



By Richard B. Lanza, CFE, CPA/CITP, PMP

Fear Not the Software

Fraud-fighting with Data Analysis Tools

ARE YOU UNKNOWINGLY WORKING WITH TERRORISTS?

Nobody today wants to find out their company is linked with terrorist organizations through a customer, supplier, or employee relationship. The U.S.

Treasury maintains a list of names and addresses of such organizations for all to download from its Web site.

With some simple queries in Microsoft Excel, Access, or other report writer, you can compare this list to your company name and address databases. Here I'll show you how to match up the lists in Microsoft Excel and then in an even easier way with an inexpensive add-in to Excel.

LIST OF TERRORISTS

The U.S. Treasury's link to "Specially Designated Nationals" is: www.treas.gov/offices/enforcement/ofac/sdn/index.shtml. Probably the best format is the ASCII delimited format for Windows computers:

www.treas.gov/offices/enforcement/ofac/sdn/sdallw32.exe. This will run an unzip file that will place a plethora of data tables in a folder of your choice on your PC. In this list are two files that are most useful: ADD.CSV and SDN.CSV. These

	A	B	C	D
1	Ent_Num	Add_Num	Address	City
2	10133	13633	40 Mint Road, Amoka Gardens, Fordsburg	Johannesburg
3	10133	13634	P. O. Box 42928	Fordsburg
4	10133	13635	16 Gold Street	Carletonville
5	10133	13636	P. O. Box 28215	Kensington
6	10132	13629	-0-	Johannesburg
7	10132	13630	71 Fifth Avenue	Mayfair 2
8	10132	13631	P. O. Box 42928	Fordsburg
9	10131	13627	386 Swanepoel Street, Erasmia	Pretoria
10	10130	13625	P. O. Box 4470	Damascus

Image 1

are comma delimited files that automatically open in Microsoft Excel. The ADD.CSV file is the addresses and the SDN.CSV are the names of the Specially Designated Nationals. Because these files don't have column headings (to tell what each column is in Excel), the names of the columns are in the description file:

www.treas.gov/offices/enforcement/ofac/sdn/dat_spec.txt. For the most part, these column headings are easily understood based on their contents. (For example, in the SDN.CSV file, column B is clearly the list of names of the Specially Designated Nationals.) These column headings have been added to image 1 (above).

Unfortunately, the names and addresses are provided in two files instead of one combined file. Therefore, if you want to relate the information from the name spreadsheet (SDN.CSV) to the address spreadsheet (ADD.CSV), you need to do so on column A (Ent_Num), which is the unique number that ties together the names and addresses.

MATCHING THE LIST IN EXCEL

Though the next steps aren't difficult, they're involved. You could prevail upon your IT department for help.

With some simple queries in Microsoft Excel, Access, or other report writer, you can compare the list of terrorist organizations to your company name and address databases.

While a wealth of additional information can be downloaded in the data files, the focus of this exercise is on the names and addresses of the Specially Designated Nationals. Therefore, the names are in SDN.CSV file (column B) and the street address line is in the ADD.CSV file (column C). These two columns can be compared using the VLOOKUP() function in Excel after a TRIM(), UPPER(), and LEFT() function is used. Here's what these functions do in a little more detail:

- VLOOKUP() – Takes a cell's value and tries to identify matches in a column of data. If there's a match, it will return a column next to the matched column.
- TRIM() – Will remove all spaces from a field except between words.
- LEFT() – Will take a specified number of characters from the cell, starting from the left-most section of the cell.
- UPPER() – Formats a cell into uppercase.

These functions can be used to align the name field from the SDN.CSV file to the address in the ADD.CSV file. You can do this by placing this function – **=VLOOKUP(A1,SDN.CSV!A\$1:B\$3310,2,FALSE)** – in cell G1 and then copying the formula down for the entire column G. (Just make sure to have both the SDN.CSV and ADD.CSV files open when creating this function.)

View the results can be viewed in image 2 (at right).

With the names and addresses in the ADD.CSV file, it can be aligned to a company's address file by following these steps:

Step 1 – Convert the field to be matched (assume B2 is one address line from the company's supplier masterfile) using the functions above in the following order:

Cell C2 in Excel = UPPER(TRIM(LEFT(B2,8)))

The nested function above will take the first eight characters of the address line field, take any spaces out of it at the beginning and end, and turn it all to uppercase text. This “normalizes” the field to make it more likely to match given that people will input information into an address line differently.

Step 2 – Follow the same process in Step 1 for converting the address field from the Specially Designated Nationals list. This could be done in column B of the ADD.CSV file (deleting the Add_Num column's contents, which aren't required for this analysis).

Step 3 – Sort both columns you plan to match in ascending order. In the above example, this would be the converted address from the supplier masterfile (column C) and the Specially Designated Nationals list (column B).

Step 4 – Use the VLookup() function to match the “normalized” address line cell (C2 in the example above) to the entire

listing of Specially Designated Nationals' address lines. Do that with this function:

Cell D2 in Company Supplier Masterfile

=VLOOKUP(C2,ADD.CSV!\$B\$1:\$C\$6820,2,FALSE)

Here are the component definitions of the function:

- VLOOKUP(C2, – Matches the cell that is the “normalized” address line from the company's supplier masterfile.
- ADD.CSV!\$B\$1:\$C\$6820, – This tells Excel to look within the ADD.CSV Excel spreadsheet and the columns B1

G1	A	B	C	D	E	F	G	H	I	J
	10	7	-0-	-0-	Panama	-0-	ABASTECEDORA NAVAL Y INDUSTRIAL, S.A.			
	15	12	-0-	-0-	Panama	-0-	ABELENUR, Nury de Jesus			
	22	14	-0-	-0-	-0-	-0-	ACECHILLY			
	24	15	-0-	-0-	-0-	-0-	ACEFROSTY			
	25	16	171 Old B	Valletta	Malta	-0-	ACEFROSTY SHIPPING CO., LTD.			
	36	25	-0-	Havana	Cuba	-0-	AEROCARIBBEAN AIRLINES			
	39	27	-0-	Managua	Nicaragua	-0-	AEROTAXI EJECUTIVO, S.A.			
	41	29	Bal Harbou	Panama C	Panama	-0-	AGENCIA DE VIAJES GUAMA			
	66	41	Avenida de	Madrid	Spain	-0-	AGUIAR, Raul			
	137	102	-0-	-0-	-0-	-0-	ALEGRIA DE PIO			
	150	111	-0-	Milan	Italy	-0-	ALOARDI, Carlo Giovanni			

Image 2

through C6820 for the address line.

- 2 – This tells Excel for the cell references B1 through C6820 to look within column C and return the value in column C (the full address of the SDN). Therefore, if this function returns a blank or “N/A” value, then there was no match. If there's a match, the information in column C for the matched row will be written in this cell.
- FALSE – This tells Excel that the only matches that are useful are perfect matches.

Because I'm limited in explaining the VLOOKUP function in detail here, please see this link on Microsoft's Web site for more information: <http://office.microsoft.com/en-us/excel/HA010563201033.aspx?pid=CL100570551033>.

USING AN ADD-IN TO EXCEL TO MINIMIZE COMPLEXITY

The previous explanation on the use of the VLOOKUP function might seem complex to some so there are alternatives to running this match.

Another approach is to review the ActiveData for Excel add-in to Microsoft Excel, available in the online ACFE bookstore that automates the matching process and will even complete a “fuzzy” match. Such close matches are based on the difference in the number of characters for the field being reviewed. (In our examples, the address field was the field used for matching.) For fuzzy matching examples, “123 Golf Course Dr.” and “123 Golf Course Drive” would match because there's only a three-character difference between the cells. Another example is “12-34 University Blvd.” and “1234 University

Fear Not the Software

Fraud-fighting with Data Analysis Tools

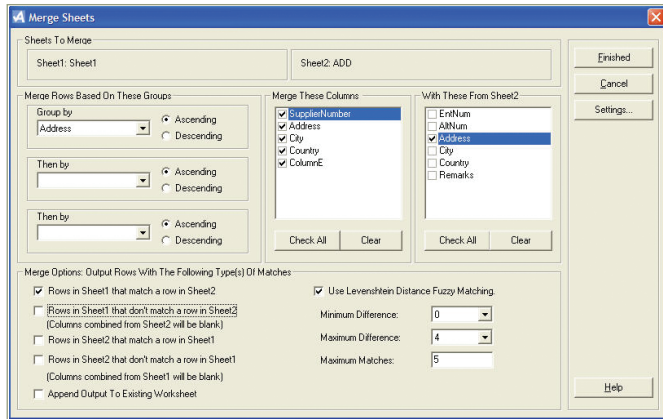


Image 3

Blvd” which shows a two-character difference with the lack of punctuation in the second address listing.

ActiveData for Excel’s Merge Sheets function (from the ActiveData menu item in Excel, select Sheet > Merge Sheets)

uses the Levenshtein “fuzzy” match to identify any matches between two fields that are between 0 to 10 characters in difference. (See image 3, left.)

This is a more ideal method to matching than the first eight characters in the address because the character differences can be anywhere in the fields (such as the beginning, middle, or end).

Whatever the tool, companies owe it to themselves to check their databases for business partners that might not be within the “good graces” of the U.S. Treasury. Better to find it proactively than for it to come as a surprise to the organization. 🔍

Richard B. Lanza, CFE, CPA-CITP, PMP, president of *Audit Software Professionals*, in Lake Hopatcong, N.J., provides audit technology and project management assistance to companies. He focuses much of his time in developing computerized audit and fraud tests. Lanza is the founder of the nonprofit Web site, www.auditsoftware.net. His e-mail address is: rich@auditsoftware.net.
