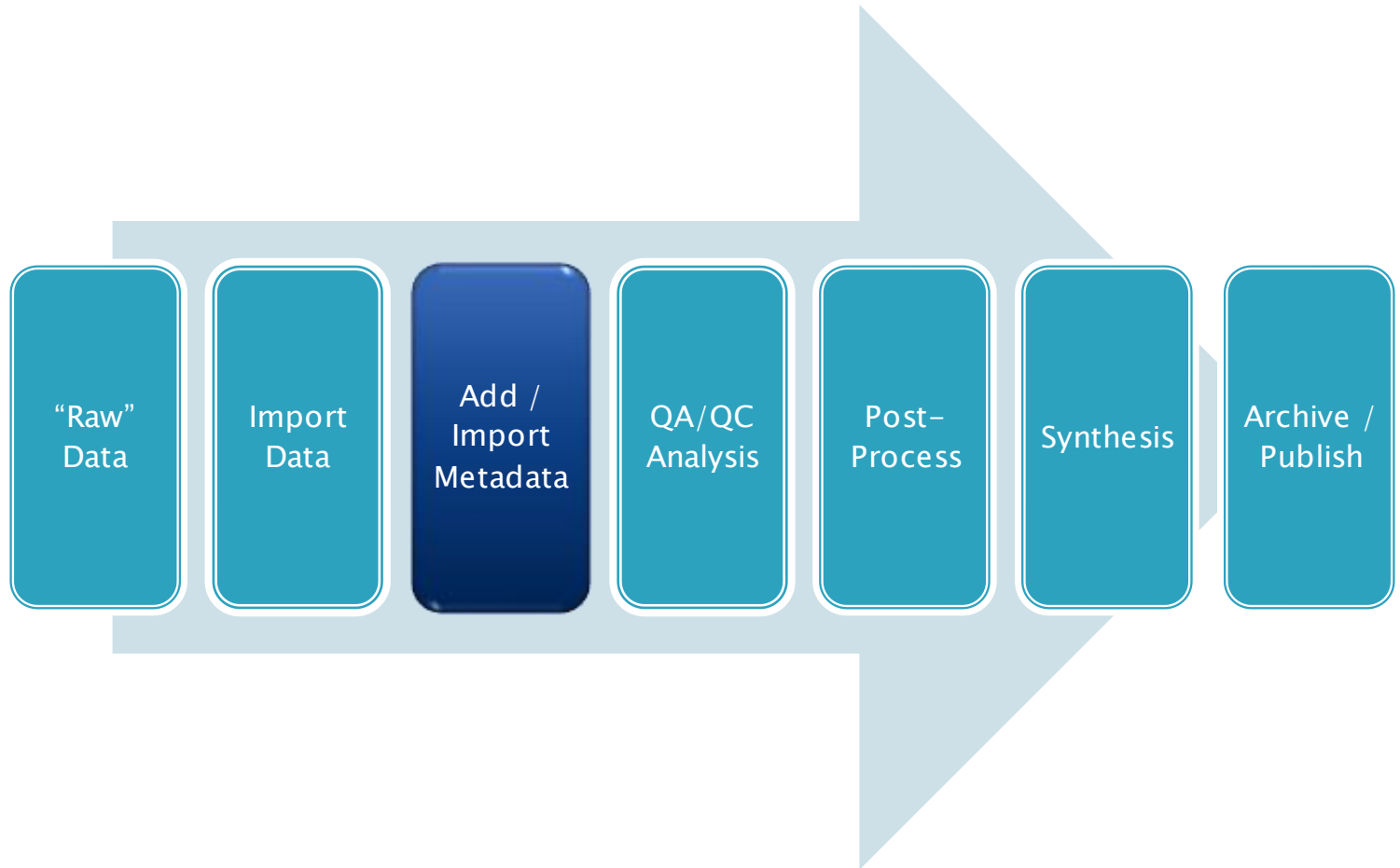
Numerical Type continuous/ratio (continuous)

Precision 3 decimal places

Flag Criteria x<=-40='I';x>60='I';x<-10='Q';x>40='Q';flag_valuecha

Month (MM)	Day (DD)	Hour (hr)	Minute (min)	Record (number)	Temp_Air (°C)	Max_Temp_Air (°C)
2008	8	7	17	0	34.730	35.509
2008	8	7	17	15	35.157	35.570
2008	8	7	17	30	35.576	35.909
2008	8	7	17	45	35.369	35.840
2008	8	7	18	0	36.094	36.606
2008	8	7	18	15	36.412	36.908
2008	8	7	18	30	36.581	37.009
2008	8	7	18	45	36.849	37.346
2008	8	7	19	0	36.632	37.312
2008	8	7	19	15	34.669	35.695
2008	8	7	19	30	31.681	33.876
2008	8	7	19	45	27.943	28.828
2008	8	7	20	0	26.417	27.002
2008	8	7	20	15	26.084	26.316
2008	8	7	20	30	25.708	26.024
2008	8	7	20	45	25.690	25.902
2008	8	7	21	0	25.964	26.138
2008	8	7	21	15	26.564	27.420
2008	8	7	21	30	26.662	27.487
2008	8	7	21	45	24.648	24.992
2008	8	7	22	0	24.795	25.237
2008	8	7	22	15	24.714	25.205
2008	8	7	22	30	24.470	24.871
2008	8	7	22	45	24.970	25.317
2008	8	7	23	0	25.116	25.283

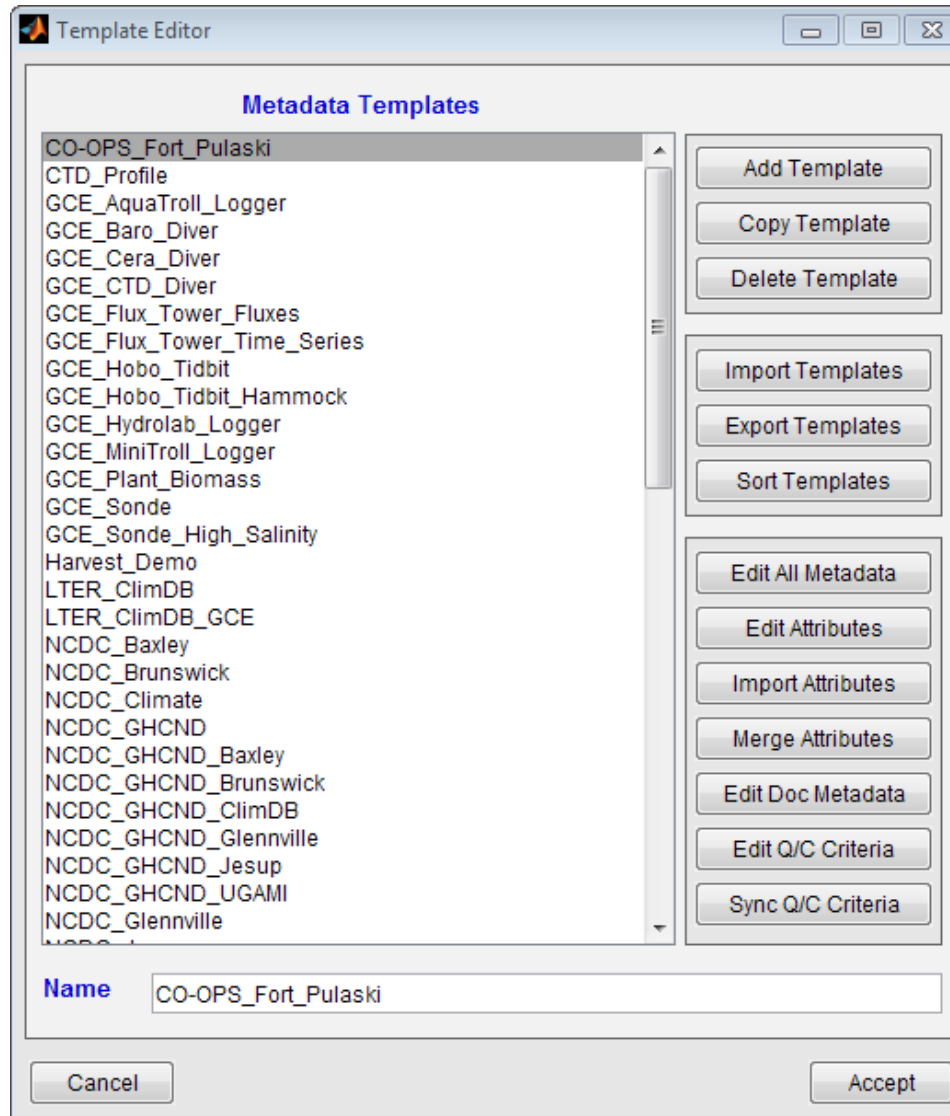
Data Management Cycle



Adding/Importing Metadata

- ▶ Metadata entry is time-consuming (and tedious)!
 - Metadata capture, re-use emphasized
 - Metadata auto-generated whenever possible
- ▶ Many pathways for building Metadata
 - Metadata can be imported along with data
 - Logger file headers (Campbell, Sea-bird)
 - Station, parameter information from USGS, NOAA
 - Tokenized headers from Data Submission Template
 - Metadata imported from other data structures
 - Metadata imported from data repository (EML/XML)
 - Metadata added from user-defined “Templates”
 - Column (attribute) metadata matched to “variables”
 - Boilerplate documentation
 - GUI tool for creating/managing templates

Adding/Importing Metadata



Adding/Importing Metadata

Metadata Editor

File Edit

Metadata Contents (click on rows to view/edit contents)

CATEGORY	FIELD	CONTENTS
Dataset	Title	Real-time climate data from the SINERR/GCE/UGAMI weather station
Dataset	Accession	none
Dataset	Investigator	[Name: Dorset Hurley Position: Research Coordinator Address: Geo:
Dataset	Abstract	Air temperature, relative humidity, barometric pressure, precipi
Dataset	StudyType	Monitoring
Dataset	Themes	Meteorology
Dataset	LTERCore	Regionalization
Dataset	Georeferences	none
Dataset	SubmitDate	not applicable
Dataset	Keywords	climate, meteorology, air temperature, barometric pressure, humi
Project	Name	Cooperative SINERR/GCE-LTER/UGA Marine Institute Climate Monitor:
Project	Leaders	[Name: Dorset Hurley Position: Research Coordinator Address: Geo:
Project	StartDate	July, 2002
Project	EndDate	unspecified
Project	Objectives	Collect baseline climate data at Sapelo Island, Georgia, to supp
Project	Abstract	Cooperative agreement between the Sapelo Island National Estuarin
Project	Funding	Jointly-funded by SINERR, GCE-LTER Project, and UGA Marine Instit
Site	Location	ML_Met -- Marsh Landing Meteorological Station Deployment
Site	Coordinates	ML_Met -- 31 25 04.2 N, 081 17 43.5 W
Site	Physiography	ML_Met -- Barrier island
Site	Landform	ML_Met -- Intertidal salt marsh

First
Up
Down
Last
Sort
Combine
Add New
Copy
Delete
Clear

Category: Dataset

Field: Abstract

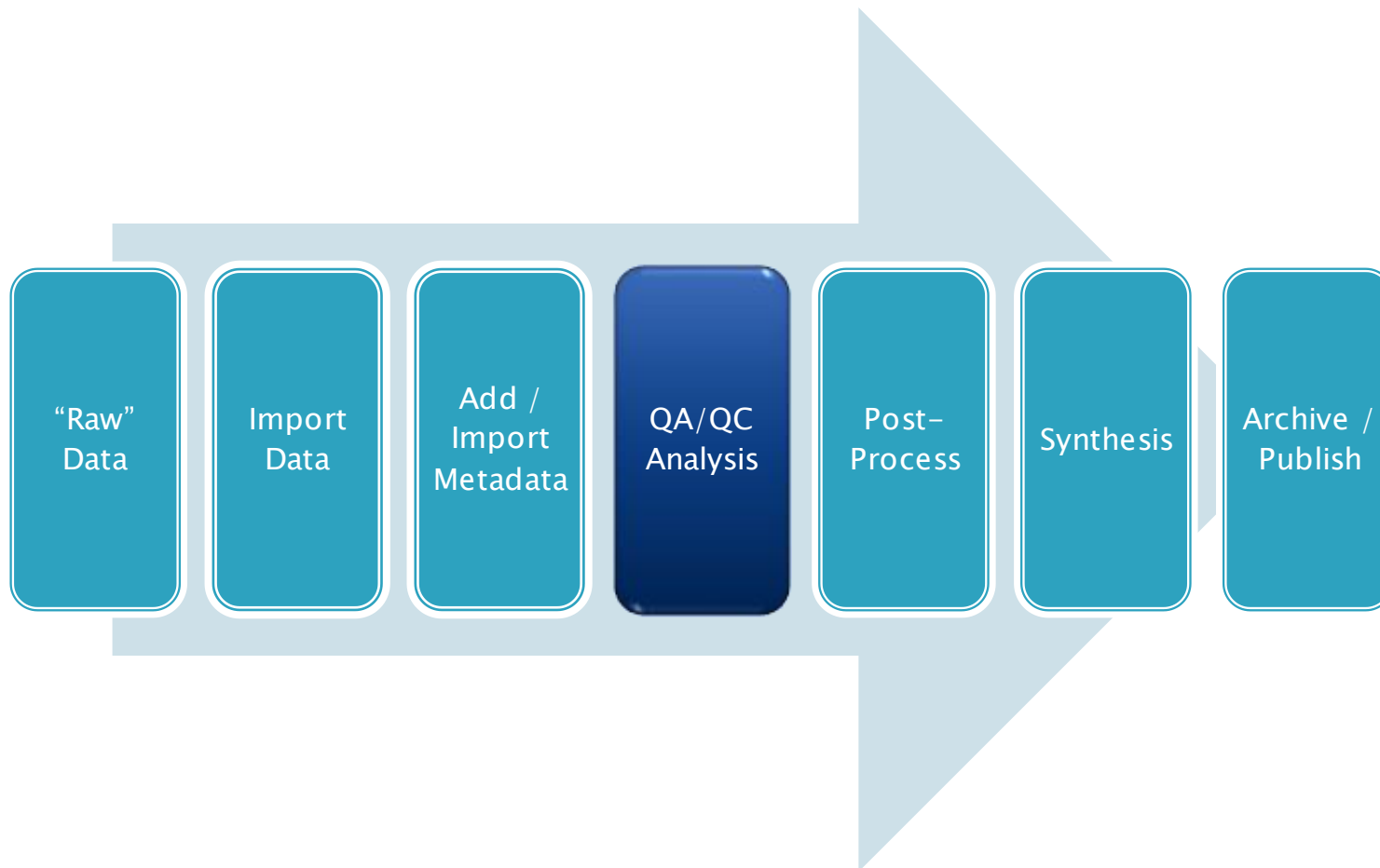
Clear Preview

Contents: Air temperature, relative humidity, barometric pressure, precipitation, photosynthetically-available and total solar radiation, and wind speed and direction were measured using an automated Campbell Scientific Instruments climate station installed at Marsh Landing on Sapelo Island, Georgia. Observations were logged at 15 minute intervals throughout the study period. The sensors were mounted on a 10m aluminum tower, with wind sensors mounted at the top, light sensors at approximately 5m, and other sensors at 2-3m to minimize interference from the surrounding landscape. This climate station was jointly operated by the Sapelo Island National Estuarine Research Reserve, the Georgia Coastal Ecosystems LTER Project, and University of Georgia Marine Institute.

(Note: use | characters to force line breaks in the formatted metadata)

Cancel Accept

Data Management Cycle



QA/QC Analysis Framework

- ▶ **Programmatic Q/C Analysis**
 - “Rules” (i.e. criteria) define conditions in which values should be flagged
 - Unlimited Q/C rules for each variable
 - Rules evaluated when data loaded and when data or rules change
 - Rules predefined in metadata templates to automate Q/C on import
- ▶ **Interactive Q/C Analysis and Revision**
 - Qualifiers can be assigned/cleared visually on data plots with the mouse
 - Qualifiers can be propagated to dependent columns
 - Qualifiers can be removed or edited (search/replace) if standards change
- ▶ **Automatic Documentation of Q/C Steps**
 - Q/C operations (including revisions) logged to processing lineage
 - Data anomalies reports can be auto-generated and annotated
- ▶ **Data analysis, synthesis tools Q/C-aware**
 - Qualified values can be filtered, summarized, visualized during analysis
 - Statistics about missing/qualified values tabulated, used to qualify derived data

QA/QC Analysis Framework

The image displays two overlapping windows from the QA/QC Analysis Framework software.

Left Window: Data from Sea-Bird Electronics 37-SM MicroCAT sondes deployed at ...

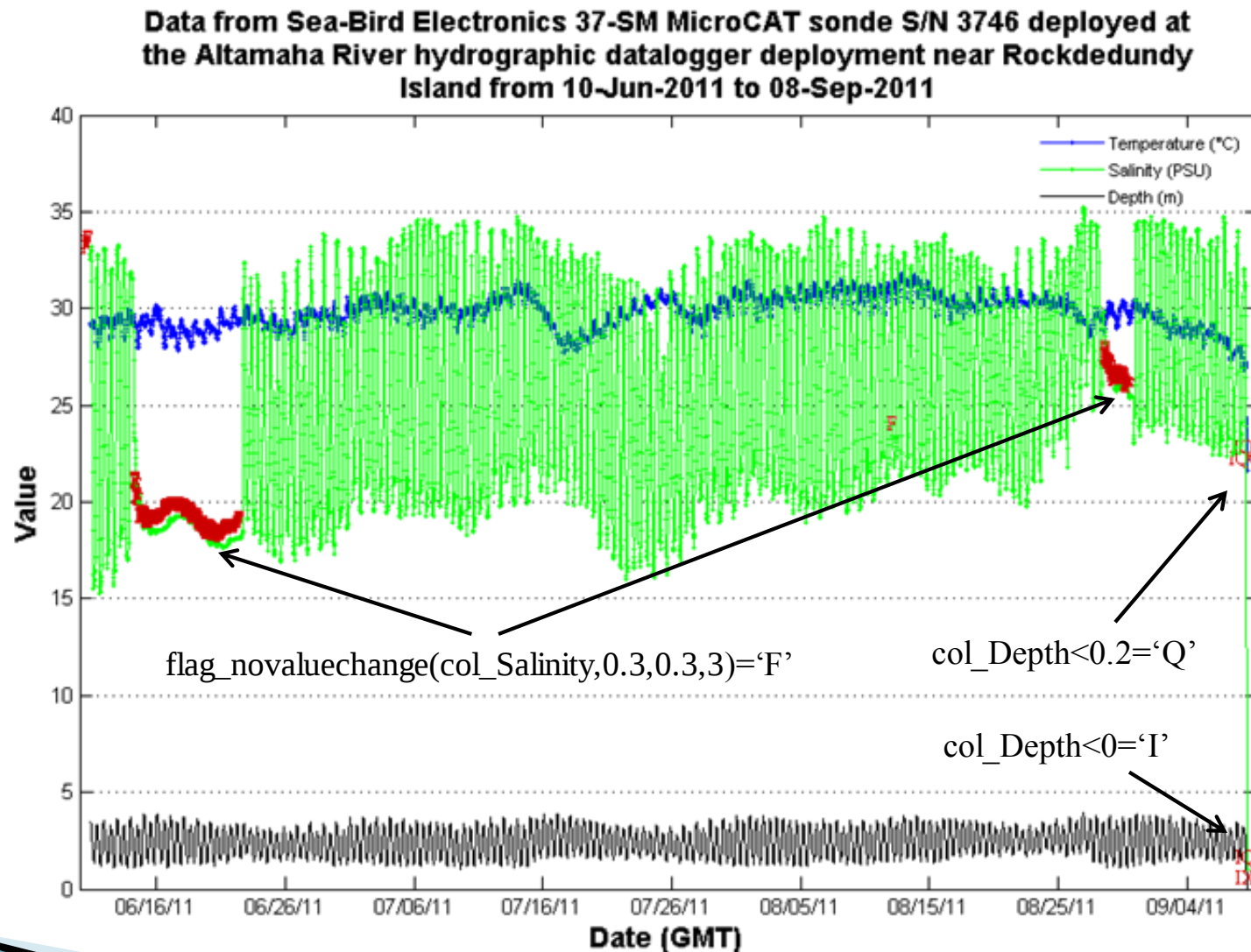
- Column List (select to display properties):** A list of data columns including Latitude (degrees), Instrument (none), Pump (none), Date (serial day (base 1/1/0000) - GMT), Year (YYYY), Month (MM), Day (DD), Hour (hr), Minute (min), Temperature (°C), Conductivity (S/m), Pressure (dbar), Depth (m), Salinity (PSU), Sigma_t (kg/m³), and Calibration (none). The 'Salinity (PSU)' column is selected.
- Column Name:** Salinity
- Column Units:** PSU
- Description:** Salinity
- Data Type:** floating-point number (f)
- Variable Type:** calculations (calculation)
- Numerical Type:** continuous/ratio (continuous)
- Precision:** 3 decimal places
- Flag Criteria:** `x<0='1';col_Depth<=0='1';x>38='Q';col_Depth<0.15='`

Right Window: Q/C Flag Criteria Editor

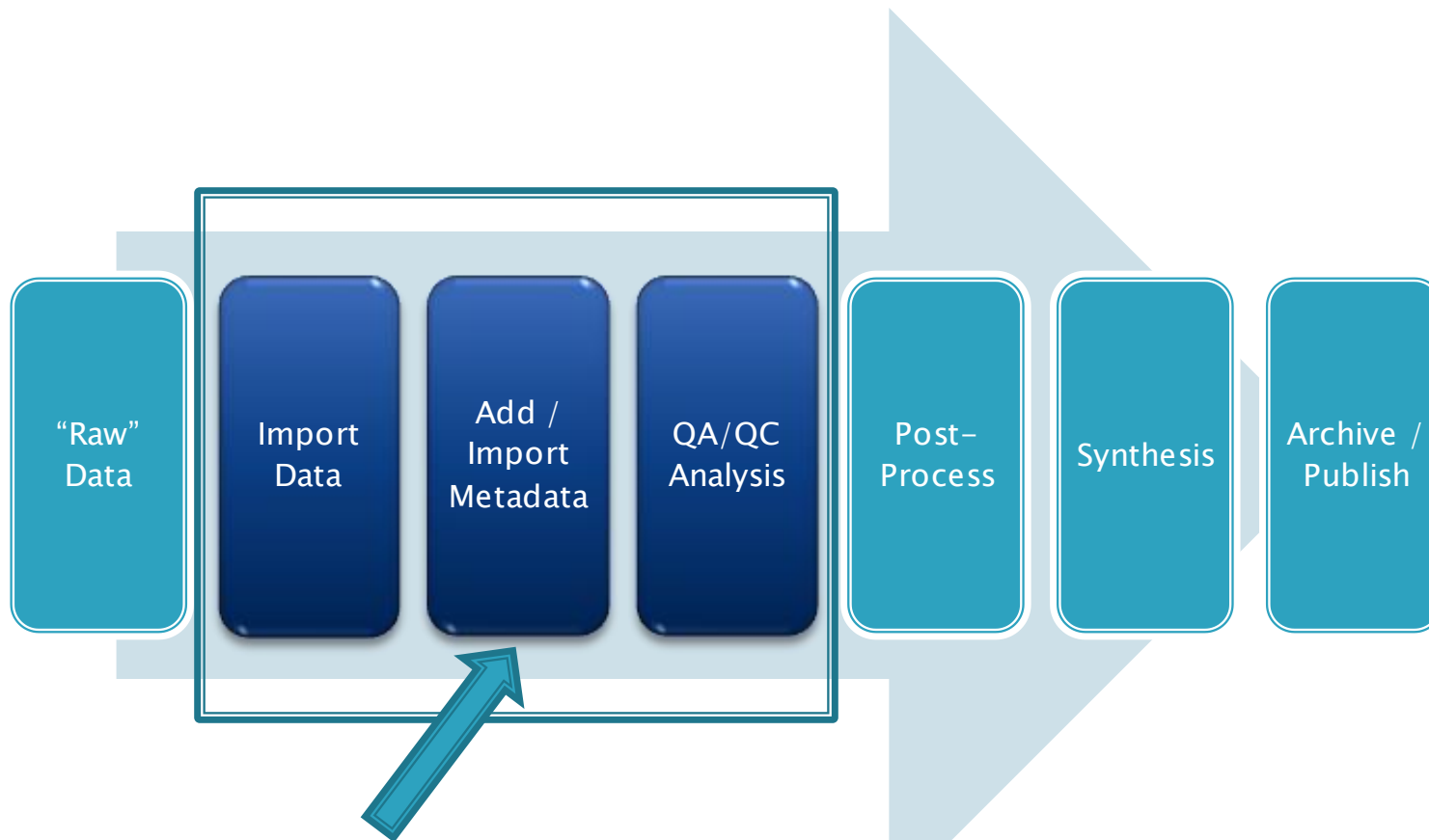
- Data Set Column:** Salinity (PSU)
- Q/C Flag Criteria Expressions:**
 - `x<0='1'`
 - `col_Depth<=0='1'`
 - `x>38='Q'`
 - `col_Depth<0.15='Q'`
 - `flag_valuechange(x,6,6,1,3)='Q'`
- Value:** > 38 = Q -- questionable value
- Custom Flags:** (Empty field)
- Buttons:** Move First, Move Up, Move Down, Move Last, Add New, Delete, Add/Edit QC Flag Function, Add/Edit Custom Criteria, Cancel, Accept.

A blue arrow points from the 'Add/Edit QC Flag Function' button in the 'Q/C Flag Criteria Editor' window to the 'Edit' button in the 'Flag Criteria' field of the 'Data from Sea-Bird Electronics' window.

Example – CTD mooring

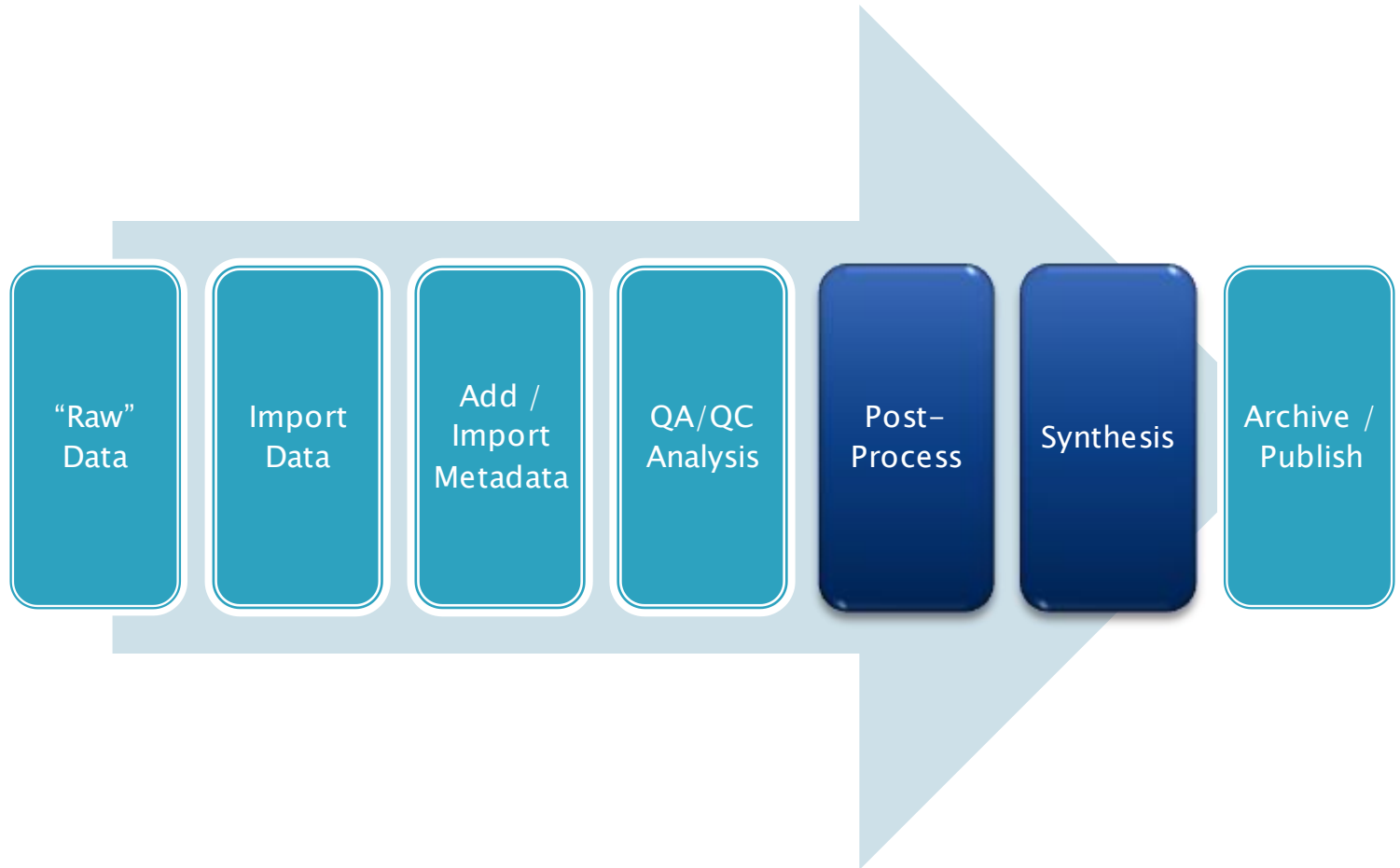


Data Management Cycle



Import Filter + Metadata Template
(One Step)

Data Management Cycle



Post-Processing and Synthesis

- ▶ Derived data sets can be created by filtering values or refactoring data table structure (e.g. combining or splitting columns)
- ▶ Data can be re-sampled or summarized by aggregation, binning and date/time scaling
- ▶ Multiple data sets can be combined by merging (union) and joining on key columns
- ▶ All derived data contain complete metadata describing the entire processing history
- ▶ QA/QC rules can be generated for derived data columns automatically based on number or percent missing/flagged values

Date/Time Interval Statistics

File Edit

Date/Time Interval for Statistical Summaries: Daily

Available Columns

- Instrument (none)
- Pump (none)
- Conductivity (S/m)
- Pressure (dbar)
- Depth (m)
- Sigma_t (kg/m³)**
- Calibration (none)

Column Selections

Group Records By (optional)

- Site (none)
- Station (none)
- Longitude (degrees)
- Latitude (degrees)**

Calculate Statistics For

- Temperature (°C)**
- Salinity (PSU)**

☐ Remove values flagged <any> List

Flag as I if percent missing > 30

Flag as Q if percent missing > 5

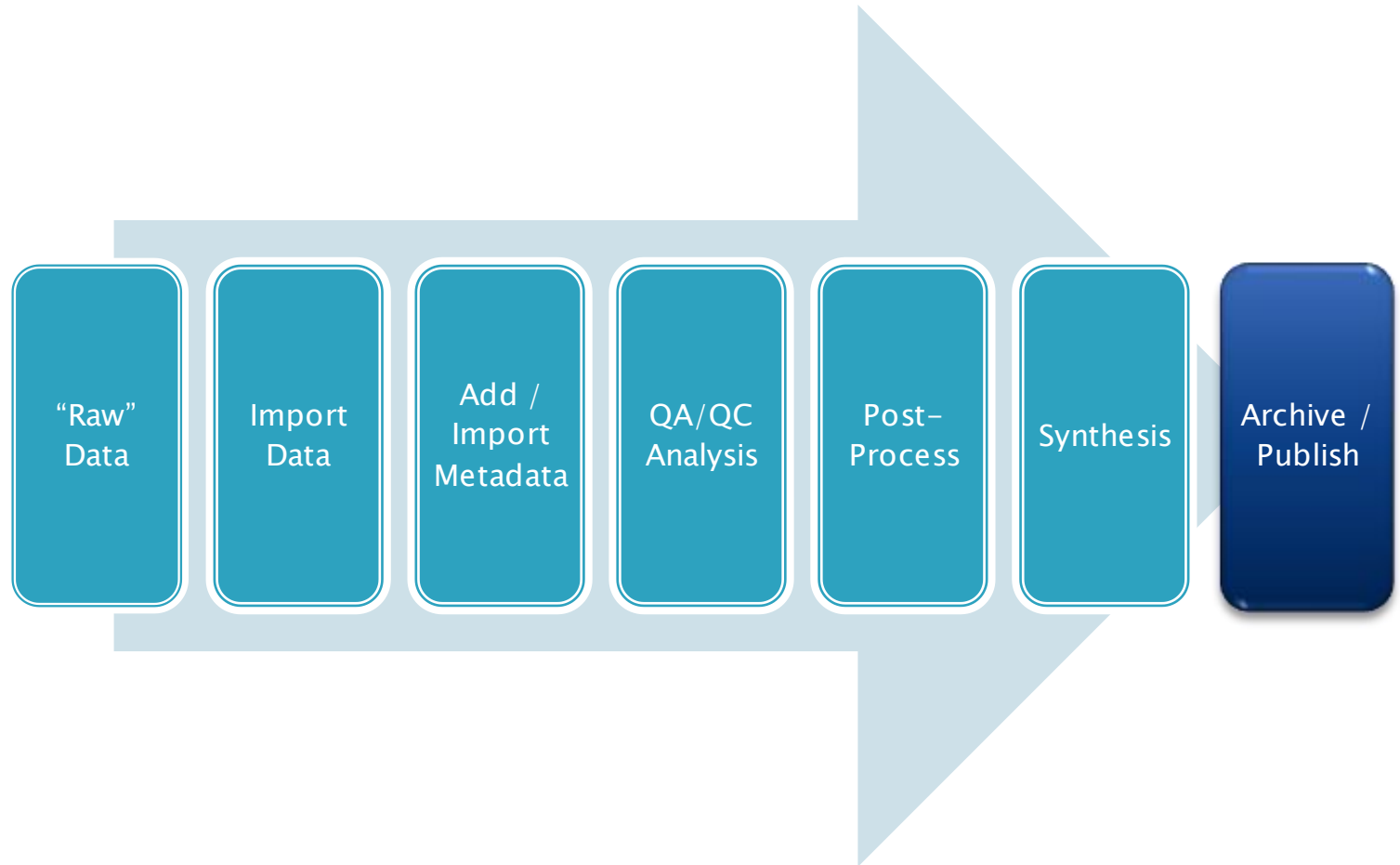
Flag as I if percent flagged > 80

Flag as Q if percent flagged > 30

Statistics: Automatic (based on numeric and variable type) ☒ Time Min/Max

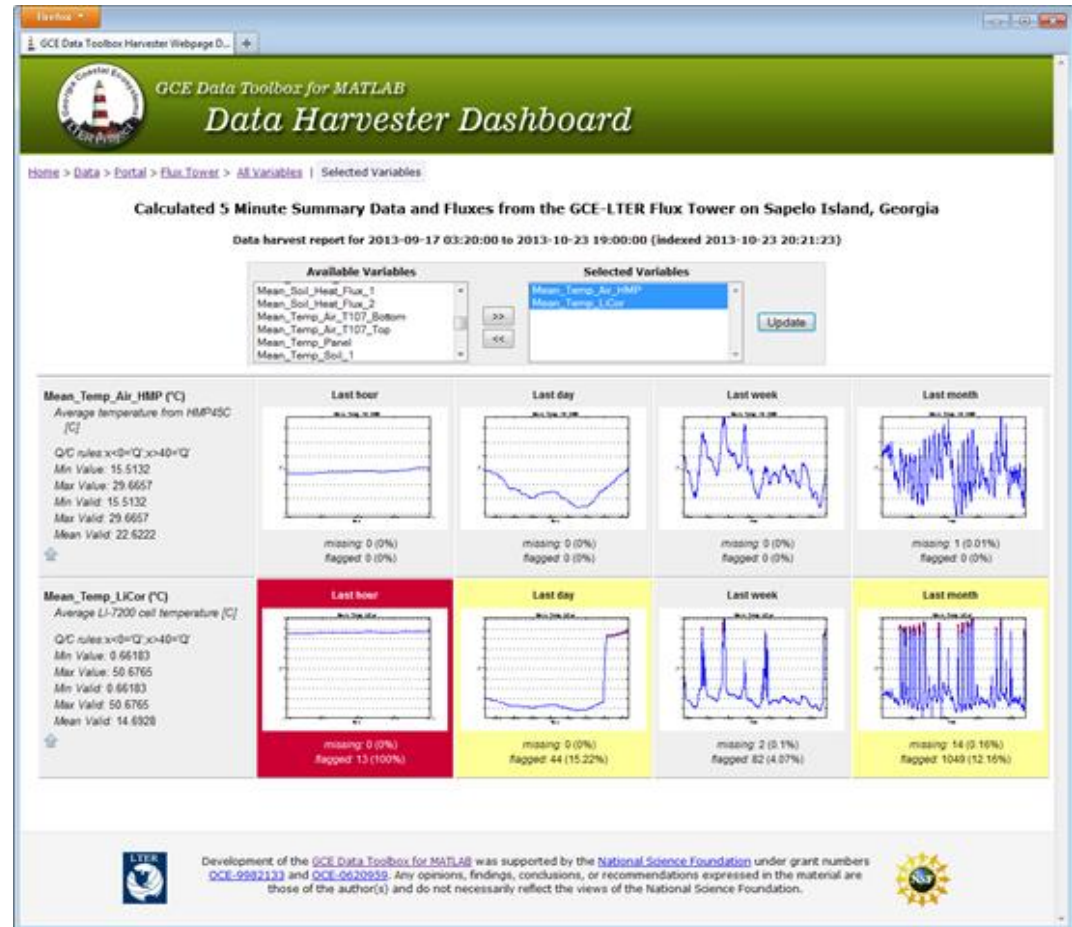
Cancel ☒ Close dialog after exporting the results Proceed

Data Management Cycle

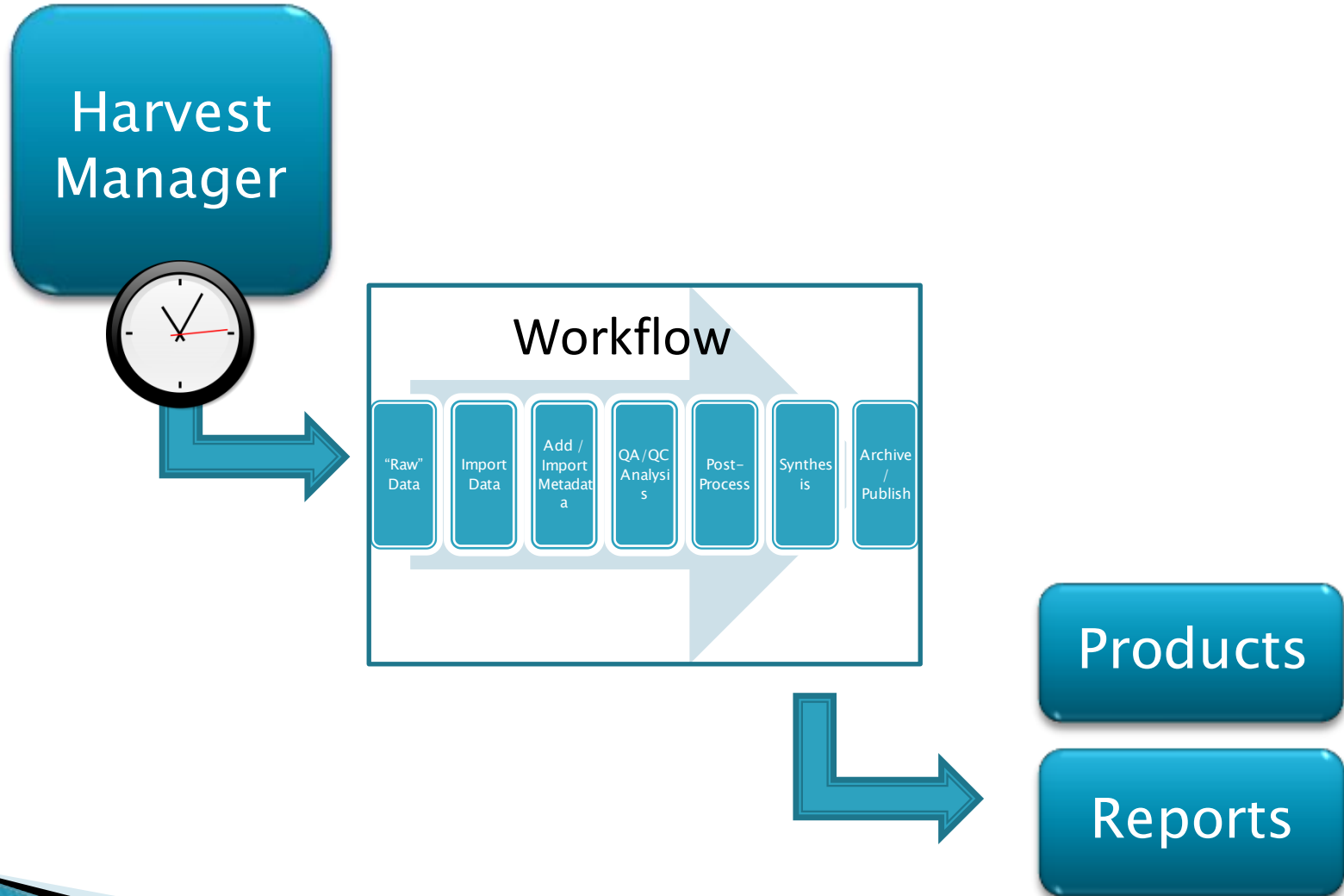


Archive/Publish Data

- ▶ Data and EML metadata can be published in scientific data archives, including LTER NIS, DataONE Member Nodes (e.g. KNB), Dryad, ...
- ▶ Data can be refactored and published in CUAHSI ODM database, Hydroserver
- ▶ Harvested data can be displayed on MATLAB-generated web pages and dashboards



Data Harvesting Workflows



Key Concepts

- ▶ Operations performed in context of a "dataset"
 - Passing data columns to a tool transports metadata as well
 - Dataset metadata used to guide transformation, plotting, analysis
 - Metadata used to auto-parameterize functions
- ▶ Workflow steps generate new datasets
 - Each step along a workflow results in a complete data set with metadata
 - Intermediate datasets can be saved or overwritten in workflows
- ▶ Processing history ("lineage") tracked
 - Each tool logs operations by date/time
 - Data revisions, deletions, flagging captured at user-specified detail
 - Lineage reported in metadata
- ▶ Metadata are "live", and updated automatically
 - Data column metadata and data revisions
 - Calculations and unit conversions
 - Code definitions
 - Metadata merged when multiple data sets are joined

Implementation Scenarios

- ▶ **End-to-End Processing (logger-to-scientist)**
 - Acquire raw data from logger, file system, network
 - Assign metadata from template or forms to validate and flag data
 - Review data and fine-tune flag assignments
 - Generate distribution files & plots, archive data, index for searching
 - Scientists can use toolbox on desktop to analyze, integrate data
- ▶ **Data Processing and Q/C**
 - Acquire, validate and flag raw data (on demand or timed/triggered)
 - Upload processed data files (e.g. csv) or values & flags to RDBMS (e.g. ODM)
- ▶ **Workflow Step**
 - Call toolbox from other software as part of workflow (tool-chaining)
 - Use toolbox as middleware between other systems (e.g. Data Turbine & ODM, Kepler, Taverna)

Real-Time Sensor Data

▶ Good Scalability

- Data only limited by computer memory (tested >1 GB data sets)
- Multiple instances can be run on high-end workstations
- Good flag evaluation performance in use, testing

▶ Good scope for automation

- Command-line API for unattended batch processing
- Timed and triggered workflow implementations easy to deploy

▶ Support for multiple formats, protocols

- Formats: ASCII, MATLAB, SQL, specialized (CSI, SBE, NWIS, HADS, ...)
- Transport: local file system, UNC paths, HTTP, FTP, SOAP

▶ Already used at 6 LTER sites (5 sites evaluating)

▶ USGS data harvester for HydroDB since 2003

GCE Data Harvesting (> 12 years)



Georgia Coastal Ecosystems LTER
Member of the NSF Long Term Ecological Research Network

Home > Data > Portal > Marsh Landing > harvest > Dataset Details | Plots

Dataset Details

Data Set ID: marshlanding_realtime_oct11

Originator: Dorset Hurley

Title: Real-time climate data from the SINERR/OCE/UGAMI weather station at Marsh Landing on Sapelo Island, Georgia, from 28-Sep-2011 to 21-Oct-2011

Abstract: Air temperature, relative humidity, barometric pressure, precipitation, photosynthetically-available and total solar radiation, and wind speed and direction were measured using an automated Campbell Scientific Instruments climate station installed at Marsh Landing on Sapelo Island, Georgia. Observations were logged at 15 minute intervals throughout the study period. The sensors were mounted on a 10m aluminum tower, with wind sensors mounted at the top, light sensors at approximately 5m, and other sensors at 2-3m to minimize interference from the surrounding landscape. This climate station was jointly operated by the Sapelo Island National Estuarine Research Reserve, the Georgia Coastal Ecosystems LTER Project, and University of Georgia Marine Institute.

Key Words: climate, meteorology, air temperature, barometric pressure, humidity, par, solar, wind, sapelo, marsh landing

Study Type: Monitoring

Study Period: 28-Sep-2011 to 21-Oct-2011

Bounding Box: West longitude: -81.29541°
North latitude: 31.41795°
East longitude: -81.29541°
South latitude: 31.41795°

Site References: [ML_MET](#) - Marsh Landing climate station, Meteorological station deployment

Downloads: [marshlanding_realtime_oct11](#) (Main data table, 2196 records)

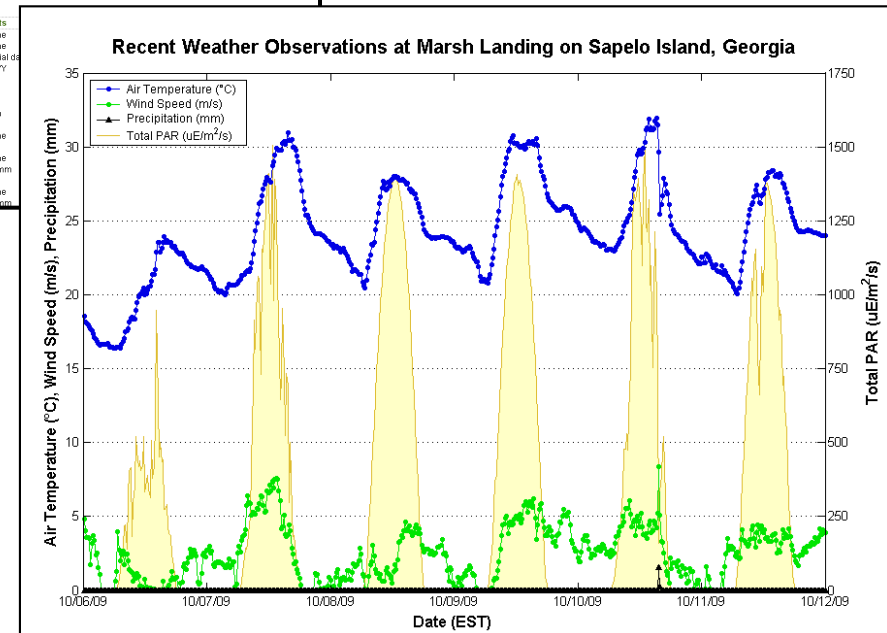
Access: Public

Metadata: [Text \(ESA FLEED\)](#)

Data Formats: [Spreadsheet CSV text file](#), [Tab-delimited text report](#), [GCE Data Toolbox file](#), [Standard MATLAB file](#), [Custom MATLAB file](#), [Custom Text file](#)

Column List:

Column	Name	Units
1	NESDRG_ID	none
2	INVSUID	none
3	Date	serial d
4	Year	YYYY
5	Month	MM
6	Day	DD
7	Hour	hr
8	Minute	min
9	Temp_Air	°C
10	Flag_Temp_Air	none
11	Max_Temp_Air	°C
12	Flag_Max_Temp_Air	none
13	Time_Max_Temp_Air	hhmm
14	Min_Temp_Air	°C
15	Flag_Min_Temp_Air	none
16	Time_Min_Temp_Air	hhmm



Ongoing Development

- ▶ **Metadata model enhancements**
 - Add “schema” support for better EML alignment
 - Improve metadata editor with dynamic form fields
- ▶ **Improve harvest management**
 - Add GUI tools for configuring harvest info, plot info
 - Add GUI for managing timers
 - Add data harvesting “dashboard” for monitoring activity
- ▶ **Workflow editor**
- ▶ **“Rebranding” as a community tool**

Resources

▶ MATLAB

- Website: <http://www.mathworks.com/products/matlab/>
- Version R13 (2002) or higher required (full or student version)

▶ Software Distribution

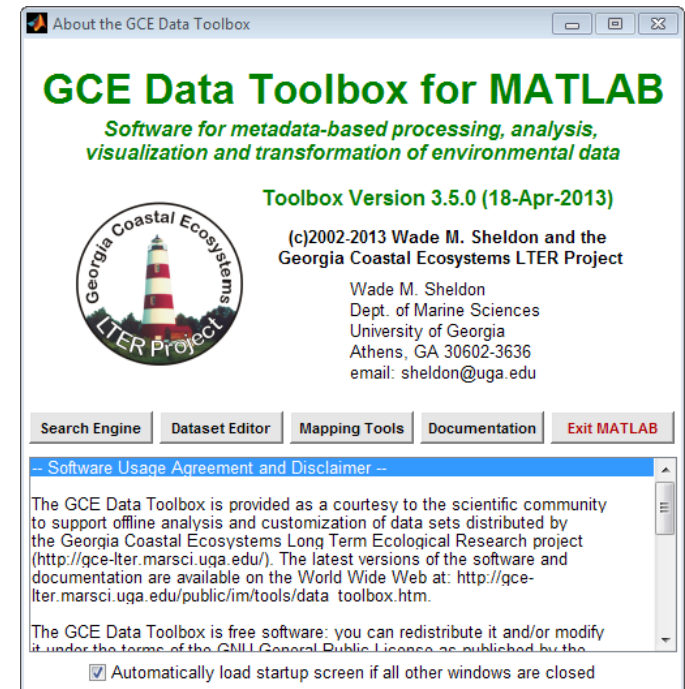
- Website: https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox
 - Documentation, FAQ
 - Distribution Downloads (stable, beta)
 - Bug reporting (tickets)
- SVN: https://gce-svn.marsci.uga.edu/svn/GCE_Toolbox/trunk
- Code is open source (GPLv3) and cross-platform

▶ User Support

- Peer-to-peer model
- Sporadic training opportunities (LTER)
- Email: gcetoolbox-l@listserv.uga.edu (<http://www.listserv.uga.edu/>)

Interactive Tutorial

- ▶ Introduction to the Dataset Editor application
- ▶ Importing and exploring data
 - Generic ASCII
 - Native logger files (Campbell, SeaBird)
 - Data services (Data Turbine, USGS NWIS)
- ▶ Metadata management
 - Defining attribute metadata
 - Documentation metadata, templates
- ▶ QA/QC framework
 - Defining flags and “rules”
 - Visual QA/QC
 - Managing flagged data
- ▶ Creating and exporting products
 - Batch processing raw data
 - Integrating data (join, merge)
 - Summarizing data
 - Exporting data sets and plots



Questions?

