

# Data Submission Training



Wade Sheldon, UGA (lead IM)  
Adam Sapp, UGA (assistant IM)

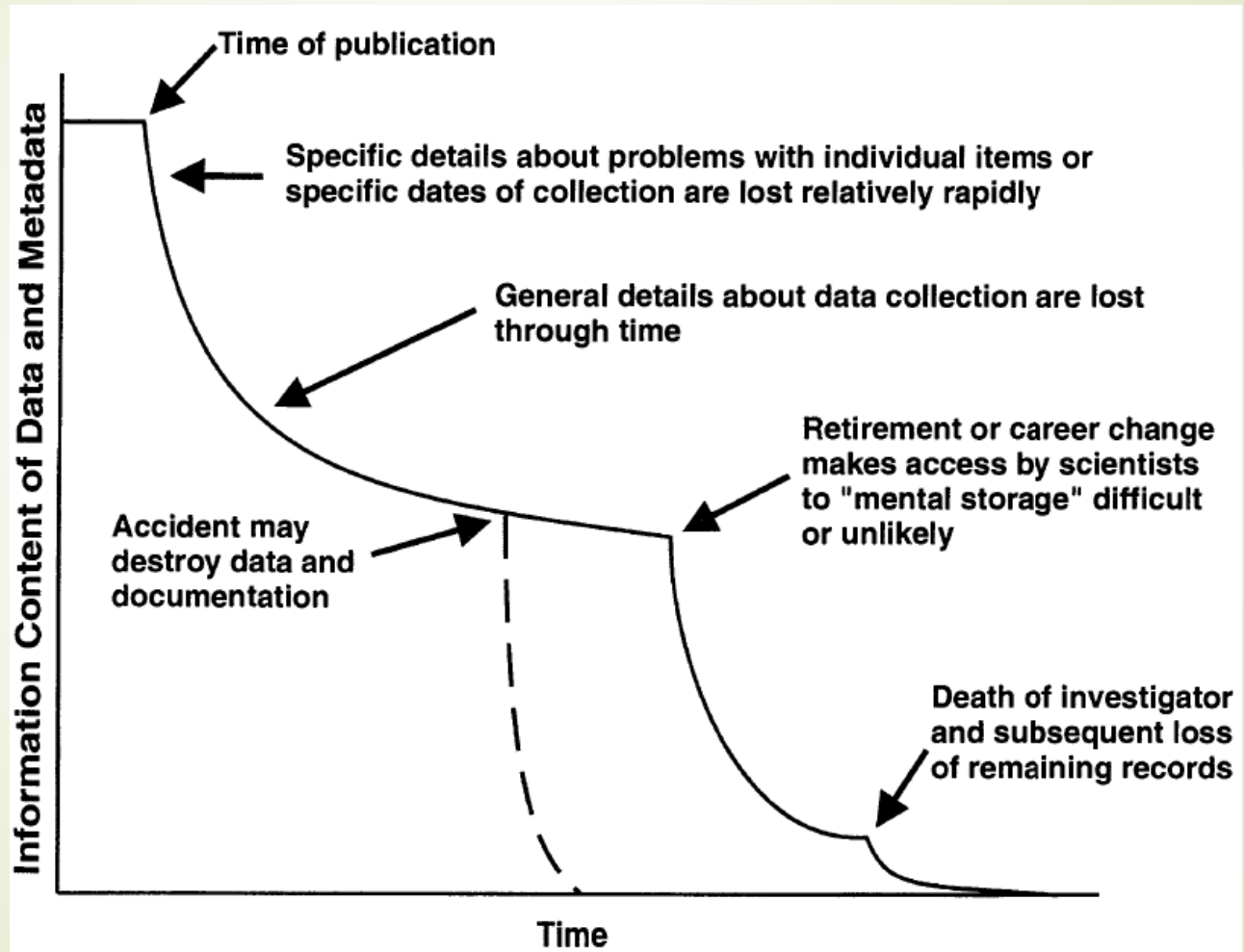
2019 GCE Annual Meeting



# Why Archive Data?

- ▶ Documenting and archiving data enables scientific discovery
  - ▶ Many ecological patterns and processes occur over expanses of time and space that cannot easily be sampled by a single investigator or team
  - ▶ Ecologists increasingly using data collected by others from numerous disciplines to address large scale questions
- ▶ Some observations cannot be repeated and would be lost if not archived
  - ▶ Climate change, sea-level rise, ocean acidification altering ecosystems
  - ▶ Land-use changes, habitat destruction
- ▶ Collaborative research requires data sharing
- ▶ If data are not documented promptly, information content “degrades” over time and cannot be used (even by the investigator)

# Information Decay



Michener et al. 1997. Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications*, 7(1):330-342.



# Why Archive Data?

- ▶ Other reason: **Because we have to!**
- ▶ NSF requires public access to primary data/samples 2 years after *collection* (not publication)
- ▶ Policy in place since 1990's, but is now being enforced
  - ▶ Data Management plans required, critically reviewed
  - ▶ Final reports not approved without data archiving
  - ▶ OCE wants relevant data listed at BCO-DMO (WHOI)
- ▶ LTER program under special scrutiny
  - ▶ 30 yr review found claims of online data overstated
  - ▶ Historic leadership role in data/metadata curation questioned
- ▶ ESA journals, etc. starting to require citable data with DOIs

# NSF OCE Data Policy

In accordance with NSF policy, Principal Investigators (PIs) working under OCE awards have additional conditions to which they need to adhere. These are specified below.

## ► DATA AND SAMPLE ARCHIVING REQUIREMENTS

The Division of Ocean Sciences requires that **metadata files, full data sets, derived data products and physical collections must be made publicly accessible within two (2) years of collection**. This includes software and derived data products (e.g., model results, output, and workflows). A brief description of [preferred data and physical collection archives and centers](#) and their criteria for submission can be found on the [OCE website](#) or through contact with the Cognizant Program Officer of the given award. Any limit on access to data, samples, or other information beyond the two-year moratorium period must be based on compelling justification, documented in the Data Management Plan of the proposal, or approved by the cognizant Program Officer.

Where no data or sample repository or archive exists for collected data and samples, the PI is required to identify a preservation plan in the Data Management Plan that complies with the general philosophy of sharing research products and data within two years of collection as described above.



# Benefits to GCE Investigators

- ▶ Automated metadata generation from GCE-IMS
  - ▶ Personnel, Study sites, Taxonomy, Project info, Site history, ...
  - ▶ Metadata generated in standardized formats (EML)
- ▶ Quality control
  - ▶ Basic: valid data type, valid value range, code checks
  - ▶ Advanced: pattern checks, historic range, variable dependencies, statistics (n-sigma), geographic precision
- ▶ Geo-referencing
  - ▶ Coordinates to place names
  - ▶ Place names to coordinates
- ▶ Data harvesting, processing, synthesis workflows for ongoing collection
  - ▶ Calibration, derived columns, data integration
- ▶ Data backup, off-site storage



# GCE Data Categories

- ▶ Quality-controlled “primary” data (*GCE-1, GCE-2, GCE-3, GCE-4*)
  - ▶ Observational data, calibrated measurements
  - ▶ Monitoring program (1 year release)
  - ▶ Investigator study data (2 year release)
  - ▶ Derived, calculated values only when essential to interpretation
  - ▶ High re-use potential
- ▶ Derived “secondary” data (*GCE-3, GCE-4*)
  - ▶ Long-term synthetic data sets from multiple primary data sets
  - ▶ Re-scaled, gap-filled “clean” data (e.g. daily, monthly, yearly)
  - ▶ Most useful for ecological analysis, cross-site comparisons
- ▶ Specialized data (*lower priority*)
  - ▶ Experimental results from controlled lab studies
  - ▶ Model calibration, output data
  - ▶ Can be archived for pubs, but require lots of metadata and qualification



# Submission Steps

1. Pre-consult with IM staff on data suitability, format (*recommended*)
2. Register study documentation (metadata) using online form
3. Prepare data and support files for submission
  - Tabular data in Data Submission Template (MS Excel)
  - Non-tabular data (e.g. GIS, raster images) in standard formats or ZIP archives
  - Protocols (PDF), custom analytical code (ZIP), ancillary metadata (TXT/PDF)
4. Upload data and supporting files using 'Add Files' link for submission
5. IM staff review submission, follow up with questions, make minor edits
6. Files returned for review (*if changes made*)
7. Tabular data processed using GCE Data Toolbox, quality-checked
8. Quality report, processed data returned for review (*if problems*)
9. Data files generated, packaged, put on server by IM staff
10. Metadata registered in the GCE Data Catalog by IM staff
11. Data catalog preview link sent for final review
12. Data set displayed in GCE Data Catalog
  - Metadata public immediately
  - Automatic data download links based on Project, Public release dates
13. After public access date, data and metadata replicated to LTER Data Portal, data set listed in DataONE and BCO-DMO



# Register Study Metadata

- Study Type (funding)
- Title
- Personnel
- Abstract
- Keywords

GCE Data Submission

https://gce-lter.marsci.uga.edu/private/app/add\_dataset.asp?type=template&id=656

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 > View Submissions > Submit Data

### GCE-LTER Data Submission

Forms for registering data sets generated with GCE-LTER support for inclusion in the [GCE Data Catalog](#) and [LTER Data Portal](#)

- Fill in all relevant fields before submitting for review (Note: fields denoted with an asterisk (\*) must be filled in before you can use Save Draft!)
- To copy metadata from a prior submission click on [View Submissions](#) above then find the relevant data set and click on the "Copy" task link
- After submitting the documentation (metadata), use the "Add Files" link to upload data files (see [Data Template](#) for details)

**Data Set ID**  (assigned by information manager) **Status**

**Study Type\***

**Related Projects**

**Data Set Title\***

**Contributor and Associated Personnel\***

**Registered Personnel**

- Adams, Tianjiao
- Alber, Meryll
- Alexander, Clark R.
- Alsbrooks, Cottie
- Anderson, Sarah

**Data Set Contributors**

- Craft, Christopher B.
- Smith, Dontrece
- Widney, Sarah
- Sheldon, Wade M.

**New Person:** Given Name  Middle Name  Surname  Suffix   
Institution  Email

(the first listed contributor will be the contact person for this resource)

**Abstract\***

The Georgia Coastal Ecosystems LTER Seawater Addition Long-Term Experiment (SALTEX) is a large-scale field experiment designed to simulate saltwater intrusion in a tidal freshwater wetland to predict how chronic (Press) and acute (Pulse) salinization will affect this and other tidal freshwater ecosystems. In order to characterize groundwater salinity, temperature, and plot flooding following experimental manipulation, unvented water pressure, temperature and conductivity were continuously measured in a PVC groundwater well installed at the SALTEX site. Measurements were made at the bottom of the well using a submerged Schlumberger CTD-Diver logger every 15 minutes from 30-May-2014 to 14-Feb-2018. In February 2016 a second CTD-Diver was deployed near the top of the well. Data were downloaded

**Key Words** (click in fields to select items from the list, click on X to remove items)

**Category**

**Context**

**Theme**

**Additional**  (separate with commas)

# Register Study Metadata

- Species
- Study Dates
- Geographic context
- Design
- Sampling
- Study plots
- Methodology

**GCE Data Submission** x +  
https://gce-lter.marsci.uga.edu/private/app/add\_dataset.asp?type=template&id=656

**Species** Species Name

**Study Dates\*** Start 5/30/2014 End 2/14/2018 (mm/dd/yyyy)

**Study Geography**  
Data set includes geographic location names or coordinates?  Yes  No (Study Sites and Locations required if "No")

Study Sites SALTEX - SALTEX Research Site x

Locations SALTEX\_Well1 (well) x

Bounding Coordinates (degrees) - optional  
North 31.339439  
West -81.466637 Clear -81.466637 East  
South 31.339439

**Study Design** Groundwater pressure, temperature and conductivity were measured continuously in a fixed-position groundwater well to characterize changes in water table, water chemistry and groundwater flow at the SALTEX research site near Champeny Island, Georgia.

**Sampling** Groundwater well water pressure, temperature and conductivity measurements were recorded every 15 minutes by autonomous data loggers installed in a PVC well. The well was also pumped every 14 days using a peristalsis pump to retrieve water samples for analysis to maintain well permeability.

**Study Plots** none

**Methods**

Method Step Well Installation

Method Details A single PVC groundwater well was installed adjacent to the SALTEX treatment plots. The well was placed in sand layers in order to prevent clogging and to allow sampling of a continuously connected groundwater source. The well was sunk to 1.24m depth, with a 1.87m riser above the sediment surface to prevent surface water contamination.

Instruments Pick

Other Instruments

Taxonomic Keys not applicable

Field Permits not applicable

**Methods**

Method Step Data Logger Installation

Method Details A Schlumberger CTD-Diver data logger was suspended by a nylon tether from the top of the PVC well riser to a height just above the bottom of the well, at a final depth of 1.16m below ground level in May 2014. In February 2016 a second CTD-Diver was added, suspended in the well to a depth of 0.58m below ground level. Both loggers were programmed to log unvented water pressure, temperature and conductivity at 15 minute intervals.

Instruments Pick Schlumberger CTD-Diver

Other Instruments

# Register Study Metadata

- Additional Methodology
- Ancillary Info
- Requested release dates
- Save Draft
- Submit

The screenshot displays a web browser window titled "GCE Data Submission" with the URL [https://gce-lter.marisci.uga.edu/private/app/add\\_dataset.asp?type=template&id=656](https://gce-lter.marisci.uga.edu/private/app/add_dataset.asp?type=template&id=656). The form is divided into several sections:

- Method Details:** A text area containing the following text: "Text files generated by Diver Office Software were parsed and imported into the GCE Data Toolbox for MATLAB software for post processing. The raw data files were documented using metadata templates, and columns for well location name, latitude, longitude and sensor elevation were added. The data sets were then integrated with barometric pressure data sets from the GCE-LTER Marsh Landing weather station on Sapelo Island, Georgia, to correct unvented water pressure readings for atmospheric pressure. Corrected depth and water level were then calculated from the corrected". Below this are fields for "Instruments" (with a "Pick" button), "Other Instruments", "Taxonomic Keys" (value: not applicable), and "Field Permits" (value: not applicable).
- Methods:** A section with a "Method Step" dropdown set to "Quality Control". Below it is another "Method Details" text area: "Quality control was performed using the GCE Data Toolbox for MATLAB software. Q/C rules were defined to assign flags to values outside of sensor range or outside of suitable deployment conditions (e.g. out of water measurements based on pressure). Flags were also automatically assigned to data values recorded during and through approximately 9 hours after pumping events (i.e. during the re-equilibration period) using the 'flag\_well\_pumping.m' function in the toolbox." This section also includes "Instruments" (with a "Pick" button), "Other Instruments", "Taxonomic Keys" (value: not applicable), and "Field Permits" (value: not applicable). An "Add Method Step" button is located below this section.
- Additional Information:** A section with the following fields:
  - Data Forms:** Copies of groundwater logs and raw data are maintained at the Georgia Coastal Ecosystems Information
  - Forms Location:** GCE-LTER Information Management Office, Dept. of Marine Sciences, University of Georgia, Athens, GA
  - Quality Control:** Quality control analysis was performed using the GCE Data Toolbox for MATLAB software. Column data
  - Validation:** not specified
  - Materials:** not applicable
  - Software:** SWS Diver-Office 5.0.25.0, MATLAB version 2017b
  - Anomalies:** (empty text area)
- Data Release\*:** Within GCE-LTER: 1/2/2019; General Public: 1/2/2021 (mm/dd/yyyy - maximum 2 years per NSF and LTER policy)

At the bottom of the form are two buttons: "Save Draft" and "Submit for Review".

*Date Submitted: (new submission)      Date Revised: (not edited)      Review Status: (not reviewed)*

# View Submission Status

- Saved Data Submissions
  - Edit
  - Add Files
  - Review (IM)
- Published Data Sets
  - Copy (template)
  - Revise
- Sortable table columns

The screenshot shows a web browser window displaying the 'View Data Submissions' page of the Georgia Coastal Ecosystems LTER Private Website. The page header includes the LTER logo and navigation links. The main content area is divided into two sections: 'Saved Data Submissions' and 'Prior GCE Data Catalog Submissions'. Both sections contain tables with columns for ID, Accession, Data Set Title, Study Type, Investigator, Submitted, Edited, Status, Files, and Tasks. The 'Saved Data Submissions' table has one entry (ID 1, Accession null, Title 'Testing test'). The 'Prior GCE Data Catalog Submissions' table has seven entries, including 'INV-GCEM-1804', 'INV-GCEM-1704', 'INV-GCEM-1604', 'INV-GCEM-1504', 'GIS-GCEL-1812a', 'GIS-GCEL-1812', and 'INS-GCEM-1708'. Each entry in the 'Prior' table includes a brief description of the study and links for 'Copy' and 'Revise'.

**View Data Submissions**

**Saved Data Submissions**

ID	Accession	Data Set Title	Study Type	Investigator	Submitted	Edited	Status	Files	Tasks
1		Testing test	GCE-III Monitoring	Pennings, Steven	2018-11-28	2018-12-23	not reviewed	2	<a href="#">Edit</a> <a href="#">Add Files</a> <a href="#">Review</a>

**Prior GCE Data Catalog Submissions**

ID	Accession	Data Set Title	Study Type	Investigator	Submitted	Edited	Status	Files	Tasks
677	<a href="#">INV-GCEM-1804</a>	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2017	GCE-III Monitoring	Angelini, Christine	2018-04-01		draft	5	<a href="#">Copy</a> <a href="#">Revise</a>
676	<a href="#">INV-GCEM-1704</a>	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2016	GCE-III Monitoring	Angelini, Christine	2017-04-01		draft	5	<a href="#">Copy</a> <a href="#">Revise</a>
675	<a href="#">INV-GCEM-1604</a>	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2015	GCE-III Monitoring	Silliman, Brian	2016-04-01		draft	5	<a href="#">Copy</a> <a href="#">Revise</a>
674	<a href="#">INV-GCEM-1504</a>	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2014	GCE-III Monitoring	Silliman, Brian	2014-04-01		draft	5	<a href="#">Copy</a> <a href="#">Revise</a>
673	<a href="#">GIS-GCEL-1812a</a>	May to July 2018 regions of interest (ROIs) of tidal marsh and tidal forest plant species to be used as ground reference data in habitat mapping	IRMA-RAPID Research Study	Hladik, Christine	2018-12-04		published	6	<a href="#">Copy</a> <a href="#">Revise</a>
672	<a href="#">GIS-GCEL-1812</a>	May to July 2018 ground control points GPS coordinates of tidal marsh and tidal forest plant species to be used as ground control points in habitat mapping	IRMA-RAPID Research Study	Hladik, Christine	2018-12-03		published	6	<a href="#">Copy</a> <a href="#">Revise</a>
671	<a href="#">INS-GCEM-1708</a>	Fall 2017 grasshopper monitoring -- mid-marsh grasshopper abundance and species diversity at eight GCE LTER sampling sites	GCE-III Monitoring	Pennings, Steven	2018-11-01		published	5	<a href="#">Copy</a> <a href="#">Revise</a>

# Data Submission Template


► Data Products > Data Submission > Data Template

The screenshot shows a web browser window displaying the 'Data Submission' page on the Georgia Coastal Ecosystems LTER website. The page title is 'File Details' and it provides information about a Microsoft Excel spreadsheet template for data submission. The page includes a navigation menu on the left, a breadcrumb trail, and a main content area with sections for Title, Archive, Description, Contributor, Citation, Key Words, File Date, and Web Link. The footer contains the date '16-May-2018', contact information, and a disclaimer about NSF funding.

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



Home > File Archive > Other Files > Forms > Data Submission > Details | Update File (GCE only)

### File Details

<b>Title</b>	GCE data-only submission form for describing tabular data files to attach to data sets submitted online
<b>Archive</b>	<a href="#">All Files</a> / <a href="#">Other Files</a> / <a href="#">Forms</a> / <a href="#">Data Submission</a>
<b>Description</b>	<p>This Microsoft Excel spreadsheet template is provided for formatting and describing tabular data files for submission to the GCE Information Management Office for archiving in the GCE Data Catalog and LTER Data Portal. Note that data set documentation (metadata) must be provided separately online prior to submission of the data files (see <a href="https://gce-lter.marisci.uga.edu/private/app/add_dataset.asp">https://gce-lter.marisci.uga.edu/private/app/add_dataset.asp</a> to register the documentation).</p> <p>For tabular data (e.g. spreadsheets, non-digital data sheets, simple comma or space-delimited logger files), the data values and column information should be entered or pasted into the "Tabular Data" worksheet unless prior arrangements are made with IM staff for parsing data from specialized formats (e.g. raw data logger files, real-time data telemetry or lab-specific storage formats). Specialized tabular or non-tabular data (e.g. GIS files, raster imagery, genomics data) can be described and uploaded directly from the View Submissions page (see below) after the metadata are submitted so use of this template is not required.</p> <p>Completed templates and data files (if provided separately) should be uploaded to the GCE IM office using the "Add Files" links for the corresponding data set metadata on the Private GCE Web Site (<a href="https://gce-lter.marisci.uga.edu/private/app/view_submissions.asp">https://gce-lter.marisci.uga.edu/private/app/view_submissions.asp</a>).</p> <p>More information about GCE data submission is available at <a href="http://gce-lter.marisci.uga.edu/public/im/data_submission.htm">http://gce-lter.marisci.uga.edu/public/im/data_submission.htm</a>. Note that a GCE Data Submission Training presentation and Zip archive of sample submissions are also available for downloading (see <a href="http://gce-lter.marisci.uga.edu/public/app/resource_details.asp?id=535">http://gce-lter.marisci.uga.edu/public/app/resource_details.asp?id=535</a>). However, this training material is now out of date, and will be updated in January 2019.</p>
<b>Contributor</b>	<a href="#">Wade Sheldon</a>
<b>Citation</b>	Wade Sheldon. 2019. GCE data-only submission form for describing tabular data files to attach to data sets submitted online. Georgia Coastal Ecosystems LTER File Archive, University of Georgia, Athens, Georgia. ( <a href="http://gce-lter.marisci.uga.edu/public/app/resource_details.asp?id=852&amp;version=3">http://gce-lter.marisci.uga.edu/public/app/resource_details.asp?id=852&amp;version=3</a> )
<b>Key Words</b>	data submission, form, IM, spreadsheet, template
<b>File Date</b>	Jan 02, 2019 (version 3)
<b>Web Link</b>	 <a href="#">view/download MS Excel file (313 kb)</a>

16-May-2018 [Contact Us](#) | [Privacy](#)

This material is based upon work supported by the National Science Foundation under grants OCE-9982133, OCE-0620959 and OCE-1237140. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



[http://gce-lter.marisci.uga.edu/public/app/resource\\_details.asp?id=852](http://gce-lter.marisci.uga.edu/public/app/resource_details.asp?id=852)



# Template – Tabular Data

Name,  
Description  
Units

Types, Precision,  
Code Defs,  
Calculations

Q/C Limits

Data  
Values

AutoSave GCE\_data\_subm\_template\_dec2018.xlsx - Excel Wade Sheldon

File Home Insert Page Layout Formulas Data Review View Help Acrobat Tell me what you want to do

B10 Date

**GCE Data Table Submission Template**

**Instructions:**

- Start by submitting documentation (metadata) for your data set online at: [https://gce-iter.marsci.uga.edu/private/app/add\\_dataset.asp](https://gce-iter.marsci.uga.edu/private/app/add_dataset.asp)
- Fill out the header section for each data column below as completely as possible. Fields in light blue are required!
- Paste or enter your data values into the 'Values' section (white cells), starting with cell B25
- Fill in missing values in the table with NaN (not a number), including text fields, and do not skip columns
- View help text for header fields by hovering the mouse pointer over the field name comment indicator (red triangle)
- Upload completed data submission template(s) using the "Add Files" link for the relevant submission at: [https://gce-iter.marsci.uga.edu/private/app/view\\_submissions.asp](https://gce-iter.marsci.uga.edu/private/app/view_submissions.asp)

Column Name:	Date	Temp_Air	Max_Temp_Air	Min_Temp_Air	Humidity	Baro_Pressure	Wind_Speed	Wind_Direction	SD_Wind_Direction	Max_Wind_Speed
Description:	Calendar date and time of observation	Air temperature, 15 minute mean	Air temperature, 15 minute maximum	Air temperature, 15 minute minimum	Relative humidity, 15 minute average	Barometric pressure, 15 minute average	Wind speed, 15 minute average	Wind vector, 15 minute vector average performed by the CSI CR1000 datalogger program; no magnetic correction	Standard deviation of wind direction	Maximum wind speed, 15 minute
Units:	yyyy-mm-dd HH:MM:SS - GMT	°C	°C	°C	%	mbar	m/s	degrees	degrees	m/s
Data type:	string	floating-point	floating-point	floating-point	floating-point	floating-point	floating-point	floating-point	floating-point	floating-point
Variable type:	datetime	data	data	data	data	data	data	data	data	data
Number type:	none	continuous	continuous	continuous	continuous	continuous	continuous	continuous	continuous	continuous
Precision:	0	2	2	2	2	1	1	2	1	2
Code values:										
Calculations:										
QC: Minimum Valid:		-20	-20	-20	-20	0	0	0	0	0
QC: Minimum Expected:		-2	-2	-2	-2	1010	1010			
QC: Maximum Expected:		40	40	40	40	110	1040	25		30
QC: Maximum Valid:		70	70	70	70	120	1100	40	360	40
QC: Custom:										
Values:	2017-01-01 00:00:00	11.29	11.46	11.12	74.0	1022.3	1.84	25.7	7.5	2.40
2017-01-01 00:15:00	11.24	11.36	11.15	75.8	1022.7	0.65	37.8	7.2	1.65	
2017-01-01 00:30:00	11.22	11.36	11.05	78.3	1023.0	0.00	0.0	0.0	0.00	
2017-01-01 00:45:00	11.02	11.12	10.88	81.2	1023.0	0.00	0.0	0.0	0.00	
2017-01-01 01:00:00	10.91	10.95	10.81	82.6	1023.2	0.00	0.0	0.0	0.00	
2017-01-01 01:15:00	10.89	11.02	10.81	82.5	1023.1	0.00	0.0	0.0	0.00	
2017-01-01 01:30:00	10.97	11.06	10.89	81.7	1023.0	0.00	0.0	0.0	0.00	
2017-01-01 01:45:00	11.01	11.09	10.95	82.1	1023.0	0.18	23.0	0.4	1.34	
2017-01-01 02:00:00	11.03	11.09	10.96	83.2	1023.1	0.91	55.0	11.2	1.20	
2017-01-01 02:15:00	11.05	11.13	10.96	83.9	1023.3	0.15	38.2	1.9	0.90	
2017-01-01 02:30:00	11.08	11.16	11.02	83.7	1023.2	0.00	0.0	0.0	0.00	
2017-01-01 02:45:00	11.08	11.13	11.02	84.1	1022.9	0.00	0.0	0.0	0.00	
2017-01-01 03:00:00	11.16	11.33	11.06	83.9	1022.8	0.00	0.0	0.0	0.00	
2017-01-01 03:15:00	11.33	11.64	11.16	84.2	1022.8	0.00	0.0	0.0	0.00	
2017-01-01 03:30:00	11.62	11.71	11.57	83.4	1022.6	0.53	126.8	6.0	1.80	
2017-01-01 03:45:00	11.74	11.91	11.60	82.5	1022.5	0.95	132.4	11.3	1.95	
2017-01-01 04:00:00	12.02	12.08	11.91	81.8	1022.4	0.00	136.3	0.0	0.60	
2017-01-01 04:15:00	11.95	12.04	11.87	82.7	1022.1	0.04	136.7	3.6	1.04	
2017-01-01 04:30:00	12.21	12.45	11.94	83.2	1021.9	0.04	138.4	2.0	1.20	
2017-01-01 04:45:00	12.36	12.55	12.17	83.6	1021.9	0.00	0.0	0.0	0.00	
2017-01-01 05:00:00	12.01	12.24	11.63	86.2	1021.8	0.00	0.0	0.0	0.00	
2017-01-01 05:15:00	11.64	11.70	11.60	88.7	1021.6	0.00	0.0	0.0	0.00	
2017-01-01 05:30:00	11.65	11.73	11.63	90.5	1021.6	0.00	0.0	0.0	0.00	
2017-01-01 05:45:00	11.65	11.70	11.60	91.7	1021.5	0.00	0.0	0.0	0.00	
2017-01-01 06:00:00	11.70	11.80	11.63	92.5	1021.3	0.00	0.0	0.0	0.00	
2017-01-01 06:15:00	11.75	11.80	11.66	93.1	1021.1	0.00	0.0	0.0	0.00	
2017-01-01 06:30:00	11.76	11.84	11.70	93.7	1021.0	0.00	0.0	0.0	0.00	
2017-01-01 06:45:00	11.78	11.87	11.73	94.2	1020.9	0.00	0.0	0.0	0.00	
2017-01-01 07:00:00	11.93	12.17	11.83	94.4	1020.8	0.00	0.0	0.0	0.00	
2017-01-01 07:15:00	12.42	12.58	12.14	94.6	1021.0	0.07	48.3	1.2	1.34	

Tabular Data

Ready

# Upload File(s) for Submission

## View Submissions

## Add Files

View Data Submissions

Saved Data Submissions

ID	Accession	Data Set Title	Study Type	Investigator	Submitted	Edited	Status	Files	Tasks
1		Testing test	GCE-III Monitoring	Pennings, Steven	2018-11-28	2018-12-23	not reviewed	1	Edit Add Files Review

Prior GCE Data Catalog Submissions

ID	Accession	Data Set Title	Study Type	Investigator	Submitted	Edited	Status	Files	Tasks
677	INV-GCEM-1804	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2017	GCE-III Monitoring	Angelini, Christine	2018-04-01		draft	5	Copy Revise
676	INV-GCEM-1704	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2016	GCE-III Monitoring	Angelini, Christine	2017-04-01		draft	5	Copy Revise
675	INV-GCEM-1604	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2015	GCE-III Monitoring	Silliman, Brian	2016-04-01		draft	5	Copy Revise
674	INV-GCEM-1504	Yearly survey of barnacle settlement in mid-marsh and creekbank plots at GCE LTER study sites in October 2014	GCE-III Monitoring	Silliman, Brian	2014-04-01		draft	5	Copy Revise
673	GIS-GCEL-1812a	May to July 2018 regions of interest (ROIs) of tidal marsh and tidal forest plant species to be used as ground reference data in habitat mapping	IRMA-RAPID Research Study	Hladik, Christine	2018-12-04		published	6	Copy Revise
672	GIS-GCEL-1812	May to July 2018 ground control points GPS coordinates of tidal marsh and tidal forest plant species to be used as ground control points in habitat mapping	IRMA-RAPID Research Study	Hladik, Christine	2018-12-03		published	6	Copy Revise
671	INS-GCEM-1708	Fall 2017 grasshopper monitoring - mid-marsh grasshopper abundance and species diversity at eight GCE LTER sampling sites	GCE-III Monitoring	Pennings, Steven	2018-11-01		published	5	Copy Revise
670	INV-GCEM-1809	Fall 2017 crab population monitoring: mid-marsh and creek bank abundance based on crab hole counts at GCE marsh, monitoring sites 1-10	GCE-III Monitoring	Pennings, Steven	2018-10-31		published	5	Copy Revise
669	INV-GCEM-1807a	Mollusc population size distribution monitoring: fall 2017 mid-marsh and creekbank, infaunal and epifaunal mollusc size distributions based on collections from GCE	GCE-III Monitoring	Pennings, Steven	2018-10-31		published	5	Copy Revise

Add Submission Files

Add Files to a GCE Data Submission

Existing Files

Filename	File Category	File Description	File Size	Date Added
li200x_20181223T93353.pdf	Research Protocol	LICOR documentation	460 kb	Dec 23, 2018
Upload_test_20190101T121311.txt	Ancillary Metadata	Test file	0.01 kb	Jan 01, 2019

Upload New File

GCE contact: Sheldon, Wade M. (required)

Data Set Title: Testing test

Research Project: [Dropdown menu with options: Not Applicable, Alber, Merryly - Altamaha River salinity modeling... (ongoing), Alber, Merryly - Altamaha watershed nutrient budgets... (ended 2006), Alber, Merryly - Analysis of historic vegetation changes... (ended 2001), Alber, Merryly - Benthic diatoms... (ended 2004), Alber, Merryly - Chlorophyll distributions... (ended 2005)]

File Category: <Select a Category>

File Type: <Select a Type>

File Description: [Text input field]

File to Upload: [Choose File] No file chosen

Notes or Analysis Requests: [Text area]

[Submit] [Reset]

(Note: Processing uploads >10 MB will take some time - please do not press 'Submit' again!)

Notes on Data Submission:

- Tabular data files should be documented and formatted using the [Data Submission Template](#) unless prior arrangements are made
- Use the form fields below to upload the submission template spreadsheet and any supporting files that should be archived with the data set (e.g. non-tabular files, protocol documents, maps)
- The total size of uploaded files must be <= 500MB or the upload may be rejected by the server - please ZIP large files or contact the [IM Office](#) for alternative file transfer options
- Files with the following extensions are not allowed for security reasons:
  - disallowed extensions: <code>.cnone>, .exe, .com, .bat, .cgi, .asp, .pl, .vbs, .scr, .java



# Data Checks (Tabular)

- ▶ Tabular data are processed in MATLAB using the GCE Data Toolbox
- ▶ Validation checks
  - ▶ Valid columns/rows
  - ▶ Missing values
  - ▶ Column data types
  - ▶ Date/time formats
  - ▶ Codes and definitions
- ▶ Quality Checks
  - ▶ Range checks
  - ▶ Time series spikes/patterns
  - ▶ Geographic coordinates in range
  - ▶ Sanity checks for replicate measurements
  - ▶ Dependency checks for related measurements



# GCE Data Toolbox Q/C

Mollusc population size distribution monitoring: Fall 2010 mid-mars...

File Edit Metadata Tools Misc Window Help

Column List (select to display properties)

- Year (YYYY)
- Month (M)
- Day (D)
- Site (none)
- Zone (none)
- Plot (none)
- Location (none)
- Location\_Notes (none)
- Longitude (degrees)
- Latitude (degrees)
- Species (none)
- Mollusc\_Length (mm)
- Condition (none)

Column Name Year  
Column Units YYYY  
Description Calendar year of ob  
Data Type inte  
Variable Type date  
Numerical Type disc  
Precision 0  
Flag Criteria x<2

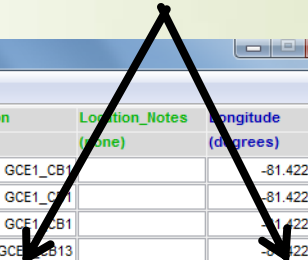
Move First  
Move Up  
Move Down  
Move Last

Data Editor

File Edit Options

All	Year	Month	Day	Site	Zone	Plot	Location	Location_Notes	Longitude
None	(YYYY)	(M)	(D)	(none)	(none)	(none)	(none)	(none)	(degrees)
<input type="checkbox"/>	1	2010	10	19	1	1	1	GCE1_CB1	-81.422539
<input type="checkbox"/>	2	2010	10	19	1	1	1	GCE1_CB1	-81.422539
<input type="checkbox"/>	3	2010	10	19	1	1	13	GCE1_CB13	-81.422539
<input type="checkbox"/>	4	2010	10	19	1	1	1	GCE1_CB13	-81.422441
<input type="checkbox"/>	5	2010	10	19	1	1	36	GCE1_CB36	-81.422287
<input type="checkbox"/>	6	2010	10	19	1	1	18	GCE1_CB8 location estimated b	-81.422205
<input type="checkbox"/>	7	2010	10	19	1	2	1	GCE1_MM1	-81.422821
<input type="checkbox"/>	8	2010	10	19	1	2	3	GCE1_MM3	-81.422546
<input type="checkbox"/>	9	2010	10	19	1	2	6	GCE1_MM6	-81.422098
<input type="checkbox"/>	10	2010	10	19	1	2	8	GCE1_MM8	-81.421750
<input type="checkbox"/>	11	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	12	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	13	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	14	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	15	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	16	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	17	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	18	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	19	2010	10	19	2	1	11	GCE2_CB11	-81.296118
<input type="checkbox"/>	20	2010	10	19	2	1	33	GCE2_CB23 location estimated b	-81.295753
<input type="checkbox"/>	21	2010	10	19	2	1	33	GCE2_CB23 location estimated b	-81.295753
<input type="checkbox"/>	22	2010	10	19	2	1	33	GCE2_CB23 location estimated b	-81.295753
<input type="checkbox"/>	23	2010	10	19	2	1	33	GCE2_CB23 location estimated b	-81.295753
<input type="checkbox"/>	24	2010	10	19	2	1	33	GCE2_CB23 location estimated b	-81.295753
<input type="checkbox"/>	25	2010	10	19	2	1	33	GCE2_CB23 location estimated b	-81.295753

Flagged Values



# GCE Data Catalog

Georgia Coastal Ecosystems LTER

Member of the NSF Long Term Ecological Research Network

Home > Data > Data Search > Data Catalog > Data Set Summary | [Update Metadata \(contributor\)](#) | [Copy as Template \(GCE\)](#)

**GCE-LTER Data Set Summary** [Jump to Files](#)

<b>Accession:</b>	<b>GND-GCED-1802</b>	<b>Research Theme:</b>	<b>Groundwater Hydrology (Directed Study)</b>
<b>Contributors:</b>	<a href="#">Christopher B. Craft</a> , Dontrece Smith, Sarah Widney, Wade Sheldon		
<b>Title:</b>	Continuous groundwater well temperature, salinity and water level measurements at the GCE-LTER Seawater Addition Long-Term Experiment (SALTEX) site from May 2014 to February 2018		
<b>Abstract:</b>	The Georgia Coastal Ecosystems LTER Seawater Addition Long-Term Experiment (SALTEX) is a large-scale field experiment designed to simulate saltwater intrusion in a tidal freshwater wetland to predict how chronic (Press) and acute (Pulse) salinization will affect this and other tidal freshwater ecosystems. In order to characterize groundwater salinity, temperature, and plot flooding following experimental manipulation, unvented water pressure, temperature and conductivity were continuously measured in a PVC groundwater well installed at the SALTEX site. Measurements were made at the bottom of the well using a submerged Schlumberger CTD-Diver logger every 15 minutes from 20-May-2014 to 14-Feb-2018. In February 2016 a second CTD-Diver was deployed near the top of the well. Data were downloaded from the loggers using Diver Office communication software, then imported into MATLAB for post-processing, quality control and documentation. Raw, unvented pressure readings were corrected for atmospheric pressure and sensor height from the bottom of the well to generate corrected pressure readings, then water level, salinity and density were calculated from the measured variables using UNESCO algorithms. These data were collected as part of the Georgia Coastal Ecosystems LTER SALTEX project ( <a href="http://gce-lter.marsci.uga.edu/public/app/send_project_email?id=73">http://gce-lter.marsci.uga.edu/public/app/send_project_email?id=73</a> ), and will be updated annually.		
<b>DOI:</b>	<a href="https://doi.org/10.6073/pasta/bdde791bb928c451261a2c3e1561ff56">10.6073/pasta/bdde791bb928c451261a2c3e1561ff56</a>		
<b>Key Words:</b>	droughts, estuaries, floods, groundwater, hydrology, pressure, salinity, SALTEX, temperature, water level, water table		
<b>LTER Core Area:</b>	<a href="#">Disturbance Patterns</a>		
<b>Research Themes:</b>	<a href="#">Q1 - External Forcing</a> , <a href="#">Q3 - Longitudinal Gradients</a> , <a href="#">GCE3 Area3 - Responses to Salinity and Inundation</a> , Groundwater Hydrology, Hydrography/Hydrology		
<b>Study Period:</b>	30-May-2014 to 14-Feb-2018		
<b>Study Sites:</b>	<a href="#">SALTEX</a> -- SALTEX Research Site, Champney Island, Georgia, USA		

» Download Geographic Coverage: [Google Earth](#), [ESRI Shapefile \(polygons, points\)](#)

**Project References:** [SALTEX - Seawater Addition Long Term Experiment](#)

**Publications:** [Herbert, E., Schubauer-Berigan, J.P. and Craft, C.B. 2018. Differential effects of chronic and acute simulated seawater intrusion on tidal freshwater marsh carbon cycling. Biogeochemistry. Published online 24-Mar-2018. \(DOI: 10.1007/s10533-018-0436-z\)](#)

**Downloads:** **Data Table:** GND-GCED-1802\_Bottom (Observations from the data logger installed at the well bottom from May 2014 to Feb 2018, 130182 records)

**Access:** Public (released 01-Dec-2018)

# GCE Data Search

The screenshot shows the Georgia Coastal Ecosystems LTER Data Search interface. At the top, the browser address bar shows the URL `gce-lter.marsci.uga.edu/public/app/data_search.asp`. The page header includes the LTER logo and navigation links: Home, Data, Data Search, Data Catalog, Use Agreement, Distribution Formats, and EML Metadata. A left sidebar contains a menu with items like Home, GCE News, Research, Study Site, Field Planning, Bibliography, Data Products, GIS Resources, Maps & Imagery, Documents, Outreach & Ed, Citizen Science, Informatics, Personnel, Affiliates, Private Site, and Site Search.

The main content area is titled "Search for Core GCE Data" and features a "Quick Search" section with a text input field for "Search metadata text for:" and a "Search" button. Below this is an "Advanced Search" section with several criteria:

- Subject:** Includes dropdowns for "LTER Core Area/GCE Theme" (set to "<any>") and "Long-term Data Series" (set to "<any>"). It also has two "Subject Search" rows, each with a dropdown for "Any Text" and a "contains" operator.
- Author:** Includes a dropdown for "Individual's Last Name" (set to "<any>").
- Temporal Criteria:** Includes a "Study Period" field with "to" and "(M/D/YYYY)" format, a radio button for "all dates within range" (selected), and a "contains dates" option. It also has a "Data Access" dropdown set to "<Any Access>".
- Spatial Criteria:** Includes a map of the Georgia coast with labels for Shellman Bluff, Sapelo Island, Carnige, Darien, Dock Junction, Brunswick, Waverly, and Jekyll Island. To the right of the map are "Geographic Bounds" fields for North, West, East, and South, a "Clear" button, a checkbox for "all locations within bounds", and a "Location" dropdown set to "starts with".
- Taxonomic Criteria:** This section is partially visible at the bottom of the page.

# General Data Guidelines

- ▶ All data sets require location information (*ideally as columns/attrib.*)
  - ▶ Registered GCE place names\* (sites, locations, plots, wells)
  - ▶ GPS coordinates (lat/lon, UTM)
- ▶ All data sets require date information (*ideally as columns/attributes*)
  - ▶ Calendar date/time
  - ▶ Year, Month, Day, ...
  - ▶ Year, YearDay
- ▶ Data tables must be column-oriented, rectangular
  - ▶ Columns contain single type of data (= variable)
  - ▶ Rows contain observations for each variable
- ▶ Data can be “packaged” as appropriate for study
  - ▶ Single table
  - ▶ Multiple tables
  - ▶ Mixed tabular and non-tabular (e.g. spreadsheet + shapefile)

\* [http://gce-lter.marsci.uga.edu/public/app/geo\\_query.asp](http://gce-lter.marsci.uga.edu/public/app/geo_query.asp)



# Metadata Guidelines

- ▶ Data Set Title
  - ▶ Where, what, when
  - ▶ Should uniquely describe data in pool of >10k data sets
- ▶ Data Set Abstract
  - ▶ Describe where, what, when plus why in more detail
  - ▶ List variables measured when applicable
  - ▶ Summarize study design, methods
  - ▶ Do NOT describe conclusions from study
  - ▶ Should answer: “Do I want to download these data?”
- ▶ Methods
  - ▶ Should provide enough detail to interpret data without calling PI
  - ▶ Instrumentation should be described sufficiently to evaluate data accuracy, determine comparability



# Data Table Guidelines

- ▶ Data Column / Attributes Names
  - ▶ Names should be descriptive, not cryptic
    - ▶ Salinity, Temperature, Depth (*not S, T, Dep*)
  - ▶ Names should include scale and statistical parameter for summarized variables (Daily\_Mean\_Salinity)
  - ▶ Names should include property measured when appropriate (Organic\_Percent, Sediment\_Mass)
  - ▶ Names should NOT include spaces, math symbols, punctuation
  - ▶ IM staff will provide advice on site standards
- ▶ Data Column / Attribute Units
  - ▶ Full unit names (celsius) or literature-standard abbreviations and symbols (°C)
  - ▶ Must be plain text (no superscripts, subscripts, LaTeX)
- ▶ Column precision should reflect significant digits

# Value Qualifiers

- ▶ Qualifiers (flags) should be in columns adjacent to data
  - ▶ NOT encoded as cell comments or text in place of numbers
  - ▶ Codes must be defined if used (e.g. E = estimated value, I = invalid out of range value)
- ▶ Bad practice:

Date	Site	Salinity	Temperature
01/01/2013 00:00	gce6_hydro	30.2	20.5
01/01/2013 00:30	gce6_hydro	30.3	20.5
01/01/2013 01:00	gce6_hydro	30.4	20.4
01/01/2013 01:30	gce6_hydro	30.5	invalid
01/01/2013 02:00	gce6_hydro	30.6	20.5
01/01/2013 02:30	gce6_hydro	30.7	20.6

- ▶ Good practice:

Date	Site	Salinity	Temperature	Flag_Temperature
01/01/2013 00:00	gce6_hydro	30.2	20.5	
01/01/2013 00:30	gce6_hydro	30.3	20.5	
01/01/2013 01:00	gce6_hydro	30.4	20.4	
01/01/2013 01:30	gce6_hydro	30.5	36.5	I
01/01/2013 02:00	gce6_hydro	30.6	20.5	
01/01/2013 02:30	gce6_hydro	30.7	20.6	

- ▶ Generally best practice to retain and qualify questionable values, delete and qualify impossible values, document rationale

# Data Format (Tabular Data)

- ▶ Data should be entered or pasted as values into the template, then formatted to reflect significant digits
- ▶ Logger files, specialized lab formats can be submitted for automated processing for ongoing data collection (*arrange with IM staff first*)
- ▶ Value formats should be consistent throughout data set

▶ Bad practice:

Date	Longitude	Latitude	Salinity
1/1/2013 1:00	-82.2345	31.1234	32.23
1/1/2013 2:00	-82.2586	31.1235	30.21
1/1/2013 3:00	-82.2843	31.1145	29.33
1/1/2013 4:00	82° 21.12'	31° 9.12'	20.1253
1/1/2013 5:00	82° 22.34'	31° 9.65'	20.1132

- ▶ Missing values should be left empty or encoded as NaN (*converted to format-specific encoding when posted*)
- ▶ Value codes must be defined if used (Spart = Spartina alterniflora, Junc = Juncus roemerianus, ...)



# Data Format (Tabular Data)

- There should generally be a single column for each variable, not multiple columns split by factor (date, species, core, ...)
- Not preferred (*repeating groups – limited analytical potential*)

Date	Site	Spartina_Biomass	Juncus_Biomass	Borrichia_Biomass
1/1/2013	1	34	5	0
1/1/2013	2	40	0	0
1/1/2013	3	25	0	2
1/1/2013	4	30	0	1
1/1/2013	5	38	0	2

- Preferred (*normalized – many ways to filter, analyze*)

Date	Site	Species	Biomass
1/1/2013	1	Spartina	34
1/1/2013	2	Spartina	40
1/1/2013	3	Spartina	25
1/1/2013	4	Spartina	30
1/1/2013	1	Juncus	5
1/1/2013	2	Juncus	0
1/1/2013	3	Juncus	0
1/1/2013	4	Juncus	0
1/1/2013	1	Borrichia	0
1/1/2013	2	Borrichia	0


# Data Format (Tabular Data)

- ▶ Data collected at different levels of replication or scales should be split into multiple data files, uploaded separately
- ▶ Not preferred (*mixed replication*)

Date	Site	Species	Height	Salinity
1/1/2013	1	Spartina	23	24.3
1/1/2013	1	Spartina	35	
1/1/2013	1	Spartina	42	
1/1/2013	2	Spartina	57	17.9
1/1/2013	2	Spartina	25	
1/1/2013	2	Spartina	32	
1/1/2013	3	Spartina	48	22.5
1/1/2013	3	Spartina	37	
1/1/2013	3	Spartina	55	

- ▶ Preferred (*split into 2 data files*)

Date	Site	Species	Height	Date	Site	Species	Mean_Height	Salinity
1/1/2013	1	Spartina	23	1/1/2013	1	Spartina	33.3	24.3
1/1/2013	1	Spartina	35	1/1/2013	2	Spartina	38.0	17.9
1/1/2013	1	Spartina	42	1/1/2013	3	Spartina	46.7	22.5
1/1/2013	2	Spartina	57					
1/1/2013	2	Spartina	25					
1/1/2013	2	Spartina	32					
1/1/2013	3	Spartina	48					
1/1/2013	3	Spartina	37					
1/1/2013	3	Spartina	55					



# Data Format (Non-tabular)

- ▶ Choose formats maximally useful to discipline
  - ▶ Vector GIS: ESRI shapefiles, File geodatabases
  - ▶ Raster GIS: ESRI Grid, GeoTIFF
  - ▶ Imagery: TIFF, PNG, JPEG
  - ▶ Genomics: FASTA, GenBank/EMBL
  - ▶ Other: consult with GCE IM staff
- ▶ Should include format-appropriate metadata in addition to GCE metadata (e.g. ESRI XML for ArcCatalog)
- ▶ Large or compound files can be combined in ZIP file
- ▶ Can be archived along with tabular data, include support files needed for interpretation (protocols, code, readme files)



# Example Data Submissions

➤ (see provided spreadsheet)

