

CONTENTdm

CONTENTdm Basic Skills 2: Working with text in CONTENTdm

Mindy D. Tran

Senior Product Trainer



Build a Collection – Six Basic Steps

1. Add a collection
2. Configure a collection
3. Create a project
4. Configure a project
5. Add and upload the digital items
6. Approve and Index

**CONTENTdm Administrator /
Collection Administrator**

**Project
Client
Operators**

**CONTENTdm Administrator /
Collection Administrator**

Agenda

- Configure a collection for textual materials
- Compound objects
 - Postcard
 - Document
 - Monograph
- PDF files

COMPOUND OBJECTS

Compound Object Types

- Postcard (i.e. 2-sided item)
- Document
- Monograph
- Picture Cube



Configuring the Collection for Text

Collection metadata fields:

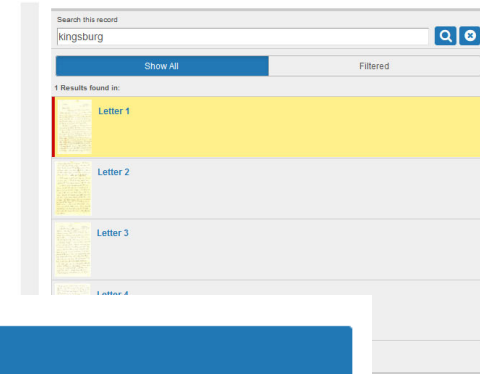
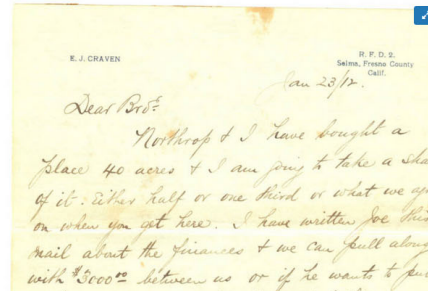
- Create a field to **tag** each related items in a compound object (e.g. Books_001)
- Add a **transcript** field in which to embed the full-text

Metadata Fields—Full Text Search

Collection field properties

View, add, edit and delete fields. Enable full text searching and controlled vocabulary. After you have added, changed, or deleted fields, index the collection to update changes.

	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab	add field
1	Title	Title	Text	No	Yes	No	Yes	No	
2	Subject	Subject	Text	No	Yes	No	No	No	
3	Description	Description	Full Text Search	Yes	Yes	No	No	No	



Rich discovery for your end-user

Transcript

Jan. 23/12

Dear Bros.

Northrop & I have bought a place 40 acres & I am going to take a share of it. Either half or one third or what we agree on when you get here. I have written Joe this mail about the finances & we can put along with \$3,000.00 between us or if he wants to put up \$8,000.00 well & good. It is up to him.

The place in question is 6 miles from here. South East & 9 m. from Northrop. Fine roads as level as a floor & about 35 min. run on a bike.

You will be 2 miles from **Kingsburg** on a main road. Mail route & telephone. There is a 3 room house, large new barn, hen houses, cow shed & tool shed. All trays sweat boxes & farming implements, mower & rake go with the place also two Black mares & work harness. The mares are 16 years old and yud & yentel (good & gentle) 17 acres peaches 12 acres Muscat Vines 7a alfalfa & 3 acres Gum trees, 1 ac farm yard & buildings. You will need to buy a cow

Optical Character Recognition (OCR)

- Allows text characters in image file to be searched
- OCR within CONTENTdm or another OCR software
- Accuracy of any OCR materials dependent upon:
 - quality of scanned image
 - quality of original document
 - presentation of characters in original document
 - font face used in original document
- OCR can be performed on JPEG, JPEG2000, PNG, GIF, and TIFF files

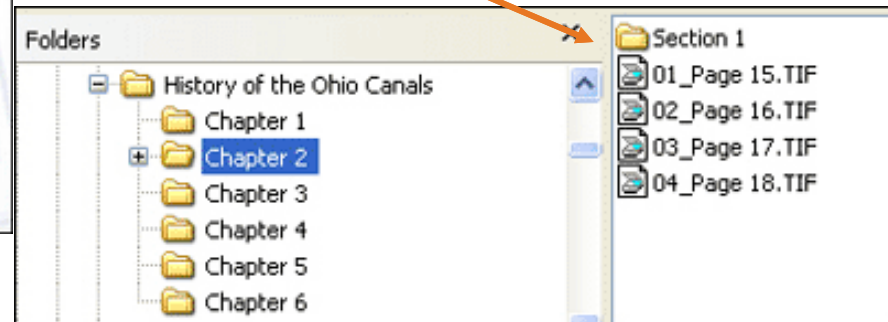
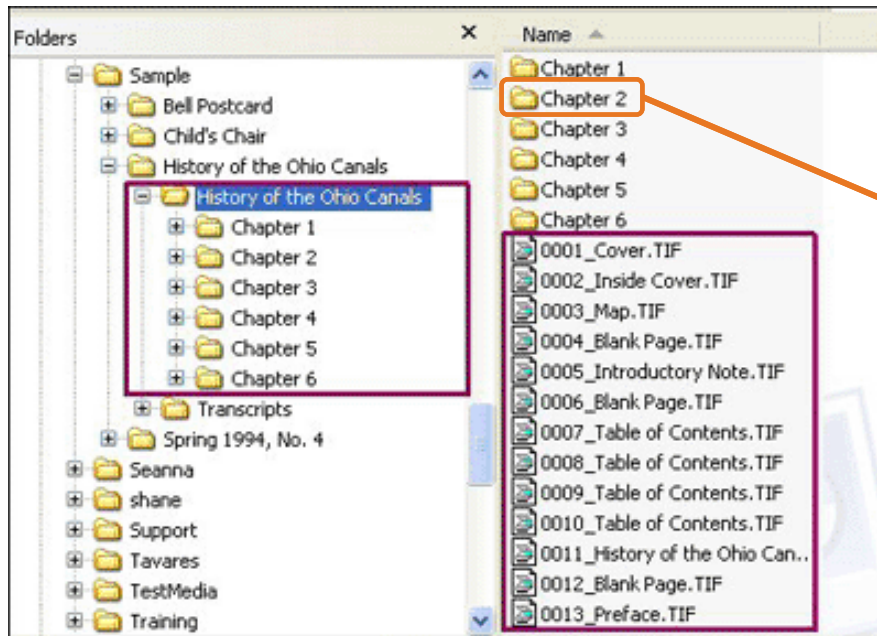
Compound Object—File Organization

Must have in place both the image files and the optional transcript files—appropriately named and organized



Image File Name	Page text file name
bell1.tif	bell1.txt
bell2.tif	bell2.txt
bell3.tif	bell3.txt

Additional File Organization for Monographs



Build a Compound Object

Process:

- ✓ Add compound object
- ✓ Select object type
- ✓ Identify the location of the “image” files
- ✓ Identify the location of the “text” files (optional)

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

DEMONSTRATION: IMPORT COMPOUND OBJECTS

PDF FILES

PDF File Characteristics

- **Good file type for material initially created in a digital form because will have embedded text**
- **Image files do not necessarily contain embedded text**
- **Adobe required to read files**



PDF Files in CONTENTdm

- **A PDF file can be treated as**
 - **Single file**
 - **Compound object**
- **Advantage of converting to compound object:**
 - **Each page of the PDF becomes a page with its own metadata record**



DEMONSTRATION: IMPORT PDF FILES

Questions?

Contact OCLC Support in your region
<http://oc.lc/support>

1-800-848-5800



A screenshot of the OCLC support website. At the top, a blue banner reads "Where to go for information or help" with the subtext "Get answers. Give feedback. Nominate a Support Star." Below this, the page title is "Contact OCLC Support in your region". There are three tabs: "Americas", "EMEA", and "Asia Pacific". On the left, a sidebar contains "Home" and "Contact Support" with sub-links for "OCLC offices" and "Regional distributors".

Thank You!

Please remember to fill out the evaluation for this class

- Instructor: Mindy Tran
- Category: Digital Collection Service
- Class name: CONTENTdm Basic Skills 2 – Working with Text in CONTENTdm

Training questions? Contact OCLC Training training@oclc.org/