

# **MULTIPAGE OBJECTS**

ntroduction	3
Creating Multipage Digital Objects	3
Jploading Multipage Objects	3
Adding A "Hero" Shot	3
Creating Records For Text Objects	. 4
Adding Pages From Multiple Files	5
Step 1: File Names	5
Step 2: Uploading Multiple Files	. 6
Adding OCR/Full Text	. 9
Adding Pages From A Single PDF	11
Step 1: Add Record With File	11
Step 2: Splitting PDFs	12
Nanaging Multipage Pages & Files	13
Batch Relabelling Of Pages	13
Reordering Pages	14
Batch Deleting Pages	16
Editing Page File Details	18
Replacing Individual Pages	20
Download PDF Options	20
Page-Level Downloads	20
Object-Level Downloads	21
Multipage Display	21

## INTRODUCTION

This manual shows how to work with multipage items like books, scrapbooks, brochures, pamphlets, etc. See the Adding and Managing records manual for guidance on uploading single-page documents; see the Indexing Articles & BMDs and Newspaper Publication & Issues to see how to manage collections of text fragments.

VITA text management features help you upload and manage the variety of complex text objects you might find in your collections for better discovery and display.

Use this manual to walk through the process of uploading multipage objects either from image & text files or readable PDFs.

## CREATING MULTIPAGE DIGITAL OBJECTS

- Scan each page of your object separately
- Scan text pages at 300dpi and save in TIFF format for master copies
- Name the scans for each object with a consistent naming convention e.g. BookTitle\_oooo1.tif, BookTitle\_oooo2.tif, BookTitle\_oooo3.tif...etc.
- See the File Naming Conventions document for more information
- Either derive JPGs for OCR processing and JPEG2000s(JP2) for zoomable viewing or use a PDF software to create readable PDFs
- Upload your JPG/JP2 and TXT files or PDF to VITA

## UPLOADING MULTIPAGE OBJECTS

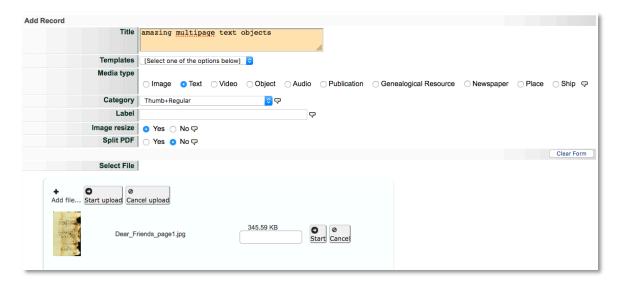
Like all records, you want to establish the record with a representative image for the object. If you are uploading a multipage PDF, the cover page will automatically be assigned this role (and can be overwritten – see below). For multipage items being loaded as image/text file pairs, add a "hero shot" first, then load the pages.

## ADDING A "HERO" SHOT

First, create a display image for the multipage object:

- Go to the Main Menu→Add record with file
- Add Title
- Select Media Type "Text" assuming this describes the primary content of the multipage object (You can add Secondary Media Types later)

- Select "Full+Reg+Thumb" or "Reg+Thumb" from the Category drop down list
- Adding a Label is unnecessary for this category
- File re-size: Select default "Yes"
- Split PDF: default is No unless you are working with a multipage single file PDF (see below)
- Browse for the file you want to represent this multipage object (e.g. a front cover or first page)



- Click Start Upload
- When file is loaded, click the "Continue" button

#### CREATING RECORDS FOR TEXT OBJECTS

As soon as an image is uploaded, you can start to create the metadata for the text object.

For full instruction on adding metadata and using the VITA templates, see the Adding & Managing Records or the VITA Metadata Application Profile documents.

## ADDING PAGES FROM MULTIPLE FILES

Go to the File/Tech Data screen. You will see your initial image listed like this



To add pages to your record, click on "Associate multiple files with this record" and follow these steps:



#### STEP 1: FILE NAMES

Here you can save a lot of time by automatically assigning page labels by "stripping" the common or "consistent" parts of the file names for all the page files you are uploading. Do not include preceding zeros from the file name as these will be stripped out automatically.

E.g. the file names for my pages files **all** include "Dear\_Friends\_" (the prefix) and zeros that precede the page number; all the files are JP2s, so the extension or suffix is ".jp2".

**Note:** we recommend uploading page files as JPG 2000s because of the improved panzoom and rotate options. See the Working with JP2 manual for more information.

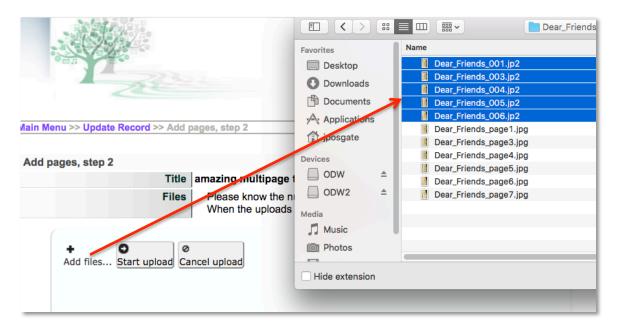
Add pages, step 1			
Title	amazing multipage text objects		
File Name constants	Dear_Friends_ Pr	refix .jp2	Suffix
File Name constants	If your pages ha and suffix (e.g. " page. You will be	ave a consistent na ".tif") in the boxes be able to further e	aming pattern with tabove. These will be dit that label in a fo
Change file size (page images only)	• Yes O No	Q	
	on to Step 2		

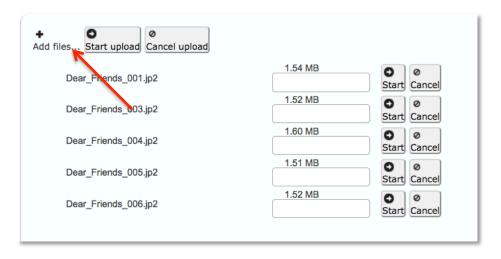
Allow the tool to resize your image files (settings for page size can be reviewed in Agency Management>Image Options)

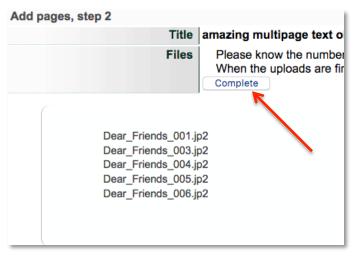
Click on "...on to Step 2" button.

## STEP 2: UPLOADING MULTIPLE FILES

This is a batch file upload screen. Choose the files from your local drive and click "Start uploading" then click the "Continue" button once all the files have completed their load.







Wait while VITA processes your page files:

Jess Posgate 3p (Test PG & Don't navigate away from this page until the "continue" message appears at the bottom.

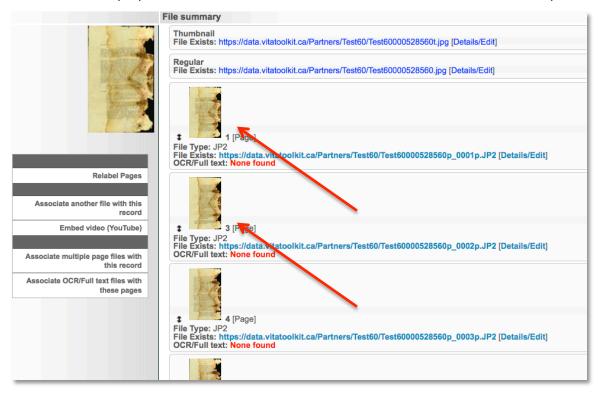
#### Then click "Continue"

Don't navigate away from this page until the "continue" message appears at the bottom.

5 processed
You completed Step 3 by adding 5 files at 6 Jan 2020 15:27:18

Continue...

New files will display in the File/tech data screen and can then be edited individually.

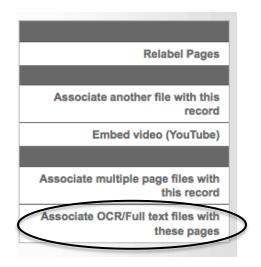


Note that the page labels match the unique variables in the file names, but with the full file name stripped away.

## ADDING OCR/FULL TEXT

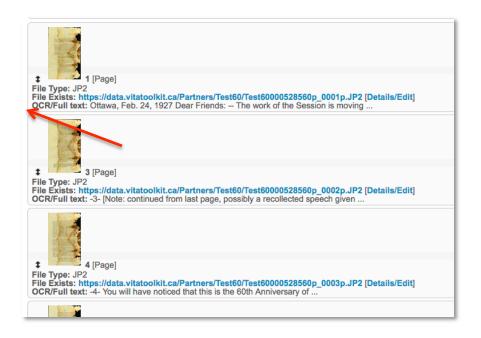
If working with JPG or JP2 and TXT files (as opposed to readable PDFs), remember the JPG/JP2 and TXT files must have **exactly** the same file names to associate with one another (only suffix should vary).

On the File/Tech data screen click on Associate OCR/Full text files with these pages



Browse and select the text files (.txt, .rtf or .pdf). Follow the same batch upload steps as for the page images.

On the File/tech data screen, you'll now see a snippet of the available text from that file:



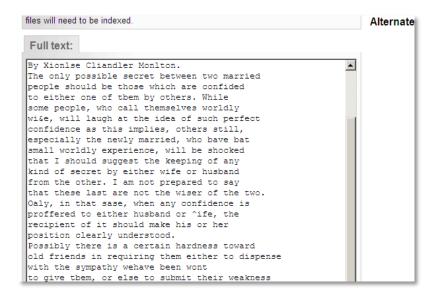
**Note:** Some PDFs and other proprietary text file formats (RTF) will "bulk up" your text content with wrapper code; whenever possible, save text content as plain text (.txt) or clean out the bulky content before uploading for better full text discovery.

## E.g. This is the text extracted from a digitally-born PDF

```
Full text:
xœc'''c''Ra'dàøÙ
                                                                        •
$ € ™Á$ŒÏñ{% X= 30,1 3ŏ600äéMâOVVŽâq`)vbð4
                                                     •‰Ý5 ¾/þl 'l
€° `éÖq[fÆNő'ÊAçrrDó(ÇöOñ8'Úéüw&* * ^Pæ¾íb¯j2 =
Äp= ĐæÚỗl ïÞ²úÅÈ »5 7?‡(µ~lçS|/-- Ô³.`Hò3£¬ñhỗUQÎÔWeBnJ> à -
-|Có ) Æ 'ÄM2YB¿DÇ g-¿£TÇ
D'Á;¦ëÒüBô<^-|ïË^Ú-..à5khTRýi¾ÿ| ÄÍüQØp<u**!Âçg=[ O¦Sq%
6jÕ'®4¬Î'lE;-SaŸx,á4ıÕðsmž|°å
¶~hë [üžÇi ëw*"÷
ü7° = "O#ÊÍ'@XÌ+ú;VÎ/,Žp,Ñ
Xû^ ãe ¾g-É
ÎOM"∢ê ±Œ
OúÏ Nî^¹" %pf /-Öà• f ÚëC9 q ê/àpfWŠý ‡" ^SÛÜýìËSM…[]òÁt)
 .'.¿'ûVöÊ ¿÷¿o {$<ÜlÂIãm +
                                      ⊗TÛ S⊗7Sœ#Lls"
Î9ðå¥l¿aï¶À|† >é³ô [¾Ÿ™ő2ú
Gîpê, ow\¶Cüê.ç? ¡ŠQ'PĭĒ.m>s*ô^¥š¼e'¥ôD?4(ö% [ \J"(Ê)å |...
|n Mœ|9ŠvUÀG pzÝ ŸN[ ^+Đ'¹Jäê«,TŏWí ï Á÷ ¡e/X•4Yøð!
=^×Lĭ &å©Ö ;f°X- ײ'~ÊÐX·"é<é.©R);R^[C`vx/ ôè,øŐ\ëŽpÊ m
Sxyü"Ô4 ;ñ°w2[?,°Ó' âÆZÁsŪ®ŪÞÈ;S‡©!Rà?$o\f:ŽD°¿kf#g
Èf©°K·2À[¬Ū' 2Ó \PŽe•Ù8, ì D ØcöVŪê8»Q™âÑ#-¦ fé4Ğ`» »
ïš4 K1¥Þ§ CVݾ ⊗Õ<£ií¼WV° 8RÒ%$f¦6À`v.Ã] S÷ÓÔ4Íž6ÑÎjH
lÄ=¶§¯+ 0¦`ÿ6Jú:ñRÄåëãñjà çq W÷: jC...€...wúÃÑ x,)Lk42ÍÙæø,K ñ?
fZf֏L£`ì - ~-m,,vš%®o¶' O",ß %M"J"';ú6'fœvàlÄl4
```

Vs.

This text extracted from the same PDF by doing a simple "select all" from the PDF, then copying and pasting into a text editor like TextEdit or Notepad, saving as .txt and uploading as outlined above.



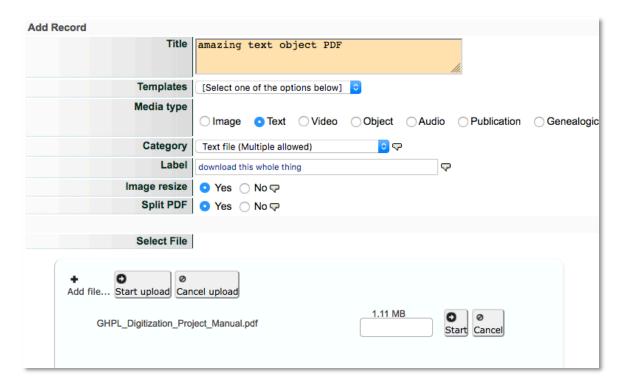
**Note:** You can edit full text on the details/edit screen for each file, but there is a character limit to that display that is often passed by newspaper page content. It is better to prepare your full text files and clean them before uploading page text files.

#### ADDING PAGES FROM A SINGLE PDF

Adding a single PDF for an issue means you do not have to strip the file names and, if it is a readable PDF, the text is automatically extracted for full text search.

#### STEP 1: ADD RECORD WITH FILE

- Go to the Main Menu→Add record with file
- Add Title
- Select Media Type "Text" assuming this describes the primary content of the multipage object (You can add Secondary Media Types later)
- Select "Text" from the Category drop down list
- Adding a Label for the full object download (Optional)
- File re-size: Select default "Yes"
- Split PDF: "Yes"
- Browse for the file you want to upload



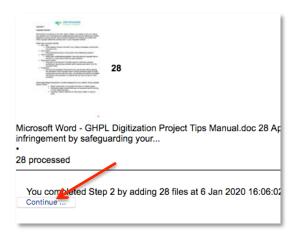
Wait for file to load, then click "Continue"

## STEP 2: SPLITTING PDFS

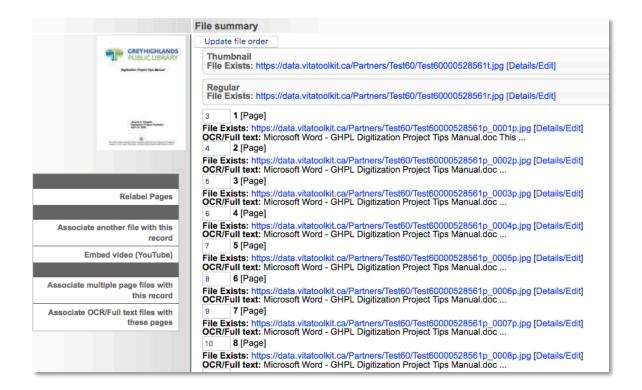
On the next screen, the tool needs to process the single PDF into multiple pages, extract the text and create the Thumbnails and other associated files for each page. Before it does that, you'll be asked to "Continue" to begin processing.



Wait for the file to be processed completely and then click the button at the bottom of the screen:



On the File/Tech screen, you'll see the files generated from your PDF, including the Thumbnail and Regular image files, every page with extracted text snippets, and a complete PDF that is automatically associated with the record.



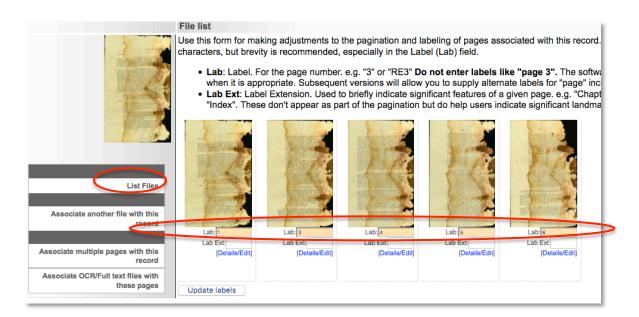
#### MANAGING MULTIPAGE PAGES & FILES

#### BATCH RELABELLING OF PAGES

Click the Relabel Pages link

Add new label numbers for all pages.

Optional (for newspaper issues): add section name (e.g. Seasonal Insert, Sports, etc)

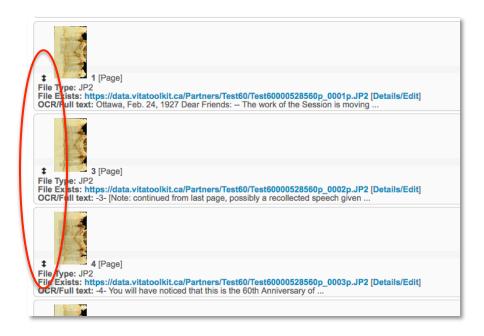


## Click **Update Labels** to commit changes

To return to File/Tech data screen, click List Files link in sidebar

#### REORDERING PAGES

For issues with less than 20 pages, use the elevator buttons by clicking and dragging to reorder pages as required.



For items with more than 20 pages, the page "order" will appear beside the page label in a numeral box. This order starts with the item Thumbnail (assigned order 1), then the Regular (assigned order 2), and then Page 1 (assigned order 3).

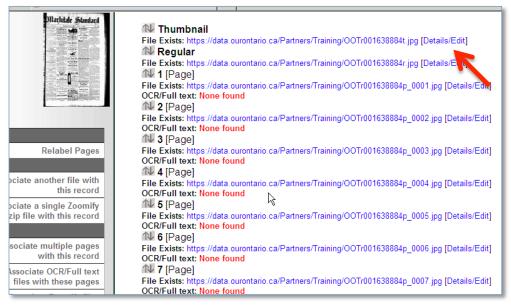
To reorder, either renumber the whole set and click update or, to insert one page between two others, use ##.5, i.e. if the first page is somehow buried in the rest of the pages, assign the page order number 2.5 and click update. The first page will then be moved into the #3 position and the rest of the pages will follow; unless the labels were correct for each page file, you will need to relabel the pages if the order is changed. See Group management for more details.

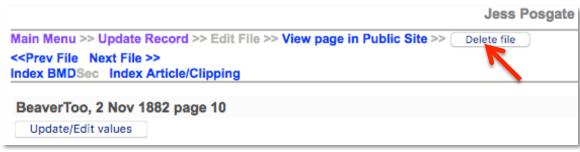
```
1 [Page]
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0001p.JP2 [Details/Edit]
OCR Full text: P, "BoOoe|1LECescTTLoaws a the we*#t; f'Te'; §s...
         2 [Page] Non public
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0002p.JP2 [Details/Edit]
OCR/Full text: (yg'ntjiii'h,','iai..ii,rti I)yriaE',l'IIIIIPrII."""" III. f . 'l'A2tti ...
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0003p.JP2 [Details/Edit]
OCR/full text: i, . - - - ,...' (5'73"! 'yr new: f .1 cJ..- A at?» , i2 - ...
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p 0004p.JP2 [Details/Edit]
OCR/Full text: n. member for Frontsuac. Hitherto the 8ir AENRY SMITH Government _ ...
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0005p.JP2 [Details/Edit]
OCR/full text: --" ' yii-T V J," if"? , -, ' . debate . we all members returned upon ...
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0006p.JP2 [Details/Edit]
OCR/Full text: SE T U PE hx Ho y q » wihie )ns Hon, J. S, MNcDONALD intimated that ...
         7 [Page]
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0007p.JP2 [Details/Edit]
OCR/Full text: NOR SUBMIT TO ARBTTRARY WEASUIET nE TT e ie atermetstrmmeepe e ...
         8 [Page]
File Type: JP2

Rije Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0008p.JP2 [Details/Edit]
OCR/Full text: irom suon old mising district, compared with th» 3 xpense of collecting ...
         9 [Page]
File Type: JP2
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528401p_0009p.JP2 [Details/Edit]
OCR/Full text: .";y 'te bey on d #, Lias 1t devoived on every memb--r of tha Hovse to ...
         10 [Page]
File Type: IP2
```

#### BATCH DELETING PAGES

**For issues with less than 20 pages,** delete the individual pages by opening the details/edit link and using the delete file button.





For items with more than 20 pages, the page "order" will appear beside the page label in a numeral box. This order starts with the issue Thumbnail (assigned order 1), then the Regular (assigned order 2), and Page 1 (assigned order 3).

To batch delete pages, enter "D" in the numeral box for the files you would like to delete and click Update. This is an irreversible decision, so be certain you want those files removed from the record and the server.

```
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528561p_0025p.jpg [Details/Edit]
OCR/Full text: Microsoft Word - GHPL Digitization Project Tips Manual.doc ...

D 26 [Page]
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528561p_0026p.jpg [Details/Edit]
OCR/Full text: Microsoft Word - GHPL Digitization Project Tips Manual.doc ...

D 27 [Page]
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528561p_0027p.jpg [Details/Edit]
OCR/Full text: Microsoft Word - GHPL Digitization Project Tips Manual.doc ...

D 28 [Page]
File Exists: https://data.vitatoolkit.ca/Partners/Test60/Test60000528561p_0028p.jpg [Details/Edit]
OCR/Full text: Microsoft Word - GHPL Digitization Project Tips Manual.doc ...

31 download this whole thing [Text file] Non public
File Type: PDF
File Exists: https://www.vitatoolkit.ca/Partners/Test60/Test600005285611T.PDF [Details/Edit]
OCR/Full text: Microsoft Word - GHPL Digitization Project Tips Manual.doc This ...
Update file order
```

# **EDITING PAGE FILE DETAILS**

On the File Details/Edit screen you can edit labels, settings, and add or replace files.

Label: Help
h .
For the page number, e.g. "3" or "RE3" Do not enter labels like "page 3". The software will
it is appropriate. Subsequent versions will allow you to supply alternate labels for "page" incl
Label (extension): Help
Used to briefly indicate significant features of a given page. e.g. "Chapter 2" or "Barber hous appear as part of the pagination but do help users find significant landmarks in the drop dow
Category:
Page
Files:
Original File name: Dear_Friends_001.jp2 Thumbnail for Page (JPG): Test60000528560pt 0001p.jpg Replace
If new files have been uploaded but old files (of any size) are still appearing here, the ima
opening the individual image and then SHIFT - Reload  Normal sized Page file (JP2): Test60000528560p_0001p.JP2 Replace
Normal sized Page file (JP2): Test60000528560p_0001p.JP2 Replace File Size: 1537821 bytes, 1600 x 999 File Type: JP2
Large file for Page: Test60000528560pf_0001p.pdf Replace File Type: PDF
Put in public display?: Help
○ Yes ○ No
Index the text in this file?: Help
○ Yes ○ No
Turn this off if you have uploaded multiple versions of the same content (say a Word .DOC a only index it once. <b>Note:</b> one of those files will need to be indexed.
and the state of t
Full text: Help
Ottawa, Feb. 24, 1927 Dear Friends:
The work of the Session is moving along rapidly, [sic] It
begins to look as though we would be out of here at Easter or shortly after. The general tone of the House is much more

**Label:** Any label information will be carried forward from the Add Record screen. Edit Labels here.

**Label (extension):** Use this to distinguish a labelled page as something else, like Title Page, or Index. The file display will include a colon and the extension label (e.g. 1 : Title Page).

**Category:** The category displayed here reflects either the category selected on the Add Record screen (e.g. Text, Detail, etc) or will be automatically populated during a batch upload of pages (i.e. Category: Page).

**Files:** Shows all the files created for and associated with that Page. The original digital file name, an option to add or replace a thumbnail and other display files for the record and/or sidebar display, the file name and file Replace option, and File Size in bytes or pixels, and File type. (See below on replacing files)

Put in Public Display? Default is "Yes". Use this option to manage files that are not



meant for public display such as specifically sized details files, or the text files associated with an object where you do not want the text file to be displayed, but want it's content to be indexed for full text discovery.

Index the text in this file? This allows you to select which of multiple files will have their content indexed for full text discovery. E.g. if you upload a PDF object, the text will automatically be extracted, but if you upload a DOC file with the corrected text, then select "No" for the PDF and "Yes" for the DOC, and only the DOC file content will be indexed.

**Full Text:** Any full text that is auto-extracted from .doc, .pdf, .txt, .xsl, and .rtf objects will appear in this input box; use this input box to manage, edit or review full text. This full text will <u>not</u> display in the public details of the object.

**Remember:** Click "**Update/Edit Values**" on this screen to commit your changes on this page.

#### REPLACING INDIVIDUAL PAGES

In the event that you want to replace a page image file or any of the derivative display files for that page, go to the file/tech data screen and click the details/edit link on the page in question.



In the Files field, you can upload a new Thumbnail, Page file, and the download PDF file for that page (the first time, these PDFs are automatically derived from your original page file, and have no full text).

Replacing the page file will not disturb any lins to index records like articles or BMDs.

If you do not see your new page, be sure to clear your browser cache and/or do a hard refresh.



#### DOWNLOAD PDF OPTIONS

#### PAGE-LEVEL DOWNLOADS

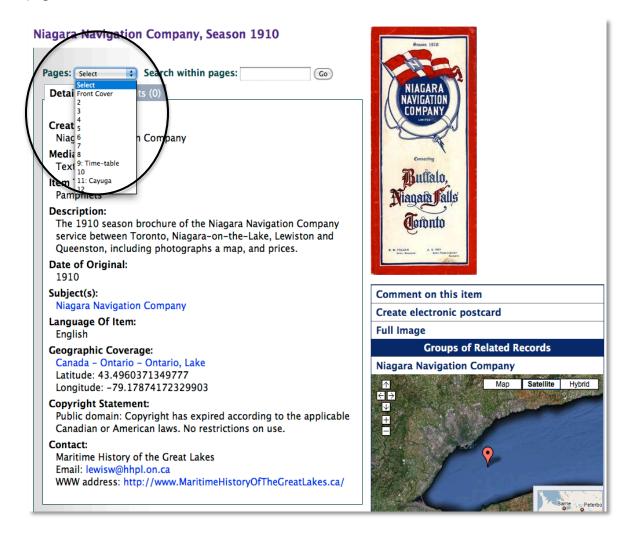
Single PDF files are automatically generated for every "Page" in VITA and offered for page-level download for the end user. If you would like to change that option, see the Agency Management manual.

#### OBJECT-LEVEL DOWNLOADS

When a single PDF is uploaded and Split, the original PDF is preserved and added to the record. The default display setting for complete PDF objects is "not public." If you want to offer the whole item for download, go to File/Tech data screen, select the details/edit for the single PDF file and change the status of that file (see <a href="above">above</a>). The download option displays in the record right sidebar; you can also change the label of that file to something non-generic.

#### MULTIPAGE DISPLAY

When multiple files are associated with one record, the public display of the internal pages will render like this:



Page-level JP2 viewer will look like this, with a search bar available for full-text search:

