**Protocol for downloading, cleaning, and selecting water quality data**

Current sources supported to varying degrees: National water quality monitoring council (“Portal”), Water Atlas. If another data source is to be used, please add the protocol for data collection to this file and notify E Levine Williams of the additional source to update the compilation and selection R code.

***Please note, mapping project are currently undergoing a reorganization. File references may be inaccurate depending on where in the phase the document is reviewed. Please address all questions to E Williams.***

1. **Determine boundaries of desired area/estuary**
   1. Refer to Table 1 for boundaries. Add updates to Table 1 as needed. Broad boundaries can be found in the “Existing data tracking” file on the “StateGrid Assignment” tab. Adjust as necessary for best cover the estuary of focus.
   2. Confirm existence of KML area file in the Water-Quality-Processing/KML folder. Create file if needed by following Site\_Section\_file\_creation\_SOP document and naming the file using the code in Table 2.
   3. Ensure files are named appropriately based on estuary, data source (Table 2), and years of data.
   4. Determine water quality data sources and refer to the following sections as needed:
      1. National water quality monitoring council (Section II)
      2. (Section III)
2. **Retrieve water quality data from the National water quality monitoring council (“Portal”)**
   1. Go to <https://www.waterqualitydata.us/portal/> to download the water quality data for the desired area.
   2. Select the “Advanced” option near the top of the screen.
   3. In the “Select Location Parameters” section:
      1. A screenshot of a computer

         Description automatically generated with medium confidenceWithin the ‘Bounding Box’ area, enter in the bounding box coordinates using decimal degrees. Do not put anything in the country, state, or county boxes.
      2. Under ‘Site Type’ select “Estuary”, “Ocean”, “Wetland”, and “Stream”.
   4. Under the “Filter Results” section:
      1. In the “Sample Media” box, select “Water”.
      2. In the “Date Range” section, enter the starting and ending dates for the time period (5-year increments is best).
      3. In the “Minimum Sample Activities per Site” and “Minimum Results per Site” enter “12”.
   5. A screenshot of a computer

      Description automatically generatedUnder the “Download the Data” section:
      1. Make sure all items under “Data Source” are selected.
      2. File format: MS Excel 2007+
   6. Select “Site Data Only” under the “Data Profiles” area and select “Download” at the bottom of the page. Select “Continue” in the pop-up window.
   7. Navigate to where the downloaded file is saved, rename the file using the estuary and data source naming convention from Table 2, “Site data”, starting year, and ending year separated by underscores (ex. CR\_Portal\_Site data\_2012\_2022). Save the file to the WQ “Raw\_data” folder.
      1. Name files with years indicated by 4 digits (e.g., 2020, 2025).
      2. If data is not for complete years, a two-digit month can be added to the file name. Do NOT include underscores within the data range (e.g., 062020\_082025).
   8. Return to the webpage and select “Sample Results (narrow)” under “Data Profiles”. Select “Download” at the bottom of the page. Select “Continue” in the pop-up window. If the results are too large, modify the date range by cutting them into 2 or more sections. Repeat 9 with the updated date ranges in separate files.
   9. Navigate to where the downloaded file is saved, rename the file using the estuary code and data source naming convention from Table 2, “Results”, starting year, and ending year separated by underscores (ex. CR\_Portal\_Results\_2012\_2022). Make sure to use the same date naming convention as mentioned in step g.i-ii. Save the file to the WQ “Raw\_data” folder.
3. **Retrieve water quality data from the water atlas** (counties included: Lake, Manatee, Orange, Polk, Sarasota, Seminole, Pinellas, & CHENP).
   1. Go to the primary water atlas web page at <https://wateratlas.usf.edu/> and select one the desired atlases. All atlases can be downloaded together in the data selection process, so it isn’t as important which is chosen in this step.
   2. Under the “Maps/Data” section, chose “Data Download”.
   3. Accept the disclaimer then select the data type:
      1. Surface Water Quality – primary data
      2. Surface Water Hydrology – flow and water level data
   4. Select the following parameters then select Submit:
      1. By Location: Water atlas, and watershed
      2. By Site Info: Water body type
      3. By Sample Info: Date range, Parameter
   5. Complete the selection page as follows:
      1. Water Atlas: select desired data sources
      2. Watershed: select desired watersheds from list
      3. Water Body Type: select Bay and Stream/River
      4. Date Range: Specify date range for data
      5. Parameter: Select parameters. Include all the following plus any additional as needed. Parameters can be searched for by name using the search box. Select submit once all parameters have been selected.

* Depth, bottom, ft
* Secchi disk depth, ft
* Salinity, PSS and PPT
* Dissolved oxygen (DO)
* Dissolved oxygen saturation (%)
* Temperature, water (deg C & deg F)
* pH
* Turbidity (NTU)
* Chlorophyll a (probe relative fluorescence)
* Chlorophyll a, corrected for pheophytin
* Total suspended solids
  + 1. Select “Give me all station data” on the next screen (Step 4 of 5).
    2. Compile data to download by specifying “Excel” as the file type and “Row” as the file format and selecting “Generate file to download”. Then “Download File” once the file has been generated.
  1. Navigate to where the downloaded file is saved, rename the file using the estuary code and data source naming convention from Table 2, “Site data”, starting year, and ending year separated by underscores (ex. CR\_WA\_Site data\_2012\_2022). Save the file to the WQ “Raw\_data” folder.
     1. Name files with years indicated by 4 digits (e.g., 2020, 2025).
     2. If data is not for complete years, a two-digit month can be added to the file name. Do NOT include underscores within the data range (e.g., 062020\_082025).
     3. Open the file to make sure the data is converted properly: Select Yes to open and read the file, then select Convert in the 2 pop-up windows.
     4. Once the data is converted, select columns A (WBodyID) through column F (Actual\_StationID) and change cell type to ‘Text’.
     5. Save the file as a .xlsx file using the same name.

1. **Clean raw data and map station locations**
   1. Project setup:
      1. If working on the Molluscs network, navigate to the “Water Quality Processing” folder.
      2. If using git, create a branch of the [Water Quality Processing repo](https://github.com/E-Levine/Water-Quality-Processing) and map locally.
   2. Save all “Site data” and “Results” files to the “Raw\_data” folder in your local drive.
   3. Open the Water Quality Processing R project located in the main folder.
   4. Open the WQ\_data\_compilation R code file and run the code following in-code annotations making sure to use the same estuary code, data source code, and data range as specified in the raw data files. Running through code will save compiled, cleaned, and filtered data as an Excel file and output a widget map (if desired) of stations included in the data file.
      1. Additional notes:
         1. Data compilation will have to be performed for each data source independently. Data from different sources can be joined together later.
         2. Items listed below are the current filtering performed on the data. If additional data or columns are required, the code can be updated and saved for estuary- or project-specific use.
            1. Minimum column requirements for location data (“keep\_site”):

* MonitoringLocationIdentifier
* OrganizationIdentifier
* OrganizationFormalName
* MonitoringLocationName
* MonitoringLocationTypeName
* MonitoringLocationDescriptionText
* LatitudeMeasure
* LongitudeMeasure
* HorizontalCoordinateReferenceSystemDatumName
* StateCode
* CountyCode
* ProviderName
  + - 1. Minimum column requirements for results data (“keep\_results”):
         1. Portal:
         * MonitoringLocationIdentifier
         * ResultIdentifier
         * ActivityStartDate
         * ActivityStartTime/Time
         * ActivityStartTime/TimeZoneCode
         * CharacteristicName
         * ResultMeasureValue
         * ResultMeasure/MeasureUnitCode
         1. Atlas: (TBD)
      2. Minimum parameter “Characters” to keep:
         * Salinity
         * Temperature, water
         * Depth, bottom
         * Depth, Secchi disk depth
         * Temperature, air, deg C
         * Turbidity
         * Conductivity
         * Specific conductance
         * pH
         * Dissolved oxygen (DO)
         * Dissolved oxygen saturation
         * Chlorophyll a, corrected for pheophytin
         * Chlorophyll a
         * Total dissolved solids
         * Total suspended solids
         * Zooplankton
         * Diatoms
         * Stream flow, instantaneous
         * Flow, severity (choice list)
         * Stream stage
         * Flow
         * Stream flow, mean. Daily
  1. Once finished with the “WQ\_data\_compilation” R code file, determine what point locations (i.e., monitoring stations, central points of areas, etc.) will be needed for mapping or data selection. Fixed monitoring stations can be displayed in maps and used for selecting data stations in the next section. Additional point locations can be added for selecting data stations in the next section. Check for required point locations in the Data/ Reference\_data/Stations\_area\_selections file and add as necessary. This file is maintained on the network or in the main branch (if working with git) for other use.
  2. Open the “WQ\_data\_selection” R code file to select desired data. This code will help output the area data and a map of selected WQ stations.
     1. Run the first three sections of code to load combined data, filter by dates, and output WQ station map.
     2. In the “Station Selection” sections, data can be limited by:
        1. Distance from specified stations based on one or two “buffer” criteria – specified locations as listed in the Excel file found in Data/Reference\_data/Stations\_area\_selections.
        2. Closest N stations to specified locations as listed in the Excel file found in Data/Reference\_data/Stations\_area\_selections. A maximum distance for consideration can also be set.
        3. Specific boundary or locations via:
           1. Station names
           2. Bounding box area

1. Combination of data from different sources – *in progress*

***Not all data sources have been added yet. Please contact E Williams for any requested updates.***

**TABLES**

**Table 1**. State grids by region. North, south, east, and west boundaries of state grid indicated. State grids in which estuary reside are indicated (Estuary), and existence of KML file noted by name. Update as appropriate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Estuary** | **North** | **South** | **East** | **West** | **Estuary KML** |
| Caloosahatchee | 27.1 | 26.1 | -81.4 | -82.4 | CR |
| St Lucie | 27.3 | 27.1 | -80.1 | -80.3 | SL |
| Lake Worth | 26.84362 | 26.527 | -80.03102 | -80.05384 | LW |
| Loxahatchee | 27.0 | 26.9 | -80.0 | -80.2 | LX |
| Tampa Bay | 28.7 | 27.45 | -82.3 | -82.9 | TB |
| St Andrews | 30.3 | 30 | -85.35 | -85.9 | SA |

**Table 2.** Naming conventions to be used for file names. Estuary or data source indicated by shorthand can be found in column 2. Update as appropriate.

|  |  |
| --- | --- |
| **Naming convention for files** | **Estuary or data source** |
| *Estuaries* | |
| CR | Caloosahatchee |
| SL | St. Lucie |
| LW | Lake Worth |
| LX | Loxahatchee |
| SA | St. Andrews |
| *Data sources* | |
| Portal | National water quality monitoring council |
| WA | Water atlas sources |