Chapter 1- Data management workflow

Introduction

* Biological data have historically been management ineffectively, i.e storing on loose hard drives, not digitizing or entering data from physical data sheets.
* Biological data analysis are not always reproducible for further analysis, especially by other organizations.
* Difficult to provide data checks and quality control manually, for large sets of data.

Objectives

* Creating and maintaining a workflow specifically for the Oyster Restoration Project, and for general data management workflow use.
* Highlighting the best ways to manage different types of data for continuous and spatial data.

Implementing a modern data workflow

* Increasing need for computer and analysis techniques to be a core skill for biologists.
* Need for flexibility in data management.

Data management tools

* Excel - computing
* R - computing
* Python - computing
* Git / Github – version control
* ArcMap- mapping and geodatabases
* MySQL - database

QA in data entry

* Water quality QA in MySQL checking for double entries.
* Field technician check on physical data sheets.
* Additional code to find outliers in data, i.e values over 40 ppt for salinity

Adding data to databases on Github

* Version control with Github, when adding new data. Able to add and remove data through “committing” and “pushing/pulling” changes.

Automated updating of supplemental data

* Creating .rmd for figures or analysis on newly imported data.

Discussion

* Many new journals are increasing their requirements for data management and sharing.

Synthesis

* Workflow management can be implemented by various conservation agencies.