

Creating Metadata in Excel

**Prepared by
Deborah Green**

January 14, 2008

**Digital Collections,
Thomas Cooper Library
University of South Carolina**

Table of Contents

Purpose	3
Before you begin	3
Setting-up the Excel sheet	3
Formatting	4
Item Level Metadata	5
Collection Level Metadata	6
Collections – Fields	6
Import into CONTENTdm	8
Single Item Metadata	8
Compound Object Metadata	11
Metadata for Monographs	13
Conclusion	16

Purpose

The purpose of this document is to give a step-by-step guideline for the creation of a metadata Excel sheet to load with images into CONTENTdm. Guidelines will be provided for single items, compound objects and monographs.

Because metadata is variable, you should consult Western States Dublin Core Metadata Best Practices 2.0, <http://ils.mdacc.tmc.edu/encompass/documents/WSDCMBP.pdf>, for more information.

Before you begin

1. **Discuss with the collection's curator the type and purpose of the collection and the important elements they want accessed by the users of the collection. (This step is the most important).**
2. Decide the collection fields that are important and that you want to make searchable. See SCDL Item-Level Metadata Guidelines at scmemory.org.
3. Decide if you are going to use Library of Congress Subject Headings (LCSH) or Keywords.
4. Discuss any copyright issues.
5. Is the collection comprised of single items, compound objects or is it a monograph?
6. Does the collection have a formal title?
7. Is it part of a larger collection?
8. Is there a web site for the collection?
9. Is there a finding aid?

Once you become familiar with the collection and have decided the collection fields you can go to CONTENTdm and set up the collection on the server. (See Managing Digital Collections with CONTENTdm pages 5-9).

Setting-up the Excel sheet

1. **Open** Excel
2. **Save As:** New name (we usually use metadata somewhere in the title and give it a number for the version of metadata it is).

- Row 1: Type the Collection Field titles in the exact order established in CONTENTdm from **Collections -- Field Properties**. Be sure that File Name is last.

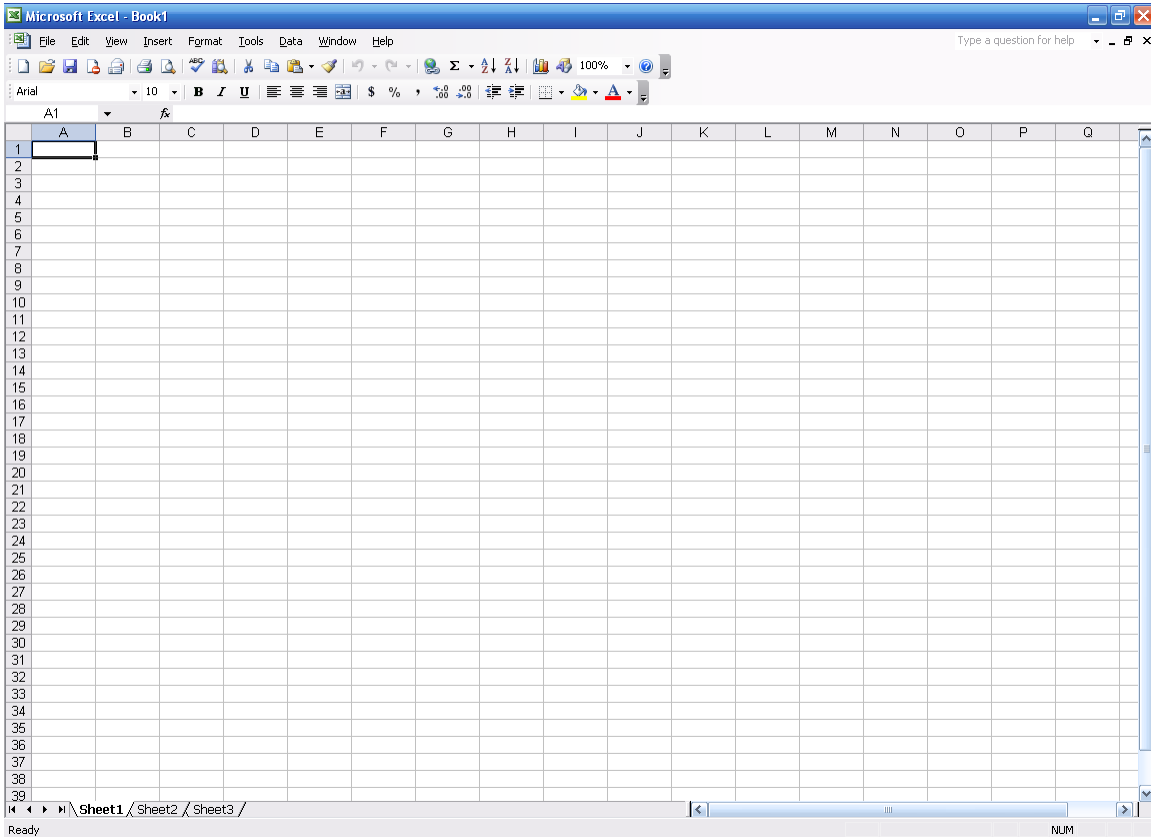
Title	Creator	Date.Original	Source	Subject.LCSH	Description	Category	

- Begin entering in metadata.
- Do not forget that metadata is a work in progress, it can be changed and rearranged and deleted.

Formatting in Excel

A few brief formatting tips, if you are already familiar with excel you can skip this section. Use the screen shot below, Figure 1, for reference.

Figure 1



- To **select the whole sheet** click the box between the "1" and "A"

2. Rows are the numbers and Columns are the Letters
3. To **select a whole row or column** click on the corresponding number or letter
4. To **automatically adjust a row or column** to the height or width or the information place your cursor on the line between either numbers or letters and double click when it becomes bold.
5. To **format information with in a specific cell**, such as date or wrap text, etc., click on the cell, then right click your mouse and choose **Format Cell**.
6. To **change the name of individual sheets, add sheets, or change the tab color**, right click on the tab at the bottom “sheet 1” and choose the desired function.

Item Level Metadata

Item level metadata is specific information that allows a specific item in the collection to be searched and retrieved from the collection. The fields most commonly used for item level metadata are the following:

1. Title (required, searchable)
2. Creator (Newman, I. DeQuincey (Isaiah DeQuincey), 1911-1985)
3. Date.Original (1963-08-12)

Date.Original should be in the YYYY-MM-DD format. To get excel to default to this format, right click on the date column or cell, choose **format cells**, choose **number tab**, then **date**, under Location choose **English (United Kingdom)**, and the 5th in the list is the correct format. Select it and click OK.
4. Source (Folder: Newman, Top., NAACP, Reports, 1962, Aug. 20 - Feb. 5, 1964)
5. Subject.LCSH / Keywords (searchable)
6. Description (searchable)
7. Category (optional, searchable, usually hidden)
8. Resource Identifier

Resource Identifier is the file name without the “.jpg” extension.

9. File Name (required)

A note to know is that CONTENTdm will not load an item without a Title and a File Name. The File Name is used to match the image with the metadata so these two

fields are required. Title should also be the first field in the Excel sheet, and File Name should be the last.

Collection Level Metadata

Collection level metadata is information that applies to the collection as a whole. The data that goes into these fields can be added at the time of loading and is usually the same for every item. Examples are provided in parenthesis.

1. Language
2. Relation.IsPartOf (Joseph E. Winter Photograph Collection)
3. Web site (<http://www.sc.edu/library/digital/collections/broadsides.html>)
4. Publisher.Digital (University of South Carolina)
5. Contributing Institution (South Caroliniana Library)
6. Rights Management (Digital image copyright 2007, The University of South Carolina. All rights reserved. For more information contact The South Caroliniana Library, USC, Columbia, SC 29208.)
7. Type (text, still image, audio)
8. Format (image/jpg)
9. Digitization Specifications (600dpi, 24-bit depth, color, Epson Expression 10000 XL Scanner, Archival Master file is a TIFF)
10. Date.Digital
Use at least the year the item or collection was digitized. Sometimes we use the month and year. (2007-12)
11. Scanner Technicians
12. Metadata Cataloger
13. Metadata Assistants
14. Collection Administrator / Maintenance

Collection Administrator / Maintenance is a hidden field and is the person who is in charge of the digital collection. For example Kate Boyd is our digital librarian so we use her name for this field.

Collections – Fields

Usually, when I create metadata I start with the Item level fields inserted in my Excel sheet. Once my item level is complete I then, **right click on the column** before File Name, and **choose insert column** to add enough spaces to add the collection level metadata into my Excel sheet. (I choose the column before File name because I want the file name to be the last field in my Excel sheet).

Before inserting columns (Item level)

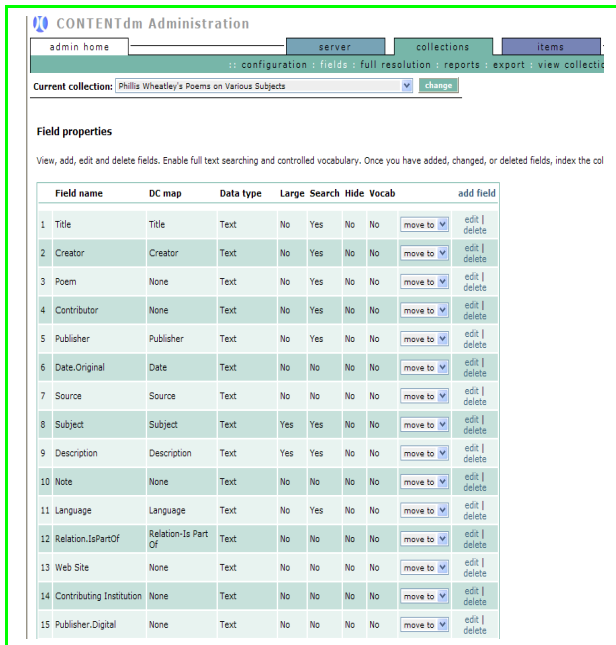
Title	Creator	Date.Original	Source	Subject.LCSH	Description	Category	File Name	

After inserting columns (adding Collection level)

Format	Resource	Digitization Sp	Date.Dig	Scanner Techn	Metadata C	Collection	File Nam

The order for the fields shown above is determined by Field Properties in CONTENTdm. It is very important to have the Excel sheet fields in the same order as the Field Properties in CONTENTdm. Figure 2 shows Field properties in CONTENTdm Administration.

Figure 2



(See Managing Digital Projects with CONTENTdm 7-9)

Single Item Metadata

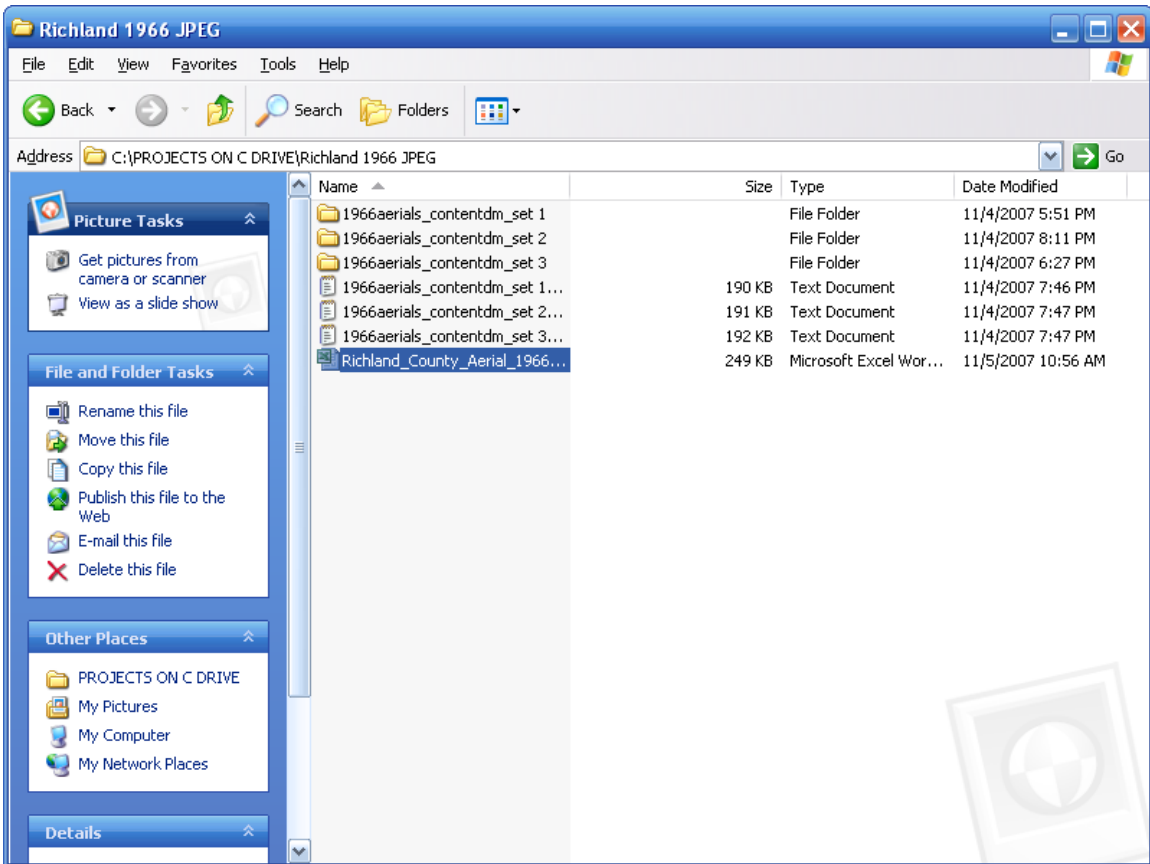
Metadata for single items depends on how many items we have to load. If there are less than 20 items we copy and paste the information from the excel file directly into CONTENTdm as we load the item. For more than 20 items, we have found it more efficient to create a tab-delimited file and load the file with the images.

To create a tab-delimited text file, click **Save As**, choose a destination for the text file, and **rename the file the same as the folder** (especially for a compound object) it is being saved in. Finally, under **Save as type**, choose **Text (Tab-delimited) (*.txt)**.

Because the file name matches the metadata with the image, we can put multiple items in a single sheet and load multiple images at one time. However, 300 single images is the max we can load at a time.

1. In a new excel sheet create separate tabs for loading sets of single objects. Figure 3 shows separate tabs for single objects. I have labeled the single objects tab “1966aerials_contentdm_set1” because I loaded the images in batches. You can also





3. Save the sheet “1966aerials_contentdm_set 1” as a tab-delimited file outside the folder “1966aerials_contentdm_set 1.”
4. You can now load the single items into CONTENTdm.

Import into CONTENTdm

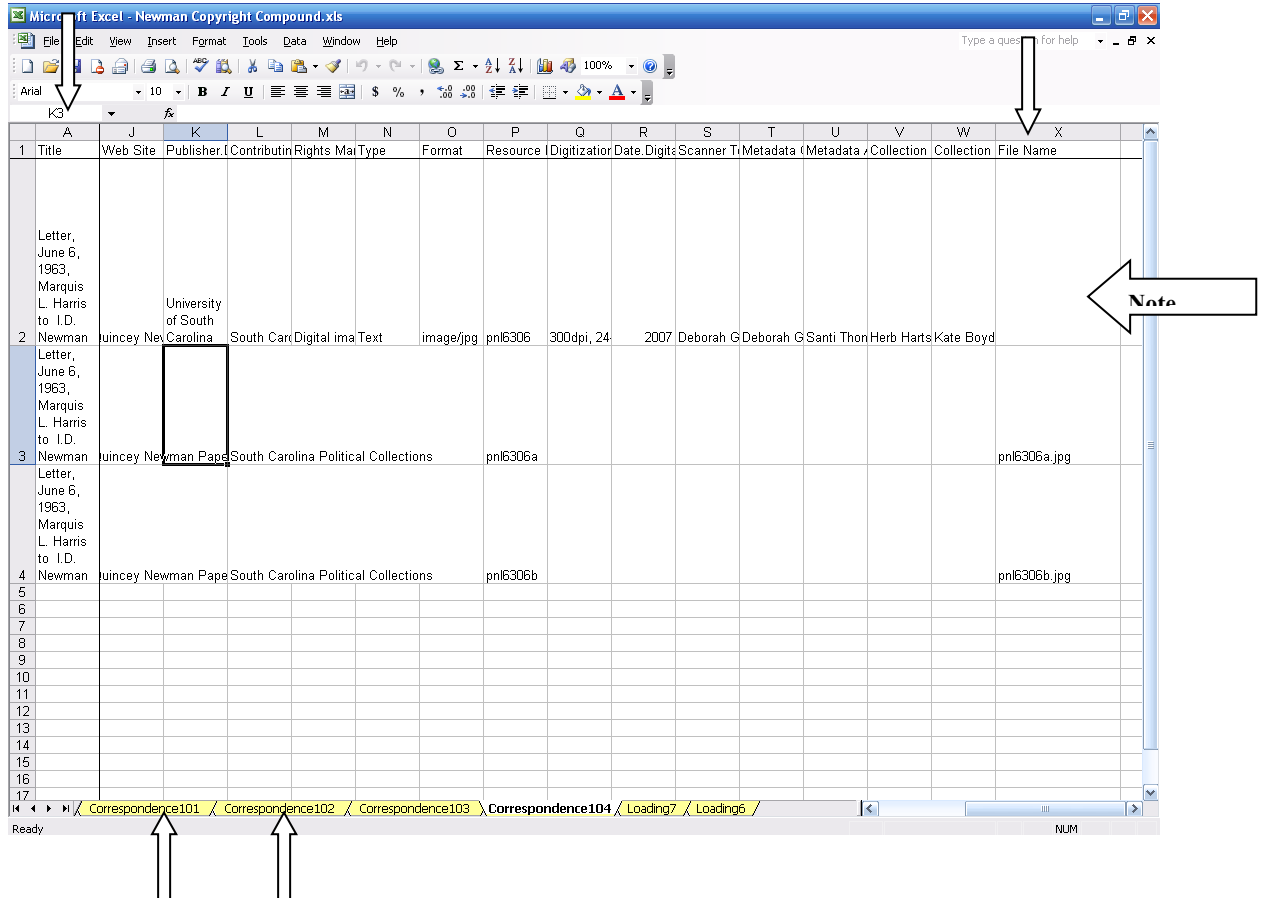
There are two ways you can enter metadata into CONTENTdm. One way is to copy and paste each field from the Excel sheet to CONTENTdm. The other way is to upload a Tab-delimited text file. (see Managing Digital Projects with CONTENTdm 16).

Compound Object Metadata

Compound objects are different from single objects in that they have two or more images for one title rather than one image for one title, for example a book or a multiple page letter. Multiple images and file names must be indicated on separate

rows of the Excel sheet, as seen in Figure 5. Each compound object gets its own excel sheet.

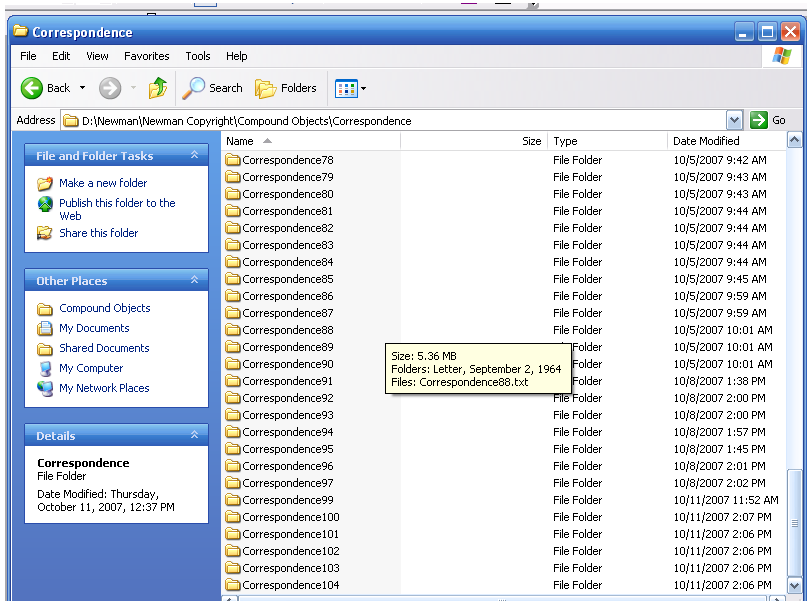
Figure 5



Column A is the title and is provided for each file name in Column X. **Note** that there is no file name in Row 2 of Column X. Row 1 is Collection level metadata (describes the whole item) and Rows 2 and 3 are page level metadata. Some versions of CONTENTdm require having a title for each file name when loading. We usually put the least amount of page level metadata as possible. For the Newman collection I gave the Title, Date.Original, Contributing Institution and Relation.IsPartOf as page level information.

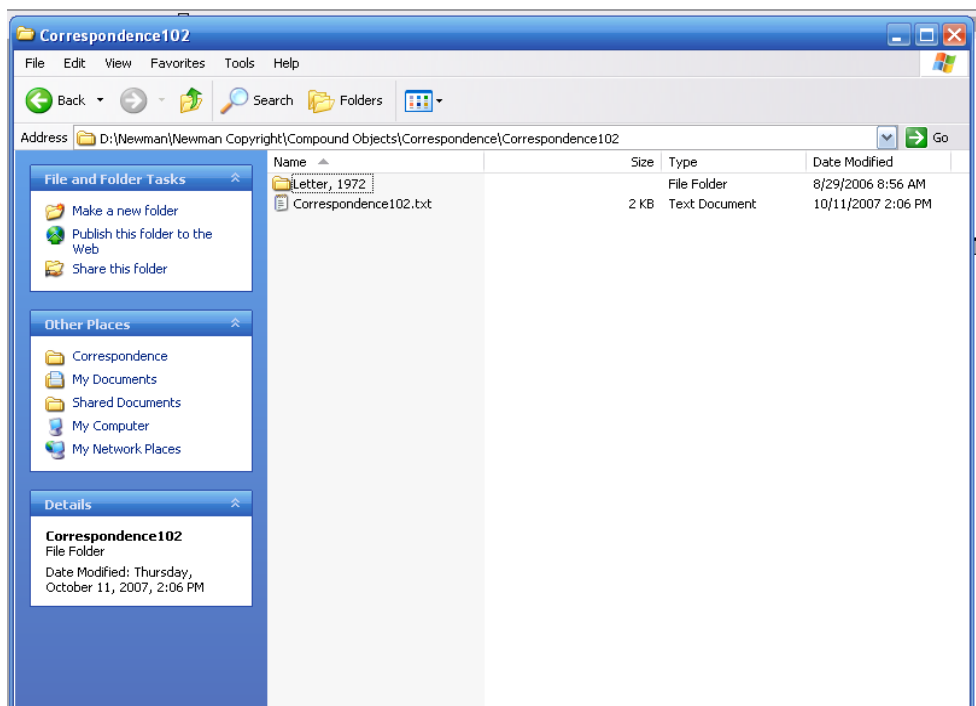
Each **tab** in the excel sheet in Figure 6 is a separate compound object. As you can see there are tabs for Correspondence101, Correspondence102, etc. Each of these tabs has a corresponding folder with the images for the compound object, as seen in Figure 6.

Figure 6



The path for Correspondence102 in Figure 7: “**D:\Newman\Newman Copyright\Compound Objects\Correspondence\Correspondence102.**” The path is significant because it shows how the collection is organized on the hard drive. **Correspondence102** is in the **Correspondence** folder because it is a letter, it is in **Compound objects** because it has multiple images, it is in **Newman Copyright** because it was written by someone other than Newman and must be loaded to the restricted part of the collection, and it is in **Newman** folder because that is the collection to which it belongs. Figure 7 also shows a tab-delimited text file with the images folder, which originated from the Excel tab.

Figure 7



For larger collections where there are both compound and single items you will have to create two tabs in your excel file. One tab will be for compound images, and the other will be for single images. In Figure 3, you can see the tabs labeled Master, Copyright Single Images, and Copyright Compound Images. This example is from the Isaiah DeQuincey Newman collection. Because the collection included items under copyright we created a Master sheet with all of the metadata, and then created separate sheets for items under copyright and for items not under copyright.

Metadata for Monographs

Although monographs are considered a compound object, they are treated differently because they often have sections or chapters.

When loading a compound object using CONTENTdm you can check a check box to have the database automatically assign a page title to the images, for example, Page 1, Page 2 or Recto, Verso. Or have the database ignore everything before the underscore, for example, pns8912a_Page 1.jpg, pns8912b_Page 2.jpg.

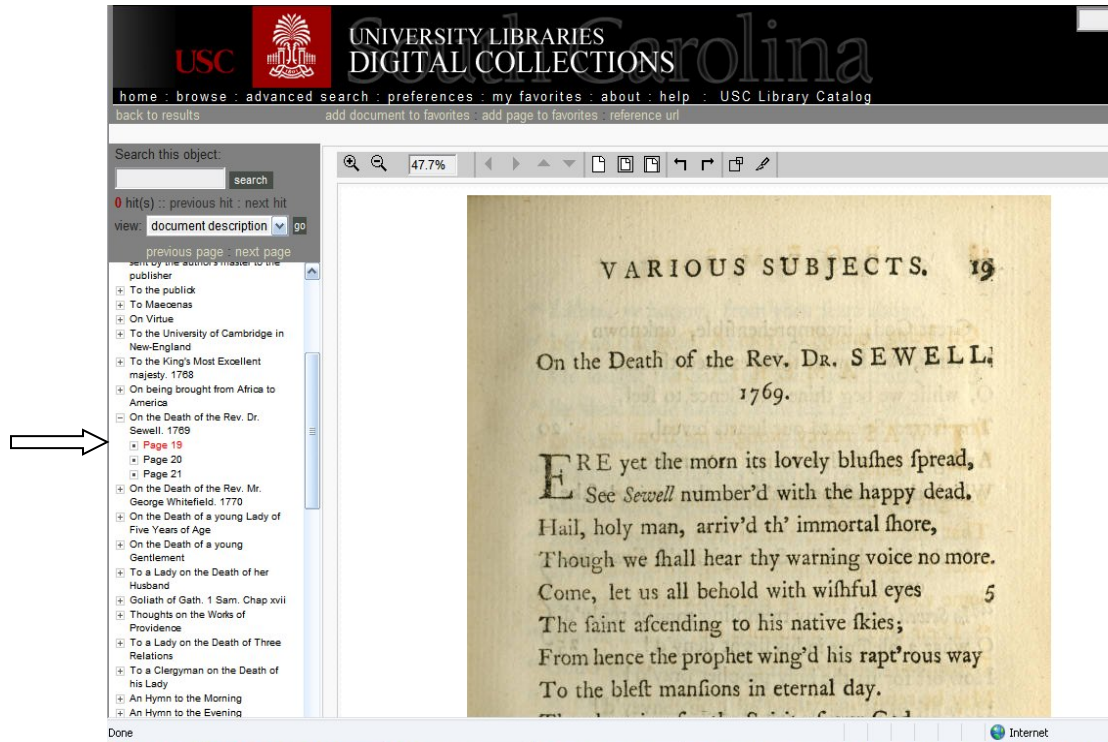
In a monograph the file names will have the underscore with page information, but because it has chapters or sections it will have a Command Level (CMD_LVL) and Command Title (CMD_LVL_Name). The command level and command title tell the database to assign chapters and nest the corresponding pages for each chapter. For example, Phillis Wheatley's *Poems on various subjects, religious and moral* in Figure 9. Notice how the CMD_LVL number changes with the CDM_LVL_Name and the Title in Column C matches the underscore in the file name in Column AC.

Figure 8

	A	B	C	AB	AC	AD	AE	AF	AG
1	0	Poems on various subjects, religious and moral	Poems on various subjects, religious	Deborah Green and Rick Fisher					
2	0	Poems on various subjects, religious and moral	Cover		rpw001.jpg_Cover				
3	0	Poems on various subjects, religious and moral	(Blank) Front pastedown		rpw003.jpg_(Blank) Front pastedown				
4	0	Poems on various subjects, religious and moral	(Blank) Front free endpaper, recto		rpw004.jpg_(Blank) Front free endpaper, recto				
5	0	Poems on various subjects, religious and moral	(Blank) Front free endpaper, verso		rpw005.jpg_(Blank) Front free endpaper, verso				
6	0	Poems on various subjects, religious and moral	(Blank) Verso of engraved		rpw006.jpg_(Blank) Verso of engraved				
7	0	Poems on various subjects, religious and moral	Frontispiece portrait		rpw007.jpg_Frontispiece portrait				
8	0	Poems on various subjects, religious and moral	Title Page		rpw008.jpg_Title page				
9	0	Poems on various subjects, religious and moral	Verso		rpw009.jpg_Verso				
10	0	Dedication	Page 3		rpw010.jpg_Page 3				
11	1	Preface	Page 4		rpw011.jpg_Page 4				
12	1	Preface	Page 5		rpw012.jpg_Page 5				
13	0	The following is a copy of a letter sent by the author's r	Page 6		rpw013.jpg_Page 6				
14	0	To the publick	Page 7		rpw014.jpg_Page 7				
15	0	Poems on various subjects, religious and moral	Page 8		rpw015.jpg_Page 8				
16	1	To Maecenas	Page 9		rpw016.jpg_Page 9				
17	1	To Maecenas	Page 10		rpw017.jpg_Page 10				
18	1	To Maecenas	Page 11		rpw018.jpg_Page 11				
19	1	To Maecenas	Page 12		rpw019.jpg_Page 12				
20	1	On Virtue	Page 13		rpw020.jpg_Page 13				
21	1	On Virtue	Page 14		rpw021.jpg_Page 14				
22	1	To the University of Cambridge in New-England	Page 15		rpw022.jpg_Page 15				
23	1	To the University of Cambridge in New-England	Page 16		rpw023.jpg_Page 16				
24	0	To the King's Most Excellent majesty. 1768	Page 17		rpw024.jpg_Page 17				
25	0	On being brought from Africa to America	Page 18		rpw025.jpg_Page 18				
26	1	On the Death of the Rev. Dr. Sewell. 1769	Page 19		rpw026.jpg_Page 19				
27	1	On the Death of the Rev. Dr. Sewell. 1769	Page 20		rpw027.jpg_Page 20				
28	1	On the Death of the Rev. Dr. Sewell. 1769	Page 21		rpw028.jpg_Page 21				
29	1	On the Death of the Rev. Mr. George Whitefield. 1770	Page 22		rpw029.jpg_Page 22				
30	1	On the Death of the Rev. Mr. George Whitefield. 1770	Page 23		rpw030.jpg_Page 23				
31	1	On the Death of the Rev. Mr. George Whitefield. 1770	Page 24		rpw031.jpg_Page 24				
32	1	On the Death of a young Lady of Five Years of Age	Page 25		rpw032.jpg_Page 25				
33	1	On the Death of a young Gentlement	Page 26		rpw033.jpg_Page 26				
34	1	On the Death of a young Gentlement	Page 27		rpw034.jpg_Page 27				
35	1	On the Death of a young Gentlement	Page 28		rpw035.jpg_Page 28				
36	1	To a Lady on the Death of her Husband	Page 29		rpw036.jpg_Page 29				
37	1	To a Lady on the Death of her Husband	Page 30		rpw037.jpg_Page 30				
38	1	Goliath of Gath, 1 Sam. Chao xvii	Page 31		rpw038.jpg_Page 31				

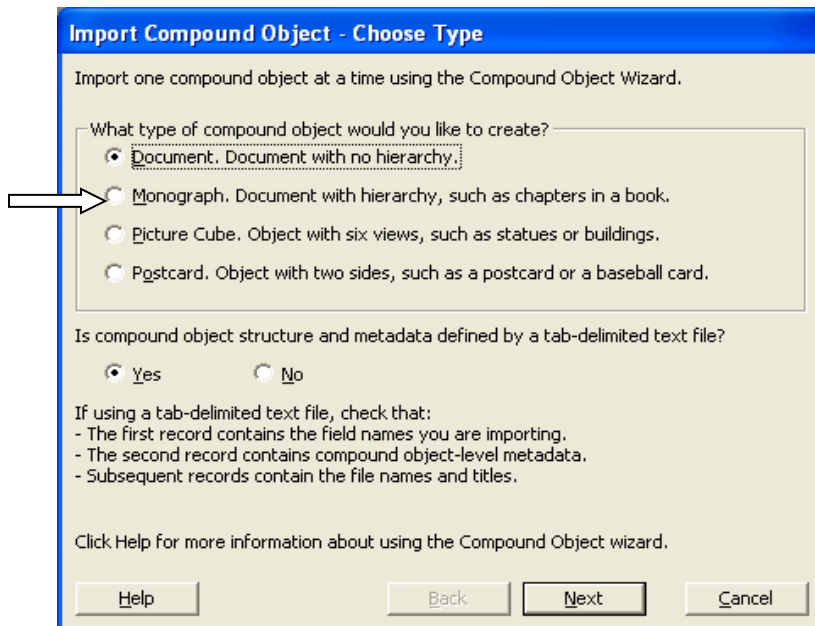
Figure 9, shows the nesting and how the excel sheet in Figure 8 is displayed in the database on the internet (<http://digital.tcl.sc.edu/u/?pwp,24>). For each section/chapter, CONTENTdm creates a + sign. Within the + sign the pages for that section/chapter are nested. Figure 9 shows how Rows 27, 28 and 29, On the Death of the Rev. Dr. Sewell 1769, in Figure 8 are revealed when the + sign for that section is selected.

Figure 9



Once you have saved the excel sheet as a tab-delimited file and are ready to load, go to Acquisition station select Monograph. Figure 10 shows the selection process for importing compound objects as monographs into CONTENTdm.

Figure 10



Conclusion

The above examples are based on my experience while working in Digital Collections at the University of South Carolina. I know the screen shots are sometimes difficult to read, but they will give you an idea of what you are supposed to be looking at while working through CONTENTdm or an Excel sheet. I have found that looking at collections that have already been loaded has helped when creating metadata. The South Carolina Digital Library has Metadata Standards posted on their web site at <http://www.digilibsc.org/guidelines/index.html>.

This document is meant to be used in conjunction with my other documentation *Managing Digital Projects with CONTENTdm*. If you have any further question you can contact me at deb_green2001@yahoo.com.