The FEC's Electronic Filing Site, Part II: Downloading, Importing

Aron Pilhofer, CFIC Director,
Investigative Reporters and Editors

Introduction:

As of the 2002 election cycle, political parties, PACs as well as candidates for president and US House are required to file campaign finance reports with the Federal Election Commission electronically. For journalists, this is a positive boon – assuming they have the skills.

Downloading the reports is quite simple. Once you have identified the candidate or committee you are interested in on the FEC electronic data download site, simply click the "download" link next to the report you want, and it will be saved to your local hard drive. (For more on the site itself, see the companion handout "The FEC Electronic Filing…Part 1: A guided tour".)

The tricky part comes next, and that's what this tutorial will cover: how to import the data into a spreadsheet, clean it and format it. Some knowledge of spreadsheets is helpful, but not at all necessary. All the skills covered here are beginner to intermediate level.

One last note: There are any number of different types of documents filed with the FEC electronically, not just campaign finance reports. All these forms can be imported in just exactly the same way described in this document, but for the purposes of this tutorial, we are going to be importing the 2001 year-end filing of House Speaker Dennis Hastert.

Downloading Data:

Again, downloading the data is very straight-forward. Locate the report you want on the FEC electronic filing site, click "download" and then save the file somewhere you can remember. For the purposes of this tutorial, I'll be using House Speaker Dennis Hastert's year-end report from 2001 (find a copy …).

First, let's take a look at the filing and see what the raw data looks like. Using wordpad, notepad (or whichever text editor you prefer), open the file. It should look a lot like this:

```
"HDR","FEC","3.00","Aristotle International CM4 PM4","Version 4.1.1","^^","",""
"F3N","C00208090","Hastert for Congress Committee","P. O. Box 625","","Batavia","IL","60510   "," ","Il
"SA11C","C00208090","PAC","Amer. Acad. of Dermatology Assoc. PAC","P. O. Box 4014","","Schaumburg","IL"
"SA11C","C00208090","PAC","AAN-Nursery Industry  PAC","1250 I Street, NW, Suite 500","","Washington","I
"SA11C","C00208090","PAC","ABATE of IL-FEDPAC","311 E. Main ST, #418","","Galesburg","IL","61401   ","
"SA11C","C00208090","PAC","ACTIVATOR","1701 Towanda Ave.","","Bloomington","IL","61701   "," ","P   ","
"SA11C","C00208090","PAC","Amer. Society of Anesthesiologists PAC","520 N. Northwest Highway","","Park
"SA11C","C00208090","PAC","Accenture PAC","800 Connecticut Ave., NW, Ste. 600","","Washington","DC","20
```

This is what we call a comma-delimited text file. Let's break that down:

- It is comma-delimited because there are commas separating each piece of information, or each column in our spreadsheet. Delimiter is just a fancy computer word for separator. When we import this file, each time the spreadsheet sees a comma, it will place a column. (Any character can be user as a delimiter, but commas are by far the most common.)

- It is a text file because the letters and numbers aren't formatted in any way. If it were formatted, you would see a lot of gobblygook in the file mixed in the text. (Try opening a Word document in another application, and you'll see what I mean.)

Lastly, notice the double quote marks around each piece of data. Those tell the spreadsheet two things: one, keep all this stuff together no matter what, and even if you see numbers here, don't format them that way in the spreadsheet.

Why would you need to do this? We're using commas to separate our pieces of information here, but what would happen in the case of the fourth line in the example above … the part that reads "1250 I street, NW, Suite 500"? There is no way for the spreadsheet to know that those commas are part of the street address, and are not delimiters. Therefore, we must put the entire address in quotes to tell the spreadsheet to treat it as a single piece of information even if there are commas inside it.

<u>Import Data:</u>
Before you close the file, we'll need to rename it. It's just a little quirk of Microsoft Excel– the file that you downloaded has the extension ".fec," which Excel won't recognize. It's also a good habit to keep a copyof your raw data just in case something goes wrong.
Click File->Save As and save the document something (and somewhere) you will remember, but make sure it ends with a .txt. I named mine "hastert.txt".Now, open Microsoft Excel, go to File->Open and use the pulldown menu at the top to navigate to the directory in which you have saved your file.
It's not there, is it? Don't panic, it is there. It's just not visible until you tell Excel to look for the right kind of file. (This is why we renamed it with a .txt on the end.) Look at the bottom of the dialog box and you'll see a pulldown menu titled "Files of Type." Pull down the handle on the right until you see one that says "Text Files (.txt, .prn)" or something similar. Your Hastert filing should magically appear. (If it doesn't, go back a few steps… you probably saved it in a different folder.)
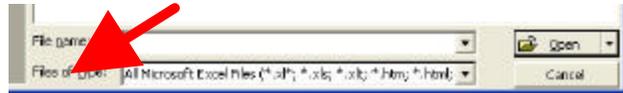
Click on your file, and click the "Open" button. A new dialog box will open with the title "Text Import Wizard." It should look a lot like this:
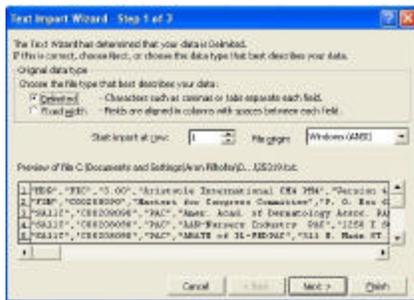
The text import wizard is one of the really nice features of Microsoft Excel and Access. It makes importing files such as these quite easy, and fast. I'm just going to cover the steps involved with importing FEC data here. But if you are interested in learning more, IRE has an extensive collection of handouts that cover data importing in detail.

To get your data into a spreadsheet, do the following:

Step One:
- You will see two radio buttons on top. Make sure the one labeled "Delimited" is selected
- If you have the "start import at row" selector, change it to 3. The first two lines are not important for us. They are a header row that indicates what software the committee used, and the second line is the aggregate data. (Not all versions of Excel have the "start import"feature. If yours doesn't, you'll have to delete the first two rows by hand later on.)
Click Next.

Step Two:
- In the "Delimiters" area, make sure you select "comma" and unselect anything else. Tab is often selected by default, and you will want to unselect it.
- Pull down the menu next to "Qualifier" and make sure the double quote mark is selected. If it isn't, all the quote marks in the data will be imported along with the inside them, and you don't want that.
- Click Next.

Step Three:
- Here it can get tricky. This is where you tell your spreadsheet how you want it to format each column of data – as a number, text or date. By default, all the columns start in "General" format, which means Excel will guess what sort of format you want based on what typeof data it thinks the column contains. When do we run into trouble with this? Zip codes. Zips should always, always be formatted as "text" and never as a number. Why? What happens if you have a zip that starts with a zero, like 05467?
Unless you specify otherwise, Excel will treat it as a number – and the leading zero will vanish. Your zip will now read 5467, which in the world of databases and spreadsheets is NOT a match for 05467.

OK, back to the text wizard.
- Scroll to the right (using the bottom scroll bar) until you find a field that looks like zip codes, and click right on the word "General." The entire column will turn black. Then find the radio button above that says "Text" and select it.

You should see the header switch from "General" to "Text" – this is good. Now you are ready to import. Click Finish, and voila! you should see your new spreadsheet full of data.

Cleanup:
You're not done yet.
Scroll down the spreadsheet until you see the "SA" codes in Column A change to "SB." SA means contribution and SB means disbursement.

This is something brand new: Not until electronic filing did we get data on individual expenditures made by candidates. These are two different sides of the ledger, and thus should be treated separately. Then scroll to the right, and you should see the problem: The columns don't line up, do they? The column layout for disbursements is different than it is for contributions, which is another reason we have to move them to a different sheet.
First, insert a new worksheet (Insert->Worksheet) and a blank sheet will open. Look at the bottom of your screen, and you should see two tabs: One will say something like "Sheet 1" (that's our new worksheet) and the other will hold the name you gave your Hastert filing, probably something like hastert.txt. Click back and forth to get used to the way worksheets function in Excel. Each workbook can contain a number of worksheets, each of which can reference data contained in another sheet-- but that is for another lesson.
Now, let's select our disbursement data and move it to our new sheet.
Click on the row number of the row in which your disbursement data begins. In the example above, I would click on row 826. Click right on the number in the gray box, and notice what happens: You have selected the entire row extending all the way across the spreadsheet.
Now, using the scroll bar on the right side of your screen, scroll to the last line of your data. Hold down the shift key, and click right on the row number in the gray box and you'll see a good part of your screen turn shaded (either black or gray or blue depending on the version of Excel you are using). You've selected everything from the first to the last row of disbursement data.
Now, click on Edit->Cut and you should see the borders of your selected area start flashing. Click on tab for Sheet 1 (our new worksheet), then click in cell A2. Then select Edit->Paste, and your disbursements will be copied right in.
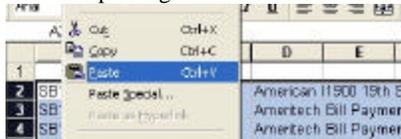
Why A2? You'll see in a minute.
Before we go further, let's go back to our main spreadsheet and add a row at the top. Click on hastert, or whatever your sheet is called, and then click right on row number 1 in the gray box. Then, select Insert->Row. This will add a blank row at the top of your sheet. Now you should have a single Excel workbook with two worksheets in it: one is expenditures (SA) and one is disbursements. Wouldn't it be nice to know what all these columns mean? Let's go get the header file, and put those in. Punch this URL http://www.fec.gov/elecfil/format.exe into a web browser, and you should be prompted to download a file called "format.exe." Do so, and then go double click on it to extract the contents. You'll find two files inside. The one we want is called "e-filing headers.xls," so double-click on it to open it in Excel.

These are headers for all the possible documents you will find on the FEC electronic filing site. Scroll down until you find one that has an "SA" in the Header column. Those are the headers for our contributions. "SB" are the headers for our disbursements. Copy the entire row and paste them into the appropriate sheets. Click on the row number of the "SA" header file to select the entire row. The select

Edit->Copy this time. Ten pull down the Window menu and find the workbook with your data in it. Click on the Hastert worksheet, and then click once in cell A1. Select Edit->Paste to paste in the header file.

Now you just need to do one more thing to make the headers line up properly: Delete the "SA" label in cell A1. To do that, simply click on the cell as shown above, then select Edit->Delete. A dialog box will pop up asking how you want the remaining cells to be adjusted. Select Shift Cells Left and click OK. Now you have headers in your file. Repeat the process for your disbursements sheet.

There is one last thing to do before we are ready to start crunching numbers, and it's a bit complicated. It is necessary if you want dates in your data. There's no way around it. Scroll over until on your contributinos data you find the column "date received." That is a date, in a standard database format year, month, day.

That's great, but it does us no good in this format. Let's clean this up. Click on the column just after the date received column (in the example to the left, it would be column P). Select Insert-> column, and like magic a new and blank column will appear. Now comes the fun part. You need to insert a formula that will make the string of numbers look like a date, and here is that formula: =MID(O2,5,2)&"/"&RIGHT(O2,2)&"/"&LEFT(O2,4). Type this into the cell just to the right of first date (as shown), and you will suddenly see a data appear. You don't have to know how it works, just that it does. NOTE: Make sure you edit the formula to match the correct row. In the example above, the old date field is column O, thus you will see "O2" in the formula. If the dates were in column D to begin with, then the formula would read =MID(D2,5,2)... etc.

This is a bit tricky, but it's the only way to get your dates to look like dates.
Once you have the formula in the right place, hit the Return key and it should look like a date (e.g., 8/12/2001). If so, congratulations! Now let's copy it all the way down the sheet. Click on the cell in which you have written your formula, and then use the right hand scroll bar to scroll to the end of your data. Hold down the Shift key and click in the last cell of the column. The entire column from the cell in which you wrote your formula to the end of your data should be selected. Select Edit->Fill->Down, and your formula will be copied all the way down.
Don't click on anything at this point, you're not done!
Now, you are going to execute several commands in a row to convert the contents of those cells from the formula you typed in and copied down to just the date that you see displayed. The cells you just copied down -- the ones with the formulas you typed in showing the real date -- should still be highlighted (selected). If not, do so now.
Select Edit->Copy, and you should see the dancing ants again. Select Edit->Paste Special, and a dialog box will open. Under the Paste section, select Values and click OK. Now your cells no longer contain formulas, but rather the result of those formulas (in this case, the formatted date).
The ants should still be dancing around your date column. If not, select the dates again, and then select

Format->Cells, and a dialog box will open.
Make sure that Date is selected in the category, and if not, select it.
Unfortunately due to a bug in Excel it probably won't make a difference what you select in the Type window, but you can select a preferred format there if you want to. Click OK.
Now you have properly formatted dates that Excel should recognize as such.
To completely finish off this part of the process, type in your own header above your new date field. New Date is a good choice. To fix the dates in the disbursements side, you can repeat the process.

Names:
There is one remaining issue with this spreadsheet, and that is the name column. You might notice that the names are sometimes delimited not delimited by commas, but rather by carrot "^" characters. There is a quick fix for this: Select the entire column of names, and then select Edit->Replace, and a dialog box will pop up. In the first text window (search for) put a carrot character. In the second text window (replace with) put a single space (hit the space bar once). Click Replace All, and all the carrots will be gone.
There are ways to split up the names so you end up with first and last names in different columns, but that is beyond the scope of this tutorial.