Data Submission Training

Wade Sheldon
GCE-LTER Information Manager

(revised January 2015)
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

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 contributed to PCO-DMO at 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
In accord with NSF policy, Principal Investigators (PIs) working under OCE awards have additional conditions to which they need to adhere. These are specified below.

A. PIs are required to submit, at no more than incremental cost and within a reasonable time frame (but no later than two (2) years after the data are collected), the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF/OCE grants to the appropriate Data Center (See appendices below or consult with the cognizant NSF Program Officer).

B. Where no data or sample repository exists for collected data or samples, metadata must be prepared and made publicly available over the Internet and the PI is required to employ alternative strategies for complying with the general philosophy of sharing research products and data as described above. Where appropriate, samples should be curated in a manner that preserves the quality and integrity of the samples. When in doubt about what to do, the PI should consult with the NSF Program Officer handling the award.

( Division of Ocean Sciences Sample and Data Policy. May 2011. National Science Foundation, publication NSF11060.)
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 processing, synthesis workflows for ongoing collection
  - Calibration, derived columns, data integration
- Data backup, off-site storage
GCE Data Categories

• Quality-controlled “primary” data \((GCE-1, GCE-2, GCE-3)\)
  ◦ 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 \((GCE-3)\)
  ◦ 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
Study Data Submission

1. Consult with IM staff on data suitability, format (recommended)
2. Fill in data submission template
   ◦ Documentation (metadata)
   ◦ Tabular data in template, non-tabular data as separate upload or link
3. Upload template and supporting files to server
4. IM staff review submission, follow up with questions
5. Template returned for review (if changes made)
6. Metadata, supporting info loaded into Metabase by IM staff
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 for downloading
10. Data catalog preview link sent for final review
11. Data set released in GCE Data Catalog (auto public/private links)
12. Metadata sync’d to LTER, KNB Metacat, DataONE, ...
## File Details

**Title**: GCE data submission form for tabular and non-tabular data files

**Archive**

This Microsoft Excel spreadsheet template is provided for submitting metadata (documentation) for both tabular and non-tabular data to the GCE Information Management Office for archiving in the GCE Data Catalog and LTER Data Portal. More information about GCE data submission is available at [http://oce-ter.marine.uga.edu/public/ndc/data_submission.html](http://oce-ter.marine.uga.edu/public/ndc/data_submission.html).

For tabular data (e.g., spreadsheets, logger files), the data values and column information can be included in the "Tabular Data" worksheet along with the documentation metadata, or provided separately if prior arrangements are made with IM staff for processing data from specialized formats (e.g., raw data logger files, nearshore data telemetry or lab-specific storage formats).

For non-tabular data (e.g., GIS files, raster imagery, genomics data, the data file(s) can be described in the template and then uploaded separately, or download links can be provided in the "Non-Tabular Data" worksheet.

Completed templates and data files (if provided separately) should be uploaded to the GCE IM office using the "Submit Data" form on the private GCE Web site ([http://oce-ter.marine.uga.edu/private/app/upload_data.asp](http://oce-ter.marine.uga.edu/private/app/upload_data.asp)).

Note that a GCE Data Submission Training presentation and Zip archive of sample submissions are also available for downloading ([http://oce-ter.marine.uga.edu/public/ndc/resource_details.asp?id=535](http://oce-ter.marine.uga.edu/public/ndc/resource_details.asp?id=535)).

**Contributor**: Wade Sheldon

**Citation**: Wade Sheldon, 2015. GCE data submission form for tabular and non-tabular data files. Georgia Coastal Ecosystems LTER File Archive, University of Georgia, Athens, Georgia. ([http://oce-ter.marine.uga.edu/public/ndc/resource_details.asp?id=535](http://oce-ter.marine.uga.edu/public/ndc/resource_details.asp?id=535))

**Key Words**: data submission, form, IM, metadata, spreadsheet, template

**File Date**: Jan 06, 2015 (version 24)

**Template - Description**

**GCE-LTER Data Submission Template**

(revised Jan 2013)

**Documentation:**
1. Please fill in all fields as completely as possible.
2. Separate items in lists using commas (e.g., key words).
3. Please do not insert or delete rows from the template unless instructed to do so.
4. Note that extremely long entries may not be completely displayed even when the cell is enlarged, but the full content will be retained and will export properly.
5. Use the 'Notes or Comments' section at the bottom of the form for ancillary information that doesn't seem to fit any metadata category or field, or to list any problems you encountered using the template.

**Data Event or File:** For data sets (e.g., spreadsheets), use the 'Tabular Data' worksheet to provide column descriptions and paste in data values. For non-tabular data (GIS shapefiles, raster imagery, genomics data) fill in file name and format information on the 'Non-Tabular Data' worksheet.

**NOTE:** When submitting updated metadata or re-using templates please highlight fields with modified content in yellow.

<table>
<thead>
<tr>
<th>Data Set Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigator</td>
</tr>
<tr>
<td>Data Set Title</td>
</tr>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Study Type</td>
</tr>
<tr>
<td>Themes</td>
</tr>
<tr>
<td>LTER Core Area</td>
</tr>
<tr>
<td>Keywords</td>
</tr>
<tr>
<td>Accession</td>
</tr>
</tbody>
</table>

**Geographic Information**
Template - Description

Instructions

Field help (comments)

Metadata

Section Tabs

Examples
### Study Design

**Description**
This study represents a regular component of the population monitoring program of the GCE-LTER. It was designed to document size distributions of common infaunal and epifaunal mollusc species and track their changes over space and time at each of the ten intertidal, marsh monitoring sites within the GCE-LTER domain. All molluscs within randomly placed, small quadrats at four plots in the mid and creekbank marshes at each of the permanent monitoring sites were collected by hand. This design fitted each sample, an area in a grassy appearance and near to (within 3 m of) a GCE permanent monitoring plot. The permanent plots were not used so as to avoid damaging the plants within them. Permanent plots 1, 3, 6, and 8 in each zone were chosen to be the anchor plots for selecting actual sample areas. When rock or other disturbances necessitated the use of alternate plots, these are indicated in the data set and noted as anomalies.

**Sampling**
For each sample, an area in a grassy appearance and near to (within 3 m of) a GCE permanent monitoring plot. The permanent plots were not used so as to avoid damaging the plants within them. Permanent plots 1, 3, 6, and 8 in each zone were chosen to be the anchor plots for selecting actual sample areas. When rock or other disturbances necessitated the use of alternate plots, these are indicated in the data set and noted as anomalies.

**Plots**
Samples were taken within 3 meters of the GCE permanent monitoring plots, which were established.

- **Begin Date**: 18-Oct-2010
- **End Date**: 21-Oct-2010

### Instruments

#### Research Methods (note: provide instrumentation details using the “Instrumentation” worksheet)

**Method Steps**
- **Method Name**
- **Method Description**
- **Instrumentation**

#### Geographic Information

- Data Table contains geographic coordinate columns.
- Data Table contains site codes, station codes or location codes.
- Data Table contains permanent plot or groundwater well IDs.
- I will provide a Google Earth file or ESRI shapefile containing research locations.
- Geographic locations are not relevant to this study.

#### Experimental or Sampling Design

**Description**
This study represents a regular component of the population monitoring program of the GCE-LTER. It was designed to document size distributions of common infaunal and epifaunal mollusc species and track their changes over space and time at each of the ten intertidal, marsh monitoring sites within the GCE-LTER domain. All molluscs within randomly placed, small quadrats at four plots in the mid and creekbank marshes at each of the permanent monitoring sites were collected by hand. This design fitted each sample, an area in a grassy appearance and near to (within 3 m of) a GCE permanent monitoring plot. The permanent plots were not used so as to avoid damaging the plants within them. Permanent plots 1, 3, 6, and 8 in each zone were chosen to be the anchor plots for selecting actual sample areas. When rock or other disturbances necessitated the use of alternate plots, these are indicated in the data set and noted as anomalies.

**Sampling**
For each sample, an area in a grassy appearance and near to (within 3 m of) a GCE permanent monitoring plot. The permanent plots were not used so as to avoid damaging the plants within them. Permanent plots 1, 3, 6, and 8 in each zone were chosen to be the anchor plots for selecting actual sample areas. When rock or other disturbances necessitated the use of alternate plots, these are indicated in the data set and noted as anomalies.

**Plots**
Samples were taken within 3 meters of the GCE permanent monitoring plots, which were established.

- **Begin Date**: 18-Oct-2010
- **End Date**: 21-Oct-2010

### Template – Sampling/Methods
Template - Supplemental

**Taxonomy & Permits**

Participants were trained to recognize mollusc species in the field by D. Bishop. Final species identifications were made (under the supervision of D. Bishop) when specimens were sorted and measured in the lab. Most species were common and easily recognizable. Useful references for marsh Collections were made as a sub-permit under the general permit granted to the University of Georgia Marine Institute, Sapelo Island, Georgia.

**Anomalies**

Marsh GCE Site 8 Plot 1 was heavily trampled due to an animal trail through permanent plot.

**Data Set Anomalies**

**Supplementary Information**

- **Data Forms**: Paper log sheets in the field, excel spreadsheets
- **Location of Forms**: GCE LTER Office, UGA Marine Institute, Sapelo Island GA, 31327
- **Validation**: Excel data spreadsheets were printed and cross-checked with bench sheets and field book records by B. B." (example)
- **Materials**: All molluscs were placed in pre-labeled plastic bags and transported back to the lab. The molluscs were stored in a cold compartment of the field car. Leaf samples were dried and stored in airtight plastic bags.
- **Software**: Microsoft Excel 2003 SP3
- **Publication History**: Not published

**Special Instructions for Information Manager (requested statistical reports, file formatting, etc.)**

Instructions
# Template - Personnel

Instructions:
- List the names and roles of all people who participated in the study (including analysis) for example roles: investigator, field technician, lab technician, field assistant, data analyst.
- Also include the institution and email for any participants who are not already listed on the GCE Personnel webpage (http://gce-lter.marisc.uga.edu/lter/asp/ltd/perpersonnel.asp).

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Role</th>
<th>Institution</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Meryl Alber</td>
<td>investigator</td>
<td>University of Georgia</td>
<td><a href="mailto:malber@uga.edu">malber@uga.edu</a></td>
</tr>
<tr>
<td>2</td>
<td>Steven Pennings</td>
<td>co-investigator</td>
<td>University of Houston</td>
<td><a href="mailto:scpennin@central.uh.edu">scpennin@central.uh.edu</a></td>
</tr>
<tr>
<td>3</td>
<td>Kathy Shulack</td>
<td>Field Coordinator</td>
<td>University of Georgia</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Nicholas Scowle</td>
<td>Research Technician</td>
<td>University of Georgia</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Kristen Anstead</td>
<td>Research Technician</td>
<td>University of Georgia</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Hongyu Guo</td>
<td>Graduate Student</td>
<td>University of Houston</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Jennifer Blake-Mahmud</td>
<td>Volunteer</td>
<td>University of Houston</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Instrument Details

### GCE-LTER Data Submission Template

**Instructions:**
- Fill out as much detail about instrumentation used to make readings, analyze or process samples as is available.
- Ignore any fields that you deem inappropriate for the instrument.

<table>
<thead>
<tr>
<th>Method Name*</th>
<th>Owner Name</th>
<th>Instrument Name</th>
<th>Variable Measured**</th>
<th>Make</th>
<th>Model</th>
<th>Serial Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size Determination</td>
<td>Alber</td>
<td>Calipers</td>
<td>Length</td>
<td>Manostat</td>
<td>5921</td>
<td></td>
</tr>
<tr>
<td>Size Determination</td>
<td>Alber</td>
<td>Fiber Optic Light</td>
<td>none</td>
<td>Postec</td>
<td>DDL</td>
<td></td>
</tr>
<tr>
<td>Size Determination</td>
<td>Alber</td>
<td>Microscope Ruler</td>
<td>Length</td>
<td>Bioquip</td>
<td>Metric</td>
<td></td>
</tr>
<tr>
<td>Size Determination</td>
<td>Alber</td>
<td>Stereo-microscope</td>
<td>none</td>
<td>Leica</td>
<td>MZ 7.5</td>
<td></td>
</tr>
</tbody>
</table>

* only required if multiple methods are listed in the primary documentation

** e.g. mass, conductivity, chloride concentration, etc.
Template – Tabular Data

**GCE Data Table Submission Template**

**Instructions:**
- Paste or enter your data values into the 'Values' section (white cells), starting with the indicated cell.
- Fill out the header section for each data column as completely as possible. Fields in light blue are required.
- Descriptions of each header field are listed in the cell comments (hover the mouse pointer over the field name to view).
- Fill in missing values in the table with 'NaN' (not a number), including text fields, and do not skip columns.
- Make sure the column formatting is appropriate for your data - values will be exported as is, so undisplayed digits will be lost.

<table>
<thead>
<tr>
<th>Column Name: Description:</th>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Site</th>
<th>GCE-LETTER sampling site number</th>
<th>Zone</th>
<th>GCE-LETTER permanent plot number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YYYY</td>
<td>M</td>
<td>D</td>
<td>name</td>
<td></td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Units:</td>
<td>YYYY</td>
<td>M</td>
<td>D</td>
<td>name</td>
<td></td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Code values:</td>
<td>YYYY</td>
<td>M</td>
<td>D</td>
<td>name</td>
<td></td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Calculations:</td>
<td>YYYY</td>
<td>M</td>
<td>D</td>
<td>name</td>
<td></td>
<td>none</td>
<td></td>
</tr>
</tbody>
</table>

| QC: Minimum Valid: | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
| QC: Minimum Expected: | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
| QC: Maximum Expected: | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
| QC: Maximum Valid:    | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
| QC: Custom:           | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |

| Values: | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
|         | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
|         | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
|         | 2010 | 10 | 19 | 1 | 1 | 1 | 1 | 1 |
## Template – Tabular Data

<table>
<thead>
<tr>
<th>Name, Description, Units</th>
<th>Precision, Code Defs, Calculations</th>
<th>Q/C Limits</th>
<th>Data Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Column Name:</strong></td>
<td><strong>Description:</strong></td>
<td><strong>QC Minimum Valid:</strong></td>
<td><strong>Values:</strong></td>
</tr>
<tr>
<td></td>
<td>Calendar year of observation</td>
<td>2010</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>Calendar month of observation</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Calendar day of observation</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010</td>
</tr>
</tbody>
</table>
Submit Data to the GCE Catalog

Notes on Data Submission:
- Data should be documented and formatted using the Data Submission Template unless prior arrangements are made with GCE IM staff.
- 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 <=50MB or the upload will be rejected by the server (contact the IT Office for help submitting large files).
- Files with the following extensions are not allowed for security reasons:
  - disallowed extensions: .coffee, .exe, .exe, .dax, .sql, .asp, .pl, .xls, .ser, .java

GCE contact: Sheldon, Wade W.  (required)

File(s) to upload: (<=50MB total)
- Browse: No file selected
- Browse: No file selected
- Browse: No file selected
- Browse: No file selected

Notes or Analysis Request:

(Processing uploads >1 MB will take some time - please do not press 'Submit' again)
GCE Data Toolbox

Flagged Values
GCE Data Catalog

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

GCE-LTER Data Set Summary

Accession: INV-GCEM-1208a2
GCE Research Theme: Aquatic Invertebrate Ecology (Monitoring)

Contributors: Meryl Alber, Steve Pennington, Jacob Shalack, Caroline Reddy, Justin Manley, Hongyu Guo, Kazik Wieski, Aaron Penn

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 mollusks 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 callipers 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 measurable.

Key Words: Aquatic Invertebrate Monitoring, community composition, creek bank, Detritus, epifauna, fall, Geukensia, Hydrobiidae, infauna, intertidal, length, Littoraria, Littorina, marshes, Melampus, mid-marsh, mollusks, monitoring, permanent plots, Polymesoda, size distribution, species diversity, Succineidae

LTR Core Area: Populations

GCE Themes: O2 - Domain Patterns, Aquatic Invertebrate Ecology

Study Period: 26-Oct-2011 to 04-Nov-2011

Study Sites:

GCE1 -- Eulonia, Georgia, USA
GCE2 -- Four Mile Island, Georgia, USA
GCE3 -- North Sapelo, Sapelo Island, Georgia, USA
GCE4 -- Meridan, Georgia, USA
GCE5 -- Folly River, Georgia, USA
GCE6 -- Dean Creek, Sapelo Island, Georgia, USA
GCE7 -- Carrs Island, Georgia, USA
GCE8 -- Alligator Creek, Georgia, USA
GCE9 -- Rockedundy Island, Georgia, USA
GCE10 -- Hunt Camp, Sapelo Island, Georgia, USA
GCE Data Search

GCE-LTER Advanced Data Search

Subject
- Core Area/Theme: <any>
- Subject Search: Any Text contains

Author
- Individual's Last Name: <any>

Temporal Criteria
- Study Period: Start date to End date (YYYY/MM/DD)
- Data Access: <Any Access>

Spatial Criteria
- Geographic Bounds: (decimal degrees)
  - North
  - West Clear East
  - South
- Location: starts with

Taxonomic Criteria
- Taxon: starts with

Display Options
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
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 Metadata 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

- Bad practice:

<table>
<thead>
<tr>
<th>Date</th>
<th>Site</th>
<th>Salinity</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>01/01/2013 00:00</td>
<td>gce6_hydro</td>
<td>30.2</td>
<td>20.5</td>
</tr>
<tr>
<td>01/01/2013 00:30</td>
<td>gce6_hydro</td>
<td>30.3</td>
<td>20.5</td>
</tr>
<tr>
<td>01/01/2013 01:00</td>
<td>gce6_hydro</td>
<td>30.4</td>
<td>20.4</td>
</tr>
<tr>
<td>01/01/2013 01:30</td>
<td>gce6_hydro</td>
<td>30.5</td>
<td>invalid</td>
</tr>
<tr>
<td>01/01/2013 02:00</td>
<td>gce6_hydro</td>
<td>30.6</td>
<td>20.5</td>
</tr>
<tr>
<td>01/01/2013 02:30</td>
<td>gce6_hydro</td>
<td>30.7</td>
<td>20.6</td>
</tr>
</tbody>
</table>

- Good practice:

<table>
<thead>
<tr>
<th>Date</th>
<th>Site</th>
<th>Salinity</th>
<th>Temperature</th>
<th>Flag_Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>01/01/2013 00:00</td>
<td>gce6_hydro</td>
<td>30.2</td>
<td>20.5</td>
<td></td>
</tr>
<tr>
<td>01/01/2013 00:30</td>
<td>gce6_hydro</td>
<td>30.3</td>
<td>20.5</td>
<td></td>
</tr>
<tr>
<td>01/01/2013 01:00</td>
<td>gce6_hydro</td>
<td>30.4</td>
<td>20.4</td>
<td></td>
</tr>
<tr>
<td>01/01/2013 01:30</td>
<td>gce6_hydro</td>
<td>30.5</td>
<td>36.5</td>
<td>invalid</td>
</tr>
<tr>
<td>01/01/2013 02:00</td>
<td>gce6_hydro</td>
<td>30.6</td>
<td>20.5</td>
<td></td>
</tr>
<tr>
<td>01/01/2013 02:30</td>
<td>gce6_hydro</td>
<td>30.7</td>
<td>20.6</td>
<td></td>
</tr>
</tbody>
</table>

- Generally best practice to retain and qualify questionable values, delete and qualify impossible values, document rationale
Data Format (Tabular Data)

- Data should be “pasted” into submission template when practical
- Logger files, specialized lab formats can be submitted for ongoing data collection/monitoring (arrange with IM staff)
- Value formats should be consistent throughout data set
  - Bad practice:

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

- Missing values should be left empty or encoded as NaN (converted to format-specific encoding when posted)
- Value codes must be defined
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)*

<table>
<thead>
<tr>
<th>Date</th>
<th>Site</th>
<th>Spartina_Biomass</th>
<th>Juncus_Biomass</th>
<th>Borrichia_Biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/2013</td>
<td>1</td>
<td>34</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>2</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>3</td>
<td>25</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>4</td>
<td>30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>5</td>
<td>38</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Date</th>
<th>Site</th>
<th>Species</th>
<th>Biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/2013</td>
<td>1</td>
<td>Spartina</td>
<td>34</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>2</td>
<td>Spartina</td>
<td>40</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>3</td>
<td>Spartina</td>
<td>25</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>4</td>
<td>Spartina</td>
<td>30</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>1</td>
<td>Juncus</td>
<td>5</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>2</td>
<td>Juncus</td>
<td>0</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>3</td>
<td>Juncus</td>
<td>0</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>4</td>
<td>Juncus</td>
<td>0</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>1</td>
<td>Borrichia</td>
<td>0</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>2</td>
<td>Borrichia</td>
<td>0</td>
</tr>
</tbody>
</table>
Data Format (Non-tabular)

- Choose formats maximally useful to discipline
  - Vector GIS: ESRI shapefiles, File geodatabases
  - Raster GIS: ESRI Grid, GeoTIFF
  - 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 combined with tabular data, include support files needed for interpretation
Example Data Submissions

(see provided spreadsheets)