Exercise 8: PDFs: Two Options for Displaying

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

PDFs can be displayed two different ways in CONTENTdm. These two ways require different settings. One way is to display the PDF as one item; the other is to convert the PDF to a Compound Object, so each page of the PDF has its own metadata and is displayed in the Compound Object viewer.

Exercise

If not already open, open your Text project.

Import a PDF as a single item

1) As you did with the adding a single image to your project, use the Add Item function to add a single PDF to your project.

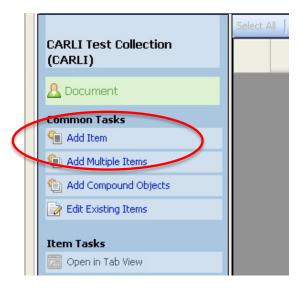


Figure 1: Add Item option in the task bar.

2) Browse to the Exercise data folder on your desktop and choose the PDF called "guidelines_for_audio.pdf".

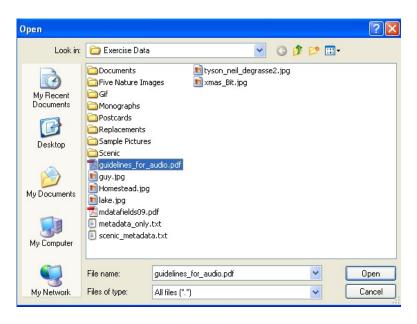


Figure 2: Select guidelines_for_audio.pdf to import.

3) Add the item and close the summary box. The PDF is now in your project spreadsheet.

4) Edit the metadata and change the title of the item to "Guidelines for Audio, Single Item" and add further metadata you wish. **Save the changes**. *Notice that some metadata has already been added for you*—this is from the Metadata Template we set up in an earlier exercise. Also notice that the text of the PDF has been automatically extracted and added to the Transcript field.

		Field Name	Field Values
	•	Title	Guidelines for Audio, Single Item
o, Single Item		Subject	
o, single item		Description	
		Creator	
	1	Publisher	Amy Maroso
Andre Statution (Friend) Control Market Statution (Statution)		Contributors	
		Date	
		Туре	Text
	1	Format	pdf
		Source	
		Language	eng
s ks bnail		Relation	
bhail		Coverage	
	1	Rights	Contact CARLI at support@carli.illinois.edu
		Audience	
lanager	1	Collection	CARLI Test Collection (CARLI)
Template			CARLI Digital Collections Users' Group Revised: 11/1/09 1 GUIDELINES FOR THE CREATION OF DIGITAL COLLECTIONS
ettings		Transcript	Digitization Best Practices for Audio This document sets forth guidelines for digitizing audio materials for CARLI Digital Collections. The issues described concern sample rates, bit-depths, file formats, and equipment for the public be differentiated concerns of audio materials. Reclargence differentiates and audio materials.
Administration	-	Identifier	guidelines_for_audio.pdf
Hummisci du011	2	Full resolution	
		OCLC number	

Figure 3: Text from PDF extracted to Transcript field automatically.

5) Upload this PDF to your approval queue.

Import a PDF as a Compound Object

Enabling the PDF-to-Compound Object Conversion Option

CONTENTdm has the ability to automatically convert PDFs into Compound Objects. Once converted, each page of the PDF will be its own page in a Compound Object and have its own page-level metadata. The text of the PDF will be extracted to the Full Text Searchable field on each page, rather than in one block as it is with a single item.

By default, your collections are not set up to do this PDF-to-Compound Object conversion. You need to turn this feature on.

NOTE: Turning on the PDF conversion option can be done in CONTENTdm Web Administration or in a project in the Project Client. At this time it is best practice to turn the feature on in the Project Client since the Project Client settings override the server settings.

- 1) If not already open, open your "Text" project.
- 2) Click on the Edit Project Settings option in the left-hand task bar.

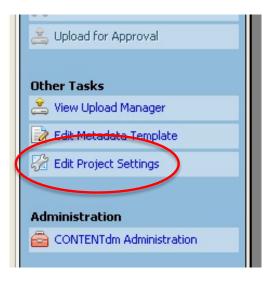


Figure 4: Select Edit Project Settings

3) Click on the Processing option in the settings menu.

Project Settings Manager - 7	fext	X
General Settings Metadata Templates Metadata Field Types Images & Thumbnails Processing OCR Project Options Find in Collection	Processing Settings - Text Specify how files are processed when imported PDF File Conversion You can import multiple-page PDF files to display in your collection in the Compound Object Viewer • Do not convert multiple-page PDF files to compound objects • Convert multiple-page PDF files to compound objects PDF print quality: PDF print quality:	
	Finding Aids When finding aids (XML files) are added to a project, CONTENTdm converts them to compound objects and also creates full views. You can format the HTML page that provides the full view by defining the Extensible Stylesheet Language (XSL) used when processing the XML files. Is Use CONTENTIAM XSL file (default) Use a custom XSL file Browse	

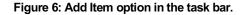
Figure 5: PDF Conversion options found under the Processing features

- 4) Change the PDF File Conversion option to "Convert multiple-page PDF files to compound objects" option.
- 5) Click OK to save your settings and enable the PDF conversion option.

Using the PDF-to-Compound Object Conversion Option When Importing PDFs

1) Import a single PDF using the Add Item option.

CARLI Test Collection (CARLI)	Select All
<u> Document</u>	
common Tasks	
1 Add Item	
Add Multiple Items	
🔞 Add Compound Objects	
📝 Edit Existing Items	
Item Tasks	
🛅 Open in Tab View	



2) Browse to the Exercise data folder on your desktop and again choose the same PDF called "guidelines_for_audio.pdf".

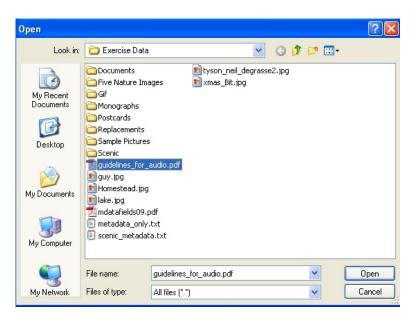


Figure 7: Select guidelines_for_audio.pdf to import.

4) Add the item and close the summary box. The PDF is now in your project spreadsheet.

5) Double-click on the item's thumbnail to open the metadata editor.

Notice that the PDF now appears as a Compound Object and the metadata can be edited as you would a Compound Object.

II Check More Actions		
Save		
View Structure +		
	Field Name	Field Values
guidelines_for_audio		 guidelines_for_audio
Page 2	Subject	
Page 3	Description	
Page 4 Page 5	Creator	
Page 6	A Publisher	Amy Maroso
Page 7 Page 8	Contributors	
Page 9	Date	
Page 10	Type	Text
Page 11 Page 12	Format	pdf
Page 13	Source	
	Language	eng
	Relation	
	Coverage	
	A Rights	Contact CARLI at support@carli.illinois.edu
	Audience	
_		CARLI Test Collection (CARLI)
		CARLI Test Collection (CARLI)
	Transcript	
	1dentifier	guidelines_for_audio.pdf
	Full resolution	
	OCLC number	
	Date created	

Figure 8: PDF has been converted to a Compound Object.

- 3) Edit the metadata and change the title of the item to "Guidelines for Audio, Compound Object". Again, some metadata has already been added for you—this is from the Metadata Template. The text from each page of the PDF has been extracted into the Transcript field of the corresponding page in the Compound Object.
- 4) Use the View option and choose to view the spreadsheet version of the PDF compound object

0	💊 Find in Collection	<u> </u> Text	1
Spell Che	ck More Actions 👻		
	🖵 Save		
	View Structure 👻		
	View Structure		
	View Thumbnails		
	View Spreadsheet		
	Page 2 Page 3 Page 4 Page 5 Page 6 Page 7 Page 7 Page 8 Page 9 Page 10 Page 11 Page 12 Page 13		

Figure 9: View the spreadsheet of the compound object

- 5) Even though the PDF has been converted into a Compound Object, in some ways CONTENTdm still treats it like a single item. Because of this, required metadata fields need to be filled.
- 6) You will see that the spreadsheet view of a compound object looks just like the main project's spreadsheet. You can view each page of the compound object as part of the spreadsheet and edit the metadata.
- 7) Use the right-click "Fill options" that we used in a previous exercise to fill all the metadata for each page's Type, Rights, and Collection fields.
- 8) Close the item and Save your changes.
- 9) Upload the item.
- 10) Approve, and index the items in your approval queue.
- **11) View** these items in the collection to note the difference between displaying a PDF as a single item and as a Compound Object.