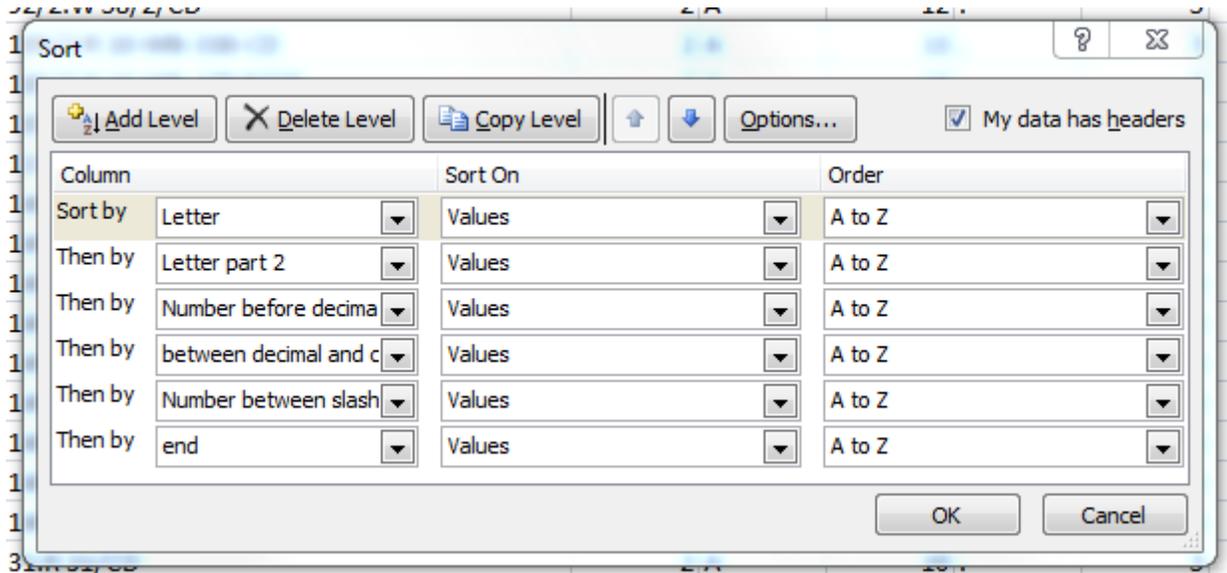


SuDoc Stem Call Number Sorting in Excel

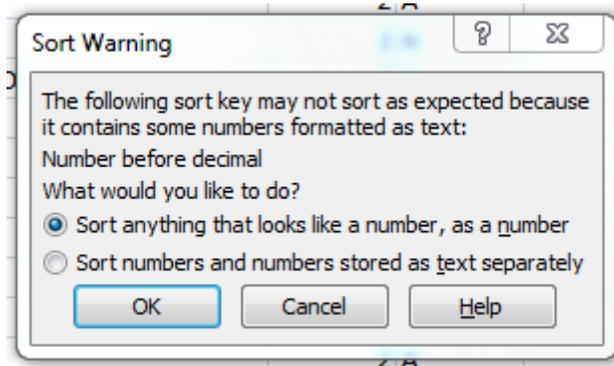
This approach is breaking the SuDoc call number stem into many constituent pieces, so we will add many columns. If you wanted a single formula approach like <http://www.libraries.psu.edu/psul/emsl/callnumbersort.html>, that formula would probably be about a page long; I think this is easier.

Throughout, I am assuming row 1 is headers, and column A has the call numbers. I'll use the call number in A2 as an example.

1. We will be finding the index of characters in the string and reusing them, so I am creating helper columns to hold those indices. It will make formulas somewhat nicer. The first of these is the initial space, which we will put in column B. Its formula is `=FIND(" ",A2)`
2. Isolate the initial letter in column C. The formula is `=LEFT(A2, B2)`.
3. The next index we want is the second space, if there is one. (To handle things like C2 CS). This doesn't always happen, so we'll use an index of 999 if it doesn't. Its formula is `=IFERROR(FIND(" ",A2,FIND(" ",A2)+1),999)`
4. Now we want to isolate the second part of the letter piece, using "." as a placeholder for call numbers without this group. I've put this in column E, with the heading "Letter part 2". Its formula is `=IF(D2-B2 > 3,".",MID(A2,B2,D2-B2))`
5. The next step gives indices for the decimal point, colon, and slash, again using an index of 999 for nonexistent decimal points, colons, and slashes. Column F is the decimal point with a formula of `=IFERROR(FIND(".",A2),999)`
Column G is the colon with a formula of `=IFERROR(FIND(":",A2),999)`
Column H is the slash with a formula of `=IFERROR(FIND("/",A2),999)`
Note that for entries with multiple decimals, colons and/or slashes, the index is the leftmost one.
6. Now we will extract the number before the decimal point. It is in column I with a formula of `=IF(E2=".", IF(F2=999, MID(A2,B2+1,MIN(D2,G2,H2)-B2-1), MID(A2,B2+1,F2-B2-1)), IF(F2=999, MID(A2,D2+1,MIN(G2,H2)-D2-1), MID(A2,D2+1,F2-D2-1)))`
7. The next piece to extract is the number between the decimal point and the colon/slash. There are a handful of entries with no decimal point at all, so they are given the value of 0. It is in column J with a formula of `=IF(F2=999,"0",MID(A2,F2+1,MIN(G2,H2)-F2-1))`
8. The next piece to extract is the number (if there is a slash before the colon) between the first slash and the colon. We will again use 0 for entries with no slash before the colon. Its formula is `=IF(G2>H2,MID(A2,H2+1,G2-H2-1),"0")`
9. The last piece is everything after the colon. I've put it in column L with the value `=IF(G2=999,RIGHT(A2,LEN(A2)-H2),RIGHT(A2,LEN(A2)-G2))`
10. Now with all the pieces broken apart, we need to do a big multiple-level sort. Select all the data. Click the "Sort & Filter" button right end of the Home ribbon in Office 2007). Pick the Custom Sort option, and Sort by Letter, then Letter part 2, then Number before decimal, then Number between decimal and colon or slash, then Number between slash and colon, then end.



It will ask you several times how you want to sort numbers formatted as text. Each time, select “Sort anything that looks like a number, as a number”.



Citation

Conley JF, Nolan LA (2013). "SuDoc Stem Call Number Sorting in Excel". Earth and Mineral Sciences Library. Online at <https://scholarsphere.psu.edu/files/x346d423k>. Last accessed February 22, 2016.