

Creating Compound Objects (Documents, Monographs Postcards, and Picture Cubes)

A compound object is two or more files bound together with a CONTENTdm-created XML structure. When you create and add compound objects to a collection they are retrieved as a single object by end users. Compound objects can be documents, books, audio files from a CD, postcards, or six-sided views of three-dimensional objects (picture cubes).

This tutorial explains how to create compound objects using the Compound Object wizard in your Project Client.

To learn about configuring the display options for compound objects on your website, see **Page Types**.

Note: CONTENTdm provides the capability to create PDF compound objects from multiple-page PDF files. To create PDF compound objects, you do not use the Compound Object wizard. For more information, see **Importing PDF Files**.

Note: With the release of the CONTENTdm Website (version 6+), there are no longer specific viewers for each of the compound object types. While the options to create postcards (now displayed as a two-page object) and picture cubes (now displayed as a six-page object) remain in the Project Client, note there are no distinctions for these objects in the end-user interface.

Learn About:

Section I: Overview

Section II: General Steps

Section III: Before You Begin

File naming and organization

Enabling full text searching

Establishing Display Image Settings

Enabling Archival File Manager

Section IV: Creating Documents

Section V: Creating Monographs

Section I: Overview

Collection owners frequently wish to store related scanned files that can be searched and retrieved as single entities. Examples include documents; books with chapters, appendices, and bibliographies; postcards; or six-sided views of three-dimensional items. CONTENTdm enables you to add these types of compound objects and associated metadata to your collections. End users can search, retrieve, and view the objects as complete entities on your website.

Section II: General Steps

Step 1. The CONTENTdm Administrator must establish any additional collection settings, such as full text searching or Archival File Manager support before the images are imported using the Project Client. You also can configure display options for compound objects on your website using the Website Configuration Tool. For more information, see **Page Types**.

Step 2. Determine scanning resolution and settings. Decide what you will include in the compound object and how the pages or scans will be named.

Step 3. Store scanned images in directories that can be accessed by the Project Client. Images and associated OCR or text transcripts require identical root files names but should be stored in separate directories. For monographs, you will use directories and subdirectories to establish organizational structure. Directory structure and file naming are explained in detail in this tutorial.

Step 4. Use the Metadata Templates to automatically add descriptive metadata (such as subject, creator, date) and administrative metadata (such as file size, height, width) that you want to appear in specified fields for all compound objects when they are added to a project. See **Using the Metadata Templates** in the Help files for more information.

Step 5. From the Project Client Home tab or Project tab, select **Add Compound Objects** from the Task pane. The wizard steps you through the process of creating the compound object which you then add to your project.

Step 6. You can view or edit the compound object in your Project.

Step 7. Upload the finished compound object for approval.

Step 8. The CONTENTdm Administrator adds the compound object to the collection and builds the text index.

Once compound objects are added and built into the collection text index, they can be searched and displayed just as any other item in the collection.

CONTENTdm Administrators may update or change compound objects after they have been added to a collection.

Section III: Before You Begin

Adding a batch of items to create a compound object requires some preliminary organization of your scanned items in a local directory. We'll explore how CONTENTdm handles directories and files during import and describe the most effective ways to organize your files before importing them.

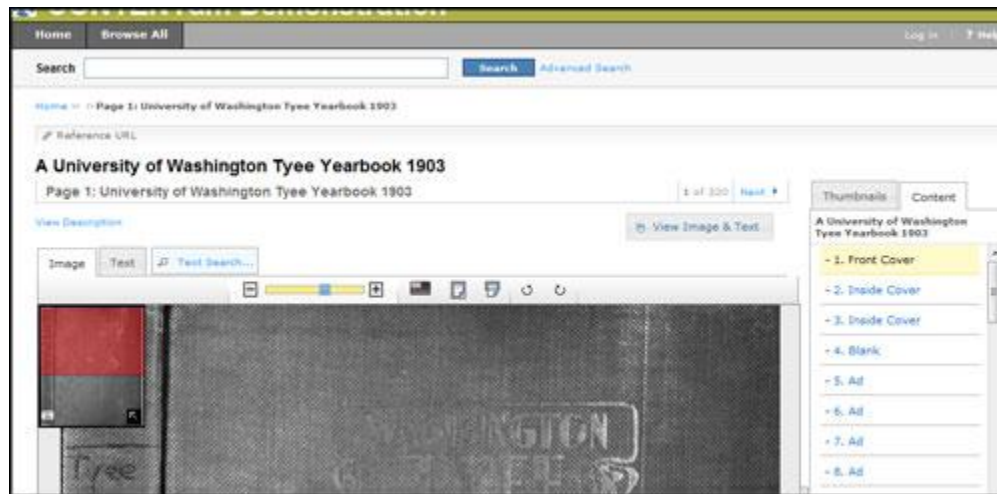
Additionally if you plan to enable full text searching of documents or monographs or want CONTENTdm to work with high resolution images, these collection settings should be established by the CONTENTdm Administrator before you import any items into the project. The appropriate steps will be described or referenced in the section below.

File naming and organization

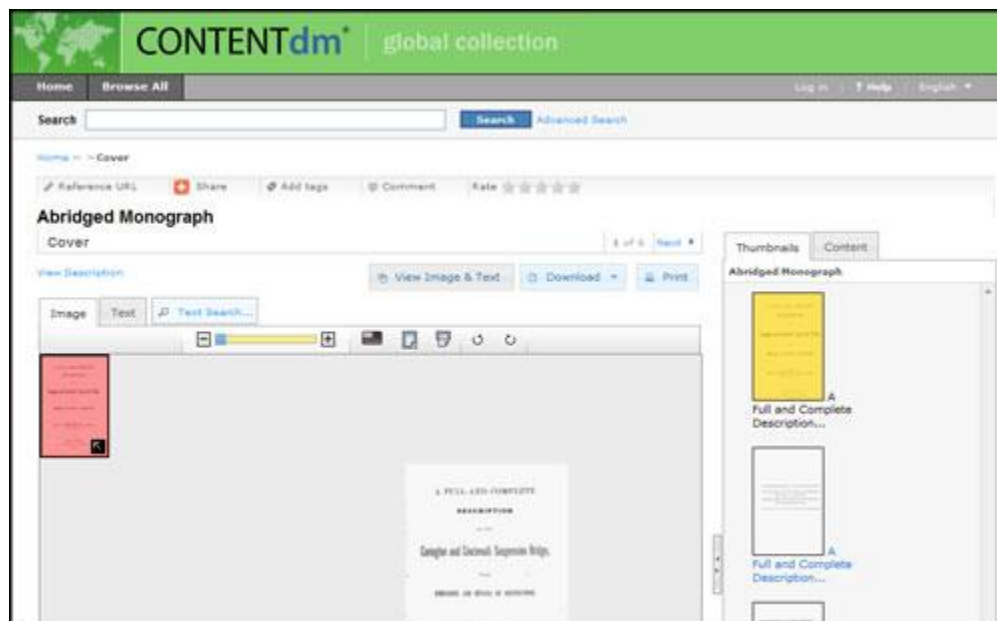
First, determine what you will scan. Will you scan blank pages, the insides of covers? How will you name or number the scans or pages such as the cover, title page, blank pages, introduction, index, or back cover?

To get a sense of how to organize and name files and directories most effectively, it will help to have a brief explanation of how CONTENTdm handles the files when they are imported to create a compound object.

- **Documents**—The wizard will import all of the files in the root directory in ascending alphanumeric order.

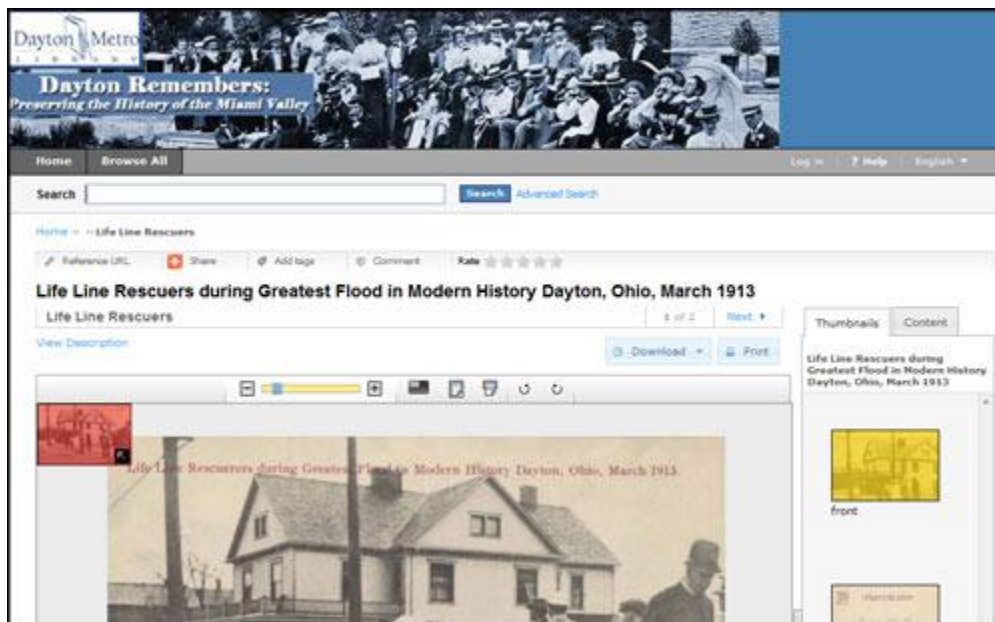


- **Monograph**—To create a monograph (a document with hierarchical structure), the wizard will always import all of the files in the root directory first and then import any sub-directories. If there are files in the sub-directories, they will be imported before any directories in the sub-directory. Order of import is ascending alphanumeric order. If using a tab-delimited text file, monograph structure is defined in the text file and the monograph files are in one directory.



- **Picture Cube**—To create a picture cube, the wizard will import six files from the root directory. If the directory contains more than six items, only six will be imported. Import up to six images for a picture cube. Items will be imported in alphanumeric order. A picture cube is displayed as a document with no more than six pages.

- **Postcard**—The wizard will import two files from the root directory. If the directory contains more than two items only two will be imported. Items will be imported in alphanumeric order. A postcard is displayed as a two-page document.



The names of your directories, subdirectories, and files can be used as title, heading, and page names within the compound object. Titles, page names, headings, page metadata, and page order can be changed in the Project Client Item Editing tab once the compound object is created, but it is usually most efficient to name the items as desired when they are in your local directory before you begin the import process.

CONTENTdm has two ways of establishing directory structure, either mimicking the existing directory structure, or using a tab-delimited text file. This matters most when importing monographs (documents with a hierarchical organization). If using directory structure, monographs require the files be stored in subdirectories within the root directory to create hierarchical organization. If using a tab-delimited text file to determine structure, the files are stored in a single root directory. For more information, see **Importing from Tab-Delimited Files**.

Any time a compound object consists of images and text files, the root file names of your images and text files must be separate but identical as shown in the table below.

Image file name	Page text file name
000_cover.jpg	000_cover.txt
001_page 1.jpg	001_page 1.txt
002_page 2.jpg	002_page 2.txt

Documents and Monographs

Documents and monographs are the most commonly created compound objects. Below are useful guidelines and tips for organizing and naming your document or monograph files

- Verify that the scanned files for your compound object are accessible from your local machine. If using files on a network, the network file must be mapped. All items for the compound object should be in a single directory. Any text files associated with the

compound object should be in another separate directory, saved as .txt files with names that correspond to the image file names.

- The order of import is ascending alphanumeric order. Determine your file names accordingly. Your page titles will be used as navigation for end users.

The compound object wizard provides several options for naming pages:

- Use the file names as titles
 - Use an organizing prefix and underscore before the file name
 - Ignore the file name altogether and specify a series label during the import process.
- If you make a mistake after creating the document or monograph, pages can be renamed or reordered in the Item Editing tab. Changes can also be made after the compound object has been added to the collection using the Edit function within the CONTENTdm Administration Web pages.
 - If you are importing OCR (optical character recognition) transcript files along with your images, review and edit the OCR text to verify that characters have been rendered correctly. Do not add any text formatting other than
 tags to the OCR text or transcript files. CONTENTdm recognizes the HTML
 tag, but no other formatting (such as carriage returns, tabs, or hypertext links) is retained by the database. If you do not already have text transcript files, one of the easiest ways to create transcripts is to use OCR software to generate text as you scan page images. CONTENTdm provides an OCR extension, powered by ABBYY FineReader, which generates text transcript files as you import scanned files into the Project Client. For more information, see **Using OCR** in the Help files.

Enabling full text searching

When full text searching is enabled, users can search the text of an item or compound object and view page images and your imported text transcripts side by side by clicking the View Image & Text button when viewing the item or object on your website.



Figure 1. Image Document with Transcript View Button

To establish full text search settings described below, you must have CONTENTdm Administration access.

One field within the collection must be designated as a full text field to enable full text searching. Designate an existing field as a full text field or create a new field. The steps below provide an example of how you might construct your fields to enable full text searching.

To enable full text searching:

1. Open CONTENTdm Administration by selecting CONTENTdm Administration from the Administration menu in the Project Client project or by pointing your Web browser to <http://your.server.address/cgi-bin/admin/start.exe>.
2. Enter your user name and password if prompted. Select the **Collections** tab.
3. From the drop-down labeled Current Collection, select the collection for which you want to add full text searching if it is not already shown. Click **Change**.
4. Click **Field properties**.
5. Create a full text field for the text transcript, by clicking Add new field at the bottom of the Field Properties page. (To modify an existing field, click edit in the row that contains the field name.) The Add field page opens.
6. Name the new field Transcript. The OCR text or transcript of each page will be displayed in this field. Configure the new field with the following settings:

Field name	Transcript
DC map	Description
Data Type	Full Text Search
Show large field	Yes
Searchable	Yes
Hidden	No
Required	No
Controlled vocabulary	No

For best performance, activate full text search for one field only. In some circumstances, such as displaying text in multiple languages, you may consider activating full text search for more than one field. However, only the first full text search field will be displayed in the image and text view on the website.

Setting Show large field to Yes enlarges the appearance of text box for data entry and display; it does not change the number of characters that can be entered in a field. The maximum number of characters permitted in a field is 128,000.

7. Click **Save changes** after establishing the new field settings. Full text searching is now enabled.

Establishing display image settings

Display image settings enable you to set whether you want to automatically generate display-quality versions of images when high resolution images are imported. You can determine the display image format, quality and compression settings. Display images optimize the end-user display by serving up a high quality copy of your full resolution image.

You can edit display image settings if you have rights to configure settings and field properties; display image settings can also be overridden in the Project Client. New collections are enabled by default to automatically generate display-quality versions of images when full resolution images are imported.

To establish display image settings for a collection:

1. Open CONTENTdm Administration by selecting CONTENTdm Administration from the Administration menu in the Project Client project or by pointing your Web browser to `http://your.server.address/cgi-bin/admin/start.exe`.
2. Click the **Collections** tab and then click **Profile**.
3. Under the **Optional Collection Settings Summary**, locate **Display Image Settings** and click **Edit**.
4. Define your display image settings. For more information see **Display Image Settings** in the Help files.

Enabling Archival File Manager

The Archival File Manager handles the storage of files in archive volumes. If it is enabled, any file imported into CONTENTdm is also copied to the storage location specified in the Archival File Manager. Additionally, each file is cataloged in a manifest file (manifest.txt) which is stored in the corresponding volume location.

Archive files can be any file type, including high-resolution images. Your archive file images may then be stored online or offline. See **Managing Archiving Files and High-Resolution Images** for a detailed explanation of creating copies of images you import using CONTENTdm.

You can edit the archival file manager settings if you have rights to configure settings and field properties.

To enable the Archival File Manager for a collection:

1. Open CONTENTdm Administration by selecting CONTENTdm Administration from the Administration menu in the Project Client project or by pointing your Web browser to `http://your.server.address/cgi-bin/admin/start.exe`.
2. Click the **Collections** tab and then click **Profile**.
3. Under the **Optional Collection Settings Summary**, locate the **Archival File Manager** and click **Edit**. Change the full resolution state from **No** to **Yes**.
4. Define your archival file manager settings. For more information see **Archival File Manager** in the Help files.

Section IV: Creating Documents

After establishing desired collection settings such as full text and display image settings and scanning your document pages, you are ready to import the pages using the Compound Object wizard.

In this example we will add a single document that supports full text searching using the Transcript field created in the **previous section**. For the purposes of this tutorial you should already have scanned images and associated text files saved in a local directory.

Below is an example of how image and text files might be named and organized within a directory so you can create a document in CONTENTdm. In this example, a portion of the names of the files will be used as page names.

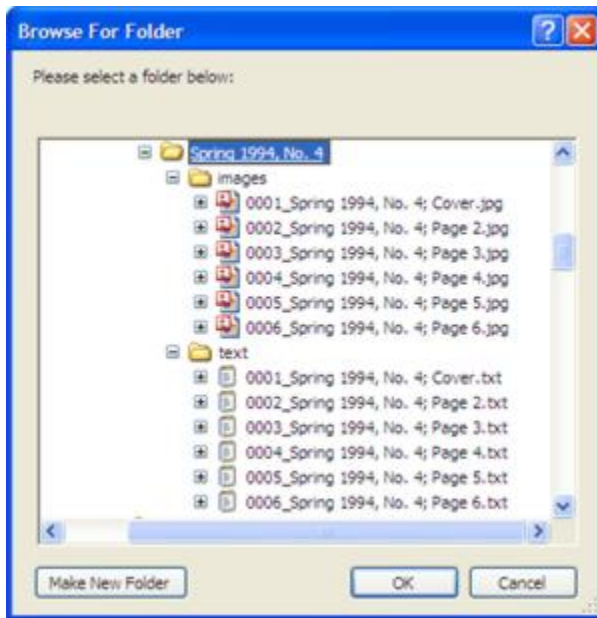


Figure 1. Document Directories

Once you have organized and named your text and image files, add the files into CONTENTdm by following the steps below.

1. Navigate to the Home tab or Project tab. (Or select **Compound Objects** from the **Add** menu.)
2. Click **Add Compound Objects** from the left task pane. The Add Compound Object wizard opens.

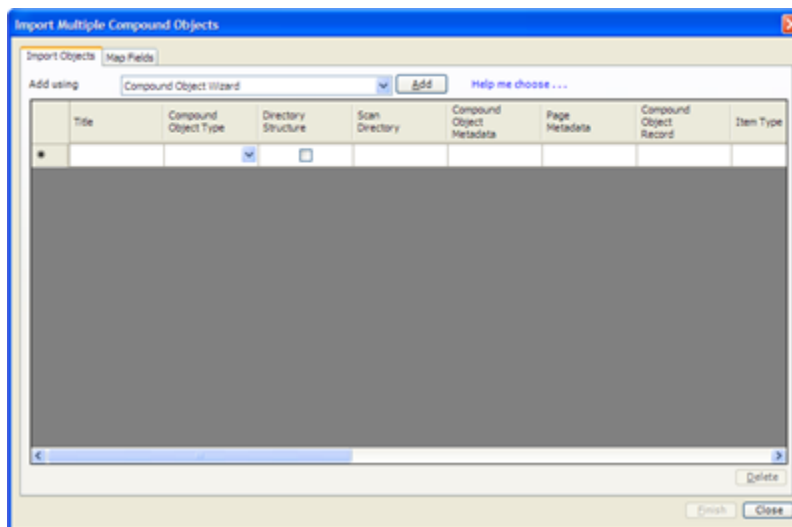


Figure 2. Add Compound Object wizard

3. Select **Compound Object Wizard** from the Add Using drop-down list and click **Add** to start the wizard.
4. From the Choose Type screen, select **Document**. This will enable you to create multiple sequential pages.

5. Select **No** in answer to the question, **Are the compound object structure and metadata defined by a tab-delimited text file.** Click **Next**. The Select Directory screen displays.
6. Click **Browse** to locate the directory that contains your scanned image files. Click **OK**.



Figure 3. Specify Image Files Directory

The directory dialog box closes and you return to the Compound Object wizard. The directory containing your images is displayed in the Directory name text box.

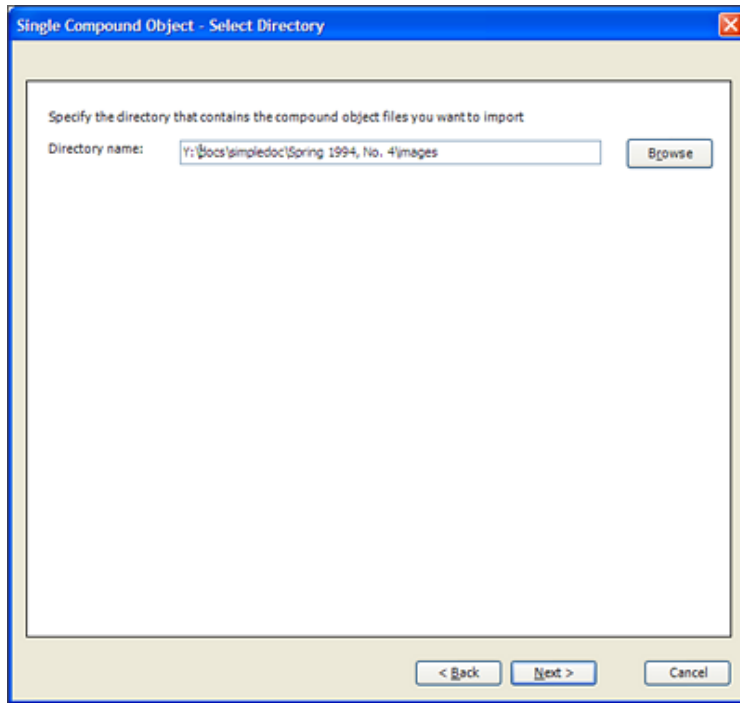


Figure 4: Image Files Directory

Click **Next**.

The Display Image Settings screen displays.

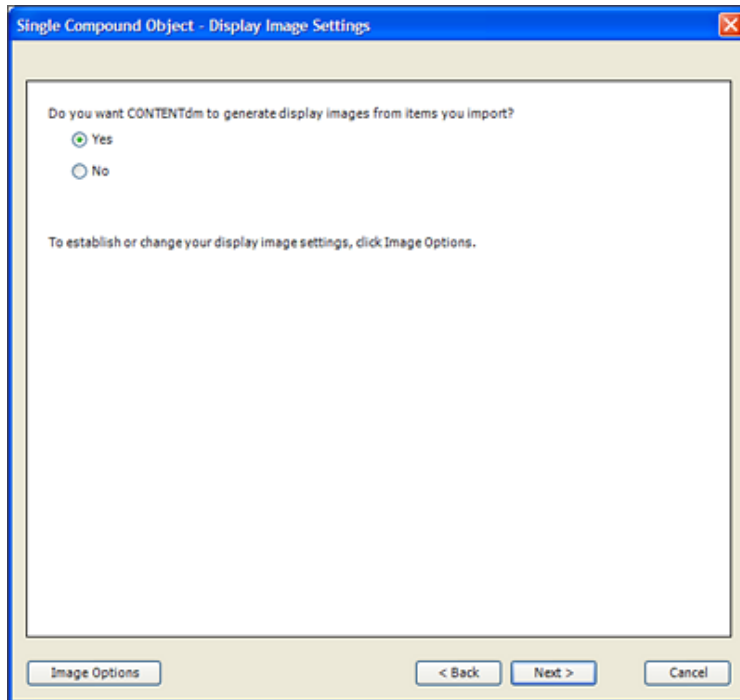


Figure 2. Display Image Settings

7. Select whether to generate display images for the files you are importing. Generating display images helps to speed delivery of high resolution images to end users. Your display image settings may be established already but you can modify them by clicking **Image Options**. When you are satisfied with your settings, click **Next**. The Page Information screen displays.
8. Specify how pages should be named. In this example the file names comprise part of the page names as we specify the following settings: **Use file names as titles** and **Ignore file name before underscore**. The file named *0001_Spring 1994, No. 4; Cover* will appear in the document as *Spring 1994, No. 4; Cover*. Specify the location of the transcript files by clicking **Browse**. Select the directory containing the text transcript files. Click **OK**. (If the transcript options are not available, this is an indication that full text searching has not been enabled in CONTENTdm Administration.) Choose whether to create a printable PDF version of the compound object and then click **Next**.

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 object for printing

Transcript

Import transcript files from a directory

Directory name:

Generate transcripts using OCR

No transcripts

PDF

Create print PDF

< Back Next > Cancel

Figure 3. Compound Object Page Information

9. In this type of import the title of your image directory will be used as the title of your compound object, but can be changed later. When you have finished entering document metadata, click **Next**. The Confirm Settings screen displays.

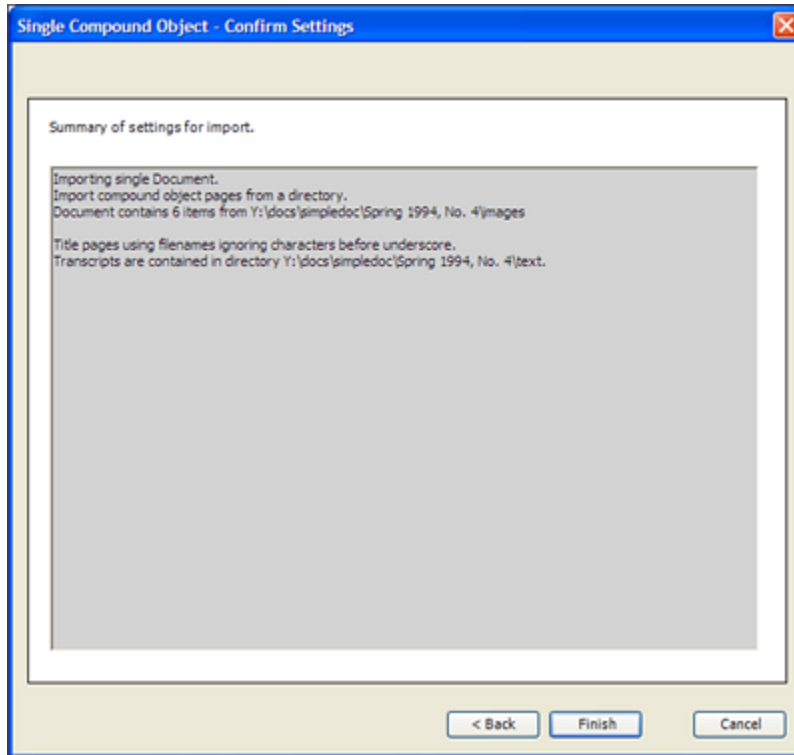


Figure 4. Confirm Settings

10. Review your settings and click **Finish**. (Or click **Back** to edit your settings.) The Summary screen displays along with any errors or warnings. Click **Close**. The Add Compound Object screen displays again, listing the compound object you have added. Click **Finish** to add the current compound object to your project. (Or click **Add** and repeat the previous steps to create another compound object.)

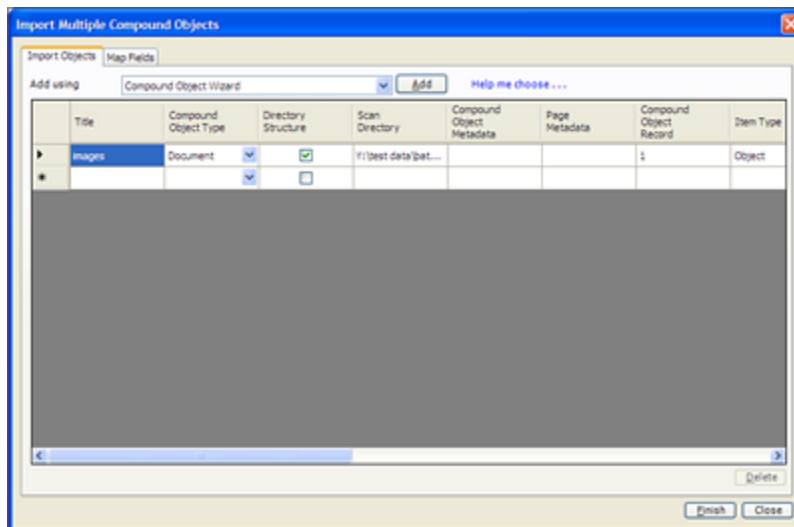


Figure 5. Add Compound Object

11. A progress screen displays as the document is processed. When processing is complete, the Summary screen displays. Click **Close**.

The compound object is displayed in the project spreadsheet.

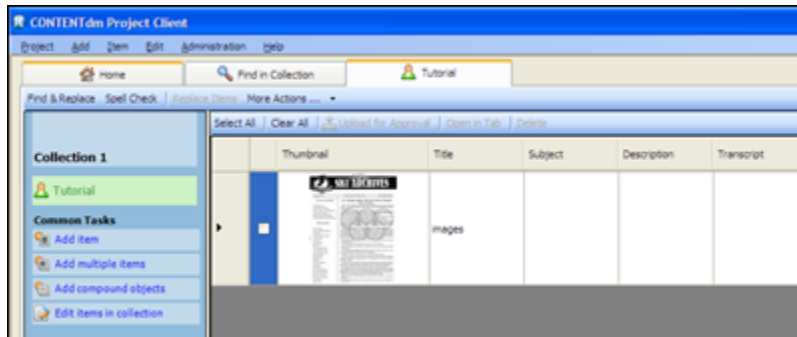


Figure 6. Finish Screen for Document

12. To review or make edits or changes to the compound object, click the thumbnail image displayed in the project spreadsheet. (Or right-click on the compound object thumbnail and select **Open in New Tab**). The compound object opens in a new tab

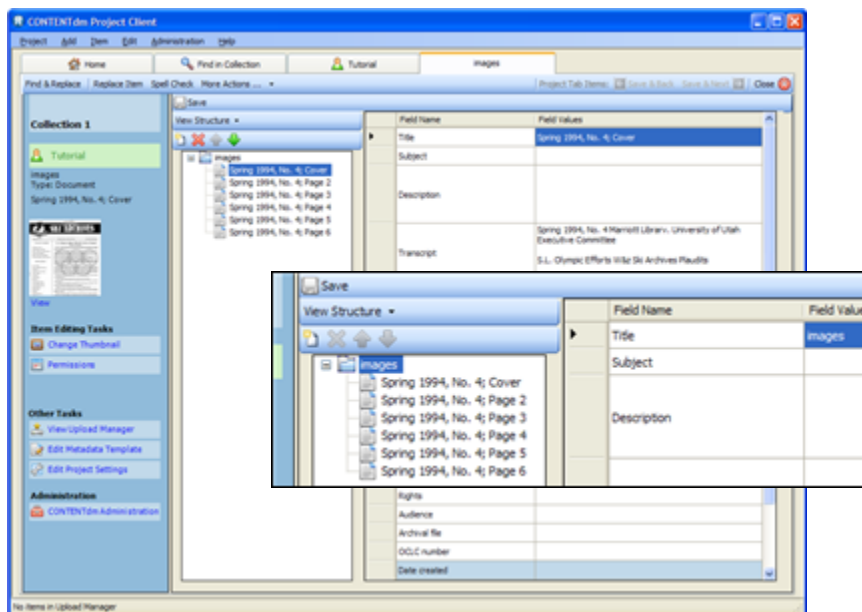


Figure 7. Preview a document

- The topmost folder of the compound object (labeled “images” in the example above) shows the compound object level record (or XML wrapper) with the compound object page navigation below it. You can move, rename or delete pages or rename the XML wrapper.
- If you have rights to edit permissions, the permissions button will be active in the left task pane. You can restrict viewing access to the compound object by user names and/or IP addresses.
- Clicking on a page name displays the page metadata at right and a thumbnail image of the page in the left task pane.
- To view a larger image of the current page, click the **View** link in the left task pane below the thumbnail image.

- Use the **View Structure** drop-down menu to change the view of the compound object navigation pane. This enables you to switch between a structural, thumbnail or spreadsheet view of the object. (Note that the spreadsheet view of the compound object provides **fill functions** and the other editing options that are available when using the project spreadsheet.)
13. In this example, we will rename the compound object level record from "images" to a more appropriate name, "Ski Archives, Spring 1994, No. 4." Click **Save**. Then click **Close** to return to the project spreadsheet.
 14. To upload the compound object from the project spreadsheet, check the box next to the document thumbnail and click **Upload for Approval** to send to the pending queue. Or, to remove the compound object from your project, click **Delete** (this does not affect your source files).

Once the document has been sent to the pending queue, the CONTENTdm Administrator must approve the document and add it to the collection before the document can be viewed or changed.

Tip: When collections are in the developmental stages, CONTENTdm Administrators can designate collections as "not published" so they are not available in a public Web interface but are still accessible to those who are working to build the collections. Learn more about **Collection Configuration** in the Help files.

Once the CONTENTdm Administrator approves and adds the document to the collection, the final document is viewable in the website.

Section V: Creating Monographs

A monograph is a compound object with more than one level of items. Monographs are similar to documents but their pages are organized hierarchically, as in the case of a book with chapters or a report with sections. CONTENTdm supports up to nine levels of hierarchy within a monograph.

Before creating compound objects, be sure to establish any desired collection settings such as the **Archival File Manger** (to add copies of any imported file into a specified storage location) and **full text searching** (to enable the page & text view and full text searching within the compound object) or project settings (such as Metadata Templates to add descriptive or administrative metadata to specified fields).

When creating a monograph, you have the option of importing metadata and defining the monograph structure using a tab-delimited text file or you can use file directories to create the structure. When using a tab delimited text file to create a monograph, all files must be in one directory. Two fields within the tab-delimited text file define the structure. For more information about creating tab-delimited text files, see **Adding Compound Objects with Tab-delimited Text** within the Help files.

If you use directories to define document structure, all image files must reside in a single directory. Organize the pages of a monograph further by using subdirectories to establish chapters, appendices, sections, or any other headings.

Text files associated with your monograph must be stored in another separate directory and saved as .txt files with root file names that correspond to the image file names.

CONTENTdm always imports all of the files in the root directory first and then imports the subdirectories. In the example below, the root directory of images is called *History of the Ohio*

Canals and the first items to appear in your imported monograph will be the files shown below the subdirectories: *0001_Cover.TIF* through *0013_Preface.TIF*. After the individual files are imported then the subdirectories or "chapters" will be imported.

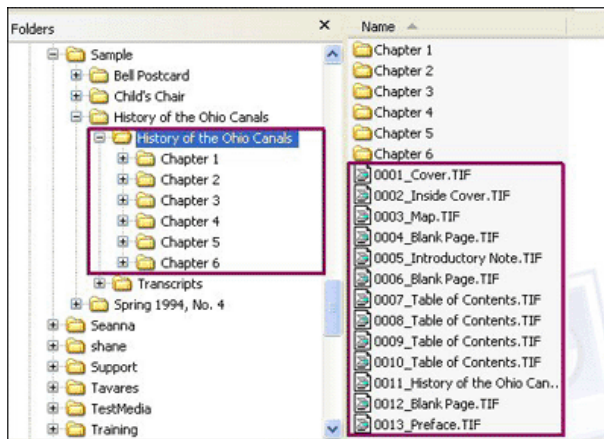


Figure 8. Monograph File Directory

- If you have files within the subdirectories, the files are imported before any directories in a subdirectory. In the example shown below, the directory *Chapter 2* contains files and a subdirectory, *Section 1*. CONTENTdm will import the files *01_Page 15.TIF* through *04_Page 18.TIF* before the subdirectory, so pages 15 through 18 will appear before Section 1 within Chapter 2 of the monograph.

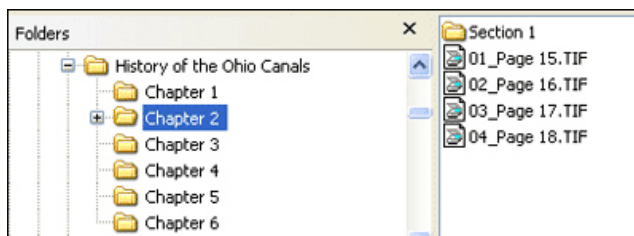


Figure 9. File Import Order

- Order of import is always ascending alphanumeric order. In these examples, numeric prefixes (01, 02, for example) help to establish the order of import.

Note: The alphanumeric sort import occurs independently at each level. So chapter prefixes need not maintain the same numbering scheme as the pages at the root level; this is true also for files in subdirectories.

To import a monograph:

1. Navigate to the Home tab or Project tab,
2. Click **Add Compound Objects from the left task pane**. The Add Compound Object wizard opens.

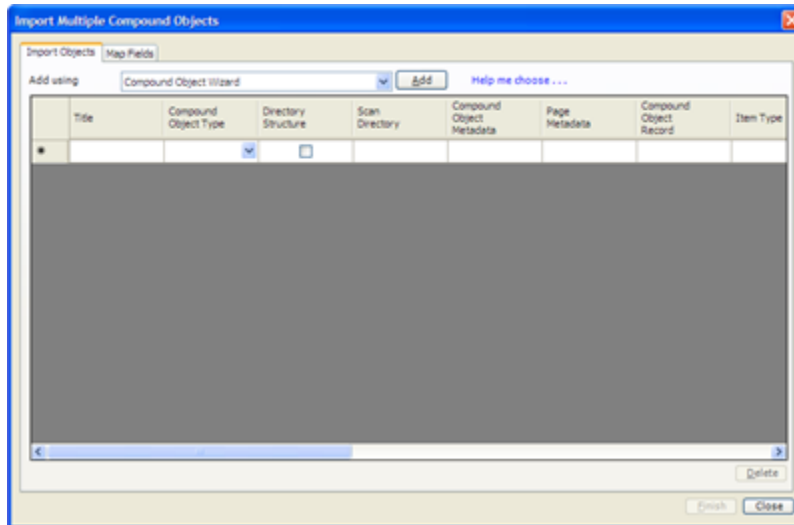


Figure 10. Add Compound Object wizard

3. Select Compound Object Wizard from the Add Using drop-down list and click **Add** to start the wizard
4. From the Choose Type screen select Monograph. This enables you to add multiple sequential pages with hierarchy.
5. Choose **No** in answer to the question **Is compound object structure and metadata defined by a tab-delimited text file?** Click **Next**.

The Select Directory screen opens.

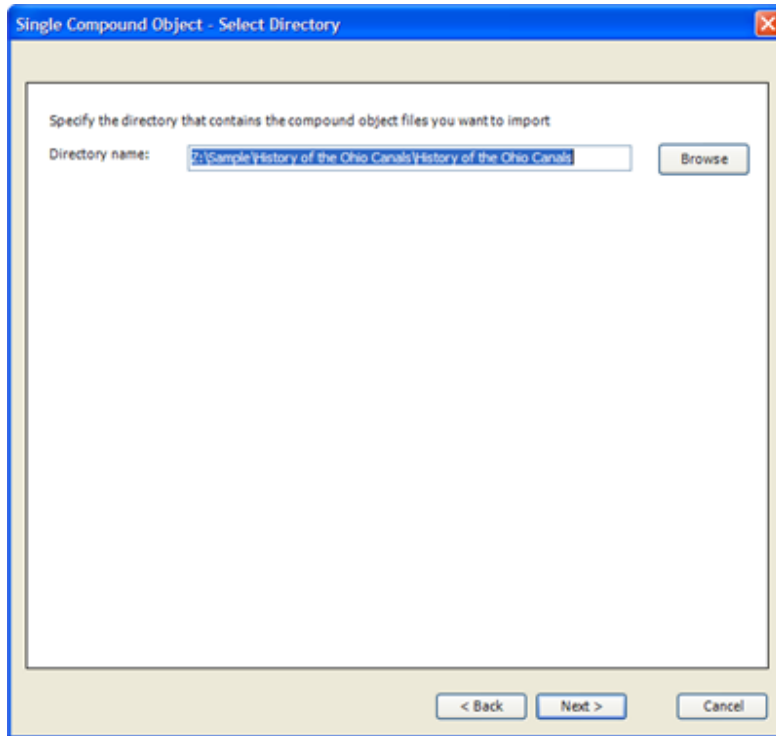


Figure 11. Select Directory

6. Select the directory that contains the compound object files that you are importing. Click **Browse** to navigate to the directory that contains your first level of files. Click **OK**. The Browse for Folder dialog closes.

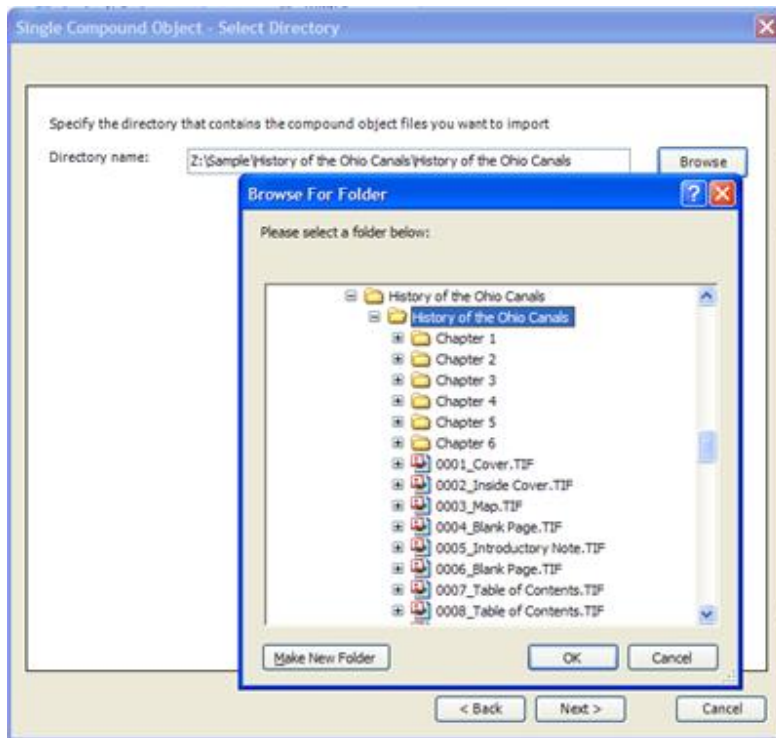


Figure 12. Choose Directory Window

7. Click **Next**. The Display Image Settings screen displays. In this example, we want to have display images generated for the high resolution TIF images we are importing.

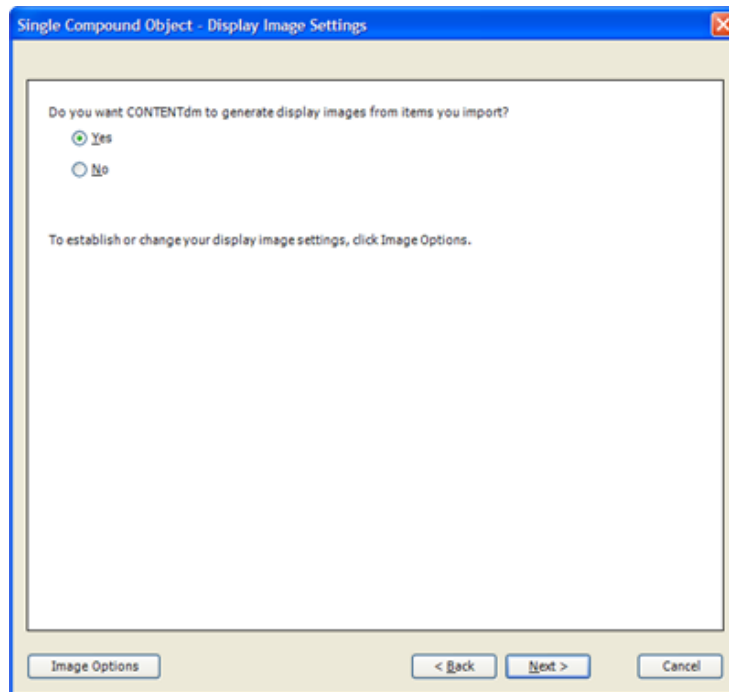


Figure 13. Full Resolution Settings

8. Select **Yes** to generate display images. (If you are adding JPEG2000 images or you want to serve up the high resolution TIF image for display instead, select **No**.)

Display image preferences may be established already by the Collection Administrator but can also be modified by clicking **Image Options**. Display image options include setting image copyrights, basic image processing and image quality. See **Images & Thumbnails** in the Help files for more information.

When you have finished establishing your image settings, click **Next**. The Page Information screen displays.

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 object for printing

Transcript

Import transcript files from a directory

Directory name:

Generate transcripts using OCR

No transcripts

PDF

Create print PDF

< Back Next > Cancel

Figure 14. Monograph Page Information

9. Select **Use file names as titles**. Also select **Ignore information before underscore**.
10. To import transcripts with the compound object, click **Import transcript files from a folder**. Click **Browse** to navigate to the transcripts directory. Click **OK**.
11. Click **Next**. The Confirm Setting screen displays. Review the setting and click **Finish** or click **Back** to make changes.

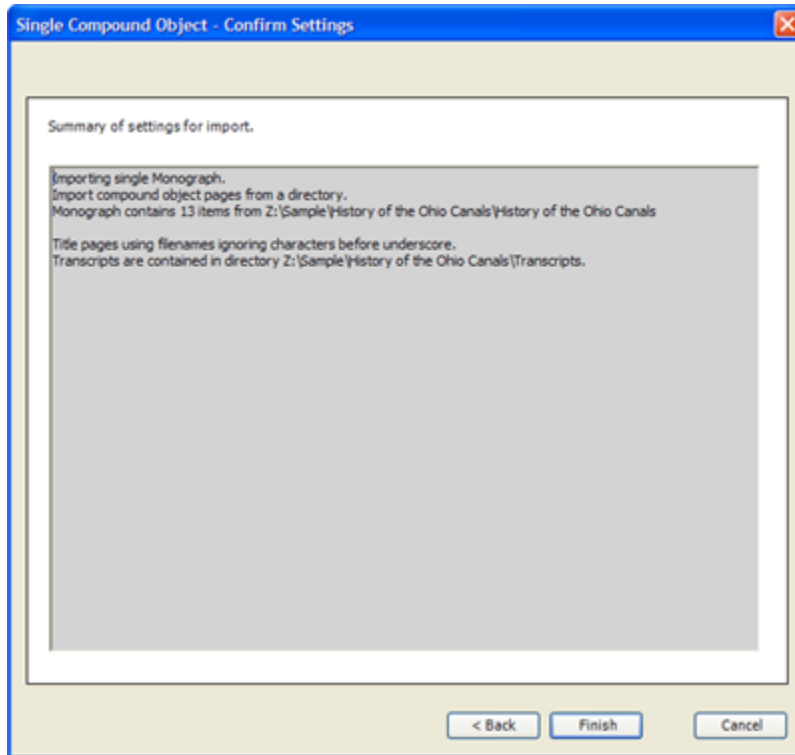


Figure 15. Confirm Settings screen

12. The Summary screen displays along with any errors or warnings. Click **Close**.

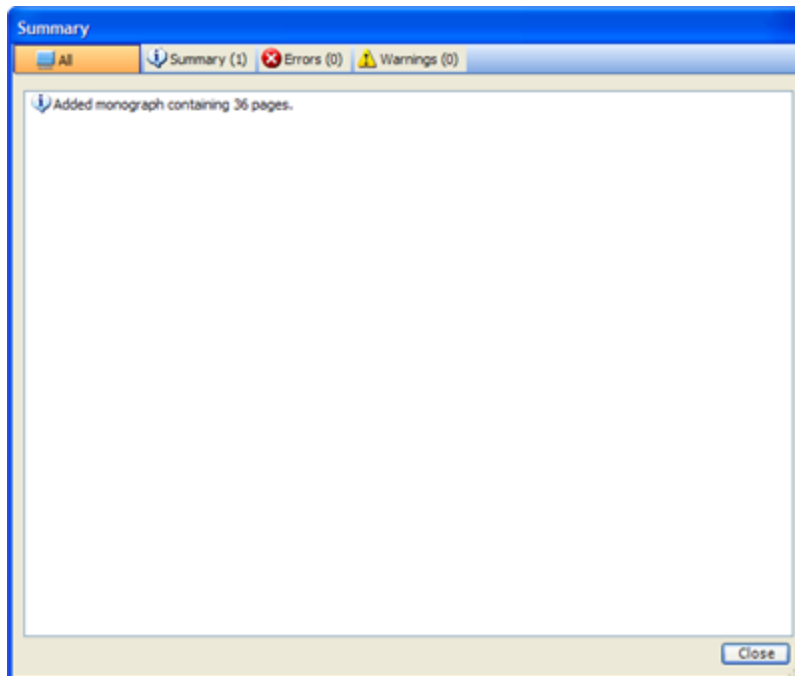


Figure 16. Monograph Summary Screen

13. The Add Compound Object screen displays the monograph in the Add list. Click **Finish** to add the monograph to the project. (Or click **Add** and repeat the previous steps to create another compound object.)

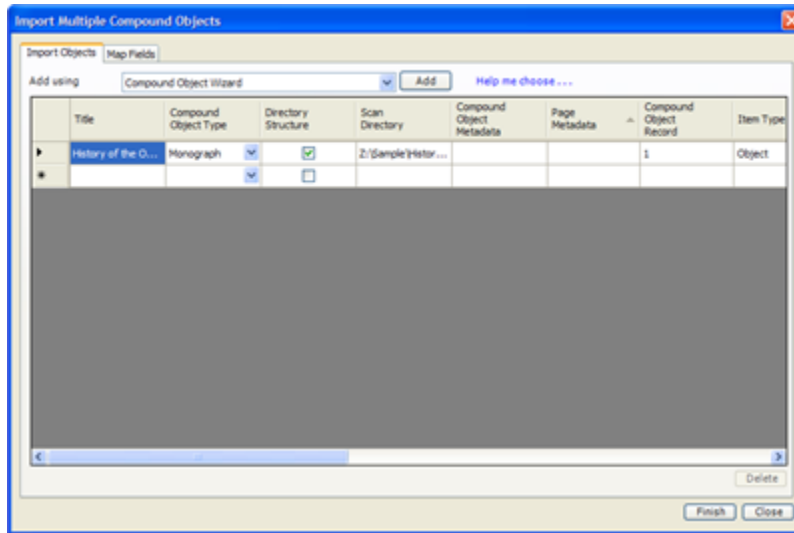


Figure 17. The Add Compound Objects wizard

14. A progress bar displays while the compound object is added to the project. When complete, the Summary screen displays along with any errors or warnings. Click **Close**. The compound object is displayed in the project spreadsheet.

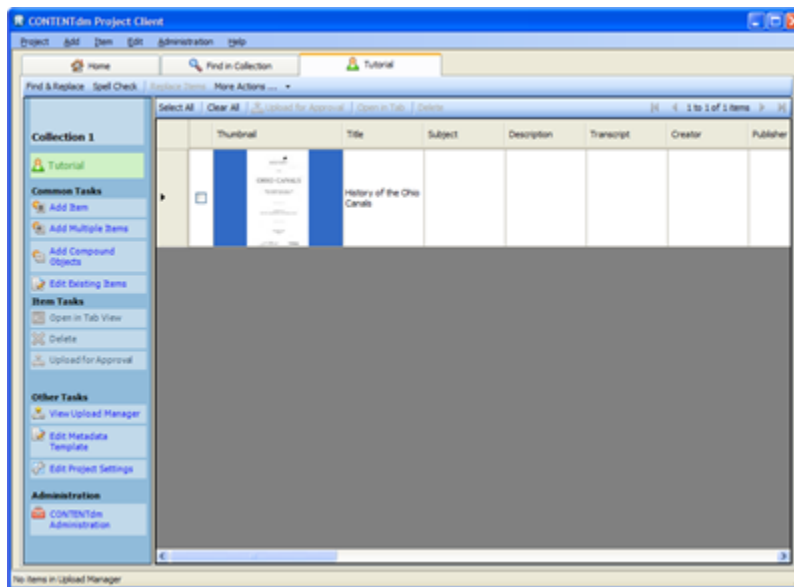
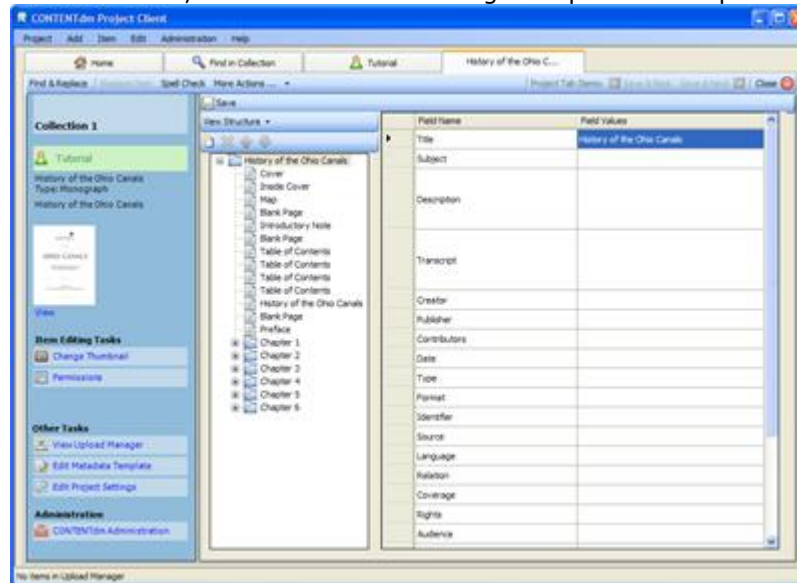


Figure 18. Project spreadsheet

You can enter or edit compound object-level metadata by typing directly into the cells of the project spreadsheet.

To edit page-level metadata, click the thumbnail image to open the compound object



in a new tab.

Figure 19. Monograph displayed in tab

- The left pane shows the structure of the compound object. The topmost item in the list is the XML wrapper that will contain the compound object metadata. All pages and subdirectories are nested within the XML wrapper
 - Open files with a + to view subitems. You can move pages up and down, rename pages, and delete items in this view.
 - If you have rights to edit permissions on an item, the permissions button will be active in the task pane at far left.
 - Clicking on a page title navigates to that page and its metadata. Any transcript files you imported will be displayed in the transcript fields.
 - Use the **View Structure** drop-down menu to change the view of the compound object navigation pane. This enables you to switch between a structural, thumbnail or spreadsheet view of the object. (Note that the spreadsheet view of the compound object provides **fill functions** and the other editing options that are available when using the project spreadsheet.)
15. Click **Save** to keep any changes, then click **Close** to return to the project spreadsheet.
 16. To upload the compound object from the project spreadsheet, check the box next to the document thumbnail and click **Upload for Approval** to send to the pending queue. Or, to remove the compound object from your project, click **Delete** (this does not affect your source files).

Once the document has been sent to the pending queue, the CONTENTdm Administrator must approve the document and add it to the collection before the document can be viewed or changed.

Tip: When collections are in the developmental stages, CONTENTdm Administrators can designate collections as "not published" so they are not available to end users in the public Web interface but are accessible to those who are working with the collections. Learn more about **collection configuration** in the Help files.

Once the CONTENTdm Administrator approves and adds the monograph to the collection, the final monograph object is viewable in the website.