

## Part 3    Uploading PDF objects

PDFs are often how text objects are stored or digitized. Use VITA to upload and manage your PDF copy of text objects to provide users with all the discovery and display options you expect for other text file types.

### Part 3.1    PDF display options

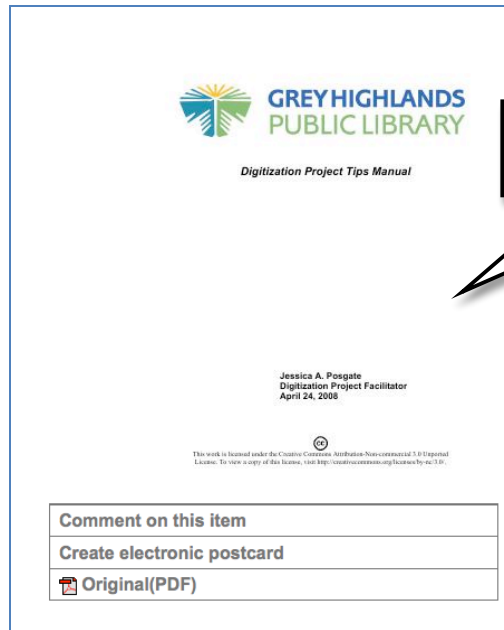
PDFs are not image files, but the VITA application will generate a regular and thumbnail display image from the first page of the PDF object during the initial upload process.

In order for the PDF contents to display, download and have full text extraction, the PDF object must be uploaded as follows:

Go to Main Menu→Add record with file

- Include Title
- Select media type “Image” or “Text” depending on what represents the intellectual content of the PDF best
- **You must select category “Text”**
- Label: Add a custom label if you prefer; default is “Download PDF”
- Allow file to be re-sized
- **Split PDF:** default is “No”. Only say “Yes” if the PDF is multiple pages in a single file, and you want the pages displayed separately, select “Yes” for Split PDF
- Browse for the PDF file—the first page will be the display image
- Click upload
- Wait for the status bar to reach 100% then click Complete

The resulting digital record will include: the image of the first page, a download link for the whole PDF object, and the full text will be extracted for object discovery. If the PDF has been split, the text from each page will be extracted separately and the page images displayed individually as well. **For more on Splitting PDFs, see Multipart PDFs Section 3.2.**



**First page display object**

**File summary**

- Thumbnail  
File Exists: <http://images.ourontario.ca/Partners/Train33/Train330001141T.pdf>
- Regular  
File Exists: <http://images.ourontario.ca/Partners/Train33/Train330001141T.pdf>
- VITA FAQ [Text file]  
File Type: PDF  
File Exists: <http://images.ourontario.ca/Partners/Train33/Train330001141T.pdf> [View] [Edit]
- OCR/Full text: Our Ontario is a service of Knowledge Ontario. For more information ...

**PDF file with full text extracted**

**If the PDF does not provide full text for your record**, open the original PDF in a reader program on your desktop, Select All (for single pages or objects where multiple pages are wrapped in a single file) or select and copy the text on individual pages, then paste the text into the Full Text box on the appropriate page edit/detail screen. Choose whether to have this text indexed, update the file and, when you have added all the text, return to the File/Tech Details screen and click the Update Index button.

### **Part 3.2 Multipage PDF files**

For PDF files where multiple pages are wrapped in one single file, use the Split PDF option on the Add Record screen. If this multipage object is uploaded as a single file, the user must download the PDF to view the internal pages and their PDF reader to search its contents.

Follow the same steps as for the single PDF (above) except

- **Split PDF:** say “Yes” when the PDF is multiple pages in a single

file, and you want the pages displayed separately.

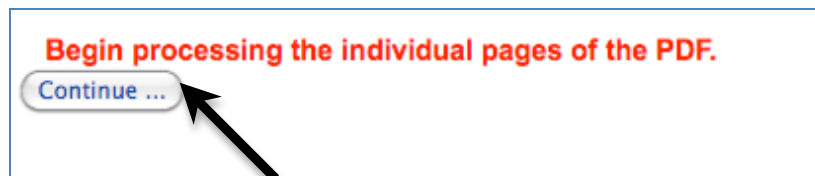
Add Record

Title	GHPL Digitization Manual
Templates	[Select one of the options below]
Media type	<input type="radio"/> Image <input checked="" type="radio"/> Text <input type="radio"/> Collection <input type="radio"/> Video <input type="radio"/> Object <input type="radio"/> Audio <input type="radio"/> Publication <input type="radio"/> Genealogical Resource <input type="radio"/> Newspaper
Category	Text file (Multiple allowed)
Label	
Image resize	<input checked="" type="radio"/> Yes <input type="radio"/> No
Split PDF	<input checked="" type="radio"/> Yes <input type="radio"/> No
Select File	

Continue ... Clear Form

GHPL\_Digitization\_Project\_Manual.pdf 1.11 MB

After clicking the Continue button, a processing message will appear. Click Continue again.



The PDF splitting will happen in front of you. Be patient. Let the whole document process, then scroll to the bottom of the screen and click Continue.

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

1

2

3

4

28


You completed Step 2 by adding 28 files at 11 Oct 2011 13:00:31  
Continue ...

You are taken to the File/Tech data screen where the display files and page files are listed, as well as snippets of the full text extracted at each page level.

**Technical data: GHPL Digitization Manual (id: 273)**

No Item Type has been assigned.  
Sortable Date has not been assigned.  
Earliest Year has not been assigned.  
Latest Year has not been assigned.  
This record will not display in the Public Sites.

**File summary**



**Thumbnail**  
File Exists: <https://data.ourontario.ca/Partners/Training/Train000000273t.jpg> [Details/Edit]

**Regular**  
File Exists: <https://data.ourontario.ca/Partners/Training/Train000000273r.jpg> [Details/Edit]

3 1 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0001.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0001.jpg) [Details/Edit]  
OCR/Full text: Digitization Project Tips Manual Jessica A. Posgate Digitization ...

4 2 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0002.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0002.jpg) [Details/Edit]  
OCR/Full text: Contents INTRODUCTION ...

5 3 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0003.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0003.jpg) [Details/Edit]  
OCR/Full text: Introduction This Project Tips Manual is designed to provide a practical ...

6 4 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0004.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0004.jpg) [Details/Edit]  
OCR/Full text: Proposing a Project Because there is so much recent interest in Proposing ...

7 5 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0005.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0005.jpg) [Details/Edit]  
OCR/Full text: Collaboration · What other individuals or organizations are willing to ...

8 6 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0006.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0006.jpg) [Details/Edit]  
OCR/Full text: · · support the requirements of processing, hosting and storing the ...

9 7 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0007.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0007.jpg) [Details/Edit]  
OCR/Full text: Scope Creep The GHPL project was challenged by the newspaper production ...

10 8 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0008.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0008.jpg) [Details/Edit]  
OCR/Full text: Collaboration The GHPL project was a direct result of collaboration. ...

11 9 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0009.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0009.jpg) [Details/Edit]  
OCR/Full text: Project Staffing · · · · Project Facilitator Project Assistants GHPL CEO ...

12 10 [Page]  
File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0010.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0010.jpg) [Details/Edit]  
OCR/Full text: Recruitment & Hiring Recruitment and hiring can be done ahead of project ...

Relabel Pages

Associate another file with this record

Associate a single Zoomify zip file with this record

Associate multiple pages with this record

Associate OCR/Full text files with these pages

Associate Zoomify files with these pages

When finished editing pages/text:

Update Index

The complete PDF will be associated automatically as well

File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0027.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0027.jpg) [Details/Edit]  
OCR/Full text: Appendix B Developing the Agnes Macphail Website Computer Language HTML ...

30 28 [Page]

File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0028.jpg](https://data.ourontario.ca/Partners/Training/Train000000273p_0028.jpg) [Details/Edit]  
OCR/Full text: Appendix C Copyright Checklist Reproduction of anything on the ...

31 **Original(PDF)** [Text file]

File Type: PDF

File Exists: [https://data.ourontario.ca/Partners/Training/Train000000273p\\_0029.pdf](https://data.ourontario.ca/Partners/Training/Train000000273p_0029.pdf) [Details/Edit]  
OCR/Full text: Digitization Project Tips Manual Jessica A. Posgate Digitization ...

Update file order

Each page is automatically downloadable as a PDF on the public side, but the whole document download is default “non-public”. Click on the Details/Edit link to make this document publicly download-able.

**Label:**  
  
 You only have 50 characters to work with for the Label.

**Category:**  
 Text file

**Files:**  
**Original File name:** GHPL\_Digitization\_Proje  
**Regular file (PDF):** Train0000002731T.PDF  
 File Size: 1111918 bytes, 0 x 0 File Type: PDF

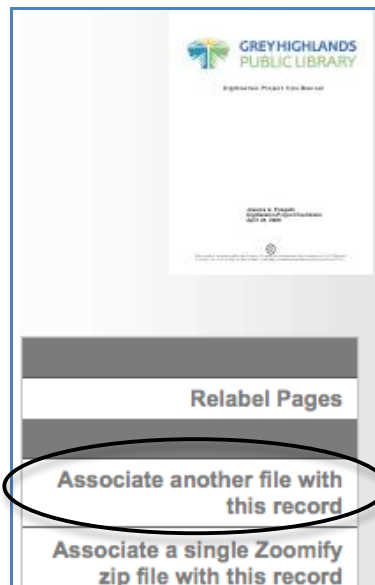
**Put in public display?:**  
 Yes  No

**Is this corrected OCR?:**  
 Yes  No

**Index the text in this file?:**  
 Yes  No  
 Turn this off if you have uploaded multiple versions of the PDF file, so that we only index it once. **Note:** one of those

**Part 3.3 Replace display image for PDF**

If the first page of the PDF object is not what you want to use to represent the object in results sets and other displays, you can upload a new display image.



Go to File/Tech Data → "Associate another file with this record"

- Browse for the file you want to represent this multipage object (e.g. an illustrated front cover or other representative page from the object itself)

Add file to record

Title	GHPL Digitization Manual
Category	Thumb+Regular
Label	
Image resize	<input checked="" type="radio"/> Yes <input type="radio"/> No
Split PDF	<input type="radio"/> Yes <input checked="" type="radio"/> No

Select File

Continue ... Clear Form

GHPL_HeroShot.jpg	114.95 KB
-------------------	-----------

- Select category “Full+Reg+Thumb” or “Thumb+Regular”
- Label is irrelevant for the display object
- Allow file to be re-sized
- Don't split PDF
- Click upload
- Wait for 100% status then click Continue.

