*Recovery and Resilience of Oyster Reefs in the Big Bend of Florida - Data Management and Access Plan*

The Lone Cabbage Oyster Reef project will generate new data and information. We have developed the following data management plan to meet basic data sharing and archiving guidelines as part of best scientific practices that requires environmental data and information collected and/or created by this project to be visible, accessible, and independently understandable to general users. The University of Florida has extensive expertise in data management and as an institutional repository (IR@UF) through the Library’s Academic Research Consulting and Services (LARCS, see http://arcs.uflib.ufl.edu/). We are working closely with LARCS in all phases of data management, access, and archiving. Our basic plan is described below and this plan will be revised as needed based on input from LARCS, NFWF, and changing data streams over the life of the project.

*1. Primary products of the research conducted by the Lone Cabbage Oyster Reef project include:*

(1a) Observational data including water quality information recorded by autonomous stations (temperature and conductivity) and biological data from oyster reef sampling (i.e., oyster density and size structure) are all captured from field observations and transferred to electronic repositories following guidelines developed in cooperation with LARCS.

(1b) Results of data/statistical analyses (e.g., parameter estimates and associated uncertainties) are derived from these observations and summarize the results relevant to the proposed research.

(1c) Presentation materials developed as part of this project including web pages, handouts, slides, multimedia presentations and assessment materials associated with the proposed broader impact and education activities.

Preliminary data (that has not been QA/QC checked), drafts of scientific papers, student proposals or theses, plans for future research, peer reviews, communications with colleagues are not included in this data management plan.

*2. Management of specific data products is done as follows:*

(2a) Water quality information: Working closely with LARCS, we have developed a MySQL relational database to support this project (see Figure 1 for physical model of database). The database provides a critical structure for tracking water quality observations within the Lone Cabbage Reef restoration site. Key elements of the database include:

(2a.1) Library for tracking which water quality sensors are deployed at each monitoring station. This is essential because individual water quality sensors at a given location will change over time as sensors are returned to the lab for cleaning, repair, or replacement. Using a library to track which sensor is deployed at a station minimizes risks of incorrectly reference data recorded from a sensor at a different location in space.

(2a.2) Custom Python script for importing raw data files retrieved from the water quality stations. This is an essential element of our database because this Python script is written to read the raw data files that are downloaded from the sensors in the field thus eliminating the need to convert raw data files into a standard structure and merge those files in a continuous record. This is important because at present we are servicing our water quality sensors every 7-10 days to minimize biofouling risks. Each servicing trip generates 9 data files (1 per sensor) and each data file has approximately 1200 observations. The use of the automated script minimizes risks of corrupting the collected data files or somehow making a mistake in organizing and reading data from each field sensor.

(2a.3) Data are automatically checked by our QA/QC protocol as part of the import process. Data that do not pass QA/QC standards are routed to an error file for further evaluation by our Research Coordinator while data that do pass are imported into the database for further examination.

(2a.4) Observational records stored within the database include date, time, temperature (Celsius), conductivity (us/S). Using standard oceanographic conversion equations (UNESCO 1983), salinity (ppt) is derived within the database and stored as a derived value.

(2b) Oyster field observations: Working with LARCS and other members of the UF data science community we have developed protocols for data entry and management of observational data on oyster populations from field sampling.

(2b.1) Field collections are spatially referenced using recreational grade GPS recording in decimal degrees. Written protocols define the gear and methods for collecting oyster observational data line transect and quadrat sampling. Data recorded on standardized datasheets using pencil and water-resistant paper.

(2b.2) A senior staff member (Research Coordinator) maintains data integrity through supervision of field activities and data capture efforts.

(2b.3) In the lab, data from the field are entered into electronic records using standard spreadsheets defined to mimic field data sheets. Data are independently entered twice to create independent data entry files. These two data entry files (known as a double-data entry system) are compared using automated scripts which also provide first level QA/QC. Discrepancies between the files or anomalous values are flagged. These flags are reviewed by the Research Coordinator to resolve errors. All paper field sheets are scanned and stored electronically and paper copies filed in the lab.

(2b.4) Once data from each field trip are validated, the master data file for oyster field assessments is amended to include these results. All records are stored in plain ASCII (e.g., .txt or .csv files). All changes to master data file are tracked using version control via Git hub.

(2a.5) Standard data analyses as graphs, tables, or statistical models are conducted using standard open source statistical analyses software (Program R) and version control of our standard data summaries for reporting purposes is maintained within our Git repository.

(2a.6) Collected and contributed spatial data will be compared to professionally surveyed field markers for QA/QC of a data sets spatial accuracy.  This is both a data management and a QA/QC product as through the comparison process we are able to coordinate with survey companies who collect data to clarify any uncertainties in survey attributes following data collection.

(2.a.7) Spatial attribute data will also be checked for completeness and accuracy through visual inspection on a quarterly basis.

*3. Quality Assurance/Quality Control:*

(3a) Quality Assurance/Quality Control (QA/QC) will generally follow guidelines outlined by the USGS Fundamental Science Practices guidelines (available: https://www2.usgs.gov/usgs-manual/500/502-2.html).

(3b) QA Before Data Collection: We have defined data standards prior to collection of the data including field methods and data recording procedures based on prior oyster reef restoration efforts. We will follow these general best practices as defined by USGS where Format - Decide the format of how the data will be collected prior to collection. Codes – We will define what each code word means on all datasheets including specifying units of measurement. Metadata - We will be sure to create metadata in unison with the data to be collected. Supervision - We will assign responsibility to a senior staff person over quality assurance.

(3c) QA/QC During Data Entry: We will use a double entry structure where we will have two people independently enter the data. We will use automated computer scripts to check for agreement between entry systems.

(3c.1) We will minimize the number of times the data need to be entered through the use of a relational database.

(3c.2) We will always document any modifications to the dataset.

(3d) QC After Data Entry: We use a variety of statistical and graphical approaches to check for outliers.

*4. Access to Data and Data Sharing Practices and Policies:*

(4a) The primary results of the proposed research, including the results of associated statistical analyses, will be disseminated primarily through publication in journals (including online-only supplements for extended tables, animations, etc.), conference presentations and student theses (long-term open-access via IR@UF).

(4b) Water quality data obtained from data sensors will be posted to the University of Florida Shellfish Aquaculture Extension Program website (<http://shellfish.ifas.ufl.edu/water-quality-monitoring/>) or similar once a standard graphical presentation platform is developed and QA/QC procedures standardized.

(4c) Final spatial data sets created specifically for the project and used in analysis will be accompanied by abbreviated Federal Geographic Data Committee (FGDC) compliant metadata.  Spatial data will be contributed to an appropriate National Spatial Data Infrastructure (NSDI) node for public access.

(4d) Final copies of all data will be made available via public repositories based on input from IR@UF.

*5. Storage and Archiving of Data Electronic*:

(5a) All data will be preserved using multiple on-site copies, with all servers using RAID hard drive arrays and offsite backups will be performed weekly using resources provided by UF Institutional Computing (<https://www.rc.ufl.edu/>).

(5b) Final versions of software source code and curriculum materials will be stored in home directories based on input from IR@UF.

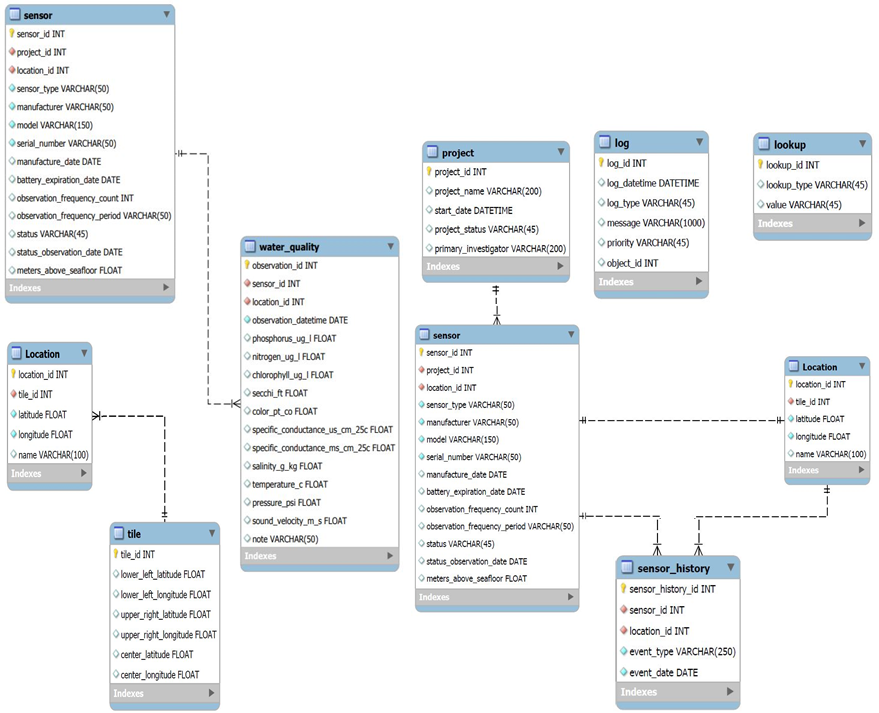


Figure 1. Physical data model developed for our MySQL relational database used to manage water quality sensor data.