Documentation for the Central Data Management Office (CDMO) Automation Procedures

Prepared for:
NOAA and the NERRS Science Collaborative

May 22, 2018

LimnoTech
Water Environment | Scientists Engineers

LimnoTech
Water Environment | Scientists Engineers

501 Avis Drive
Ann Arbor, MI 48108
734.332.1200
www.limno.com

Documentation for the Central Data Management Office (CDMO) Automation Procedures

Prepared for:
NOAA and the NERRS Science Collaborative

May 22, 2018

TABLE OF CONTENTS

1 Introduction

The Central Data Management Office (CDMO) automation procedures are intended to minimize the amount of effort CDMO staff must spend to generate reporting plots for reserves within the National Estuarine Research Reserve System (NERRS) each year. This documentation outlines all of the steps necessary to install R., the required R packages and additional software needed, and how to run the automation powershell script. After extracting the files within the zipped folders, the processing folder should contain the following:

- 29 reserve-level templates: one for each reserve and one for future reserves
- One national-level template
- Two comma-delimited files for use with the powershell script
- One folder for annual system-wide monitoring program (SWMP) data updates
- One powershell script
- One batch file
- One "00_Additional_Materials" folder that contains scripts and executables that will assist with installation.

The batch file will run the powershell script that will then run R scripts associated with each of the reserve-level templates and the national-level template. After executing all template R scripts, the powershell script will then zip all template folders so that they are ready for distribution to the reserves.

2 Initial Installation/Set-up

The section below outlines the steps needed to perform the initial installation and and set-up that is required to successfully run the reserve-level templates and national-level templates from the Windows command line.

1. Copy the following files to a dedicated folder on the desktop/server machine where the NERRS sites will be processed:
    - Create folder for processing ("C:\NERRS\" is used as the example here)
    - Files to place in main processing folder:
        - NERRS_run_reserves.ps1
        - NERRS_run.bat
        - NERRS_sites.csv
        - NERRS_years.csv
    - Supporting folders (extract to processing folder):
        - NERRS_data_folder.7z
        - NERRS_reserve_folders.7z
        - 00_Additional_Materials (folder)
    - Note: the NERRS_data_folder.7z file currently contains a copy of the all data that was used to create the 2015 annual reserve-level reports. This data was included so that CDMO can run the reserve processing script and verify that set-up was successfully completed.
2. Install and configure R software (version 3.4.4 or later):
    - Uninstalling previous versions of R is recommended (if feasible)
    - Download version 3.4.4 or later from: https://www.r-project.org/
    - Check Windows PATH for R "bin" folder reference, and add if necessary (one time step):
        - System → Advanced System Settings → Environment Variables
        - Select "Path" entry listed under "System Variables", then "Edit"
        - Add "C:\Program Files\R\R-3.4.4\bin" if not present (with semicolon delimiter)
    - Note: RStudio is not required to run the automated sequence for NERRS processing and reporting.

3. Install supporting R packages:
    - Ensure that the current user has “write” access to the R ‘library’ folder: “C:\Program Files\R\R-3.4.4\library”
    - Open a Windows command line (with administrative privileges) and navigate to the “00_Additional_Materials” subfolder.
    - Run the “00_initial_R_package_install.R” script (4/25/18 version) to install required packages:
        - R.exe --vanilla -f 00_initial_R_package_install.R
    - Install the pandoc library from the separate installation file (“pandoc-2.1.3-windows.msi”)

3 Pre-Execution Annual Checks

The section below outlines the manual steps that will need to be completed each year before running the reserve processing script.

1. Add new system-wide monitoring program (SWMP) data to the "00_Data" folder. Data that is placed in this folder will be moved to the appropriate reserve-level template "data" folder. If duplicate data files are present within the "00_Data" folder and a reserve-level template "data" folder, the reserve-level template data will be overwritten by the files from the "00_Data" folder.
2. Create templates for new reserves (as necessary):
    - Create a copy of "Reserve_Level_Template" folder
    - Rename "Reserve_Level_Template" using the 3-letter reserve identification code for the new reserve.
    - Repeat these steps for every new reserve.
3. Configure CSV input files (as necessary):
    - "NERRS_sites.csv" – should already contain a full listing of reserve sites, but can be modified as needed to incorporate additional sites by adding the 3-letter reserve identification code for all new reserves
    - "NERRS_years.csv" – specify the "target_year" as well as the year range (via the "start_year" and "end_year" fields).

4 Script Execution & Output

The section below outlines the steps needed to execute the reserve processing script.

1. Execute the reserve processing script:
    - Open a Windows command line window (with administrative privileges)
    - Navigate to the main processing folder (populated under step #1 in Section 2)
    - Run the "NERRS_run.bat" batch file, which will run the PowerShell command with an "ExecutionPolicy" bypass (to work around potential restrictions for running scripts).
2. Results for reserve and national summaries:
    - Reserves: a zip file containing the entire file set for each reserve will be written to the "00_Reserve_distrib_files" folder (created by the script)
    - National summary: a zip file containing the entire file set for the national summary will be written to the "00_Reserve_distrib_files" folder (created by the script)