

# FEC Bulk Download – Getting Started

November 29, 2019

```
[1]: import datetime
      startTime = datetime.datetime.now()
      print(startTime.strftime("%Y-%m-%d %H-%M-%S"))
```

2019-11-29 20-53-43

## 1 Historical Background

The *Wayback Machine* shows the contents of [www.fec.gov/finance/disclosure/ftpdet.shtml](http://www.fec.gov/finance/disclosure/ftpdet.shtml) back to 2004 when in early years the FTP site was at <ftp://ftp.fec.gov/FEC/>. Anonymous FTP and tools like `wget` were quite convenient to grab any or all of the contents.

The `ftpdet.shtml` web page was quite useful until roughly June 2017 when the link went stale, and by July 2017 was redirected to the now “classic” version: <http://classic.fec.gov/finance/disclosure/ftpdet.shtml> but with files still on the same FTP server.

### 1.1 Amazon-hosted bulk downloads

The “classic” `ftpdet.shtml` page still exist as of this writing (2019-11-29), but by Feb. 2018 the original FTP server was replaced with hosting on Amazon with quite long base links: <https://web.archive.org/web/20180221210046/https://cg-519a459a-0ea3-42c2-b7bc-fa1143481f74.s3->

Anonymous FTP and `wget` commands no longer worked to transfer files from the AWS-hosted site. A Feb. 2018 posting on the [FEC GitHub page](#) described [Retrieving S3 bulk download files](#). The posting described the AWS CLI tool for downloading bulk data, but the [AWS Command Line Interface](#) link suggest the original `aws` command-line tool has been superceded with a newer `aws2` tool by Nov. 2019.

### 1.2 AWS CLI version 2

Here is the AWS link for [Installing the AWS CLI version 2](#). Linux probably works best for this, but I installed the tool in Windows 10. The default installation put the tool in the `PATH`.

```
[2]: !where aws2
```

C:\Program Files\Amazon\AWSCLIV2DevPreview\aws2.exe

To learn more about the tool: `aws2 help`

```
[3]: !aws2 --version
```

aws-cli/2.0.0dev1 Python/3.7.5 Windows/10 botocore/2.0.0dev1

The instructions on the FEC GitHub page worked fine replacing `aws` with `aws2`.

## 2 FEC Bulk Downloads

### 2.1 AWS setup info

```
[4]: REGION = "us-gov-west-1"
FECBULK = "s3://cg-519a459a-0ea3-42c2-b7bc-fa1143481f74/bulk-downloads"
```

### 2.2 Show high-level view of FEC Bulk Download folders and files

Let use the Linux `ls` command to list the files in the 2020 folder:

```
[5]: !aws2 --region $REGION s3 ls $FECBULK/ --no-sign-request
```

```
PRE 1978/
PRE 1980/
PRE 1982/
PRE 1984/
PRE 1986/
PRE 1988/
PRE 1990/
PRE 1992/
PRE 1994/
PRE 1996/
PRE 1998/
PRE 2000/
PRE 2002/
PRE 2004/
PRE 2006/
PRE 2008/
PRE 2010/
PRE 2012/
PRE 2014/
PRE 2016/
PRE 2018/
PRE 2020/
PRE 2022/
PRE 2024/
PRE Presidential_Map/
PRE data.fec.gov/
PRE data_dictionaries/
PRE electronic/
PRE fecviewer/
PRE paper/
PRE presidential/
```

```

PRE presidential_matching_funds/
PRE press/
2019-11-29 04:01:08 549601 ec_table.inc
2019-11-29 04:01:10 570190 electioneering.csv
2019-11-28 23:31:49 33458 foia_files_summary.xml
2019-11-29 04:35:54 65899 hybrid_pac.inc
2019-11-29 04:35:52 205350 ie_nm.inc
2019-11-29 04:35:52 205350 ie_tp.inc
2019-11-29 04:35:37 235949 ieoc_alpha.inc
2017-08-28 20:02:53 541 index.html
2018-03-31 07:51:12 524 index2.html
2019-04-01 09:43:48 16087997 leadership_pac_disbursements_2013-2019
20190329.xlsx
2019-04-01 09:54:58 16087997 leadership_pac_disbursements_2013-2019
20190329.xlsx
2019-11-29 04:35:35 957988 leadership_pacs.inc
2017-11-03 09:37:24 8123 list.js
2019-11-29 04:35:30 801244 lobbyist_table.csv
2019-11-29 04:35:32 2107769 lobbyist_table.inc
2019-11-29 04:35:34 2367007 lobbyist_table.xml
2019-11-29 04:35:53 178773 lobbyist_table_xml.zip
2019-11-29 04:35:44 5835759 new_f1dt.inc
2019-11-29 04:35:49 5835759 new_f1nm.inc
2019-11-29 04:35:50 5835759 new_f1tp.inc
2019-11-29 04:35:43 104 pres_f1_date.html
2019-11-29 04:35:42 1451 pres_f1dt.inc
2019-11-29 04:35:42 1434 pres_f1nm.inc
2019-11-29 04:35:43 104 pres_f2_date.html
2019-11-29 04:35:38 495630 pres_f2dt.inc
2019-11-29 04:35:40 495613 pres_f2nm.inc
2019-11-29 04:35:40 495629 pres_f2pty.inc

```

## 2.3 Show list of files in the 2020 directory

```
[6]: !aws2 --region $REGION s3 ls $FECBULK/2020/ --no-sign-request
```

```

2019-11-29 04:28:03 4350962 2020_HOUSE_SENATE_CAMPAIGNS_DOWNLOAD.csv
2019-11-29 04:28:08 1189380 2020_INDEPENDENT_EXPENDITURE_DOWNLOAD.csv
2019-11-29 04:28:08 11150748 2020_PAC_DOWNLOAD.csv
2019-11-29 04:28:11 1580624 2020_PARTY_DOWNLOAD.csv
2019-11-29 04:28:11 270837 2020_PRESIDENTIAL_CAMPAIGNS_DOWNLOAD.csv
2019-11-29 04:11:32 2827 CommunicationCosts_2020.csv
2019-11-29 04:11:33 2876 ElectioneeringComm_2020.csv
2019-11-29 04:11:33 746429 Form1Filer_2020.csv
2019-11-29 04:11:35 534861 Form2Filer_2020.csv
2019-11-29 04:10:32 360613 candidate_summary2020.zip
2019-11-29 04:11:36 1306425 candidate_summary_2020.csv

```

```

2019-11-29 04:10:33      120759 candidate_summary_grid2020.zip
2019-11-28 23:32:25       59595 ccl20.zip
2019-11-29 04:26:51    3243203 ccsummary2020.zip
2019-11-29 04:26:55     771225 ccsummary_grid2020.zip
2019-11-28 23:32:22     629854 cm20.zip
2019-11-28 23:32:16     209463 cn20.zip
2019-11-29 04:10:33    1703635 committee_summary2020.zip
2019-11-29 04:11:37    5893451 committee_summary_2020.csv
2019-11-29 04:10:35     420155 committee_summary_grid2020.zip
2019-11-29 04:11:39       5523 ec_exp_2020.csv
2019-11-29 04:11:39     953902 independent_expenditure_2020.csv
2019-11-24 14:44:42   492945822 indiv20.zip
2019-11-29 04:11:40       89836 leadership2020.csv
2019-11-24 14:47:23    16468994 oppexp20.zip
2019-11-24 14:46:28    14531508 oth20.zip
2019-11-24 14:45:34    3540394 pas220.zip
2017-08-31 13:28:34         71 readme_future_election_cycle_folders.txt
2019-11-28 23:32:02     123445 weball20.zip
2019-11-28 23:32:12     316693 webk20.zip
2019-11-28 23:31:56     76539 webl20.zip

```

## 2.4 Save files from 1980 directory locally in K:/TEMP/1980

Use Linux-like cp (copy) command.

Let's use 1980 since it's a fairly small folder. Let's save log of transfers.

```
[7]: !aws2 --region $REGION s3 cp --recursive $FECBULK/1980/ --no-sign-request K:/
    ↪Temp/1980/ > 1980-Download.log
```

## 2.5 Save files from selected directories locally in K:/TEMP

Save files from two selected folders locally in K:/TEMP/

```
[8]: directoryList = ['1978', 'data_dictionaries']

[9]: for dir in directoryList:
    print(dir)
    !aws2 --region $REGION s3 cp --recursive $FECBULK/$dir/ --no-sign-request K:
    ↪Temp/$dir/ > $dir-Download.log
```

```

1978
data_dictionaries

```

```
[10]: stopTime = datetime.datetime.now()
    print(stopTime.strftime("%Y-%m-%d %H-%M-%S"))
    print(stopTime - startTime)
```

```

2019-11-29 20:53:54
0:00:11.003393

```

**Earl F Glynn, Olathe, KS**