DWRAT DataScraping GitHub Repository README

Payman Alemi

2023-08-01

I have set up 3 folders for this project, each of which has several subfolders. We have intentionally added some of the subfolder paths to the .gitignore file for this repository because they contain massive files that GitHub cannot handle. While the files themselves are ignored, the folders are referenced by Git to maintain the repository’s structural integrity.

**Supply**  
This folder contains the files necessary for automating the Santa Rosa Plains (SRP) GS Flow and PRMS (Precipitation-Runoff Modeling System) hydrology models. As of 2023-08-01 this folder has 5 subfolders. The pre-processing and post-processing of the *PRMS* model has been nearly entirely automated. By contrast, much of the *SRP GS Flow* model still needs to be automated.

* **Documentation:** This folder contains relevant emails, PDFs, Word Documents and other files containing instructions or information about the project.
* **InputData:** This folder contains the datasets used for loops and functions, e.g. station lists. Maintaining these datasets as CSVs is easier than creating then as dataframes. *Gag files in this folder are ignored by the .gitignore file*
* **ProcessedData:** This folder contains datasets that have been manipulated in some way by R Scripts, spreadsheets, etc. *Files in this folder are ignored by the .gitignore file*
* **Scripts:** This folder contains all scripts associated with this project
* **WebData:** This folder contains unaltered datasets scraped or downloaded from the Internet, e.g. weather station data. *Files in this folder are ignored by the .gitignore file*

**Demand**  
This folder contains the files necessary for converting the raw diverter demand data from eWRIMS into a processed, QAQC’d demand dataset ready for importation into DWRAT. As of 2023-08-01, this folder has 4 subfolders, which just contain placeholder scripts–the actual scripts have yet to be written.

* *Documentation*: This folder will contain relevant emails, PDFs, Word Documents and other files containing instructions or information about the project.
* *InputData:* This folder just contains the output CSV from the GIS pre-processing steps, which serves as the input file for the R pre-processing steps and ultimately the demand module scripts.
* *Intermediate Data:* This folder contains datasets that have been modified from the RawData folder. These datasets are considered intermediate because they will be used as inputs for other R Scripts. *Files in this folder are ignored by the .gitignore file.*
* *Module and Script Comparison*: This folder compares the outputs of the Excel Demand Data Modules and the outputs of the corresponding R scripts. Each Excel module has a corresponding R Script.
* *OutputData:* This folder will contain the final demand dataset CSVs that are ready for importation to the Upper Russian River (URR) and Lower Russian River (LRR) Drought Water Rights Allocation (DWRAT) models.
* *RawData:* This folder contains the downloaded flat files from eWRIMS.*Files in this folder are ignored by the .gitignore file.*
* *Scripts:* This folder will contain the scripts that convert the raw diverter demand datasets into the final datasets to be used by the DWRAT models.

**Allocation**  
This folder will contain the files necessary for running the URR and LRR DWRATs, each of which will have a separate folder.

* ***URR DWRAT***
  + *Input****:*** This folder contains the input files for the URR DWRAT run.
  + *Output:* This folder contains the output files for the URR DWRAT run.
* ***LRR DWRAT***
  + *Input:* This folder contains the input files for the LRR DWRAT run.
  + *Output:* This folder contains the output files for the URR DWRAT run.