**Graduate Proposal**

**Mel Moreno**

**July 18, 2019**

**Thesis overview**

This thesis uses ongoing research and restoration efforts in the Big Bend region of Florida as a case history to develop data products that follow best practices of being open and reproducible. This is done in two chapters. Chapter 1 describes the development of a data management plan for the restoration of Lone Cabbage oyster reef in Suwannee Sound. This chapter includes a description of data workflow products related to counts of oysters and water quality measurements from field collection, data analyses, and long-term storage. Chapter 2 uses publicly available spatial data to assess changes in geographic features in the Suwannee Sound region. This chapter catalogs available imagery, and image characteristics, and then develops a full reproducible approach that provides a template for selecting geographic features of interest, identifying image characteristics needed to evaluate this geographic feature over time, and then identify the available images. Combined these chapters provide insight to resource managers related to managing living data from ongoing field studies and public image repositories.

**Chapter 1**

**Data Management Workflow**

#### 

#### Background and Context

Traditional field biology programs, many of which are designed to monitor animal populations and their environments, have experienced substantial changes in data collection, management, and storage technology in recent years. Changes include new sensor technology, data collection methods, and data observing platforms that are being used in large-scale monitoring programs including SECOORA (Southeast Coastal Ocean Observing Regional Association) and NEON (National Ecological Observing Network). As an example, advancements in sensor technology have allowed for significant changes in water quality monitoring such as transitioning from discrete single location and single point in time sample collections to real-time continuous observations at multiple locations. While the scale and technological capacity of many monitoring programs has increased these monitoring programs are still most often conceived, planned, and used by personnel trained as biologists and not data scientists. The lack of training in basic data management, curation, and workflow of data generated from these types of data collection platforms was highlighted in a recent NSF (National Science Foundation) survey (Lowndes et al. 2017) which highlighted that of the 704 scientists who participated in the survey, “data skills” was identified as the largest unmet need (Barone et al., 2017; Lowndes et al., 2017).

**Continuous Data, Management and Analyses**

The US Gulf of Mexico region is undergoing a large restoration effort to reverse observed declines in key ecosystem components including seagrass, fish communities, and oyster reefs using funding from the consolidated Deepwater Horizon settlements (see <https://www.nfwf.org/gulf/Pages/home.aspx> as an example). These restoration projects vary in spatial scale and funding, but, like other restoration efforts, these projects will have data collection and evaluation efforts that occur frequently throughout the project.

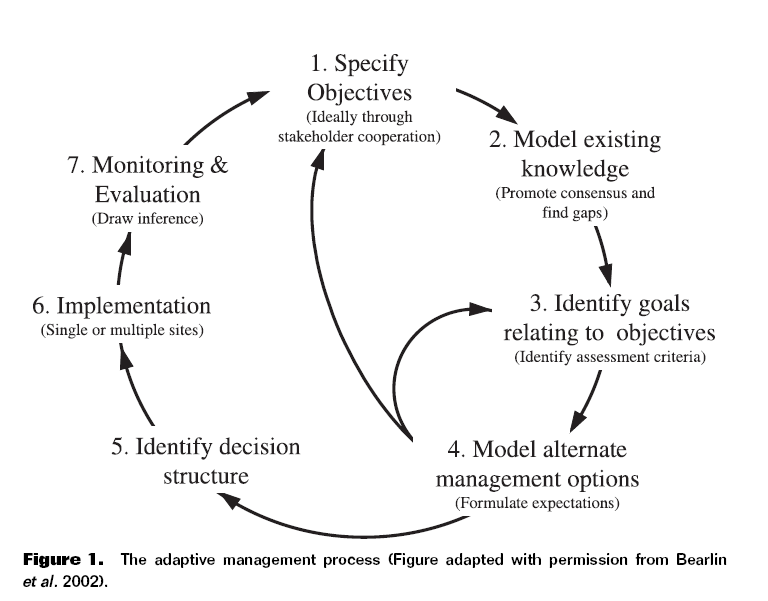
Several of the ongoing restoration programs in the Gulf of Mexico have similar objectives such as restoring oyster or seagrass habitats. For programs with similar objectives, restoration actions may be improved by adaptively informing ongoing restoration while the restoration is taking place, or informing restoration based on results from a similar study elsewhere, to help improve restoration outcomes. For this type of system to work efficiently, a workflow must be developed that captures data as it is collected, and guides this information from collection, to analyses, and data storage. Developing this type of data system is essential to improve data quality by reducing the likelihood of data collection errors or reducing the availability of data for use by other researchers conducting similar restoration efforts.

**Adaptive Management**

“Living data” are defined as data which are continuously collected and updated. These types of data are critical to this type of adaptive learning to inform restoration and management actions (Yenni et al., 2018). These informed adaptations during a restoration project can be small such as shifting the location of an autonomous sensor, to larger changes including restoration practices or revamping of sampling programs because of low statistical power. Living data are challenging to work with from a data management perspective because the data (by design) change as new data are collected. In a restoration or management context as these data are being collected, they must be processed, and analyses of these data completed to help draw inferences on how the system of interest is responding to the restoration action. This idea of iteratively integrating new data, analyses, and comparing these outcomes with previously stated objectives is not new and is a central aspect of the “adaptive management” process for natural resources first described in the 1970’s (Holling 1978; Walters 1986).

By design, an adaptive management program requires rapid feedback between data collection, analyses, and interpretation to drive the process of updating knowledge, examining management and restoration options, making decisions and implementing actions. This process is repeated (Figure 1) to improve management actions such as identifying the best restoration approach. To carry out a restoration project adaptively, data used in these continuous efforts meet quality assurance/quality control (QA/QC) protocols to identify and correct inconsistencies and errors in field or sensor observations before these data are used in an analysis. Errors in these data, or delays in producing the data in a usable framework, can quickly lead to a breakdown in the adaptive learning process either in terms of slowing the analyses limiting their utility for timely decision making, or worse, erroneously informing the decision-making process because of errors in data management or analyses.

An example of programs that require an adaptive management approach to restoration are projects funded by the National Fish and Wildlife Federation as part of the Gulf Environmental Benefit Fund (NFWF-GEBF). These projects explicitly require an adaptive management plan to guide the restoration process. Extensive data management plans are also mandated to capture data collected and analyzed with the overall purpose of creating opportunities to improve future restoration actions by maximizing learning from previous and ongoing restoration efforts. The Lone Cabbage Reef (LCR) restoration project is a large restoration effort in the eastern Gulf of Mexico funded by NFWF-GEBF. The project’s primary goal is to restore specific historical oyster reefs so that they may be resilient to changing sea level and river discharge. This project generates data from multiple sources including continuous autonomous water quality data from sensors and observations of oyster populations from field biologists. These data are generated at different time frequencies with sensor data obtained at hourly time intervals from multiple spatial locations and biological data collected at discrete time intervals from multiple spatial locations. For both cases, there is a need to prepare data, meet data quality standards, and complete routine analyses of data to ensure data collected are useful for project objectives. Because this is a long-term restoration project with numerous uncertainties in how the ecosystem will respond to restoration actions, developing a data management and workflow system that automates as many aspects of the workflow including QA/QC, measurement errors, and inconsistency in naming conventions is essential. Creating this workflow will allow for rapid analyses of data to inform decision making related to sensor deployment or modifying the reef restoration process through additional construction efforts.



### Figure 1- (Schreiber et al., 2004) The adaptive management process (Figure adapted with permission from Bearlin et al., 2002).

### Objective

In this chapter, I will develop the basic elements of the LCR restoration project data management plan for water quality and biological data. The objective is to develop and implement a data management workflow, which starts at the data collection point (i.e., field collection) and ends at the visualization/ interpretation of collected data from different data streams. I will document how these data are recorded, data QA/QC procedures, data checking (anomalous values), data visualization, and data releases for analyses using multiple software tools. This chapter provides an example of how a living data project can function to inform an ongoing, long-term restoration project and serve as an example for other projects with data collection efforts.

### Implementing a modern data workflow

Creating and automating a data management workflow for living data is an emerging skill for natural resource professionals. More than ever, data management is recognized as a core skill for biologists and ecologists (Hampton et al., 2017). Even though the design of workflow will be specific to the LCR restoration project, the steps outlined can be broadly used for many conservation efforts. The tools used to implement the data management workflow, are also readily available online and most tools offer tutorials and workshops for more in-depth training. The approach for this workflow requires basic knowledge of computer coding and version control structure (to track changes in data and computer code). I will use freely available open source tools including program R (<https://www.rstudio.com/>) for data QA/QC, analysis, and visualizations, and GitHub (<https://github.com/>) for version control. Steps 1, 3, 4, and 7, in Figure 1, are directly used in my proposed workflow, with applicable modifications.

## 

## Methods

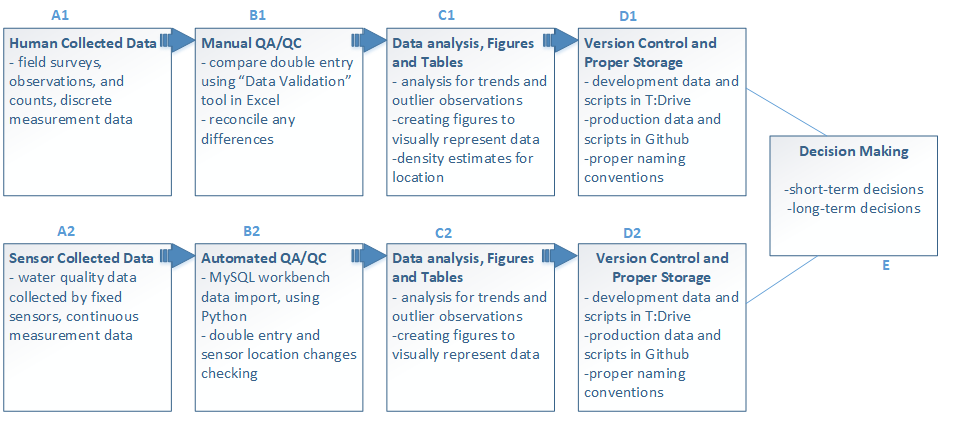


Figure 2- Data management workflow designed for the LCR project.

### Field Collections

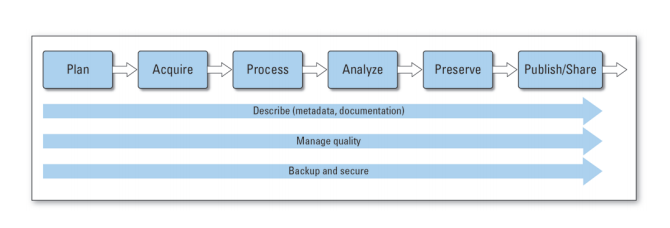
One of the goals of a successful data management plan is to minimize errors in collected data. Often, the first step in the data collection process is transcribing an observation in the field to paper data sheets for analyses back in the lab. This simple effort of recording the data in the field is the first opportunity to introduce errors in the data collection process. These errors can come from a variety of sources such as the wrong date or site name on a sheet or the person recording the data may be unfamiliar with terminology or protocols. To minimize these types of mistakes it is best to follow proven practices for data management such as those from the United States Geologic Survey (USGS) who recommend development of a standard set of data guidelines before field collections begin (Figure 3).

Figure 3- USGS Science Data Lifecycle Model (<https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf>).

These types of data guidelines define the basic types of data that are to be collected including date and time, site naming conventions, and units of measurements for specific observations. This type of predetermined information is a key first step in reducing the risk of this type of data error in the field. As an example, simple differences in how dates are recorded by different people such as YYMMDD or MMDDYY formats can create confusion as to when a sample was physically collected. Errors in site names can place the data observations in the wrong location spatially. To minimize this risk, whenever possible data sheets can be pre-populated with as much information as possible before going into the field.

#### Human Collected Data

For the LCR restoration project, observational data collected in the field primarily consists of oyster counts and size measurements from line transects among randomly selected oyster reefs delineated into strata based on specific research questions, which are then recorded on waterproof paper data sheets (Figure 2, Box A1). To reduce chance of field errors and save time while in the field, I will work to develop and improve data workflow by providing guidance on pre-populating data sheets when possible with basic information including, date and location following data naming standards and field protocols (Figure 4).

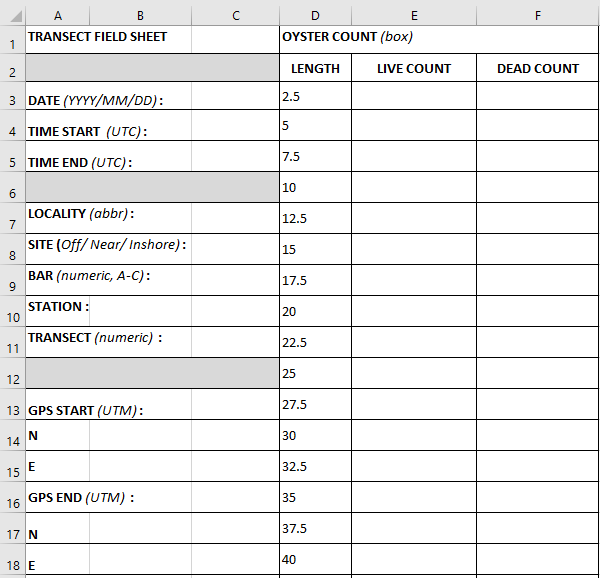


Figure 4- Physical data sheet template created and managed in Excel with labels and units for each applicable field.

#### 

#### Sensor collected data

Sensor collected data differs from human collected data, in that sensor data are measurements recorded by an instrument automatically. These types of data are a common component of many large-scale observational platforms that may record environmental or biological data continuously, and then make these observations available for use at set time intervals or through “live” feeds. Examples of these types of data include river discharge information provided by USGS or wind observations from a NOAA (National Oceanic and Atmospheric Administration) weather buoy, which can be accessed online by APIs or software such as R.

The LCR restoration project has a small array of sensors (N=9) that track temperature (°C) and conductivity (μS/m) of water near the oyster reef restoration site. To retrieve the data from these sensors, the sensor must be physically removed from the water and the associated data files are downloaded from the receiver (Figure 2, Box A2). Sensors are serviced bi-weekly to ensure functionality.

An individual sensor data file with 14-days worth of observations contains about 900 lines of data and a total of about 450 observations. While the observations are collected automatically, there are still opportunities to introduce errors when these data are collected. This can include incorrect naming of files once downloaded to a laptop in the field, copying over files on the laptop erroneously, or failing to “start” the sensor once redeployed. Reducing these error opportunities will ensure a continuous sensor stream of interrupted measurements. I will develop existing protocols to minimize these procedural errors related to data retrival. I will review these protocols and revise as necessary as part of my data workflow development.

### 

### QA/QC during data entry

#### Paper data sheets to electronic records

The process of transferring data from paper data sheets to electronic form that will make it compatible to a computer for data analyses, is the most common source of potential errors. I will work to minimize this risk of errors as part of my workflow design. For data entered by hand, I will first use a data entry system that reduces the likelihood of introducing errors via data entry. This will be done using a standardized template so data sheets and digital spreadsheets are input similar ways. This follows USGS Data Management guidelines which suggests that the most effective way to ensure data quality, is to prevent the creation of defective data. I will use a Data template structure based on USGS Data Management Standards ([https://www.usgs.gov/products/data-and-tools/data-management/quality-design-recommended-practices?qt-science\_support\_page\_related\_con=0#qt-science\_support\_page\_related\_con](https://www.usgs.gov/products/data-and-tools/data-management/quality-design-recommended-practices?qt-science_support_page_related_con=0" \l "qt-science_support_page_related_con)).

For the LCR restoration project, I will design an Excel workbook intended as a Data Template for easy and efficient data entry (Figure 5). This workbook will be modified for data entry using “Data Validation” features in Excel that restrict the types of data that can be typed into each predefined column (Figure 2, Box B1). These restrictions include the use of “drop down” style menus that require the user entering data to choose a value for entry based on a pre-populated list of values. These pre-populated lists of values, such as site name abbreviations, are based on the terms defined by the data abbreviations guide for the project. Other types of restrictions include specific formatting for date or time values, as well as “limits” on observational data entered in each cell. By restricting the choice of the user when selecting locations, dates, units, and measurement ranges this limits the potential for data entry errors such as capitalization or use of zeros instead of the letter “O”. To simplify entry, each data column matches an entry on the physical data sheet used in the field.

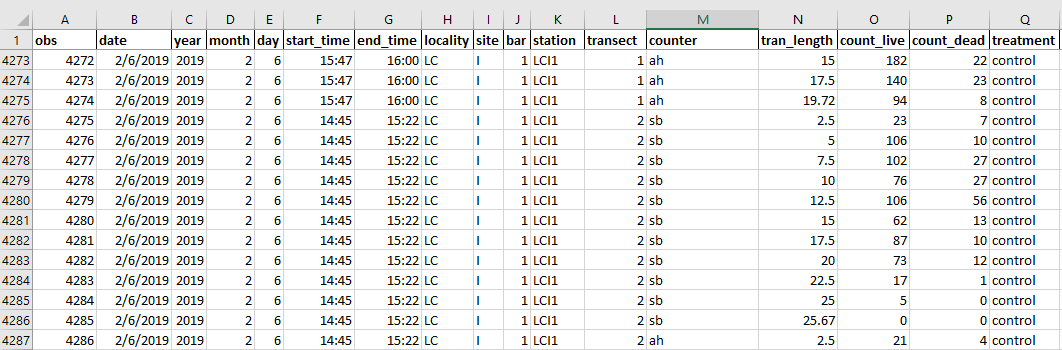
As an example of the capability of “Data Validation” features, oyster length measurements will be restricted from being entered at a size greater than 125-mm. While oysters greater than this size are observable, to enter a value above this level requires manual override from someone with supervisory control. This data entry system will also require a “double entry” system where each line of data will be entered into the workbook twice, typically by separate users, and then these data will be compared electronically. If the entered data do not match exactly, the original data sheets will be examined to determine why discrepancies exist. Using different people for each round of data entry is preferred because different people may interpret the handwriting on the field data sheets differently. A third tab will then be used in Excel to compare the two user data entry tabs for discrepancies. Any identified errors are then reconciled against the field data sheets and by a project supervisor.

Figure 5- Data entry workbook in Excel to enter in data collected from the field. Each column is restricted on what information can be entered into it.

#### Transfer electronic records from sensor to database

When individually collected sensor data files are transported back to the lab, these files must be checked for errors and the data amended to an existing database to provide a continuous record of the water quality observations of interest (Figure 2, Box B2). I will develop a three-step process where:

Step 1. Working with UF Library team, I will help to develop Python code that will distinguish files from each of the two types of sensors that make up the water quality sensor array (Star-Oddi or Diver), based on proper file naming convention.

Step 2. Python code will then check data for errors including duplicate observations or to check the sensor is identified properly in our database. This checking will be done by comparing observations downloaded from sensors to pre-defined minimum and maximum values determined for each sensor. An additional example is that all active and functioning sensors which are deployed in the field are stored in a data table in our MySQL database, where the start day, time, and location are recorded. If the data file list of sensors does not match the list of active sensors known in the database, then an error message will be reported.

Step 3. MySQL will import all checked and correct observations in their appropriate tables.

Once imported, a second set of QA/QC protocols will be performed within the MySQL database, which will be examining observations for non-sense values based on expected temperature and conductivity values for the array location. While I will not develop the MySQL database as part of my thesis, I will work closely with University of Florida Library staff to define database relationships, error checking routines, and workflow within the MySQL database. I will also develop basic Python skills to allow me to conduct routine maintenance on the database such as error checking and adding additional water quality stations as needed.

### Data Analysis, Figures and Tables

Once data are standardized and available for use in the computer, basic visualization of the data via graphs and figures is a key next step for data checking and the beginning of the analyses (Figure 2, Boxes C1 and C2). I will develop a group of data visualization products to be used both to check data from field collections and water quality sensors. These figures will be integrated with the living data such that as data are entered into the database, and after they pass initial QA/QC, the figures will be automatically updated to allow visual assessments of the recorded data. I will focus my efforts on creating these visualization products for the water quality data collected by the LCR restoration project.

I will also develop a set of summary tables as part of the data workflow to provide basic information on water quality variables at different time intervals. These summary tables and figures will follow data reproducibility guidelines from USGS where the tables will be created from the living data using standard code that reproduces the same table and adding newly updated data when needed. By developing code for tables, figures, and any other reproducible analyses as the data are updated, it will reduce the total time for data feedback loop.

### Version Control

Version control is defined as a software that allows for the saving and management of changes in content, documents, and other developmental information. The key purpose of using version control software is to document and confirm that changes in content are intended and planned . Version control can be incorporated into a data workflow using software such as Github, (Figure 2, Boxes D1 and D2). The USGS Data Management Guidelines encourage the use of version control software and repositories for data and code used for projects, which will allow the project data analysis to be accessible and reproducible (<https://www.usgs.gov/products/data-and-tools/data-management/repositories>).

Version control can be critical to ensuring that data are not duplicated, lost, or time is not wasted by not working with the proper files. The Data Carpentries provide detailed reasons for using version control (<http://swcarpentry.github.io/git-novice/>) that can be generalized as (1) a version control system saves all versions of a file, (2) version control records who made what changes to specific files and (3) allows these changes to be undone if needed, (4) version control software notifies each user when there is a conflict between different people’s work such as code.

The LCR restoration project will use a GitHub structure for version control. I will manage the Excel workbooks used for data entry and initial QA/QC in Git to allow each user to see when new data are available. I will also use GitHub to track changes in routine R files used for data summaries that are pushed to the web and included in standard reports to funding agencies.

#### 

#### Proper Storage

I propose that the data workflow for both data and code scripts be separated into two modes. The first mode is “development” mode, meaning that data that are currently undergoing a QA/QC process. The second mode is “production” mode, where the processed data are ready to be analyzed. Github repositories will only have publicly available production data and scripts (Figure 2, Boxes D1 and D2). Raw sensor data files will not be found in these repositories.

For the data and scripts that are in development mode, the proper storage for these documents will be in our projects internal server, commonly referred to as the T:Drive (Figure 2, Box D2). This server is only available to members of the LCR restoration project is not publicly available. Raw sensor data files would be stored and archived in this server. However, for other projects it would be advised to look into a protected and secure server to store raw data files.

##### Naming conventions for files

I will develop a naming structure which will require files start with the date of creation, in the format YYYYMMDD. Each file will have additional information, that will usually have a prior set of approved abbreviations, after the date. One advantage of this naming structure is that all files will be ordered chronologically when sorted by name, so there is very little confusion on when the files were created (Table 1). Following guidelines from USGS Data Standards suggest that file names should be in all uppercase or all lowercase letters, instead of a combination of both. For the LCR project, I will propose as part of the naming convention standards that all files are lowercase, and the context of the file names are separated with an underscore. If files are not named correctly, they will be renamed to follow our guidelines. Files that are not named correctly, also have the risk of being overlooked, or re-organized in an incorrect folder. Correct naming conventions are critical to create the correct interface between the field collected water quality sensor data and the Python code that reads and stores these data.



Table 1- Example of file naming structure

As per USGS Data Standards, naming conventions are necessary to make data easier to use, to integrate and to share. This is especially true because data that are represented will be in a format that has already been established and planned ([https://www.usgs.gov/products/data-and-tools/data-management/data-standards#examples](https://www.usgs.gov/products/data-and-tools/data-management/data-standards" \l "examples)). Creating a table beforehand, on how each data type will be named, formatted, and defined will provide data integrity and accuracy (Table 2).

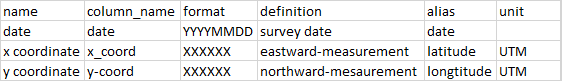


Table 2- Data Standard examples of creating a definition list for naming conventions.

### 

### Discussion

Using the Lone Cabbage Reef restoration project as a case history, I will develop a data management workflow that is adaptable for multiple types of data and meets best practices of data validation and reproducibility (Table 3). This data workflow will integrate living data from observations recorded from paper data sheets and autonomous sensors that monitor water quality.

|  |  |
| --- | --- |
| Organization/ Person | Deliverables |
| ARCS (Academic Research Consulting and Services) | 1. Data management to UF faculty through the lifecycle of their data 2. Providing researchers, a competitive edge to secure funding 3. Creating plans that follow funding agencies compliance |
| LCR Project | 1. Executing Data Management and Access Plan:  * Observational data including autonomous stations and oyster reef sampling * Management of specific data products * Quality Assurance/Quality Control * Access to Data and Data Sharing Practices and Policies |
| Proposed research | 1. Investigate data management needs for LCR project 2. Design and create data management workflows as it pertains to the LCR project, for each data sampling type 3. Continually updating workflow for maximum efficiency 4. Ensure data management plan is compliant with USGS Data Standards |

Table 3- Deliverables comparison of ARCS, LCR Project and my proposed research.

The main goal of my proposed workflow is to make data available for rapid analyses to adaptively assess the LCR restoration project and ongoing water quality and oyster monitoring efforts. This will help to meet the adaptive management requirements for this project by providing the data in a structure that allows rapid assessment and evaluation to inform decision making related to the ongoing monitoring efforts (Figure 2, Box E). To do this, these data must be properly processed and managed to support reproducible analyses. My project will ensure that best practices are established and followed for data input, management, and basic summaries and visualization. This information will be useful for (1) increasing efficiency in the LCR restoration project. The LCR restoration project involves a large restoration project as well as integration of historical data from two other sampling epochs. Because a single data management workflow was not used across these epochs, significant effort has been required to standardize existing data. By establishing a data workflow at the beginning of the LCR restoration epoch, the data will be managed in a common structure over the life of the project. These data will then be available and used to make decisions related to future conservation and restoration efforts like the LCR restoration project. Having precise knowledge of biological data interpretations, will ensure both time and money are used efficiently. (2) This data workflow will inform a variety of short-term decisions that must be made to adaptively improve the ongoing LCR monitoring efforts. As an example, sampling frequency, sampling locations, and sampling times of both the oyster populations and water quality can be informed by rapidly processing existing data. This can prevent data gaps from occurring from events such as biofouling of water quality sensors. (3) Long-term decisions as part of the adaptive management process of this project can also be informed by this data workflow. Overall well-designed data workflow programs are critical to meeting basic requirements of an adaptive management plan. When combined this approach can be highly effective in maximizing the effectiveness of conservation actions such as the LCR restoration in a cost-effective manner.

**Chapter 2**

**Case Study - Spatial Analysis of Deer Island**

**Introduction and Background**

Sea level rise represents one of the most significant management challenges to resource managers, residents, and urban planners in coastal regions globally. As a coastal state with low relief geomorphology, Florida will experience significant coastal change due to rising sea levels in both the built (ref) and natural environments. ~~Restoration efforts in coastal zones may be strongly influenced by landscape level processes. The Gulf of Mexico coastline, with its low relief geomorphology, especially in Florida, is also vulnerable to coastal erosion (Geselbracht et al., 2011).~~ Much of the Florida coastline consists of a 1-meter elevation contour that extends inward anywhere from 3 to 10 kilometers. This low elevation leaves the Florida coastline susceptible to frequent coastal changes. Other types of landscape changes occur at different time scales and may have different (and unknown) effects including conversion from wetlands to shallow shores.

The Big Bend region of Florida, located approximately ~~60 miles~~ 100-km west of Gainesville, Florida along the Gulf of Mexico is generally defined from Apalachee Bay to Anclote Key . The Big Bend is largely undeveloped and about 30% of the Big Bend land area and over 60 miles of coastline are under conservation protection (Main and Allen 2007). Human population density around the Big Bend is the lowest of any other coastal Florida city and the percentage of intact natural habitat is high (Geselbracht 2007). Due to, in part, low human densities, coastal areas have not been heavily impacted by boat traffic, dredging, heavy industrial pollution, eutrophication, or other anthropogenic impacts (Seavey et al., 2011).

Despite the lack of human influence, many observable declines in ecosystem and habitats have been documented (Seavey et al., 2011). Now you can describe some of the changes that have been documented. You can say things like Seavey et al. (2011) used aerial imagery and field-based surveys to document X% change in Y between A and B in XYZ. Rabbe and Stumpf (ABCD) used Landsat imagery from 1996-1995 to assess whether trends in coastal marsh in the Big Bend were evident and found no measurable change in A and B. Rabe and Stumpf (2016) compared 19th century surveys with satellite imagery collected

**Reason for Research**

Because of how little the Big Bend coastline has been influenced by outside forces, there is a high interest to protect the coastal areas that have not been colonized. There are several restorative and conservation projects in the Big Bend, which are funded through National Fish and Wildlife Foundation (NFWF), who have been allocating money from the 2010 Deepwater Horizon oil spill, since 2013 (https://www.nfwf.org/gulf). The agreement of the settlement is directed to fund projects that will benefit areas of natural resources in the Gulf Coast that may have been impacted by the spill. The awards are invested into projects to conserve and enhance coastal habitats. The LCR restoration project is a program funded through NWFW to restore and monitor oyster populations. The LCR restoration project has been working with other agencies such as Florida Fish and Wildlife Conservation Commission (FWC) and Nature Coast Biological Station (NCBS) to unify available biological data including water quality and species density monitoring.

These biological data are important to illustrate a larger picture of the natural impacts that have occurred in the Big Bend. Spatial analysis in the Big Bend is a monitoring evaluation that has not been fully explored, despite large conservation interest in the area. An example of spatial analysis efforts can been seen in Raabe (2004), who digitized mapping imagery from surveys of the coastline, from approximately the Suwannee River mouth to Tampa Bay, and collected topographic sheets from the 1800’s and compared these surveys to available satellite imagery from 1995 to characterize changes in coastal habitats between these two time periods. Research such as Raabe (2004) are useful because they provide resource managers with long-term perspective on how resources are or are not changing. Examining these trends will provide necessary information to the efforts in the area to conservation efforts, including the LCR restoration project.

**Objectives**

In this chapter I will develop a data management plan, catalog of images and associated metadata for the Suwannee Sound region of the Big Bend (area of key management interests) and a highly documented, reproducible analyses to assess trends in a geographic feature in the Big Bend region (Deer Island) that may have changed in recent decades due to sea level rise, coastal erosion, subsidence, or other factors. This information is of interest because the majority of coastline in Suwannee Sound is public land managed as part of the Lower Suwannee National Wildlife Refuge. Identifying a set of images, and a reproducible approach to assessing change in these images of this region of management interest, can provide insight into how this region may be changing. This information can then be used by resource managers to inform management actions that are proposed such as hydrologic restoration with the LSNWR or prioritization of land areas for purchase to allow for land transition and habitat migration as sea-level rises.

I will review all available mapping imagery and materials of the Big Bend coast and Suwannee Sound to A) organize and store mapping materials for future conservation projects as per USGS data management standards, B) conduct a geospatial analysis on coastal changes, gained and/or lost, from the earliest appropriate mapping data of Deer Island C) outline methods of geospatial analysis for future use and analyses of the LCR restoration project for maximum reproducibility.

**Study Area – A Case History with Deer Island**

The area of study that will be analyzed is the coastline of Deer Island, which is located in Suwannee Sound south of the east pass of the Suwannee River and north of the town of Cedar Key, Florida (coordinates). Deer Island is a barrier island consisting of 90 acres in total area, which comprises of 25 acres of upland habitat and 20 acres of wetland habitat (https://www.privateislandsonline.com/united-states/florida/deer-island). The island coastline features a sandy beach facing the open Gulf of Mexico. Deer island is not inhabited, but there are some man-made structures still remaining from the late 1800s. Deer Island is located 8 miles north of Cedar Key, Florida. Deer Island is an important feature in this region because it is one of the few barrier islands along this coastline which may offer some protection from erosion and storm surge to adjacent upland habitats which are primarily part of the LSNWR. Deer Island has a long cultural history in the region as a recreation area, and long-time local residents have described large changes in the areal extent, shape, and vegetation coverage of the island. This follows observed losses of other coastal islands including Derrick Key, approximately 5-km south of Deer Island.



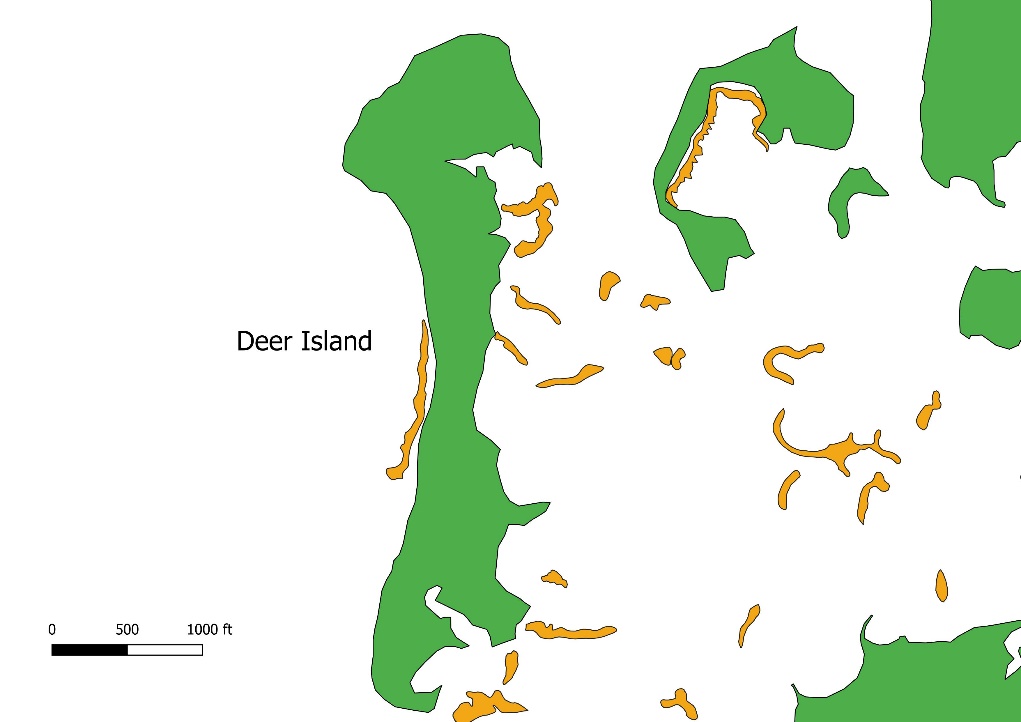
Figure 1- Zoomed out view of study area, Deer Island, for spatial context in relation to Lone Cabbage Reef. Land mass is colored in green, and oyster clusters are colored in orange, for reference.

Figure 2- Zoomed in view of study area, Deer Island. Land mass is colored in green, and oyster clusters are colored in orange, for reference.

**Methods**

Defined methods for this case study are not set. Many software programs and packages are available through the University of Florida and open source resources. There are general methods and techniques that will be mentioned in this section.

*Cataloging available imagery*

*Data Organization and Storage*

I will follow USGS Data Management standards (ref) for processing spatial data by using open file formats such as geoTIFF and use open-source solutions whenever possible. I will also follow USGS metadata guidelines to document the who, what, where, when, why, and how is important so that data can be understood, re-used, and integrated with other datasets (ref). Specifically USGS Survey manual section SM 502.7 states:

“Metadata must accompany all USGS scientific data and other information products. Metadata records are to be developed in a standardized way that enables users to understand the context and to evaluate the usefulness of the data or information product. Metadata records for scientific data must comply with standards such as the FGDC Content Standard for Digital Geospatial Metadata, the International Organization for Standardization suite of standards, or other USGS endorsed FCDC standards. A minimum of one metadata review by a qualified reviewer is required for all USGS scientific data and other information products approved for release.”

Some metadata software recommended by USGS are USGS Metadata Wizard (<https://www.sciencebase.gov/catalog/item/50ed7aa4e4b0438b00db080a>) and USDA Metavist (<https://www.nrs.fs.fed.us/pubs/2737>). These software allow for the user to create FDGC (Federal Geographic Data Committee) Metadata for geospatial datasets. The USGS suggests that these metadata to be incorporated in published geospatial datasets to standardize ways groups are storing and recording their geospatial data sets.

As far as my graduate research analysis, I will be using the T:Drive storage of the LCR restoration project to store my datasets. I will use best practices to store and backup my geospatial datasets as per USGS Data Management standards with input from UF Academic Research Computing team. These details will be finalized during imagery processing and analysis. Completed analysis and geospatial datasets will be located in GitHub, a version control online software, for easability to download and reproduce.

After gathering, storing and documenting all available mapping imagery of the case study area, a few images will be selected for analysis. These mapping images will have associated metadata to ensure that the map analysis will be compatible between all maps. It is unsure how many maps at this time will be used to produce the final analysis, but I propose that each decade, ranging from two to four decades, will have at least one selected map, to show the greatest date range of coastline change of Deer Island.

Skills and techniques are still being learned and solidified through recommended GIS courses. Geospatial analysis will be accomplished by using ESRI software products such as ArcMap. ArcMap tools such as “create features”, will allow the user to construct points, polylines, and polygons at on the spatial imagery, allowing for digitization of any area of the map. These features can then be compared and analyzed to observe short-term or long-term trends between spatial imagery of Deer Island. I propose to evaluate if this method can be easily reproduced for future available mapping needs for this restoration effort.

*Documenting Workflow*

As in Chapter 1, it will be important to document the workflow of a complete and accurate product. Having a completed analysis of Deer Island will be available for any biologist interested in the Big Bend area at the end of my graduate research. As previously mentioned, the Big Bend is an area of interest for many conservation agencies and groups, so it is imperative to document the workflow in an easy and reproducible way that is approachable to many people with many different skill sets.

The USGS Data Management best practices for sharing data are to:

1) document the process thoroughly

2) create an easy to find data storage

3) put the information “out there” for people to locate

Objective A, of my proposed graduate research, covers practices 2 and 3 of the USGS Data Management best practices, and Objective C covers practice 1. The recommendations for sharing datasets are to clearly define the purpose of the research, describe attributes and geography, include associated links, specify a required data citation and acknowledgements, and create a second public version containing all appropriate metadata. Workflow documentation will contain step by step guide, screen shots, and descriptive text. Final documentation will be pushed to Github in a .doc or .pdf format.

**Expected Outcomes**

The results I intend to find are a decline in land mass of Deer Island in the recent decades, despite lack of much human influence. At this time, there are only story accounts, from local residents, of smaller islands existing around Deer Island, that are no longer visible or available, which suggests that shoreline changes have occurred since at least one generation. This case study will provide a good starting point for future spatial analysis in the region.

## Bibliography

Barone, L., Williams, J., & Micklos, D. (2017). Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLoS computational biology, 13(10), e1005755. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005755

Bearlin, A. R., Schreiber, E. S., Nicol, S. J., Starfield, A. M., & Todd, C. R. (2002). Identifying the weakest link: simulating adaptive management of the reintroduction of a threatened fish. Canadian Journal of Fisheries and Aquatic Sciences, 59(11), 1709-1716. [https://www.nrcresearchpress.com/doi/abs/10.1139/f02-140#.XSuJYOhKiUk](https://www.nrcresearchpress.com/doi/abs/10.1139/f02-140" \l ".XSuJYOhKiUk)

Geselbracht, L., Freeman, K., Kelly, E., Gordon, D. R., & Putz, F. E. (2011). Retrospective and prospective model simulations of sea level rise impacts on Gulf of Mexico coastal marshes and forests in Waccasassa Bay, Florida. Climatic Change, 107(1-2), 35-57. <https://link.springer.com/article/10.1007/s10584-011-0084-y>

Geselbracht, L. 2007. Conservation action plan for marine and estuarine resources of the Big Bend Area of Florida. The Nature Conservancy, Florida Chap-ter. conpro.tnc.org/968/big\_bend\_conservation\_ action.../download\_public

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., ... & Fernández, D. S. (2017). Skills and knowledge for data-intensive environmental research. BioScience, 67(6), 546-557. <https://academic.oup.com/bioscience/article/67/6/546/3784601>

Holling, C. S. (1978). Adaptive environmental assessment and management. John Wiley & Sons. <http://pure.iiasa.ac.at/id/eprint/823/>

Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O’Hara, C. C., ... & Halpern, B. S. (2017). Our path to better science in less time using open data science tools. Nature ecology & evolution, 1(6), 0160. <https://www.nature.com/articles/s41559-017-0160/>

Main,M.B.andG.M.Allen.2007.Florida’ s environment: North central region. Wildlife Ecolo-gy and Conservation Department, Florida Cooper-ative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, Gaines-ville, Florida, USA.

Raabe, E. A., Streck, A. E., & Stumpf, R. P. (2004). Historic topographic sheets to satellite imagery—A methodology for evaluating coastal change in Florida's Big Bend tidal marsh(No. 2002-211). US Geological Survey. <https://pubs.er.usgs.gov/publication/ofr2002211>

Seavey, J. R., Pine III, W. E., Frederick, P., Sturmer, L., & Berrigan, M. (2011). Decadal changes in oyster reefs in the Big Bend of Florida's Gulf Coast. Ecosphere, 2(10), 1-14. https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/ES11-00205.1

Schreiber, E. S. G., Bearlin, A. R., Nicol, S. J., & Todd, C. R. (2004). Adaptive management: a synthesis of current understanding and effective application. Ecological Management & Restoration, 5(3), 177-182. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1442-8903.2004.00206.x>

Walters, Carl J. 1986. Adaptive Management of Renewable Resources. Macmillan Publishers Ltd. <http://pure.iiasa.ac.at/id/eprint/2752/>

Walters, C. J. (2007). Is adaptive management helping to solve fisheries problems?. AMBIO: A Journal of the Human Environment, 36(4), 304-308. [https://bioone.org/journals/AMBIO-A-Journal-of-the-Human-Environment/volume-36/issue-4/0044-7447(2007)36[304:IAMHTS]2.0.CO;2/Is-Adaptive-Management-Helping-to-Solve-Fisheries-Problems/10.1579/0044-7447(2007)36[304:IAMHTS]2.0.CO;2.full](https://bioone.org/journals/AMBIO-A-Journal-of-the-Human-Environment/volume-36/issue-4/0044-7447(2007)36%5b304:IAMHTS%5d2.0.CO;2/Is-Adaptive-Management-Helping-to-Solve-Fisheries-Problems/10.1579/0044-7447(2007)36%5b304:IAMHTS%5d2.0.CO;2.full)

Yenni, G. M., Christensen, E. M., Bledsoe, E. K., Supp, S. R., Diaz, R. M., White, E. P., & Ernest, S. M. (2018). Developing a modern data workflow for evolving data. BioRxiv, 344804. <https://www.biorxiv.org/content/10.1101/344804v2.abstract>