

Excel & EAD: A Quicker Means of Encoding Finding Aids

By Donald C. Force (2007)

Updated by Valerie Higgins (2009)

Updated by Matt Carruthers (2010)

In some cases, Microsoft Excel can be used to speed the often repetitive process of marking up finding aids in EAD. The strategies in these manual are most useful for encoding finding aids which are consistently displayed and include fairly uniform folder lists. At Indiana University, this means that Excel is primarily useful for marking up finding aids at the University Archives, where the folder lists tend to be fairly uniform. This contrasts to the Lilly Library, which possesses a wider variety of materials and hence has finding aids that greatly differ in scope, content, and presentation. However, Excel can often be used to encode sections of finding aids at the Lilly library, such as a list of items or names that needed <item> or <persname>, respectively. Generally speaking, the process works best for long lists of names and folders located at the same level, though this is not to say that it cannot be used when lists contain regular shifts between levels.

Furthermore, it should be noted that the following procedures are only for encoding container lists. Finally, the process described below is not definitive, nor does it account for every instance that will present itself in finding aids. It merely covers the basic, most common examples, and provides a base from which to work and build. Future interns are encouraged to experiment and tweak the procedure to make it even more efficient and effective.

The following procedures have been divided into four sections.

- I. Introduction to the Excel and the template
- II. Using the Excel template
- III. Converting Excel text into validated Oxygen text
- IV. Macros
- V. Encoding Multiple Levels

The first section provides a discussion of the components involved in encoding finding aids using Excel. The second section discusses how to encode folder lists that contain uniform dates using Excel. The third section discusses how to convert the text so that it validates in Oxygen. The fourth section provides a very brief discussion about macros. The final section provides instructions for encoding two levels at once.

I: Introduction to Excel and the Template

The power of Excel resides in three of its features. First, its columns allow for the creation of a template enabling most tags to be copied using the quick drag-and-drop feature. Second, its delimit feature that allows for strings of text to be separated into different columns. Third, its ability to use macros to help change and move text at the press of a “hot key” cuts down on the number of key strokes needed to encode text.

Excel cannot do everything. Text copied from Excel directly into Oxygen will not validate because of spacing issues resulting from the different columns. As a result, an intermediate program, such as Word or TextPad, that use macros to reformat the spacing, is also required. It should be noted, macros are not need to accomplish the work but make the process quicker by being able to handle multiple commands at once.

To begin with, let's take an example found in the average folder list at IU's University Archives. This folder list can be at any container level (C02, C03, etc.) and merely includes a list of names with dates after them. So, in the Word document, we may see a group of folder names such as:

Series: Correspondence, 1918-1978

Subseries: Professional, 1930-1975

Downer, Alan, 1964

Faison, Walter C. (Estate Executor), 1974

Farm, 1959-circa 1973

Illinois Wesleyan University, 1963

Imperial Equipment Corporation, undated

In Oxygen, the actual coding should look similar to the this:

```
<c01 level="series">
  <did>
    <unittitle>Correspondence, <unitdate type="inclusive" normal="1918/1978">1918-1978</unitdate></unittitle>
  </did>
</c01>
<c02 level="series">
  <did>
    <unittitle>Professional, <unitdate type="inclusive" normal="1930/1975">1930-1975</unitdate></unittitle>
  </did>
</c02>
<c03 level="file">
  <did>
    <unittitle>Downer, Alan, <unitdate normal="1964">1964</unitdate></unittitle>
  </did>
</c03>
<c03 level="file">
  <did>
    <unittitle>Faison, Walter C. (Estate Executor), <unitdate normal="1974">1974</unitdate></unittitle>
  </did>
</c03>
<c03 level="file">
  <did>
    <unittitle>Farm, <unitdate type="inclusive" normal="1959/1973">1959-circa 1973</unitdate></unittitle>
  </did>
</c03>
<c03 level="file">
  <did>
    <unittitle>Illinois Wesleyan University, <unitdate normal="1963">1963</unitdate></unittitle>
  </did>
</c03>
<c03 level="file">
  <did>
    <unittitle>Imperial Equipment Corporation, <unitdate type="inclusive"
normal="1918/1978">undated</unitdate></unittitle>
  </did>
</c03>
</c02>
</c01>
```

While the coding of a list of three folder names may easily be done by copying and pasting, it becomes quite the tedious process when the list is ten pages long. Using Excel, one can create a template with nearly all the codes already in place. From here, the user has to copy the text from a Word or HTML document, move it into Excel and perform some minor alterations to have the necessary coding.

Using Excel, this will look like this:

A	B	C	D	E	F	G	H	I	J	K	L
<c03 level="file"><did><unittitle>	Folder Name: Level 3	<unitdate normal="	<unitdate type="inclusive" normal="	<unitdate type="inclusive" normal="1919/1978	Normal Date	>	Date 1	</unitdate>	</unittitle>	</did>	</c03>
<c03 level="file"><did><unittitle>	Downer, Alan,	<unitdate normal="	<unitdate type="inclusive" normal="		1961	>	1961	</unitdate>	</unittitle>	</did>	</c03>
<c03 level="file"><did><unittitle>	Faison, Walter C. (Estate Executor),	<unitdate normal="	<unitdate type="inclusive" normal="		1974	>	1974	</unitdate>	</unittitle>	</did>	</c03>
<c03 level="file"><did><unittitle>	Farm,	<unitdate normal="	<unitdate type="inclusive" normal="		1959-1973	>	1959-circa 1973	</unitdate>	</unittitle>	</did>	</c03>
<c03 level="file"><did><unittitle>	Illinois Wesleyan University,	<unitdate normal="	<unitdate type="inclusive" normal="		1963	>	1963	</unitdate>	</unittitle>	</did>	</c03>
<c03 level="file"><did><unittitle>	Imperial Equipment Corporation,	<unitdate normal="	<unitdate type="inclusive" normal="	1919/1978		>	undated	</unitdate>	</unittitle>	</did>	</c03>

Trying to incorporate the series and subseries names and dates makes the process more complicated than what it is worth. I suggest encoding these in Oxygen and then using Excel to start with the folder names, in case, the <c03> level.

Looking at the columns individually, the picture becomes (slightly) less overwhelming...

Column A (<c03 level="file"><did><unittitle>)

A
<c03 level="file"><did><unittitle>
<c03 level="file"><did><unittitle>
<c03 level="file"><did><unittitle>
<c03 level="file"><did><unittitle>
<c03 level="file"><did><unittitle>
<c03 level="file"><did><unittitle>

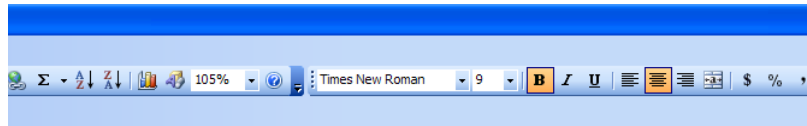
This contains the container level tag, the mandatory <did> tag, and the <unittitle> tag. All three of these tags must be in this order and are dependent on each other when filling out container lists. The only thing that may change with this tag is the level of the container but rather than changing this manually each time, as will be mentioned later, the previous or next levels may easily be placed as other columns.

Column B (Folder Name)

B
Folder Name: Level 3
Downer, Alan,
Faison, Walter C. (Estate Executor),
Farm,
Illinois Wesleyan University,
Imperial Equipment Corporation,

This contains the actual folder name. This is how the names will appear in the finding aid.

Columns C-E (<unitdate normal=", etc....)

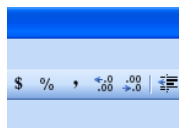


C	D	E	
<unitdate normal="	<unitdate type="inclusive" normal="	<unitdate type="inclusive" normal="1919/1978	Norm
<unitdate normal="			1964
<unitdate normal="			1974
	<unitdate type="inclusive" normal="		1959/1
<unitdate normal="			1963
		<unitdate type="inclusive" normal="1919/1978	

Similar to Column A, these are the tags used for the <unitdate>. There are three columns due to the different types of dates that most typically appear in a finding aid – single year (C), span dates (D), undated folders (E). Dates that contain days are more complex and harder to normalize using Excel. These will be discussed later. In Column E, the dates will be determined per the institution’s policy for handling undated material (typically, they correspond

either to the span dates of the collection or series in which the folder(s) resides). The dates must be manually updated.


Column F (Normal Date)



F	G
978 Normal Date	">
1964	">
1974	">
1959/1973	">
1963	">
978	">

This is the normalized version of the folder’s date. In the fourth row, notice the slash rather than the hyphen. Additionally, the sixth row is blank because it is an “undated” folder and “undated” is not needed in the normalized tag.

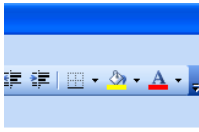
Column G (">:)



Date	G	D
	">	
	">	1964
	">	1974
13	">	1959-c
	">	1963
	">	undate

The closing characters for the <unitdate> tag. These must be placed here because if they are merged with the previous cell (Normalized date), it would be extremely difficult to normalize the dates.

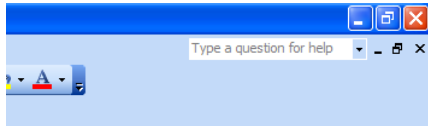
Column H (Date 1)



G	H	I
</unitdate>	Date 1	</unittitle>
</unitdate>	1964	</unittitle>
</unitdate>	1974	</unittitle>
</unitdate>	1959-circa 1973	</unittitle>
</unitdate>	1963	</unittitle>
</unitdate>	undated	</unittitle>

Like Column B (Folder Name), this is the date of the folder and how it will appear in the finding aid. Unlike Column F (Normal Date), the fourth row contains a hyphen and the word “circa” and the sixth row contains the word “undated”.

Columns I-L (</unitdate></unittitle></did></c03>)



I	J	K	L
</unitdate>	</unittitle>	</did>	</c03>
</unitdate>	</unittitle>	</did>	</c03>
</unitdate>	</unittitle>	</did>	</c03>
73 </unitdate>	</unittitle>	</did>	</c03>
</unitdate>	</unittitle>	</did>	</c03>
</unitdate>	</unittitle>	</did>	</c03>

These are the necessary closing tags for the tags that were opened in Columns A and C-E. These are placed in individual columns because additional tags may be placed between them, such as </c02>, <note><p>, or <pysdesc> tags.

While this template only has 12 columns, the template may account for many more tags and attributes. I have created templates that include over twice as many columns to account for different container levels and tags. Typically, a portion of those columns will be hidden but they are there in case they are needed. While 12 columns may be the bare minimum needed to produce a valid document, I have created smaller versions for other purposes, such as encoding a list of names using the <persname> tag or a list that requires the <item> tag. The Excel sheet can become as simple or complicated or as needed.

Now, since we are familiar with the various columns, the question becomes how do we move text that is all on one line into different columns without having to retype everything? The answer is by using a few tools that you probably have not used before in Excel but which become extremely useful for encoding finding aids.

II. Using the Excel Template

To refresh, we are working with the following folder group:

Series: Correspondence, 1918-1978
Subseries: Professional, 1930-1975

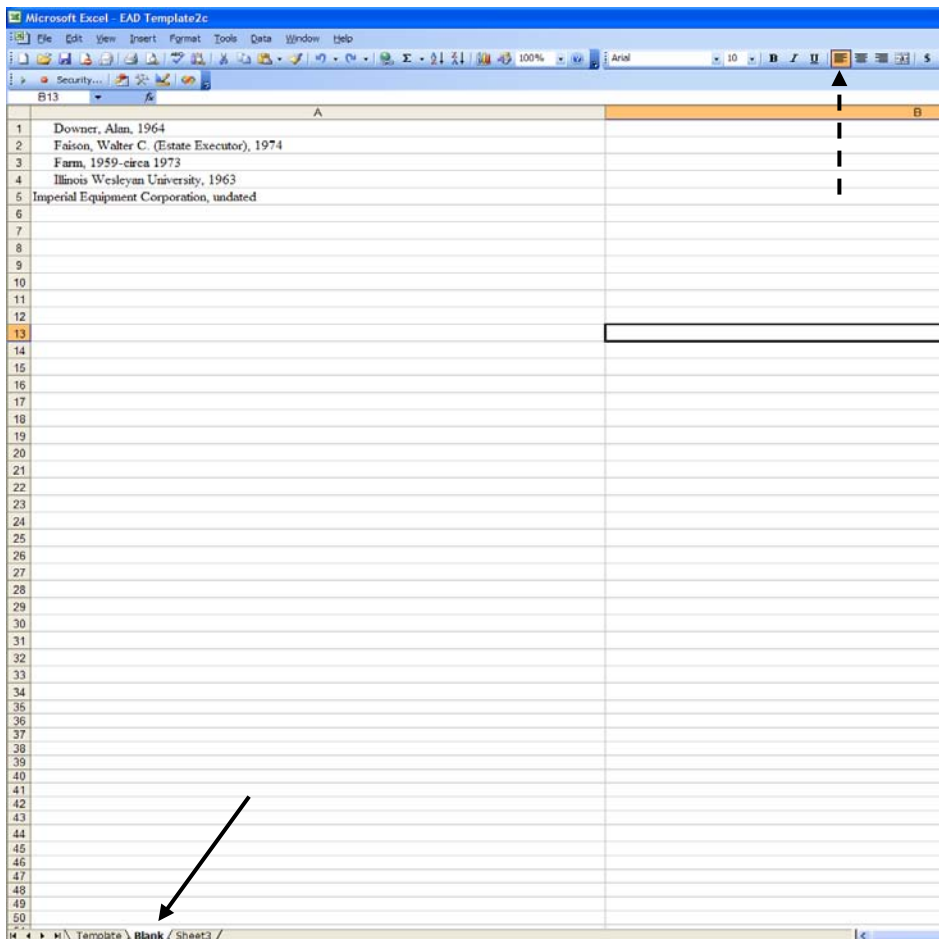
Downer, Alan, 1964
Faison, Walter C. (Estate Executor), 1974
Farm, 1959-circa 1973
Illinois Wesleyan University, 1963
Imperial Equipment Corporation, undated

While the following is a step-by-step process that one can use each time with groups of folder entries, it is highly recommended that macros be created that, once set-up, will do nearly all the steps involved at the press of a couple shortcut keys.

Separating the date from the folder name is one of the most daunting challenges to encode finding aids using Excel. The date needs to be separated and copied so that it may be normalized. Using Excel's "Text to Columns" feature, it is an easy process but involves a few steps.

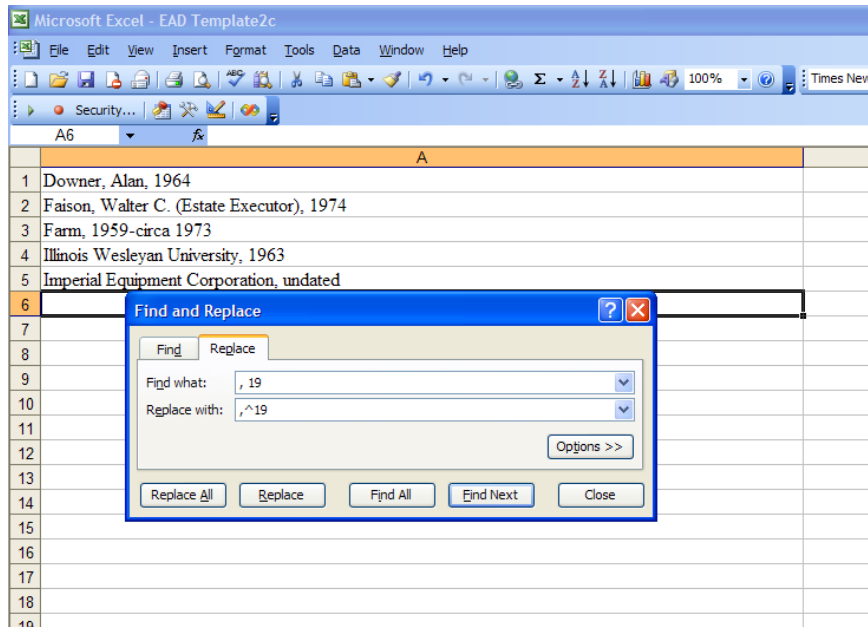
I have always preferred to do the first couple steps in a blank worksheet because it is easier to see if any mistakes occur during the first few steps, but I do not see why these steps cannot be incorporated using the template.

First, copy the entries you want to encode into a blank worksheet (solid arrow), separate from the worksheet that contains the Template. Also, depending on how the Word or HTML document is tabulated, when you paste your entries your text may be indented. This will have no effect on the process, though if you wish everything to be flush left, highlight those entries you want aligned, and click the "Align Left" button on the toolbar (dashed arrow). If you are using Excel 2007, select the "Home" tab and click the "Align Text Left" button located in the "Alignment" group.

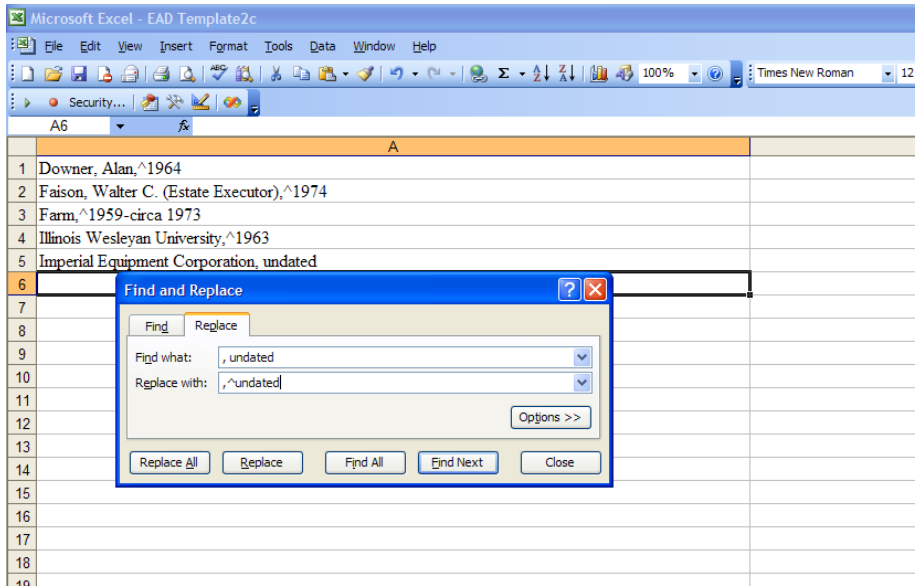


To separate the dates from the text, we need to insert a character that will serve as a temporary placeholder and serve as a marker for Excel. Excel will use this character to determine where to separate the text. Any character may be used, but it is important to use a unique character so other parts of the entry are not erroneously divided into subsequent columns. The following examples use the carrot (^). To accomplish this, we need to do a search and replace.

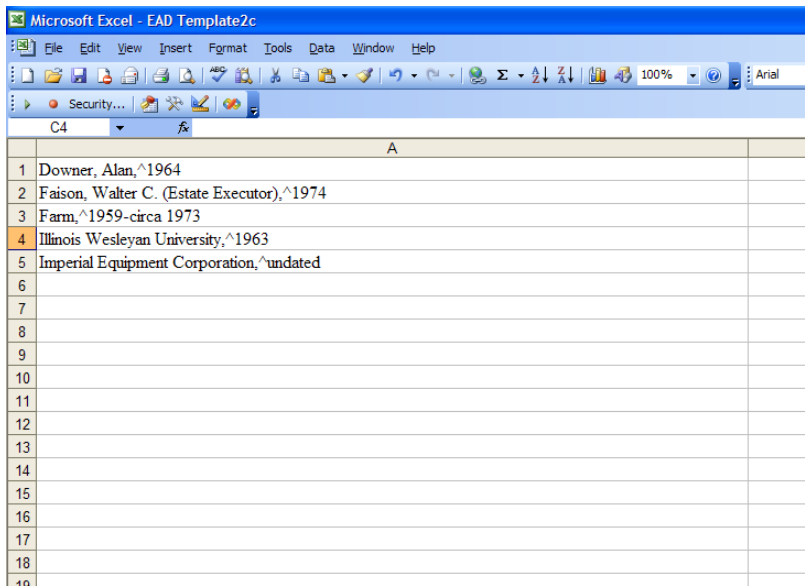
First, do a “Find and Replace” (CTRL + H or, from the main toolbar, Edit > Replace. If you are using Excel 2007, go to the “Editing” group under the “Home” tab, click the “Find & Select” button, and choose “Replace...”). In the **Find what:** space, insert “, 19” and **Replace with:** “,^19” (no quotes).



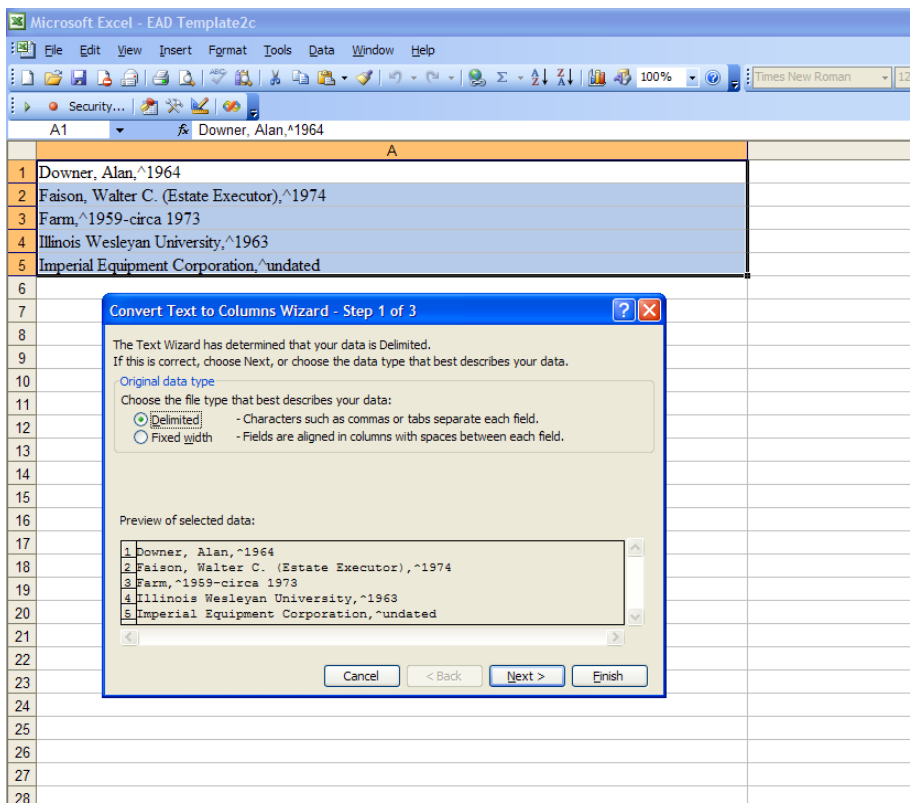
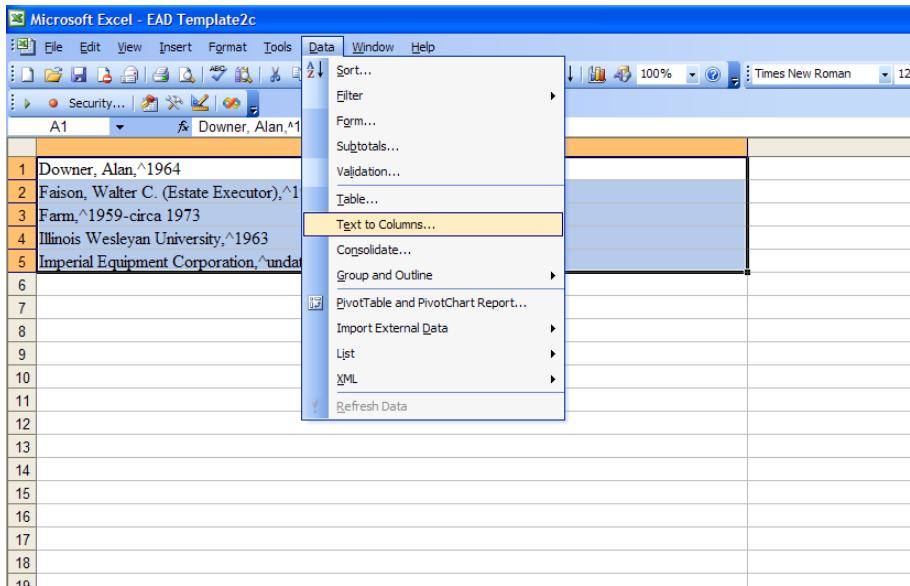
Click **Replace All**. The carrot now appears between the comma and the date. You will need to repeat the process for the undated entry, only in the **Find what:** line you will place “, undated” and **Replace with:** “,^undated”. Also, you will have to repeat the search and replace if you have any dates in the 1800s or 2000s (“, 18” --> “,^18” and “, 20” --> “,^20”, respectively). You do not have to search for each year because you only want to replace the space that follows the comma; the digits that follow that space are only there to ensure Excel replaces the proper space. In other words, if you omitted the numbers, a carrot would appear between the comma and Alan (first entry).



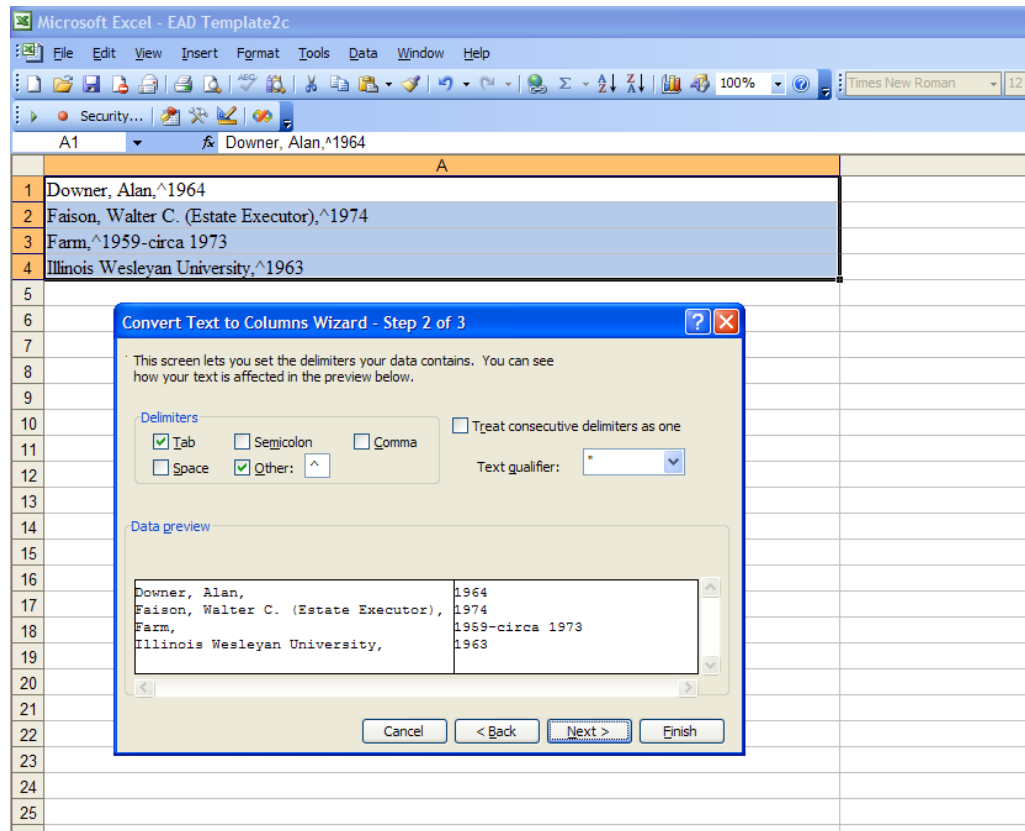
After doing your **Find and Replace**:



Next, highlight your entries (if they are not highlighted then nothing will happen). In order to separate the years from the rest of the text, we will use the **Text to Columns** feature. From the main tool bar, under the **Data** menu, select **Text to Columns**. If you are using Excel 2007, go to the “Data Tools” group under the “Data” tab and click the “Text to Columns” button.

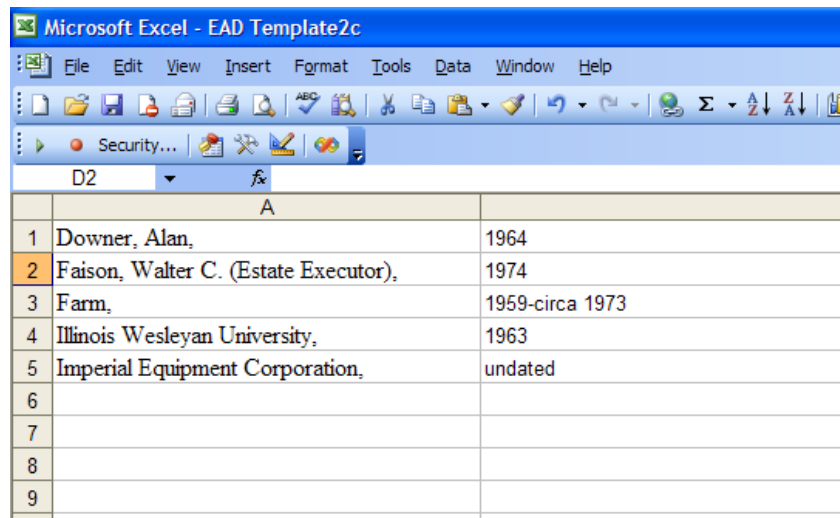


Make sure **Delimited** is selected. (Use **Fixed with** if all your entries are the same length. Using the preview section, you will be able to select where you want the divide to occur.) Click **Next**.

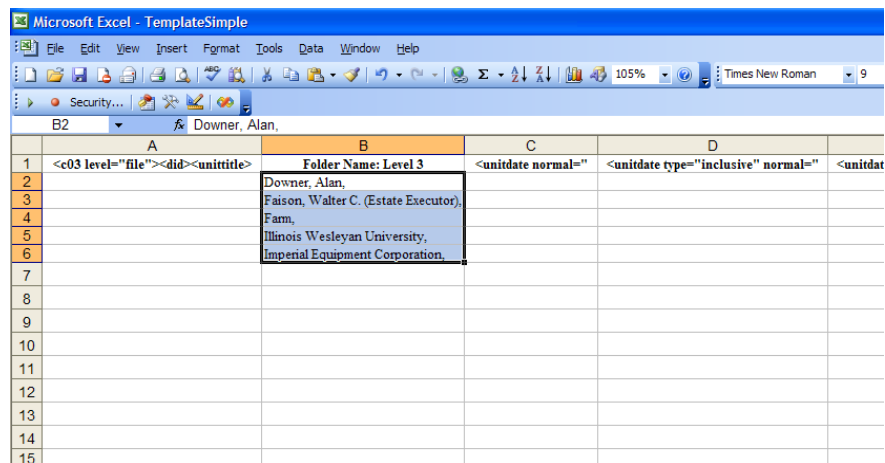


On this screen, make sure **Other** is selected and that in the box you place your carrot or character you have used. To be safe, it is best to unselect **Tab**. Nothing should happen if it is selected, but if you do have a tab (typically, five consecutive spaces) in your text, then Excel will separate the text at this point as well as at the carrot. The **Data preview** section will show you what is going to be separated into a new column. Select **Finish**.

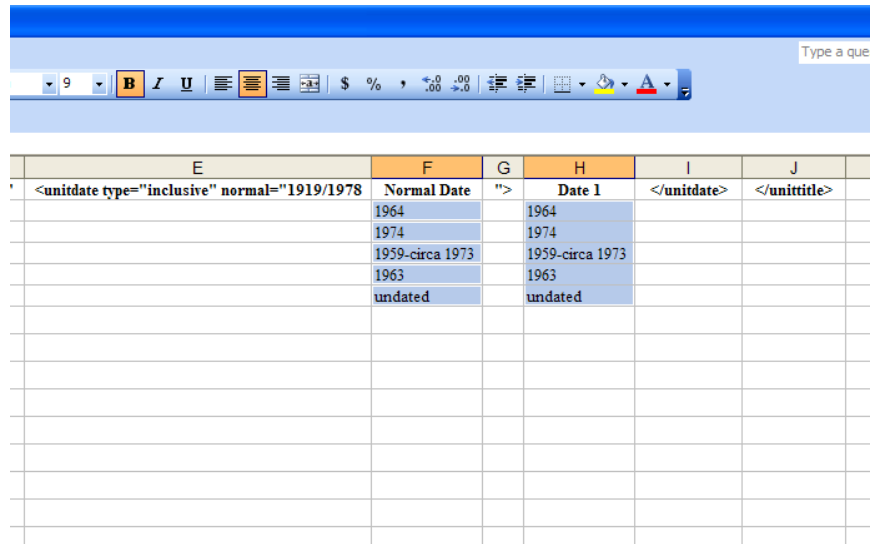
The dates have moved to the next column.



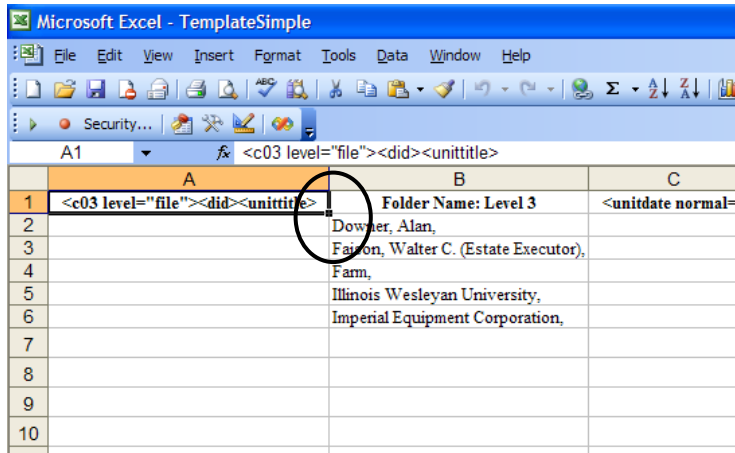
Now copy and paste the two columns into their appropriate columns into the worksheet with the template.



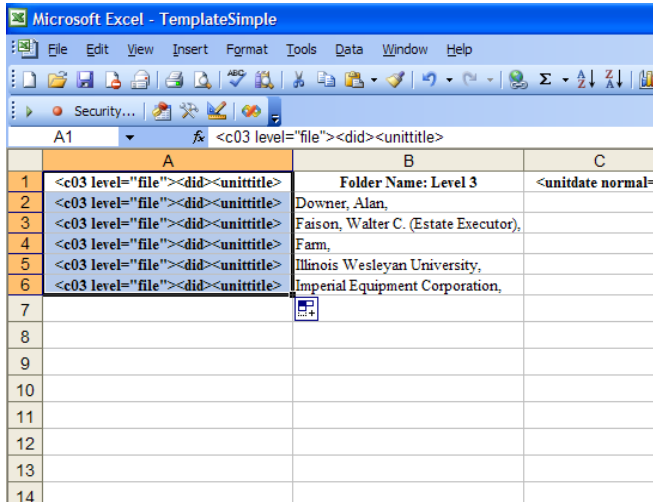
The date column will need to be copied twice. Once for the date as it will appear in the finding aid and once for its normalized version. Since the tagging for the dates is a bit more complex than adding the other tags, we'll deal with the dates last.



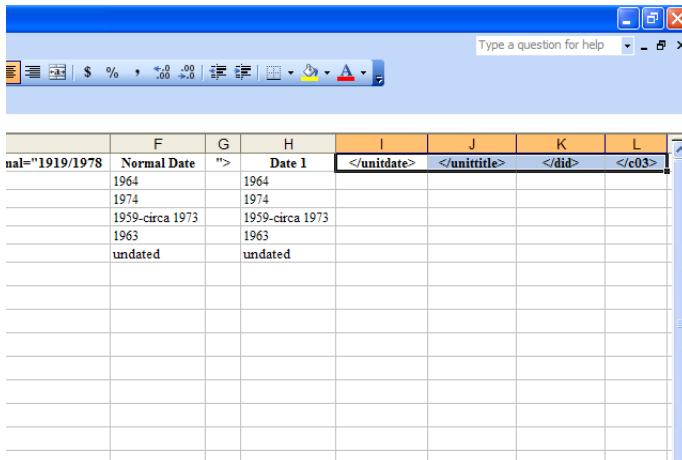
From here, you can click and drag the appropriate column tags using Excel's click-and-drag copy feature. To do this, click the cell you want to copy, left-click and hold on the little black box located in the bottom-right corner of the cell (circled). The cursor will become a solid plus sign (not shown).



By clicking and dragging, you should arrive at:



You can also do multiple columns at once if they are adjacent to one another. In the following example, I have highlighted columns I, J, K, and L and copied their contents in one quick motion.



Use the small black box in the lower right-hand corner of the cell in column L to copy the tags.

inclusive" normal="1919/1978. Column F is 'Normal Date', G is '>', H is 'Date 1', and I through L contain XML tags: </unitdate>, </unititle>, </did>, </e03>. A small black box is visible in the bottom right corner of the cell in column L."/>

E	F	G	H	I	J	K	L
=" <i>inclusive</i> " normal="1919/1978	Normal Date	>	Date 1	</unitdate>	</unititle>	</did>	</e03>
	1964		1964	</unitdate>	</unititle>	</did>	</e03>
	1974		1974	</unitdate>	</unititle>	</did>	</e03>
	1959-circa 1973		1959-circa 1973	</unitdate>	</unititle>	</did>	</e03>
	1963		1963	</unitdate>	</unititle>	</did>	</e03>
	undated		undated	</unitdate>	</unititle>	</did>	</e03>

Now that the basic tags are done, let’s move on to the dates. This may be one of the most challenging parts of this procedure, but it makes normalizing dates a lot easier.

(While the following only addresses single year, span dates, and “undated” entries, it is possible to use Excel to normalize more challenging dates, especially those with a day or days associated with them (e.g. 5 June 1980 or 6-10 June 1980 or June 5, 1980). This type of formatting is a bit trickier using Excel, but it is possible. Future student workers should continue to develop this workflow and add their strategies to this manual.)

Make sure you have copied the dates from the blank sheet into the two appropriate columns -- in this case columns F and H. As noted above, “Date1” (H) will be used as the date for how it appears in the finding aid and “Normal Date” will be the normalized date for the <unitdate> tag.

	B	C	D	E	F	G	H	I	J
itititle>	Folder Name: Level 3	<unitdate normal="	<unitdate type="inclusive" normal="	<unitdate type="inclusive" normal="1919/1978	Normal Date	>	Date 1	</unitdate>	</unititle>
itititle>	Downer, Alan,				1964		1964	</unitdate>	</unititle>
itititle>	Faison, Walter C. (Estate Executor),				1974		1974	</unitdate>	</unititle>
itititle>	Fam,				1959-circa 1973		1959-circa 1973	</unitdate>	</unititle>
itititle>	Illinois Wesleyan University,				1963		1963	</unitdate>	</unititle>
itititle>	Imperial Equipment Corporation,				undated		undated	</unitdate>	</unititle>

Because normalized dates cannot have any text in them, the words “circa” and “undated” will need to be deleted from the normalized column (F). You may either do this by doing a search and replace or manually deleting the text.

	B	C	D	E	F	G	H	I
itititle>	Folder Name: Level 3	<unitdate normal="	<unitdate type="inclusive" normal="	<unitdate type="inclusive" normal="1919/1978	Normal Date	>	Date 1	</unitdate>
itititle>	Downer, Alan,				1964		1964	</unitdate>
itititle>	Faison, Walter C. (Estate Executor),				1974		1974	</unitdate>
itititle>	Fam,				1959-1973		1959-circa 1973	</unitdate>
itititle>	Illinois Wesleyan University,				1963		1963	</unitdate>
itititle>	Imperial Equipment Corporation,						undated	</unitdate>

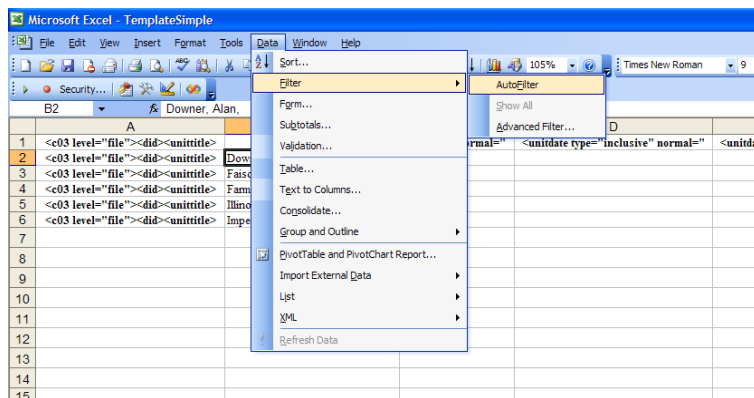
Also, if you have a lot of span dates (e.g. 1959-1973), you need to turn the hyphen into a slash (/) [circled]. This may be done by a simple search and replace. You will be able to create a macro that looks for words such as “circa”, “undated” and the hyphen and replaces them appropriately.

	B	C	D	E	F	G	H	I
Folder Name:	Level 3	<unitdate normal="	<unitdate type="inclusive" normal="	<unitdate type="inclusive" normal="1919/1978	Normal Date	">	Date 1	</unitdate>
Downer, Alan,					1964		1964	</unitdate>
Faison, Walter C. (Estate Executor),					1974		1974	</unitdate>
Farm,					1959-1973		1959-circa 1973	</unitdate>
Illinois Wesleyan University,					1963		1963	</unitdate>
Impenal Equipment Corporation,							undated	</unitdate>

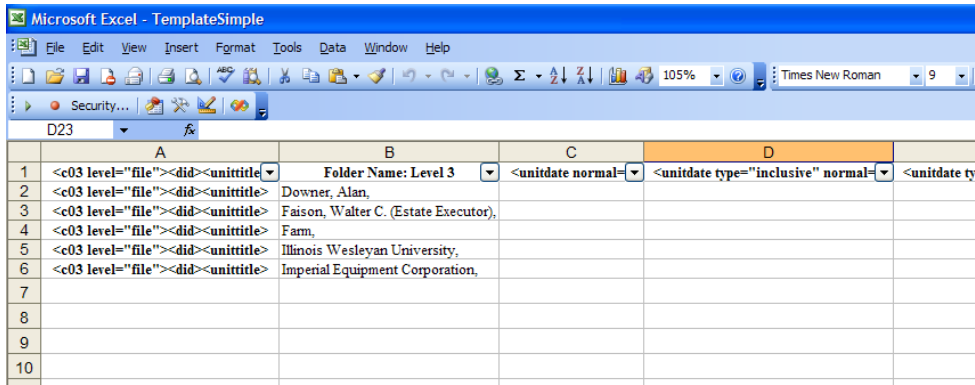
The type of date determines the type of tag and attributes it may or may not require. As you can see from the previous examples, the template has three different types of tags for the dates: single year (C); span dates (M); and entries that have no date, typically identified as either n.d. or undated (E). For column F, note how the normalized dates have already added. Your repository’s policy toward normalizing entries without dates determines these years (typically they are either the span dates of the series or collection). These dates must be changed manually.

There are two ways to fill in the appropriate tags. You can copy and paste the needed tags next to the dates one-by-one or you can filter the dates so you can use the click-and-drag feature. While the former may be easier to use in cases where you only have a few entries (such as like in the example images provided), when you have a large list of entries that intermix single years with span dates and “undated” entries, filtering the columns is faster.

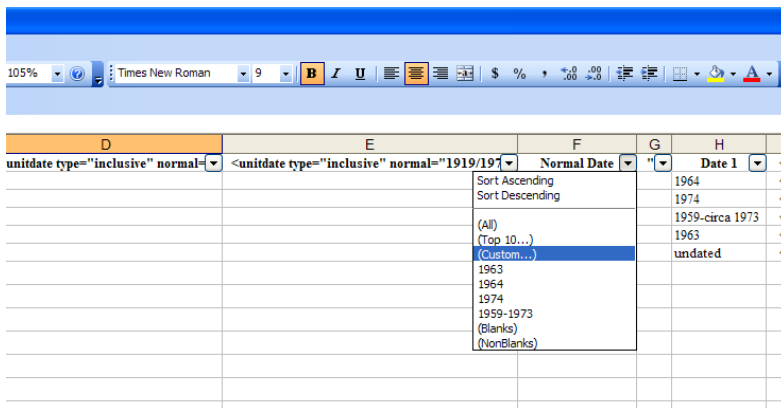
In order to use the filtering system, from the main menu, select **Data > Filter > AutoFilter**. If you are using Excel 2007, highlight the column you wish to filter, go to the “Sort & Filter” group under the “Data” tab, and click the “Filter” button.



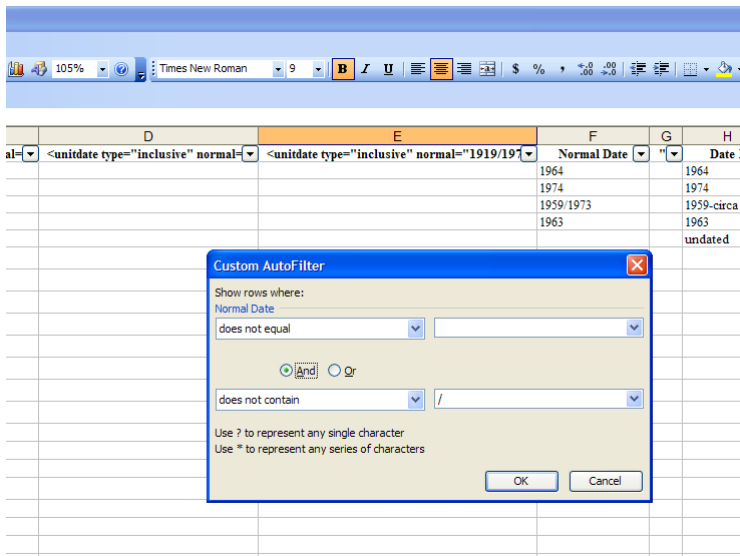
Your first row will now appear with small gray boxes with arrows in them, which, when clicked, open a drop-down menu. A note about the filtering feature: It is best to re-filter the columns (turn it off and then turn it back on) any time you add or remove text from the columns. Chances are, if you perform the following filters to the data and the wrong text appears, it is because the AutoFilter feature needs to be reset.



To filter the data so you only view the entries with single years, you need to remove all the other data from sight. First, click the gray arrow and then select (**Custom...**). If you are using Excel 2007, click the gray arrow and go to “Date Filters.” In the submenu that appears, select “Custom Filter...”



A new window will appear. In the first drop-down menu select “**does not equal**” and then leave the next drop-down box blank. This will account for the “undated” line. Make sure “And” is selected. In the second drop-down menu, select “**does not contain**” and in the adjacent box place a slash (/). This will account for the span dates.



Select **OK**.

Your data should now look like the following:

	A	B	C	D	E	F
1	<c03 level="file"><did><unititle>	Folder Name: Level 3	<unitdate normal=	<unitdate type="inclusive" normal=	<unitdate type="inclusive" normal="1919/197	Normal Date
2	<c03 level="file"><did><unititle>	Downer, Alan,				1964
3	<c03 level="file"><did><unititle>	Faison, Walter C. (Estate Executor),				1974
5	<c03 level="file"><did><unititle>	Illinois Wesleyan University,				1963

The “Farm” and “Imperial Equipment Corporation” are now hiding. (If you are ever in doubt if cells are hiding, you can tell one of two ways: 1) The row numbers on the left-hand side are blue; 2) The filter arrow of the column that you filtered will be blue.) From this point, you can click and drag the tag just needed for the single year; in this case, column **C**.

To re-filter the data so you just see the span dates, go back to the same filter menu. In the second drop-down menu, change the “**does not contain**” to “**contains**”.

D	E	F	G	H
<unitdate type="inclusive" normal=	<unitdate type="inclusive" normal="1919/197	Normal Date	"	Date 1
		1964		1964
		1974		1974
		1959/1973		1959-circa 1
		1963		1963
				undated

Custom AutoFilter

Show rows where:

Normal Date

does not equal

And Or

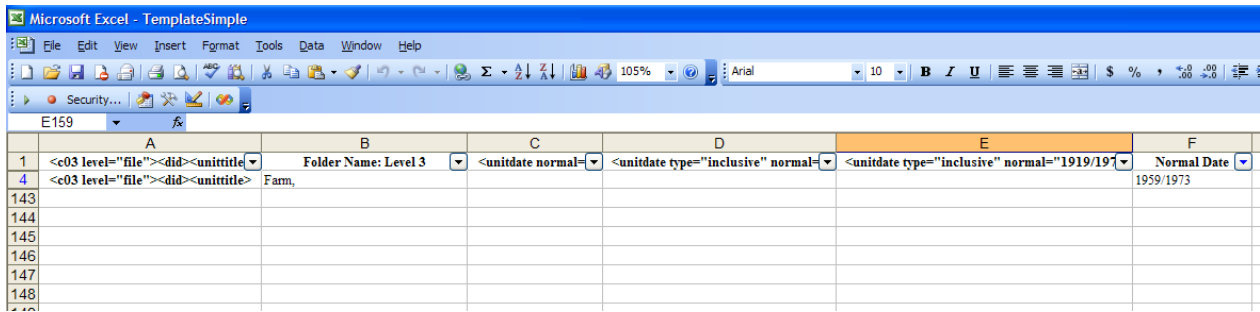
contains /

Use ? to represent any single character
Use * to represent any series of characters

OK Cancel

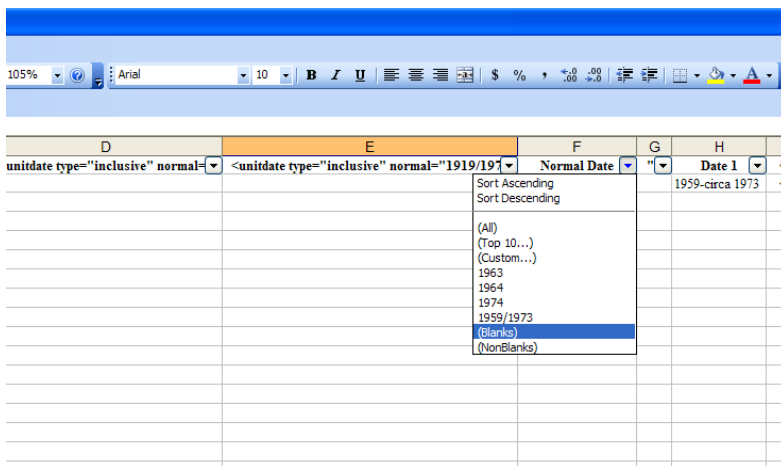
Select **OK**.

You should now see:

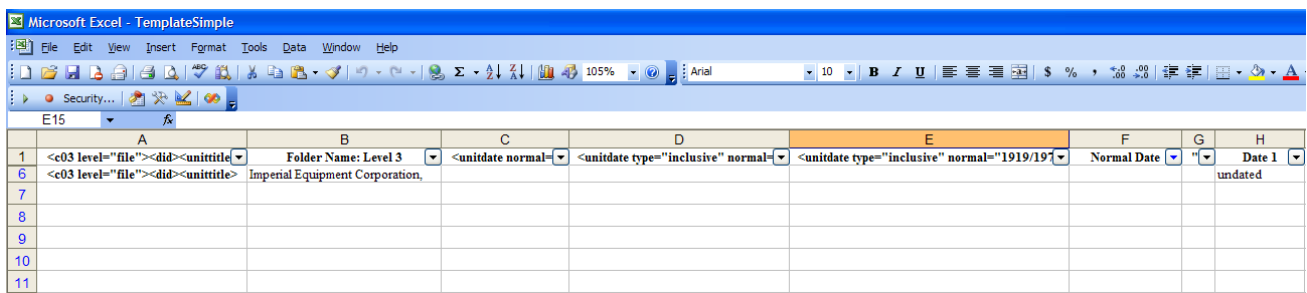


Only the “Farm” entry appears with all the other entries being hidden. From this point, you can click and drag the tag just needed for the span dates; in this case, column **D**.

In order to display just the “undated” entries select “(Blanks)” from the filter menu. In Excel 2007, you can also accomplish this by going to the custom filter, setting the dropdown menu to “does not equal,” and leaving the text box blank.

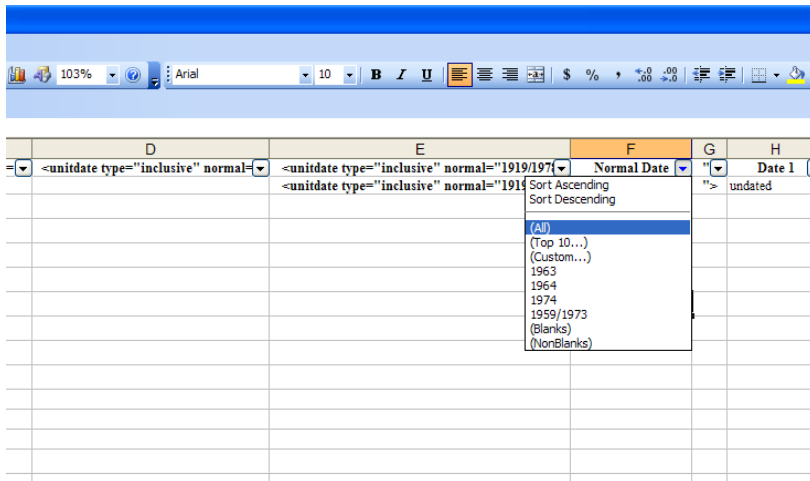


You should now see:

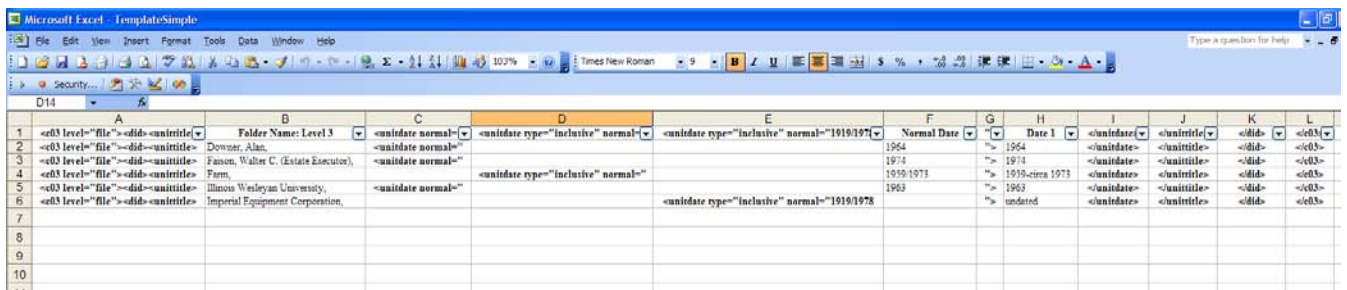


Only the “Imperial Equipment Corporation” entry appears with all the other entries being hidden. From this point, you can click and drag the tag just needed for the span dates; in this case, column **E**.

After inserting all your tags, you can make all your entries reappear by selecting “(All)” from the filter drop-down menu.



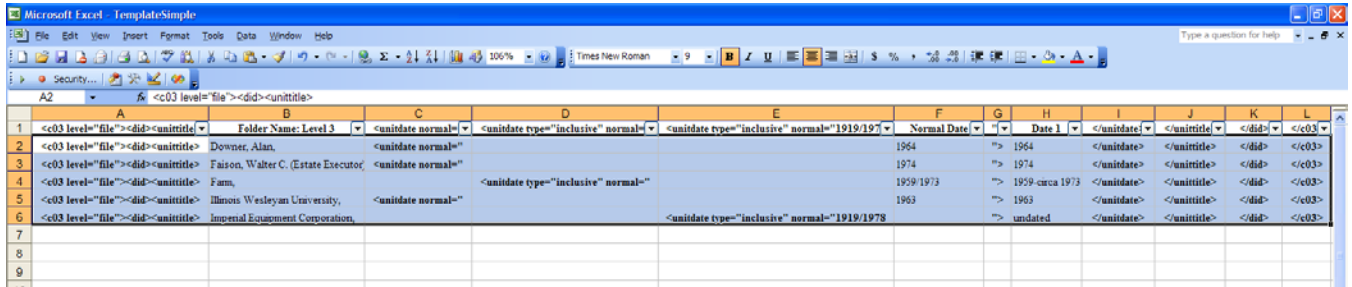
Your final product should look very similar to the following:



You are now ready to move these entries into Word so they may validate in Oxygen.

III. Converting Excel text into validated Oxygen text

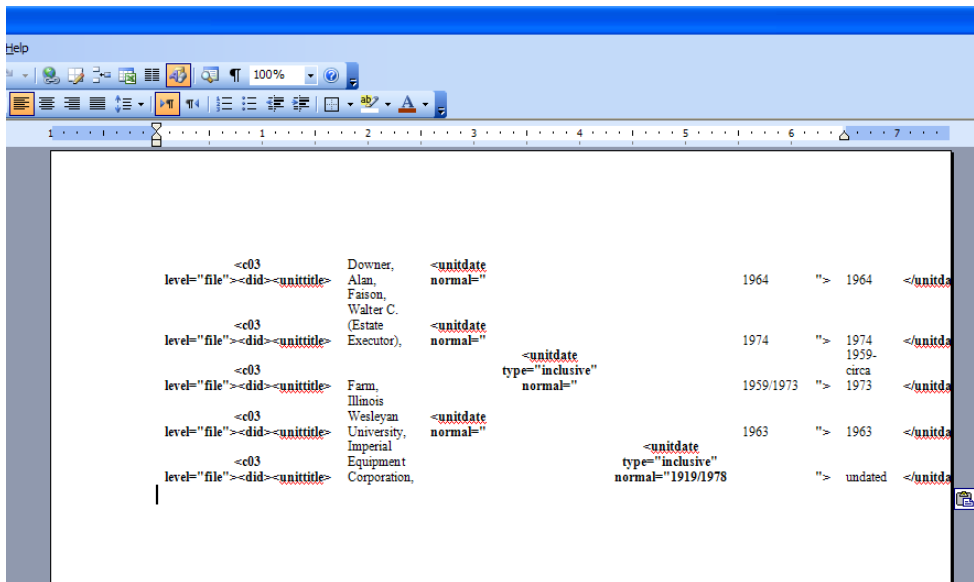
First, copy only the tags and text that should appear in Oxygen.



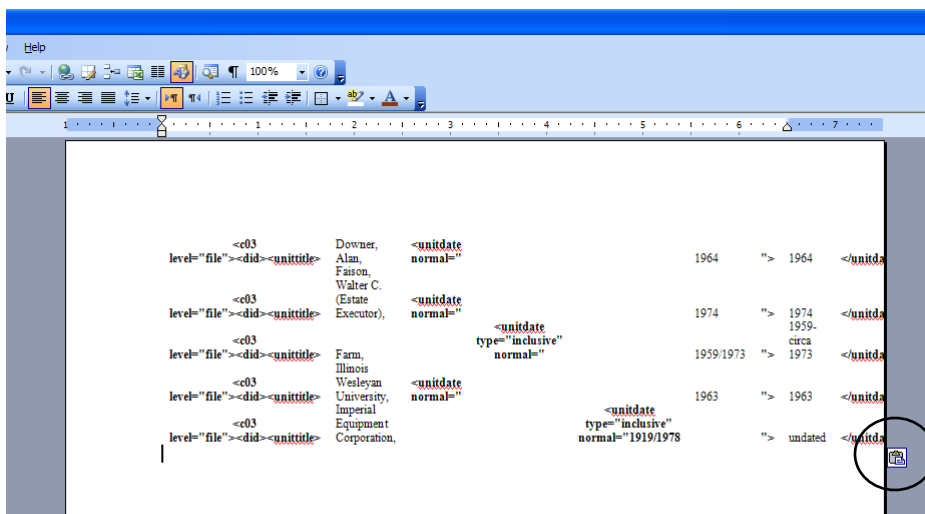
	A	B	C	D	E	F	G	H	I	J	K	L
1	<c03 level="file"><did><unittitle>	Folder Name: Level 3	<unodate normal="	<unodate type="inclusive" normal="	<unodate type="inclusive" normal="1919/1978	Normal Date	>	Date 1	</unodate>	</unittitle>	</did>	</c03>
2	<c03 level="file"><did><unittitle>	Downer, Alan,	<unodate normal="			1964	">	1964	</unodate>	</unittitle>	</did>	</c03>
3	<c03 level="file"><did><unittitle>	Faison, Walter C. (Estate Executor,	<unodate normal="			1974	">	1974	</unodate>	</unittitle>	</did>	</c03>
4	<c03 level="file"><did><unittitle>	Farm,		<unodate type="inclusive" normal="		1959/1973	">	1959-circa 1973	</unodate>	</unittitle>	</did>	</c03>
5	<c03 level="file"><did><unittitle>	Illinois Wesleyan University,	<unodate normal="			1963	">	1963	</unodate>	</unittitle>	</did>	</c03>
6	<c03 level="file"><did><unittitle>	Imperial Equipment Corporation,			<unodate type="inclusive" normal="1919/1978		">	undated	</unodate>	</unittitle>	</did>	</c03>
7												
8												
9												

Open Word and place the copied text into a blank document. For the next couple steps, I recommend creating a macro in Word that will fix the spacing automatically. See the end of this document for some information about macros.

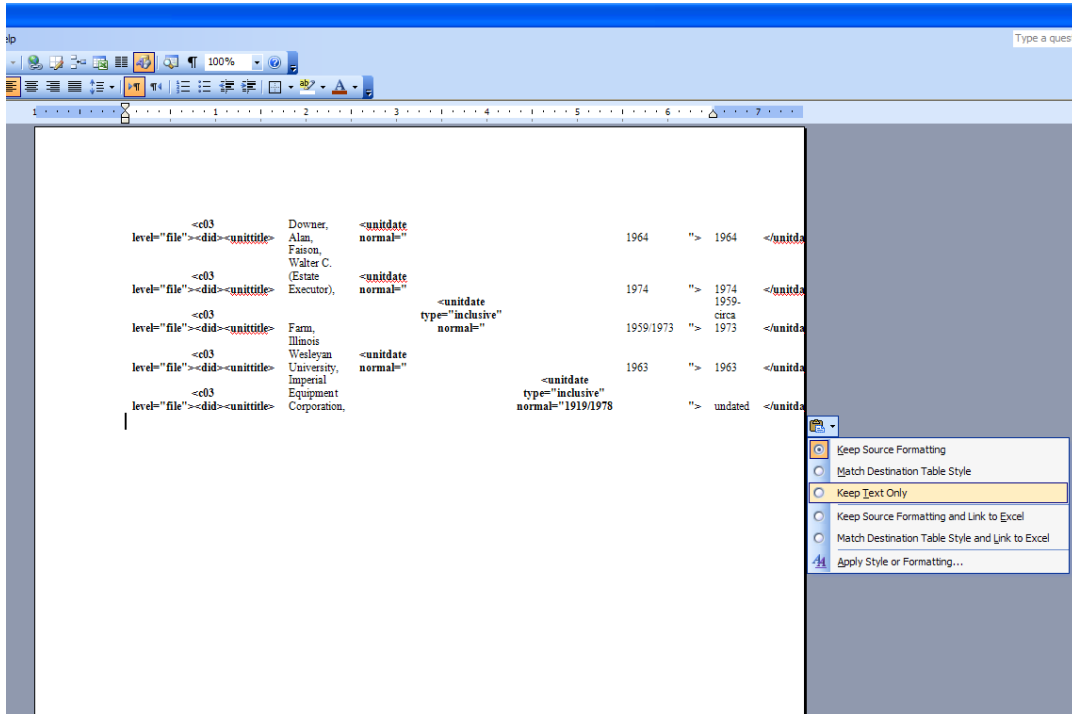
Don't panic, at first it will look very similar to this:



At the end of the text, there will always appear a very small clipboard (circled). Sometimes it may be difficult to locate this small icon because of the spacing of the text, but it should always appear immediately after you paste something into Word; your next keystroke will cause it to disappear. Click this icon and new menu will appear.



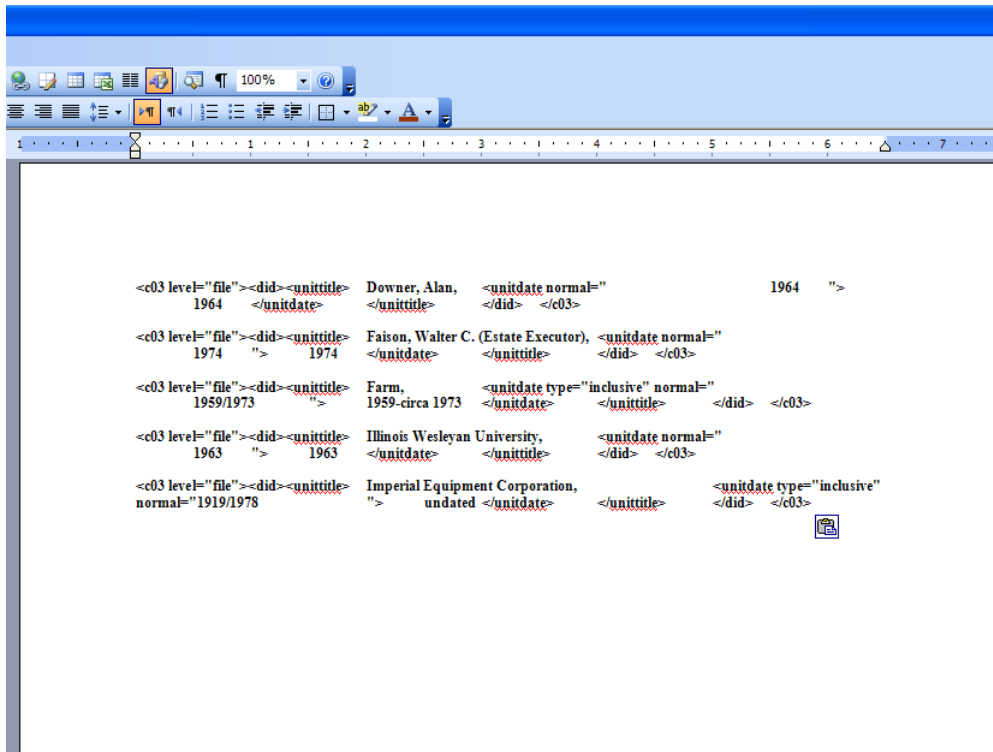
From the menu that appears, select **Keep Text Only**.



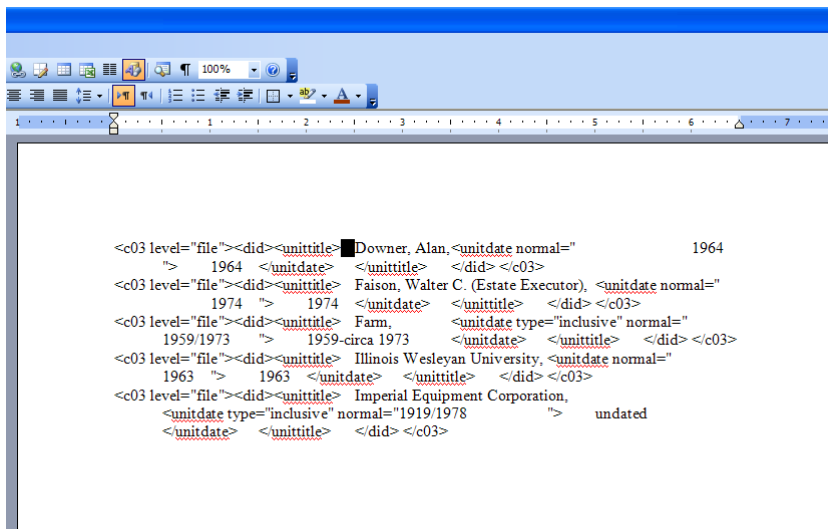
The text should condense though large gaps will remain. By keeping only the text, you have removed the table borders that prevent the text from validating in Oxygen. (Note: for particularly large amounts of text, Microsoft Word will sometimes attempt to paste the text as a table, and the “Keep Text Only” option will not be available. Though this is fairly rare, simply use Notepad instead of Word. Notepad does not try to automatically format text. Once you have pasted the text into Notepad, simply follow the same directions given below.)

Alternatively, if you are using Word 2007, you can click the “Paste” drop-down menu in the “Clipboard” group under the “Home” tab, select “Paste Special,” and select “Unformatted Text” to achieve the same results.

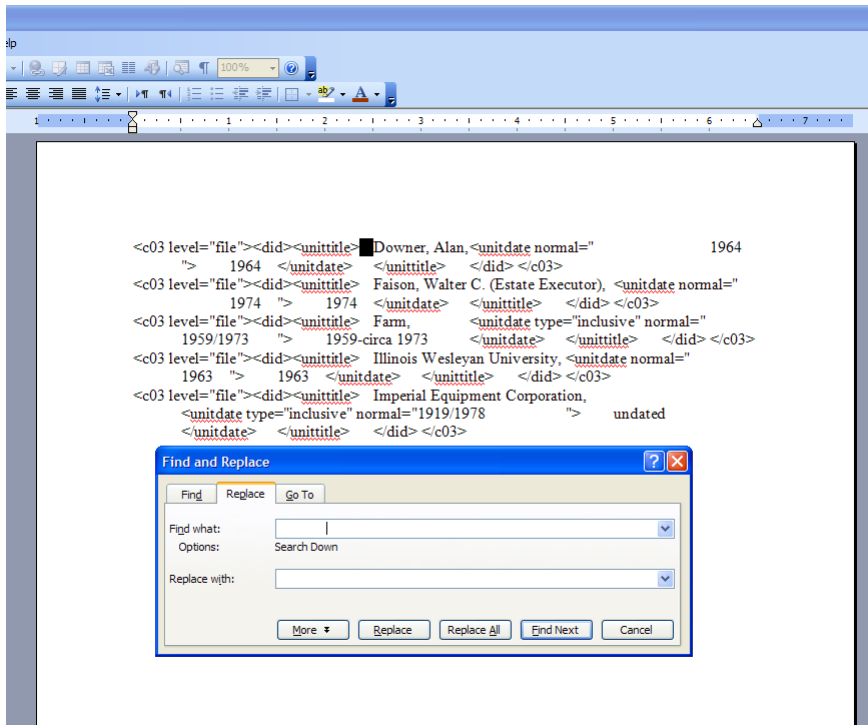
You could place this text into Oxygen and it should validate, but it would not reformat cleanly, not to mention be very discouraging to look at.



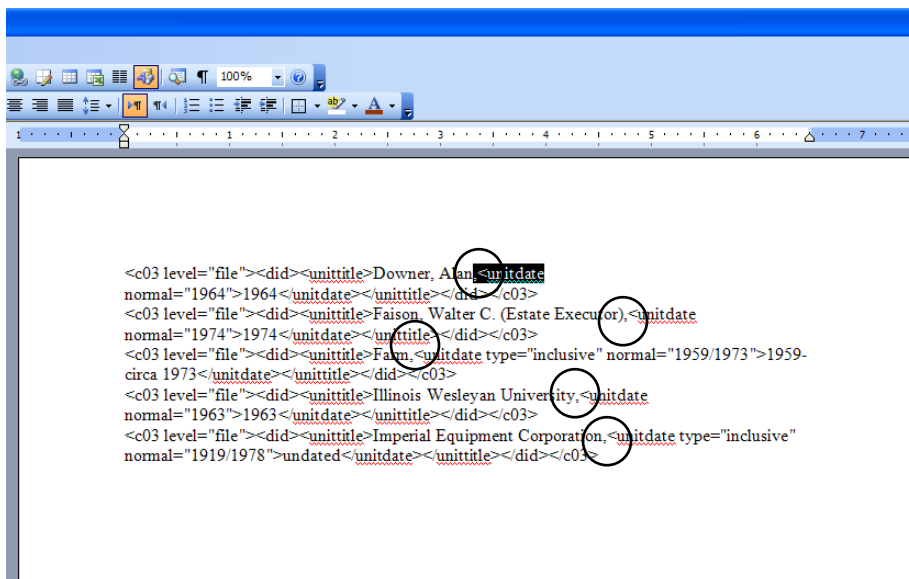
To remove the spacing and to make it appropriate for Oxygen: First, select the spacing between <unittitle> and the first letter of the first entry. Copy this space (CTRL + C).



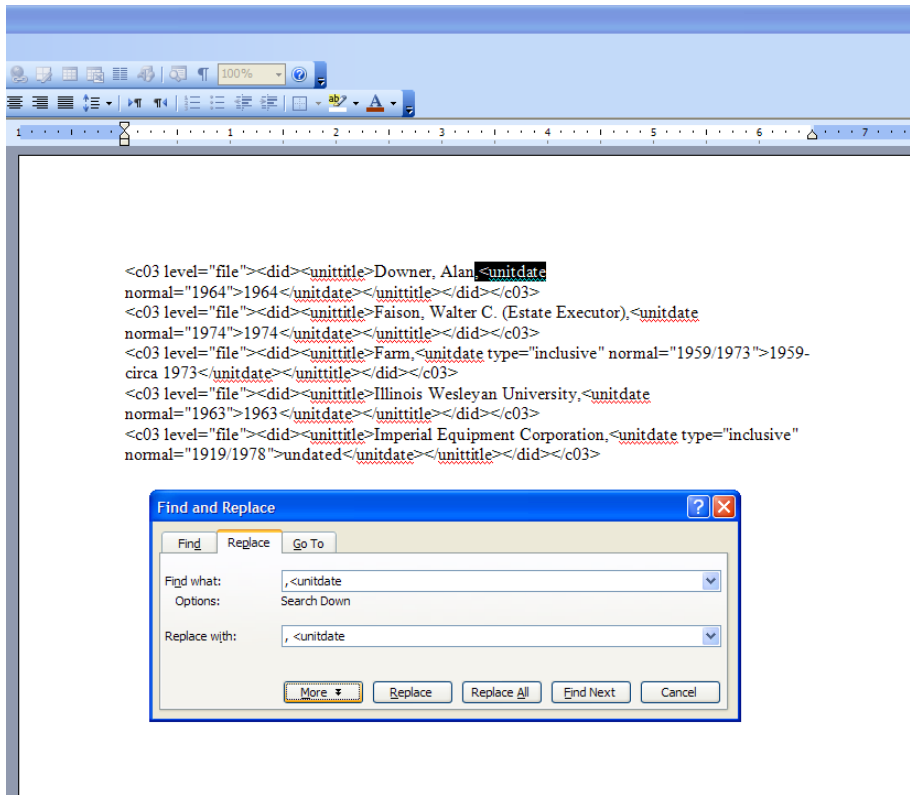
Do a Search and Replace (CTRL + H). Paste the space you have copied into the **Find what:** line (*You have to paste using CTRL + V*). Though nothing will appear, you will notice that the cursor is about five spaces from the left-side of the box. Leave the **Replace with:** box blank. You are replacing the spaces with nothing; this will close up all your spaces. Click **Replace All**. Depending on where your cursor is in the document, Word may ask you if you want to “continue searching the document”, click **Yes** to ensure that you replace all the spacing.



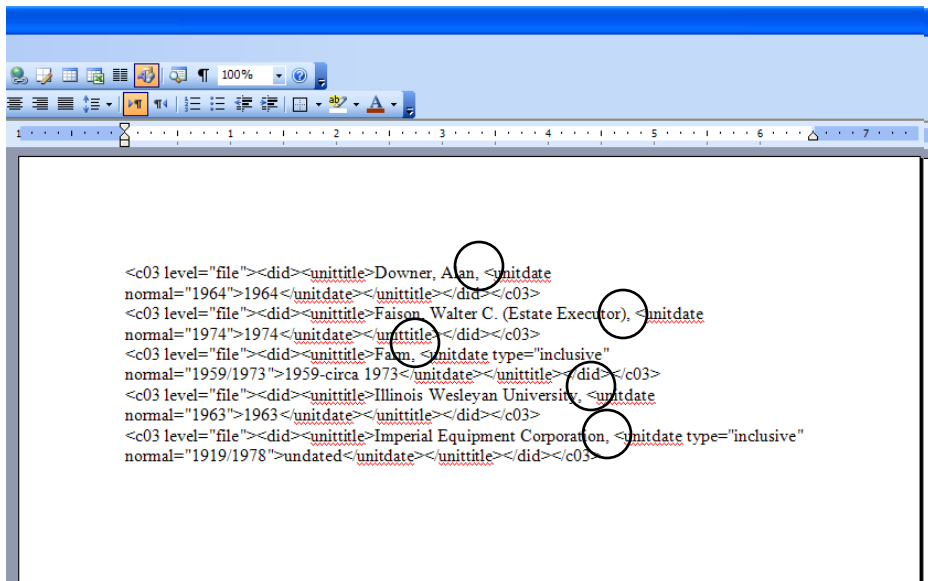
As you can see, all the big spaces have been removed. You will notice that there is also no longer a space between the comma and <unitdate> tag (circles). This is bad and would result in a space not appearing between the comma and date in the finding aid (e.g. Alan,1964). So, you need to perform another copy and paste action to add this space.



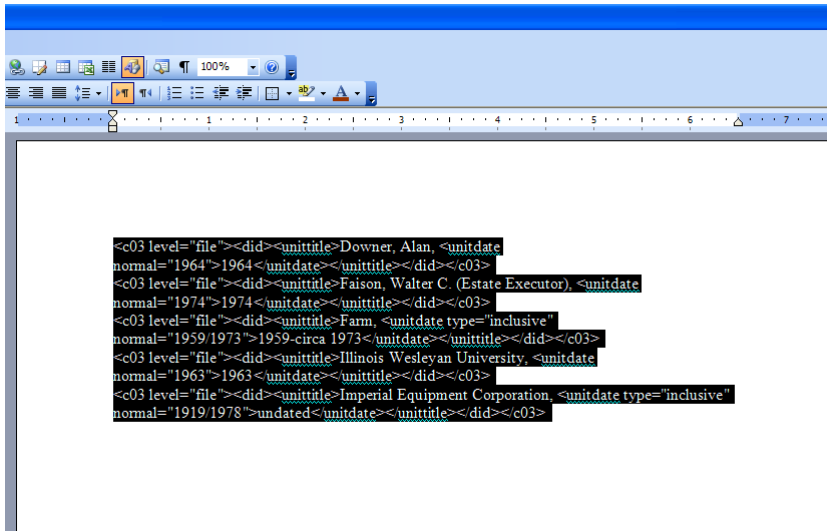
Copy the “,<unitdate” from the text. (You can actually get away with just copying the comma, bracket, and first couple letters (“,<un”) but this example copies the entire word for clarity purposes.) Place the copied text in the “Find what” section and add a space to the copied text in the “Replace With” section.



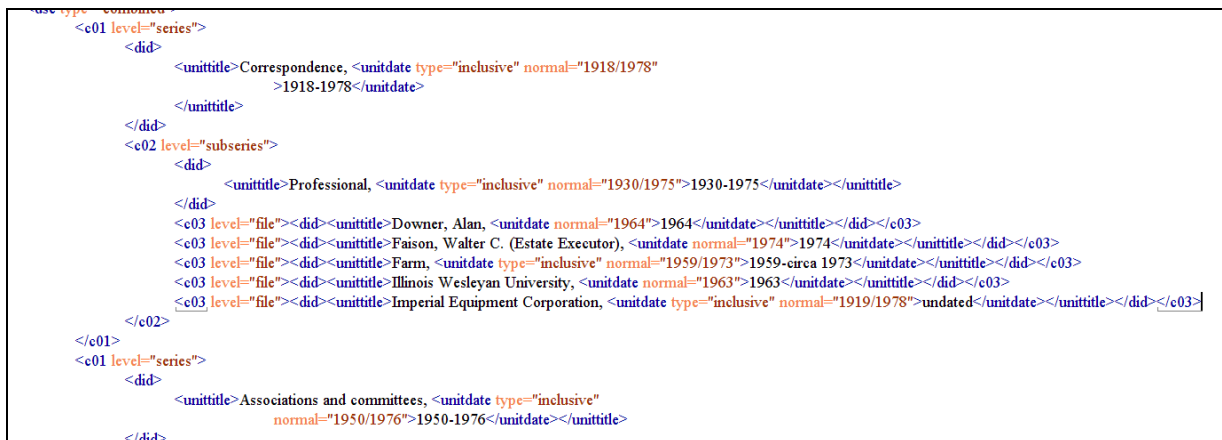
Finally, click “Replace all”. A space will now appear between the comma and the <unitdate> tag (circles).



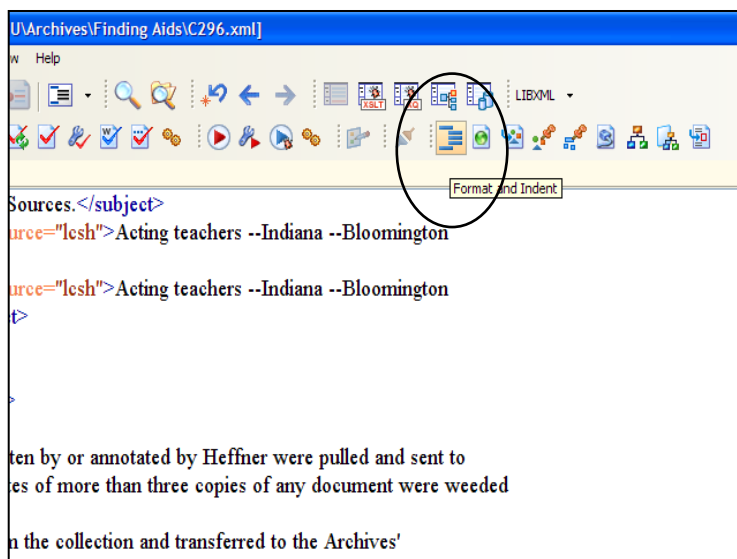
In order to move the text from Word into Oxygen, first, select *all* the text (CTRL + A) and copy (CTRL + C). Select Oxygen.



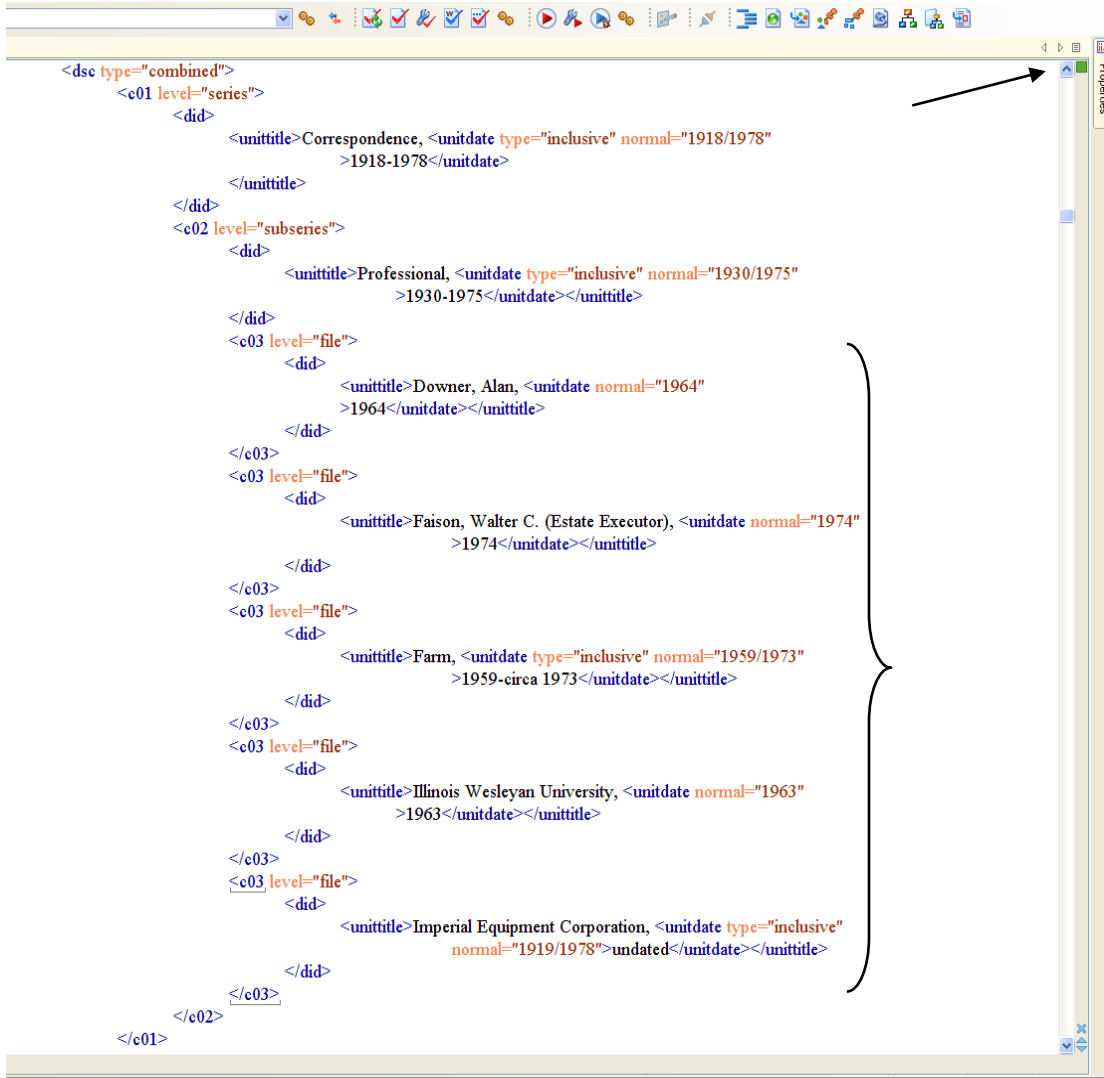
Paste the section into the appropriate place in Oxygen. The document should validate. If any errors were made (extra punctuation, a missed column of tags, etc.), chances are, the document will not validate and red squiggly lines will appear; you will not be able to format the document either.



Click the Format and Indent button (circled) on the menu bar.



Voila! It looks like you have just encoded it the regular way. You can see that it validates by the green square in the top right (arrow).



```
<dsc type="combined">
  <c01 level="series">
    <did>
      <unititle>Correspondence, <unitdate type="inclusive" normal="1918/1978"
        >1918-1978</unitdate>
      </unititle>
    </did>
    <c02 level="subseries">
      <did>
        <unititle>Professional, <unitdate type="inclusive" normal="1930/1975"
          >1930-1975</unitdate></unititle>
      </did>
      <c03 level="file">
        <did>
          <unititle>Downer, Alan, <unitdate normal="1964"
            >1964</unitdate></unititle>
          </did>
        </c03>
        <c03 level="file">
          <did>
            <unititle>Faison, Walter C. (Estate Executor), <unitdate normal="1974"
              >1974</unitdate></unititle>
            </did>
          </c03>
          <c03 level="file">
            <did>
              <unititle>Farm, <unitdate type="inclusive" normal="1959/1973"
                >1959-circa 1973</unitdate></unititle>
            </did>
          </c03>
          <c03 level="file">
            <did>
              <unititle>Illinois Wesleyan University, <unitdate normal="1963"
                >1963</unitdate></unititle>
            </did>
          </c03>
          <c03 level="file">
            <did>
              <unititle>Imperial Equipment Corporation, <unitdate type="inclusive"
                normal="1919/1978">undated</unitdate></unititle>
            </did>
          </c03>
        </c02>
      </c01>
```

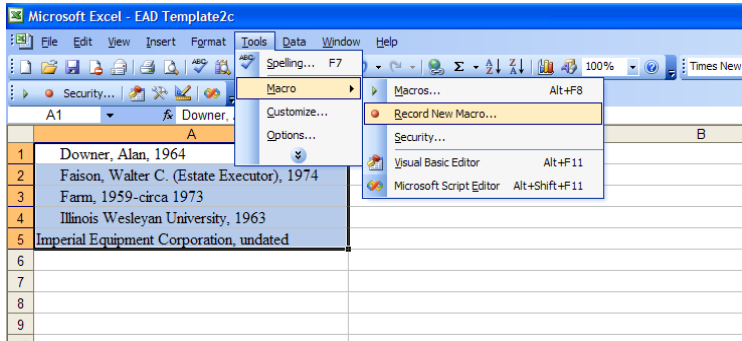
The image shows a screenshot of an XML editor interface. The main area displays a hierarchical XML document structure. The root element is <dsc type="combined">. It contains a <c01 level="series"> element, which in turn contains a <did> element. This <did> element has a <unititle> element with the text "Correspondence, <unitdate type="inclusive" normal="1918/1978">1918-1978</unitdate>". Below this, there is another <did> element, which contains a <unititle> element with the text "Professional, <unitdate type="inclusive" normal="1930/1975">1930-1975</unitdate></unititle>". This <did> element also contains a <c02 level="subseries"> element. The <c02> element contains several <did> elements, each with a <unititle> element. These include: "Downer, Alan, <unitdate normal="1964">1964</unitdate></unititle>", "Faison, Walter C. (Estate Executor), <unitdate normal="1974">1974</unitdate></unititle>", "Farm, <unitdate type="inclusive" normal="1959/1973">1959-circa 1973</unitdate></unititle>", "Illinois Wesleyan University, <unitdate normal="1963">1963</unitdate></unititle>", and "Imperial Equipment Corporation, <unitdate type="inclusive" normal="1919/1978">undated</unitdate></unititle>". A large curly brace on the right side of the editor groups these five <unititle> elements. In the top right corner of the editor, there is a small green square icon, which is pointed to by a black arrow from the text above. The editor's toolbar and a "Properties" panel are also visible.

Despite the copious amount of steps it took to get to the final product, if you use macros, the process is quicker than copying and pasting between Word and Oxygen. The following is a quick crash course in creating macros.

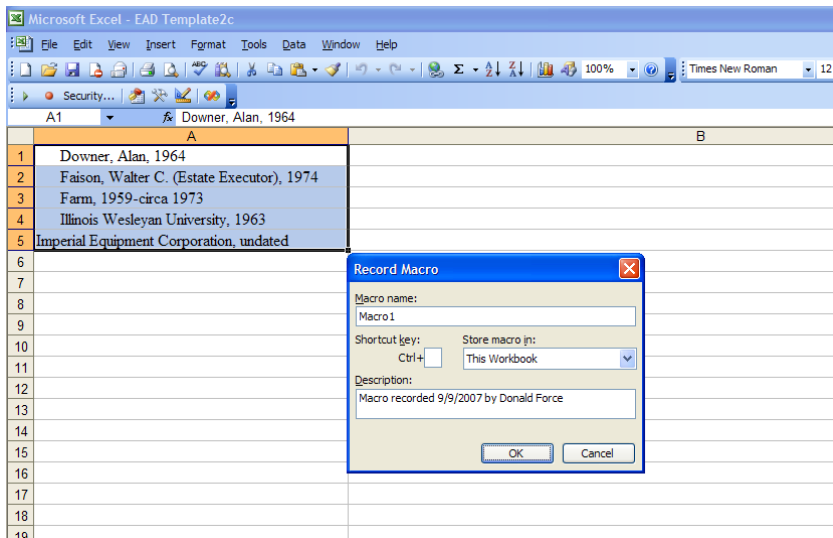
IV. Macros

As noted earlier, one of the strengths of Excel is its ability to use macros. Macros are very small programs that may be used to do repetitive work at the press of a couple keys. In other words, if you know you will be doing the same thing over and over again, you can create a macro, assign it to a shortcut key and then, when it is needed again, press that key and have the process run. While macros can be daunting to learn, Excel makes them fairly easy to set up. Additionally, there is a substantial amount of how-to literature on them and the Books 24X7 database has several good Excel books that discuss macros in further detail.

One of the first macros we need to create for encoding finding aids is how to automate the process of searching and replacing. Macros are located under the Tools menu. We will be starting with **Record New Macro...** but if you already have a couple macros set up and you forget what they do or which keys to press to activate them, you can select **Macros...** (or press Alt+F8).



A new window will appear.



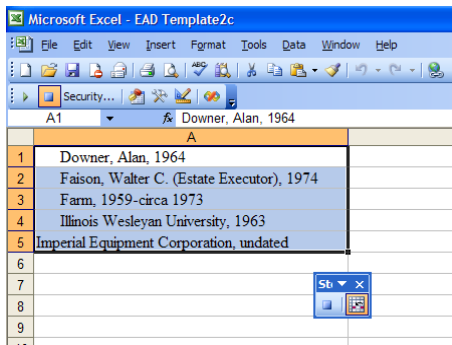
You may give your macro a name, such as “delimiter” or “normalizedate” – something to give you an idea about what it does. Assign it a shortcut key. You will be able to use any non-numeric key and if you select a key that is already used by Excel, the macro command will supersede it. In other words, you

probably do not want to use **C**, **X**, **V**, **Z**, or **A** since these are the keys most frequently used in combination with the CTRL key. The shortcut key is case sensitive so CTRL + SHIFT + **C** is not the same thing as CTRL + **c**. For the Delimit macro, I suggest using CTRL + **d**.

You can store the macro in a couple different places: This Workbook (default), Personal Macro Workbook, or a New Workbook. I have always used the default. Just remember, if you want to use the macro on a different machine, you must save your Excel document to either a flash drive or upload it to server space.

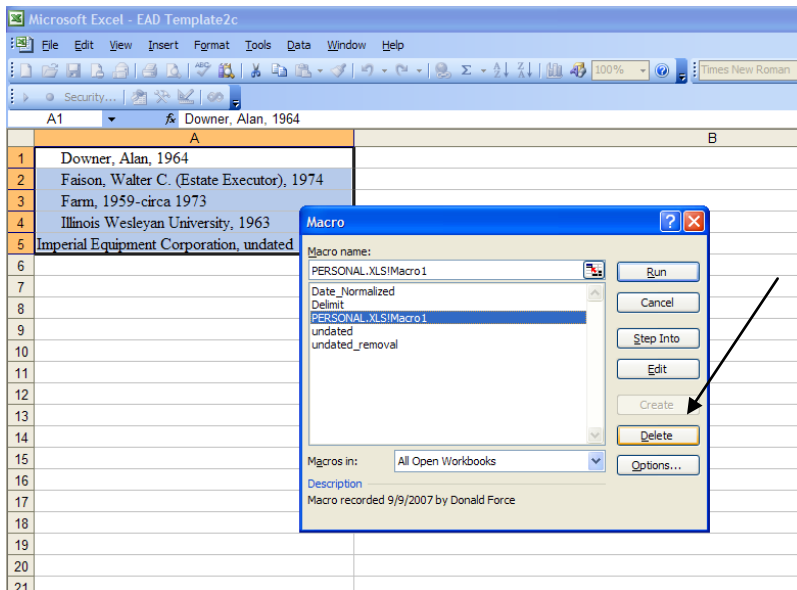
Finally, you can create a description of the macro, briefly describing what it does, when you created it, and anything else that may be important about it.

Once you have done this, select **OK**. This window will close and Excel will revert back to your current working document but there will be another icon/window and toolbar on the screen.

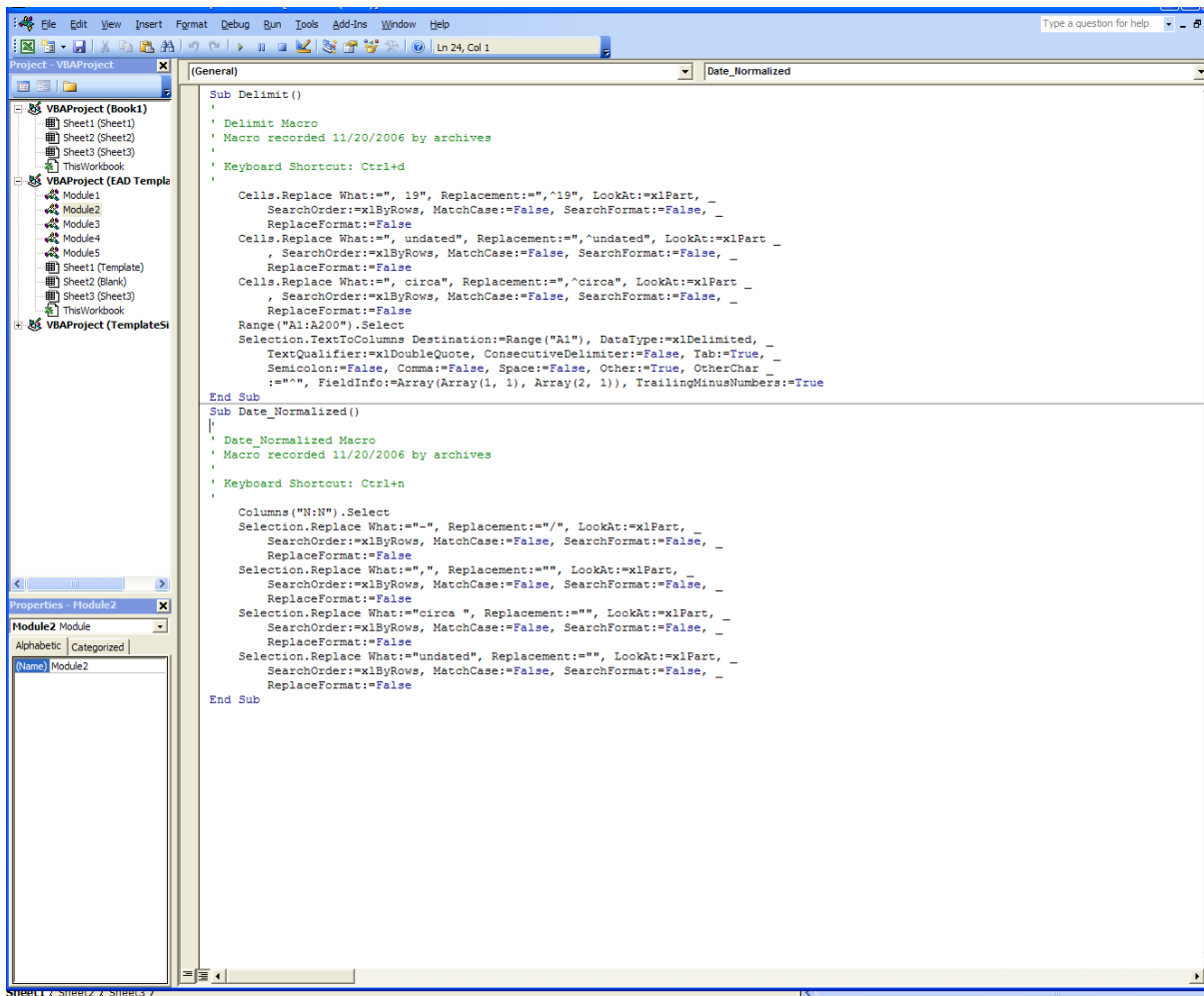


Seeing this icon/window means the macro you are creating is recording. It is recording your keystrokes. You now want to go through the steps outlined above to delimit the year from the folder name to record this as your macro. Start from the point where you do a search and replace for the carrots and continue until you are done delimiting the entries. Once you are done, to stop the macro from recording simply press the square button. To run the macro, once it is created, press the shortcut key you created. *Note: The first time you create the macro you will not have to run it on the text you have in Excel or Word because you will be making the changes as you establish the macro(s).*

A few things to observe. If you make an error in the process of creating a macro, sometimes it is easier to start over again than try to correct it with additional keystrokes. You can easily delete the macro you created by going to **Macros...** from the **Tools** menu, selecting the macro you just created, and then deleting it (arrow).



Also, it may take a few times of trial-and-error to get the macro to do exactly what you want it to do. If you can edit the coding of any macro you create by visiting the **Macros...** window and selecting **Edit**. This will open another Window.



The center sections are what you need to be most concerned with. The green text is general information about the macro (if you were to add a comment, it would appear here). From the end of the green until the line where “End Sub” appears, this information is the macro. Here you can see what exactly Excel is doing when it applies your macro. For example, if we look at the first example a bit closer.

```

Sub Delimit()
'
' Delimit Macro
' Macro recorded 11/20/2006 by archives
'
' Keyboard Shortcut: Ctrl+d
'
Cells.Replace What:=", 19", Replacement:=",^19", LookAt:=xlPart, _
    SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, _
    ReplaceFormat:=False
Cells.Replace What:=", undated", Replacement:=",^undated", LookAt:=xlPart _
    , SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, _
    ReplaceFormat:=False
Cells.Replace What:=", circa", Replacement:=",^circa", LookAt:=xlPart _
    , SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, _
    ReplaceFormat:=False
Range("A1:A200").Select
Selection.TextToColumns Destination:=Range("A1"), DataType:=xlDelimited, _
    TextQualifier:=xlDoubleQuote, ConsecutiveDelimiter:=False, Tab:=True, _
    Semicolon:=False, Comma:=False, Space:=False, Other:=True, OtherChar _
    :="^", FieldInfo:=Array(Array(1, 1), Array(2, 1)), TrailingMinusNumbers:=True
End Sub
Sub Date_Normalized()
'
' Date Normalized Macro

```

The first line: **Cells.Replace what:=", 19", Replacement:= ",^19", ...** this is the search and replace to insert the carrot between the comma and 1964 in the first entry (Downer, Alan, 1964). The coding that follows it is for Excel and how it processes the command. The next two **Cells.Replace** lines are also search and replace lines in order to insert the carrot. The **Range("A1:A200").Select** is informing Excel to only replace the instances of the search and replace that are found in cells A1 through A200. You can also set this up to perform the search and replace on those cells that you have highlighted (to do this, just make sure you highlight the cells before you start the macro). Finally, the **Selection.TextToColumns** is the coding that tells Excel to separate the text at the carrot and move the text that is to the right of the carrot into a new cell.

I found that one of the easiest ways to create macros that do multiple actions at once is by creating the macros in individual steps and then in the editor, you can merge them by cutting and pasting them. For example, you may first want to create a macro that only replaces “ , 19” with “ ,^19”; and then create one that replaces “ , 20” with “ ,^20”, etc. Then, in the editor, these will appear separately, but you can merge them together. Typically, the order of the commands does not matter, but I imagine in some circumstances it does. For example, you would want to make sure all your search and replaces are before the command to delimit the text.

Finally, once a macro is performed, you cannot undo it! Constantly save your work and revert to the “good” version if the macro does not do what you wanted it to do.

V: Encoding Multiple Levels

In most cases, you will only use Excel to encode a container list at a single level. However, when you have a container list consisting of two levels that is displayed in a very regular manner, it is sometimes possible to encode more than one level at once. This process is most useful when the Word or HTML document you are working with is set up in such a way that Excel automatically places the <c02>s into column A and the <c03>s into column B if you use the paste command while cell A1 is highlighted. If they all paste into the same column and you must manually move the <c03>s into column B, it may be simpler to encode it either by hand or in smaller pieces using the basic template.

I have used this process most frequently for lists of <c02>s which sometimes have <c03>s nested within them. I have not found a way to use Excel to encode more than two levels at once, and I have not attempted to use macros when encoding multiple levels since the process of encoding multiple levels is not as uniform as the process of encoding a single level.

For this process, I use a slightly different template than that shown above. While it may initially appear complicated, it is mostly just an expansion of the usual template. On this template, the columns are as follows:

Column A: </c02>

(If the closing tag for the <c02> is placed at the end of the spreadsheet, it will close the <c02> too soon, leaving the <c03>s outside the <c02>s that should contain them. The </c02> tag must be the first column in the spreadsheet.)

Column B: <c02 level="file"><did><unittitle>

Column C: c02 Folder Name

Column D: <c03 level="file"><did><unittitle>

Column E: c03 Folder Name

Column F: <unitdate type="inclusive" normal="

Column G: <unitdate normal="

Column H: c02 Normal Date

Column I: ">

Column J: c02 Display Date

Column K: </unitdate>

Column L: <unitdate type="inclusive" normal="

Column M: <unitdate normal="

Column N: c03 Normal Date

Column O: ">

Column P: c03 Display Date

Column Q: </unitdate>

Column R: </unittitle></did>

Column S: </c03>

The process is very similar to that used to encode a single level except with additional filtering steps. For instance, when using column B to open the <c02>, <did>, and <unittitle> tags, make sure to first filter column C so that only those rows containing c02 folder names are displayed when you are using the drag-and-fill feature. The same must be done for the c03 folder name and when filling in the </unittitle>, </did>, and </c03> closing tags.

The only tag which must be treated differently is the </c02> tag at the beginning of the spreadsheet. When filling in this tag, filter the spreadsheet to display only those rows which contain <c02> folder names and drag the </c02> to fill in all the cells before those containing <c02 level="file"><did><unittitle>. Then, *remove the </c02> from before the very first one listed.* Since there are no previous <c02>s, it is not necessary to close one before opening a new one. Unfilter the spreadsheet, and then *manually insert the final </c02> after the last <c02> and any <c03>s nested within it.* By placing the closing tags for the <c02>s in this way, the previous <c02> is always closed right before a new one opens, which ensures that any <c03>s nested within that previous <c02> will be contained within the <c02>.

Follow the same steps outlined in the sections above to complete the encoding.

