Contents

[Introduction 1](#_Toc422324777)

[data\_clearinghouse 1](#_Toc422324778)

# Introduction

Acquiring large sets of reports from PeMS can be very time consuming, and require a lot of manual clicking. For example, downloading a year 5-minute station reports from Data Clearinghouse requires manually clicking 365 different links.

These scripts automate the repetitive clicking process. Folders are named after the type of reports the contained scripts download.

# Ignoring data folders

The PeMS downloads are too large to push to GitHub. Any file or folder with “data” prepended to the name will be ignored when pushing.

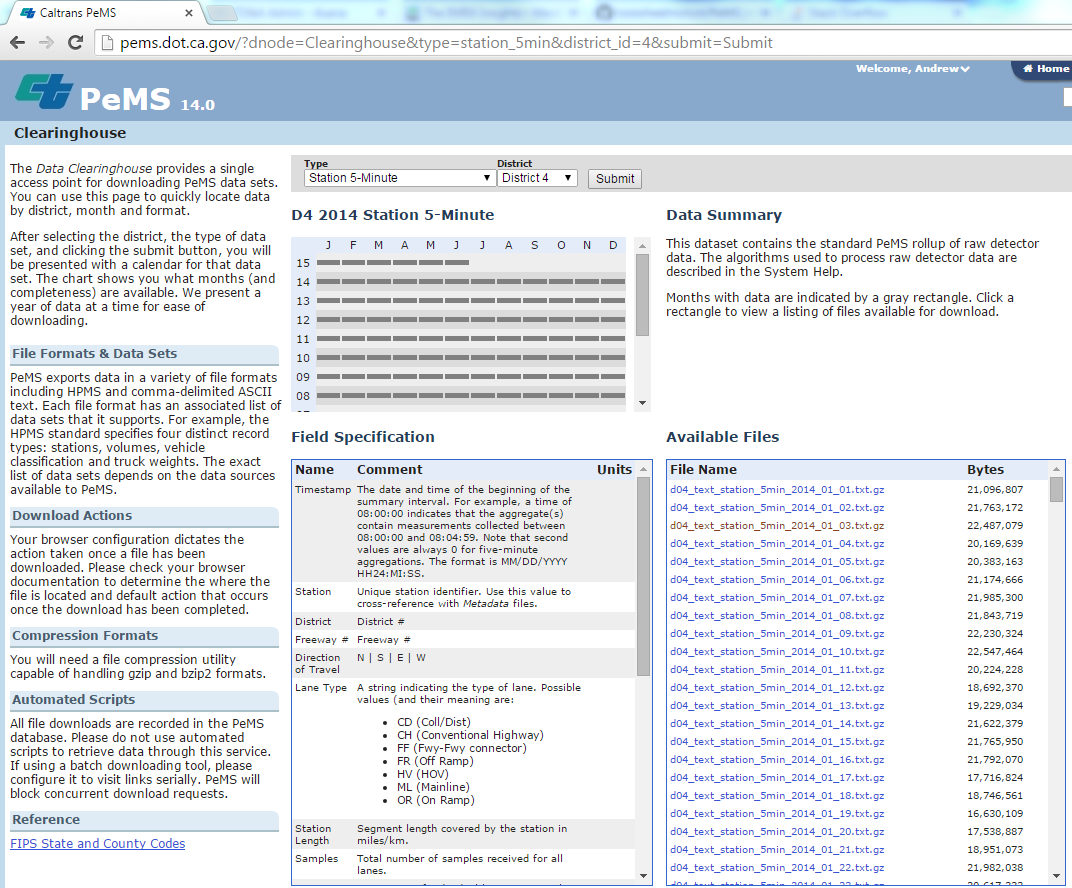
# Scripts

## clearinghouse

The Clearinghouse is the main point for downloading bulk data sets.

Step 1: Generate the right html page to download

Let’s say I want to download all *Station 5-Minute* files for 2014. You need to go to log in to the PeMS website, go to the clearinghouse, and **point and click** until you have the correct page:



Notice that the url query string specifies the data set, *station\_5min*, but does not specify the year 2014. That is because the actual download links (in the Available Files box) are served up by an internal Javascript that does not update the url.

For now, I am not enough of a ninja to bend that Javascript to my bidding, so my script requires you to point and click until you see the files you want in the Available Files box.

Now you need to **download the raw html file**. The clearinghouse\_downloader.py script will parse out the 365 download links from this html file and then download them all. This will take many hours.

Step 2: Configure the config file and run

Update the fields in clearinghouse.ini. This is the config file and it tells the script where to read and write things and also provides your PeMS credentials.

The [Paths] are pretty self-explanatory. Notice that *out\_dir\_path* is the only one the references a directory, so it need to end with a ‘\’ character. This is where the downloaded files will be saved.

## detector\_health

This script is still very hacky. You need to manually edit lines:

12, out\_path

19, start\_date

20, end\_date

23-27, insert your PeMS credentials

Andrew will update this to use a nice config file in the near… future (06/17/15).

Generally speaking, we download the detector health reports by generate query strings for the dates we want. So there is no need to download and parse a HTML file like with the Clearinghouse downloader.