Data Pull User Guide

Tetra Tech

March 27, 2024

## Introduction

The following user guide pertains to the [*data\_pull.R*](https://github.com/KateriSalk/Alaska_IR_Automation/blob/main/Code/3_Data_Pull/data_pull.R) script developed by Tetra Tech for the Alaska Department of Environmental Conservation (AK DEC). This guide assumes that the user has relatively recent versions of [R](https://www.r-project.org/) and [RStudio](https://posit.co/download/rstudio-desktop/) installed and that they are familiar with the R coding language. Please direct any questions regarding the usage of the *data\_pull.R* script to Amber Bethe Crawford, AK DEC (amber.crawford@alaska.gov).

### Purpose

The purpose of the *data\_pull.R* script is to remotely download Water Quality Portal ([WQP](https://www.waterqualitydata.us/)) data for different waterbody types within the state of Alaska over a defined period of time. The script was developed to support AK DEC’s Integrated Reporting effort, and therefore, downloads all relevant data for the past five years. That being said, the script can be adapted to download a subset of data for other purposes.

## Required Packages

The required packages for the *data\_pull.R* code are [*TADA*](https://github.com/USEPA/TADA) and [*tidyverse*](https://www.tidyverse.org/). To install the *TADA* package, you will also need the [*remotes*](https://cran.r-project.org/web/packages/remotes/index.html) package. Note, the *remotes* package is not required beyond installation of the *TADA* package. Below is a code chunk that demonstrates how to install and load these packages.

# install.packages('tidyverse')  
library(tidyverse)  
  
# install.packages("remotes",  
# repos = "http://cran.us.r-project.org")  
library(remotes)  
  
# remotes::install\_github("USEPA/TADA",  
# ref = "develop",  
# dependencies = TRUE)  
library(TADA)

## Define Date Parameters

The only inputs that the user is required to provide are the desired start and end sampling dates. Both the start (startDate object) and end (endDate object) dates must be specified in the ‘YYYY-MM-DD’ format.

startDate <- '2018-02-01'  
endDate <- '2023-02-01'

## Data Pull/Download

The data download section is broken into five different “pulls”. Each pull is for a specific site type that is relevant to AK DEC’s Integrated Reporting process. The site types as categorized in the [WQP](https://www.waterqualitydata.us/) (see ‘Site Type’ under the ‘Basic’ option) include: “Lake, Reservoir, Impoundment”; “Stream”; “Estuary”; “Ocean”; and “Aggregate surface-water-use”. Data are obtained from the WQP using the *TADA\_BigDataRetrieval* function. Documentation on this function can be found by running the code below:

?TADA\_BigDataRetrieval

The inputs into the *TADA\_BigDataRetrieval* are the dates defined in the previous section, the WQP site type, the state code (i.e., AK), and applying auto clean. Note, at the time of devleopment, the *TADA\_BigDataRetrieval* function will not return a ‘TADA object’ without the applyautoclean = T.

#Estuary  
est\_data\_pull <- TADA\_BigDataRetrieval(startDate = startDate,  
 endDate = endDate,  
 siteType = 'Estuary',  
 statecode = 'AK',  
 applyautoclean = T)

The auto-cleaning applies the functions: T*ADA\_ConvertSpecialChars*, *TADA\_ConvertResultUnit*, *TADA\_ConvertDepthUnits*, and *TADA\_IDCensoredData*. For more information on the auto-cleaning functions run the following lines of code to review the documentation:

?TADA\_ConvertSpecialChars  
?TADA\_ConvertResultUnits  
?TADA\_ConvertDepthUnits  
?TADA\_IDCensoredData

The output of the *TADA\_BigDataRetrieval* function is a WQP data table with an additional 27 columns that contain the *TADA*-specific edits and flags.

## Exporting as CSV

The resulting tables from *TADA\_BigDataRetrieval*, one for each site type, can either be combined using rbind into one large table for export and/or be exported individually. Exporting to csv is performed using *dplyr::write\_csv* (loaded in with *tidyverse*). The user can change the export file location if needed.

write\_csv(est\_data\_pull, 'Data/data\_pull/data\_pull\_estuary.csv')