

## Exercise 7: Build a document

### Before you begin

To start a Project in the Project Client, you will need your CONTENTdm:

- User Name and Password
- URL for the CONTENTdm Server (<https://server17214.contentdm.oclc.org>)

### Purpose

Multiple images or scans can be assembled to form a single object within the CONTENTdm collection, creating an item known as a Compound Object. A compound object consists of multiple images or pages that are added to a collection and retrieved as a single object by end-users. Compound objects can be documents, books, the front and back of postcards, or six-sided views of three-dimensional objects.

In this way, you can store and retrieve related items including the front and back of postcards, multi-page documents, and up to six views of multi-dimensional items, such as sculptures or furniture.

### What you will learn

At the end of this exercise you will have learned how to build a fully searchable compound document using the **Compound Object Wizard** in the Project Client. You will provide searchable text in the Description field to enable full-text searching of the transcript of a handwritten document. Students are expected to have experience building CONTENTdm Collections and Projects using simple objects and metadata templates.

## The Exercise

- 1) If not already open, open your **Text** project.

Note: Full-text searchability is possible when there is a text transcript of the imaged document. In order to make your document's text fully searchable, both the image files and the text files corresponding to them have been installed on your workstation. In this exercise, you will make use of the **Transcript** field, which you added in Exercise 1.

### Edit your Metadata Template

Although this is a Compound Object, we will use the **Project template (general)**.

*Why not the Compound Object template?* The project template (general) will act on each individual page of the compound object and the system-created STRUCTURE of the item. The compound object template would act on ONLY the system-created structure of the item. We want to fill in information for both the pages and structure.

#### In the Project Template (general)

- 1) Enter your name in the **Publisher** field.
- 2) Enter the word **Text** into the **Type field**.
- 3) Use the **Format field** to store the **File Extension**
- 4) Enter the three-letter ISO language code **eng** into the **Language field**.
- 5) Enter **Rights information** into the **Rights field**.
- 6) Enter your **full collection name** into the **Collection field**.
- 7) Use the **Identifier** field to store the **File Name** of the imported images (already done for previous exercise)
- 8) Once everything is entered on the template, click **OK**.

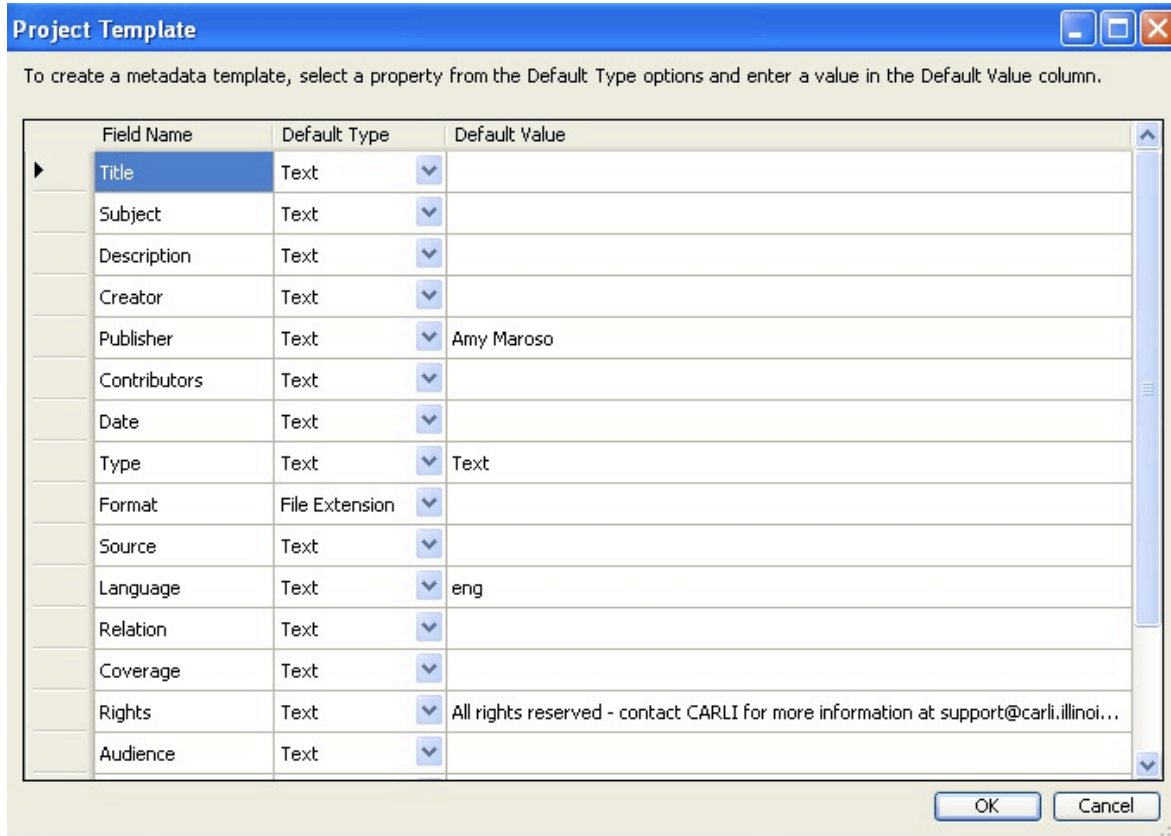


Figure 1: General project template

- 1) Click **OK** to close the Project Settings.

## Import a Compound Object

1) Now you are ready to import files for this exercise from the folder indicated by the workshop instructor. For the **Bell Handwritten Letter**, this will be the Exercise data/Documents/Bell Handwritten Letter/Tiff folder. Click on **Add Compound Object** in the task bar or via the Add menu to open the Compound Object wizard.

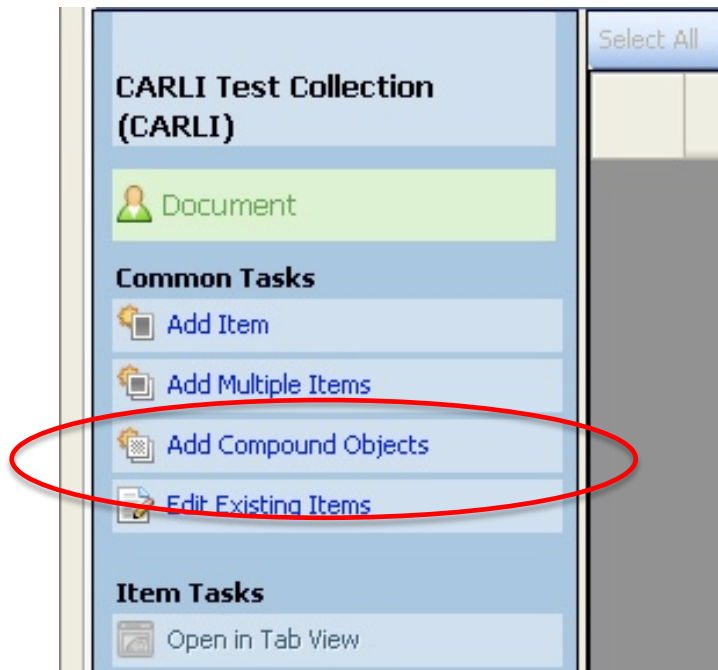


Figure 2: Add a Compound Object option in the task bar.

2) Note that **Compound Object Wizard** is the default option. This is the import method we will use since we do not have a tab-delimited text file of multiple compound objects to import. There are other options of compound object import available. If you have questions about the best method to use for your project, contact CARLI at [support@carli.illinois.edu](mailto:support@carli.illinois.edu)

3) Click the **Add** button to start adding a new compound object.

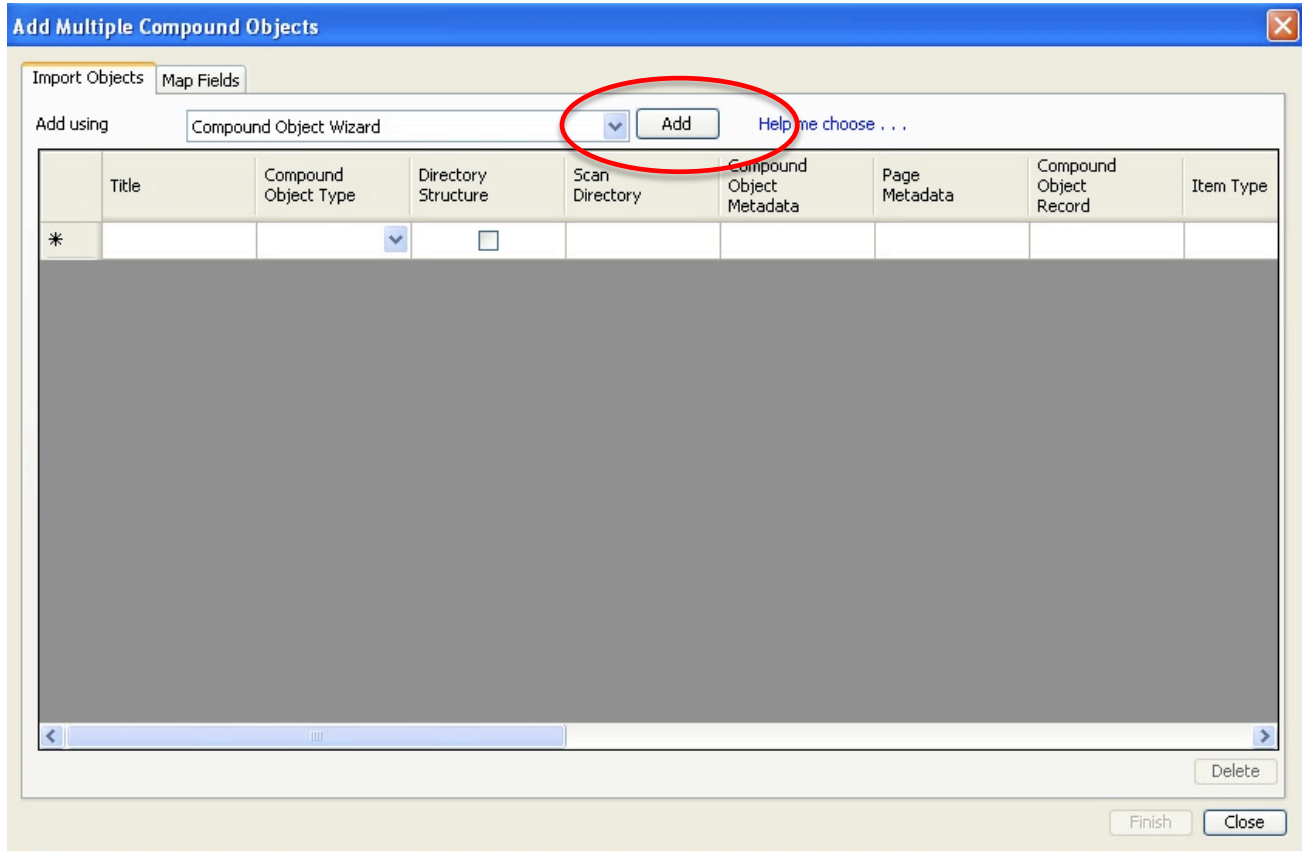


Figure 3: Compound Object wizard is the default.

4) The single compound object import wizard will appear and lead you through the steps of importing a compound object. We will import a **Document**, as it is a non-structured multi-page compound object.

5) We do not have a tab-delimited text file containing the metadata for this item, so the make sure the structure is set to **No**.

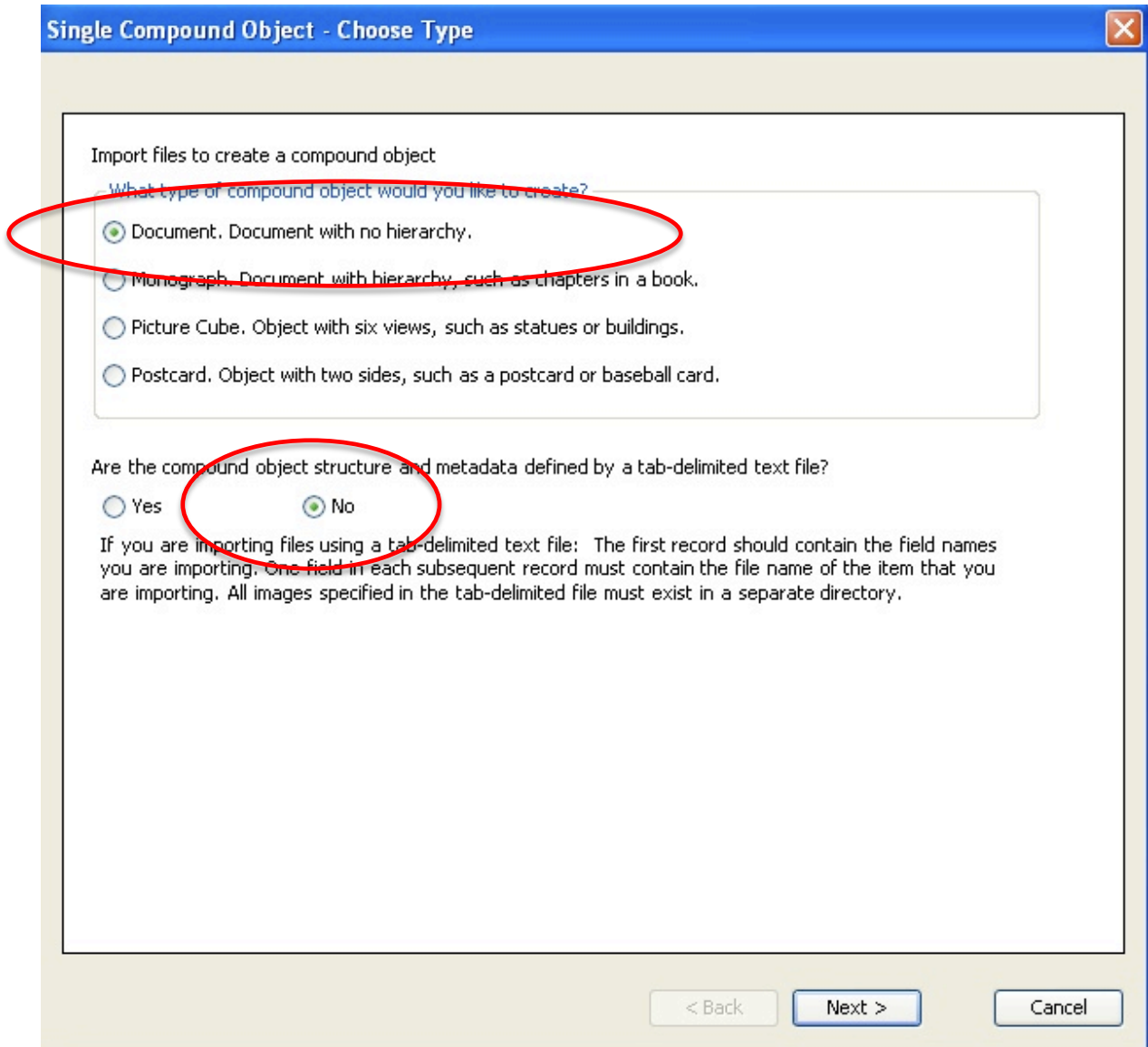


Figure 4: Importing a Document without using a tab-delimited text file of metadata.

6) Click **Next** to continue through the import wizard.

7) On the next page you will select the location where the images are located.

Click on **Browse** to locate the document's images. They are in the following location: Desktop > Exercise Data > Documents > Bell Handwritten Letter > Tiff. Make sure the "Tiff" folder is highlighted and opened and you can see the three files within it.

8) When the folder is opened, click **OK**.

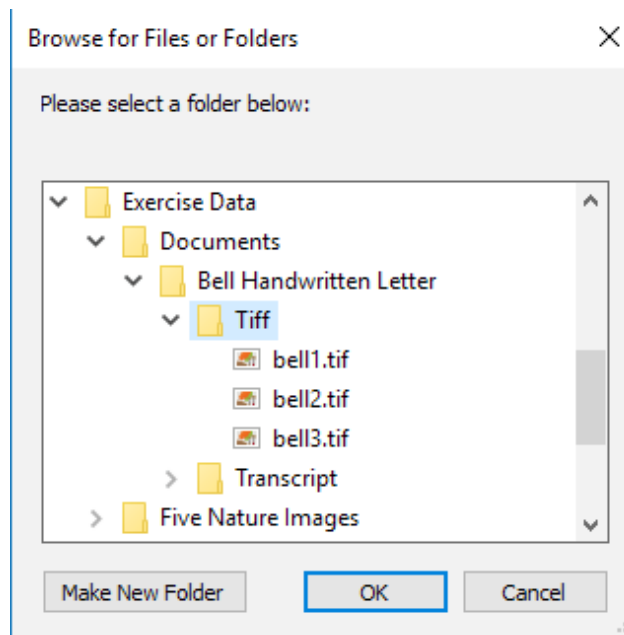
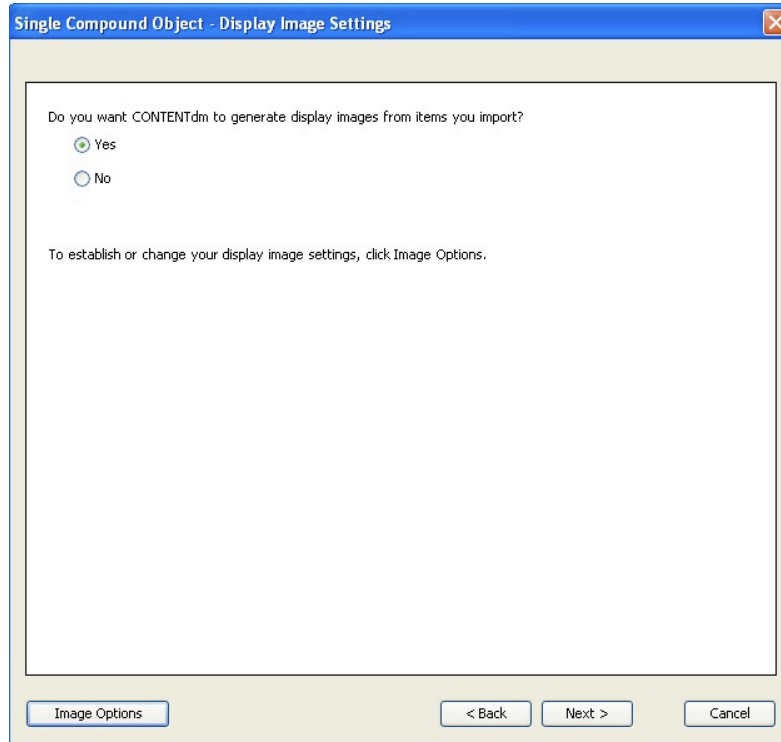


Figure 5: Opening the folder of letter images correlated to their texts

9) The location of the images is entered into the wizard. Click **Next**.

10) Allow CONTENTdm to generate display images for the items (this feature is turned on by default). Click **Next**.



**Figure 6: CONTENTdm will generate display images for you.**

**11)** Next you will determine how to name the pages and import transcripts for your document.

In the **Specify page names** section of the window, choose "**Label pages using sequence**". We will use the default option of "Page," but you could use other descriptors, like "Leaf," "Plate," etc.



Single Compound Object - Page Information

Specify page names

Use file names as titles

Ignore information before underscore. Example: 001\_Page1 becomes Page1

Label pages using sequence

Name:  Begin with:

Label pages using tab-delimited text file

Transcript

Import transcript files from a directory

Directory name:

Generate transcripts using OCR

No transcripts

PDF

Create print PDF

< Back Next > Cancel

Figure 7: Label the individual pages.

## Adding a Transcript

### A few things to remember about Transcripts for your items:

- To make them fully searchable, you need to add them to a field that was set in your collection's Field Properties options to "Full Text Search."
- You may have only ONE Full Text Search field per collection.
- Each image file of your item must have a corresponding individual transcript file in text format (file extension of .txt). This file must have **EXACTLY THE SAME NAME AS THE IMAGE FILE** – the only difference should be the file extension. For example:

Image file: bell1.tif  
Corresponding transcript file: bell1.txt

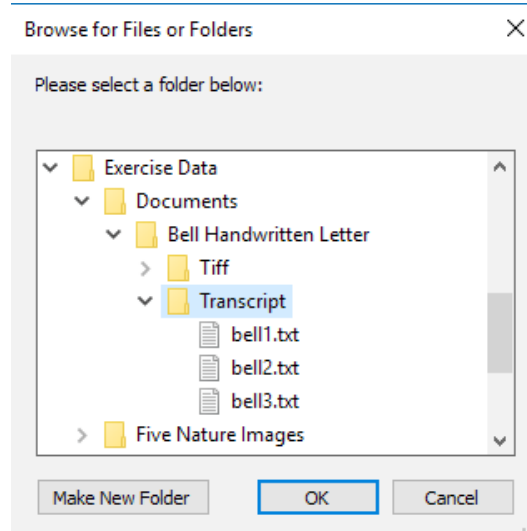
- Transcript files should all live together in the same folder with no other files. Do not put them in the same folder as the image files. Doing so will cause problems when you create a compound object.

1) Choose the appropriate option in the Transcript section of the window. Since we have transcripts, we will use the option "Import transcript files from a directory."

The screenshot shows a dialog box titled "Single Compound Object - Page Information". The "Transcript" section is highlighted with a red border. In this section, the option "Import transcript files from a directory" is selected with a radio button. Below this option is a text field for "Directory name:" and a "Browse" button. Other options in the section include "Generate transcripts using OCR" and "No transcripts", both of which are unselected. The "Specify page names" section above it has "Label pages using sequence" selected, with "Name:" set to "Page" and "Begin with:" set to "1". The "PDF" section below has "Create print PDF" unselected. At the bottom of the dialog are buttons for "< Back", "Next >", and "Cancel".

Figure 8: Import transcript files from a directory option.

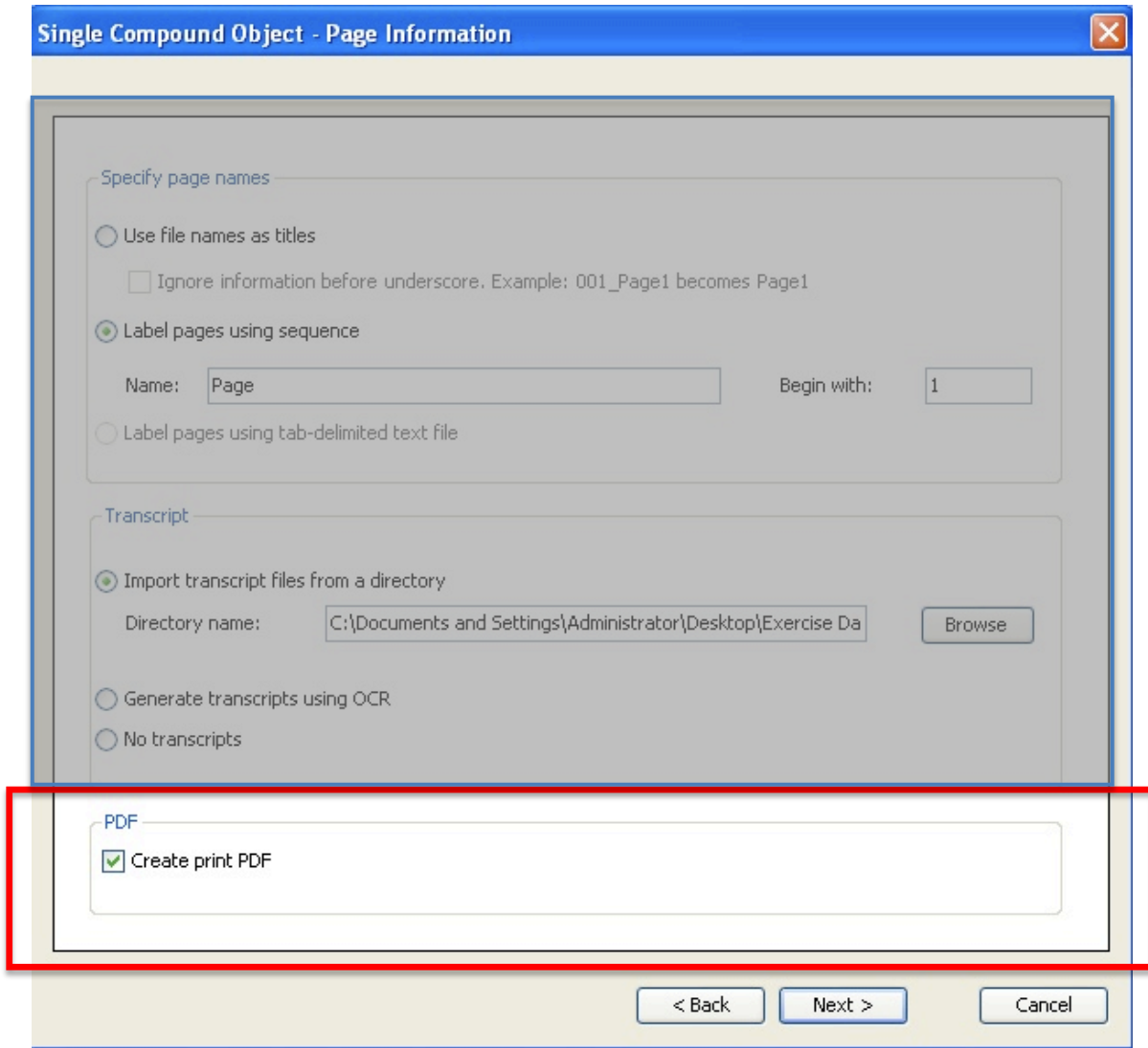
2) To locate the transcripts for the three image pages, click on **Browse**. The files are in the following location: Desktop > Exercise Data > Documents > Bell Handwritten Letter > Transcript.



**Figure 9: Locate and open the folder holding the transcript files**

- 3) Make sure the ***Transcript*** folder is opened with the three file names displaying within it.
- 4) Click **OK**, to add the files to the wizard.

**OPTIONAL:** Create a print PDF that will allow end-users to more-easily print a copy of the document.



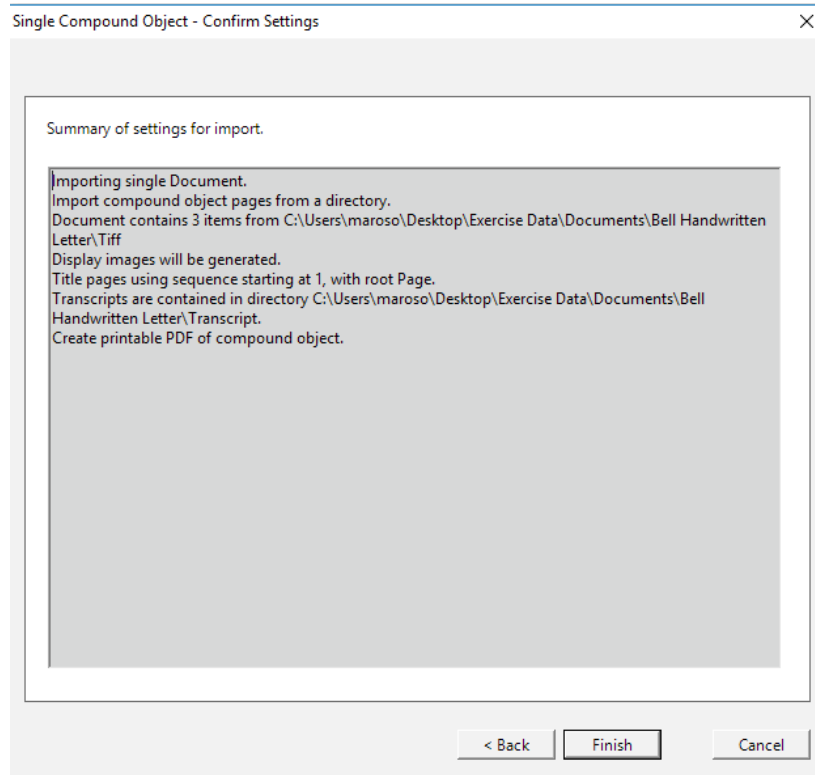
The screenshot shows a dialog box titled "Single Compound Object - Page Information". It has three main sections: "Specify page names", "Transcript", and "PDF".

- Specify page names:** Contains three radio button options: "Use file names as titles" (unselected), "Label pages using sequence" (selected), and "Label pages using tab-delimited text file" (unselected). Under "Use file names as titles", there is an unchecked checkbox for "Ignore information before underscore. Example: 001\_Page1 becomes Page1". Under "Label pages using sequence", there is a "Name:" text box containing "Page" and a "Begin with:" text box containing "1".
- Transcript:** Contains three radio button options: "Import transcript files from a directory" (selected), "Generate transcripts using OCR" (unselected), and "No transcripts" (unselected). Under "Import transcript files from a directory", there is a "Directory name:" text box containing "C:\Documents and Settings\Administrator\Desktop\Exercise Da" and a "Browse" button.
- PDF:** This section is highlighted with a red box and contains a checked checkbox for "Create print PDF".

At the bottom of the dialog box are three buttons: "< Back", "Next >", and "Cancel".

**Figure 10: Check box to create print PDF automatically**

**5)** Click **Next** takes you to a confirmation screen. Read this carefully as it describes where the files are located and confirms your choices for page labeling.



**Figure 11: Confirming file location and choices**

**6)** Click **Finish** to add the completed item to your import wizard.

**7) Close** the final summary page and return to the Add Multiple Compound Object Wizard. You will now see the document you just created listed in the pending Compound Object creation spreadsheet.

If you wanted to add more compound objects you could do so now, following the same procedures. Each one would be added to this spreadsheet.

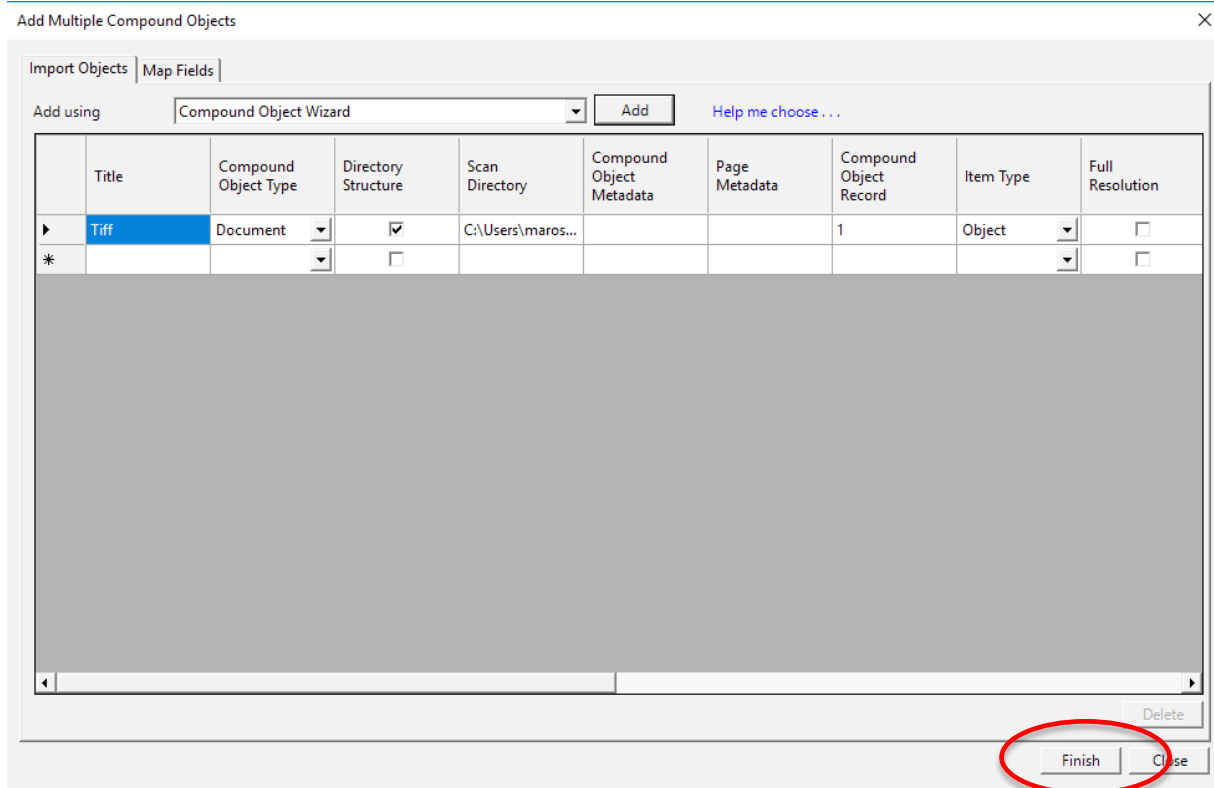


Figure 12: Spreadsheet of compound objects waiting to be created.

8) Click **Finish** to import the compound object into your project.

9) **Close** the summary window. The item is now in your project spreadsheet and can be edited much the same way individual items can be.

## Editing Compound Object in the Project Client

1) Double-click the thumbnail to open the document in a new tab.

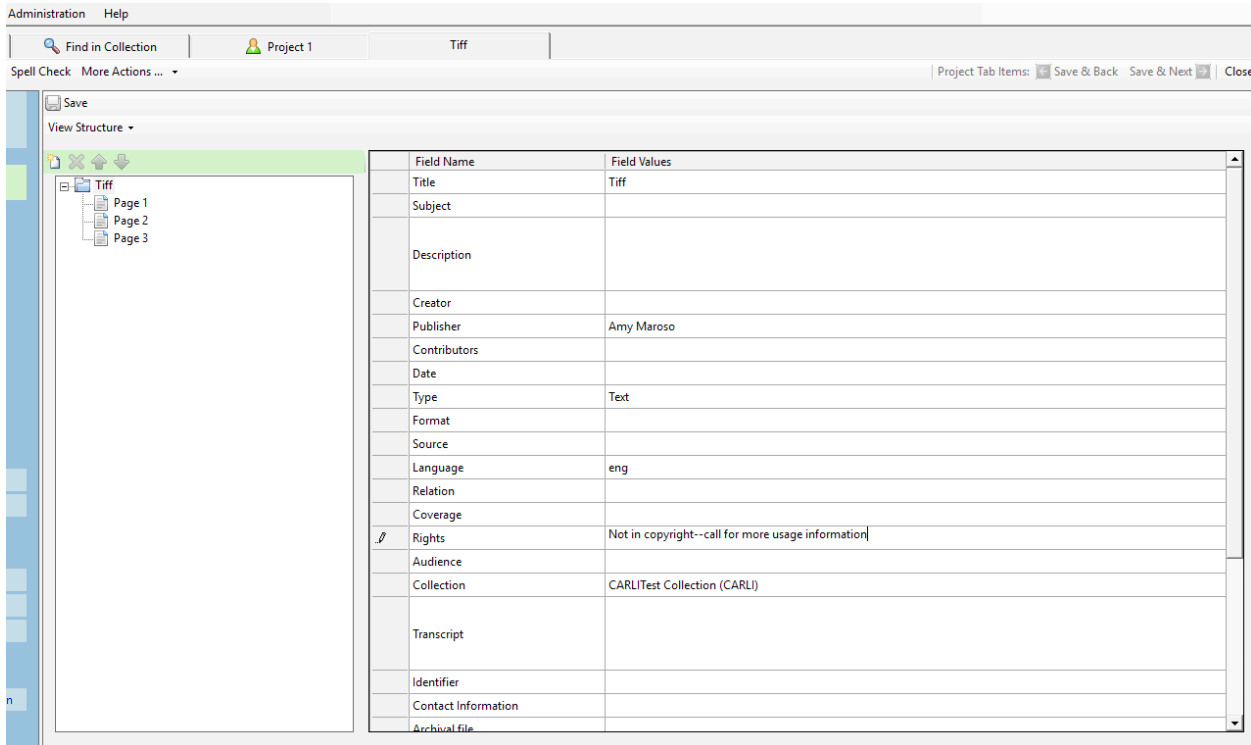


Figure 13: Viewing and editing the assembled compound object.

From this tab, you can edit the structure and metadata of the compound object much as you would a single image.

From the structure view of them item, you see that CONTENTdm has created the STRUCTURE of the compound object and the three pages are housed within it.



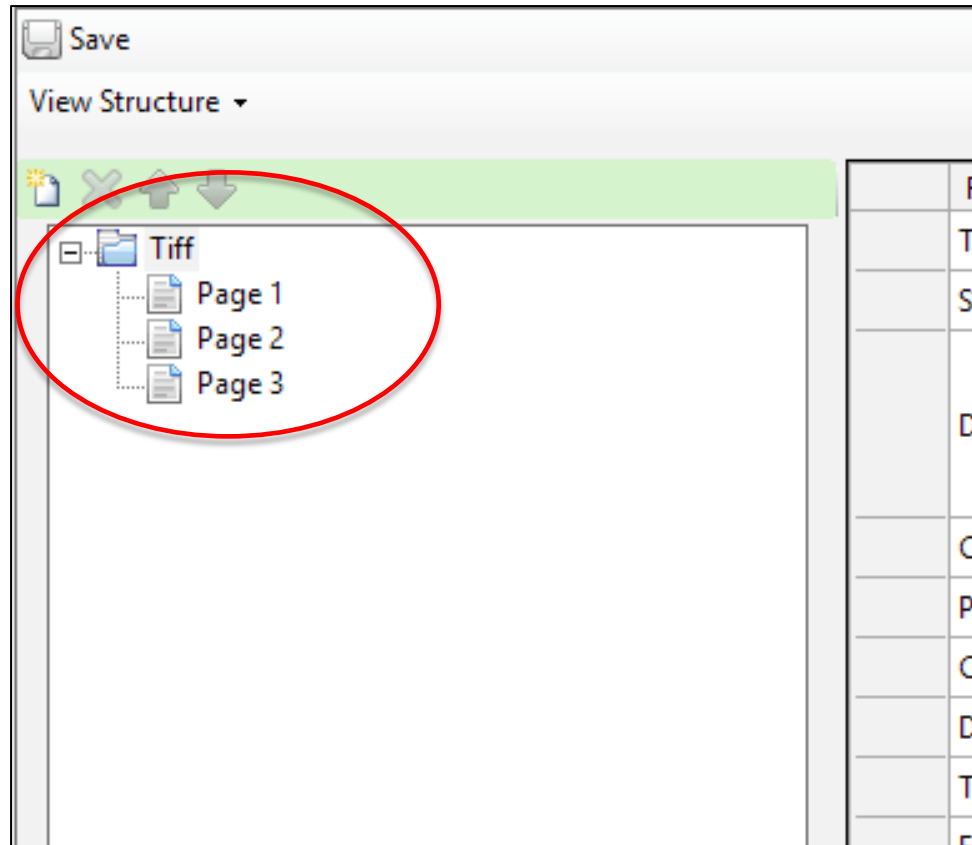


Figure 14: Structure of the compound object.

2) Click on **Tiff** in the structure window, if not already highlighted. Note that the metadata entered in general metadata template has been added to the fields.

3) Change the **Title** to read **Early Days of Invention**

4) Add a better **description** to the **Description field**.

5) Click on **Page 1** in the compound object structure window. Note that the metadata entered in the general metadata template has also been added to the fields, and that the transcript files have been automatically added to the Transcript field.

6) Add a better **description** to the **Description field**. Note that this will populate the Description field on Page 1 ONLY. It will not add any additional metadata to the compound object structure or to any of the other pages!

**Other things you can do when editing Compound Objects in the Project Client:**

You can reorder the pages of your document, or add or remove a page to the document. Use the submenu options that are right above the Compound Object structure window.

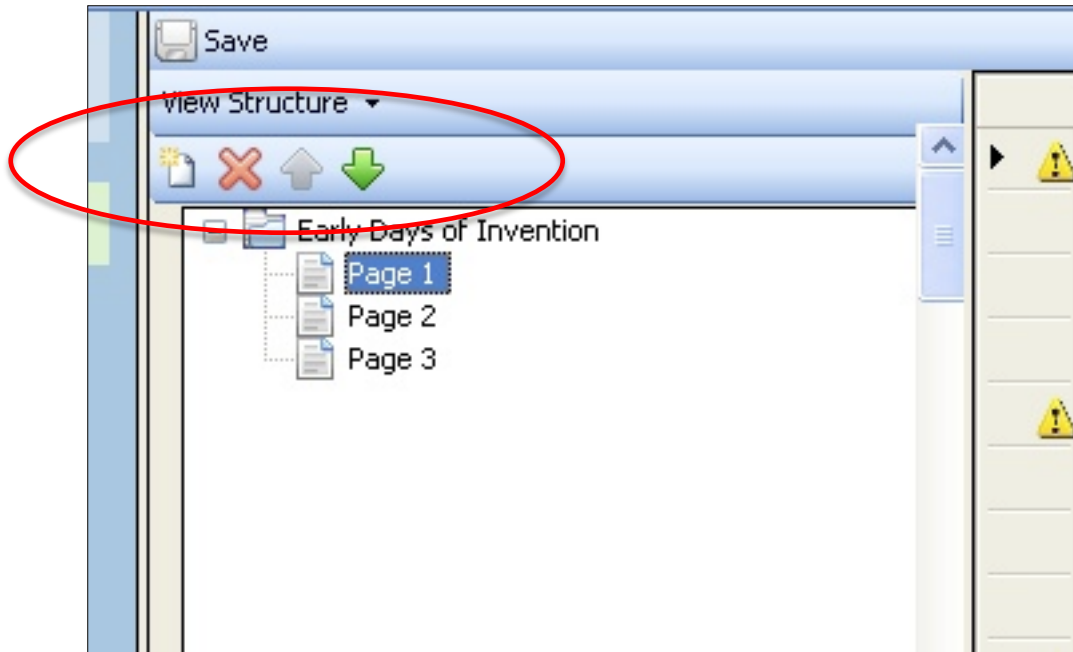


Figure 15: Options for adding, removing, or reorganizing pages of a Compound Object.

**7) Move** Page 1 of the document to the **end of the document** by selecting the page and clicking the green down arrow.

**8) Close** the document and **Save your changes**.

**9) Upload** the document, then **Approve** it and **Index** the collection.

If you wish, you may now close your Project Client for the time being.