

**Title:** Accessible PDF Guidelines,  
University of Notre Dame, Hesburgh Libraries

**Creator:** Digital Standards Team

Alexander Papson, Metadata and Digital Services  
Librarian, (Chair)  
Julie Arnott, Manager, Preservation  
Tracy Bergstrom, Interim Co-Program Director,  
Digital Library Initiatives and Scholarship  
Curator, Italian Imprints and Dante Collection  
Dan Brubaker-Horst, Digital Library Applications  
Lead  
Susan Good, Webmaster and Desktop Consultant,  
Kresge Library  
Adam Heet, Library Assistant II, Architecture Library  
Rick Johnson, Interim Co-Program Director, Digital  
Library Initiatives and Scholarship – E-Research  
And Digital Initiatives  
Patricia Lawton, Catholic Research Resources Alliance  
Digital Projects Librarian  
Denise Massa, Curator, Visual Resources Center  
Joe Reimers, Technology Support Specialist, Kresge  
Library  
Sara Weber, Special Collections Digital Project  
Specialist

**Date Issued:** 2012-03-18

**Identifier:** Accessible PDF Guidelines final

**Replaces:**

**Is Replaced**

**By:**

**Latest Version:** 1.0

**Status of**

**Document:**

**Description of**

**Document:**

This is a Digital Standards Team recommendation.

This best practices document recommends the implementation of a core set of guidelines for creating accessible PDF documents. The document gives directions on how to use optical character recognition on a text document to aid in accessibility.

**Change Log:**

## **PDF and OCR/Basic Accessibility Best Practices**

Rev20130131app

PDF documents are a useful way to share items with others with less concern about software issues. You can convert a wide range of items into a PDF document including images, text files, word processing documents, spreadsheets, slide documents, webpages, and screen captures.

### **PDF Creation:**

Converting items to a PDF document is fairly easy and can be done in a couple of ways. You can scan the item directly to PDF, save a single file as a PDF, or save multiple files to make one PDF.

If you have a document that spreads over two pages such as a magazine or journal article it is best practice to create the PDF as single pages. As a single page the user can scroll up and down through one page and still know where they are in the order of the document. With two pages side by side it becomes more difficult to keep track of where in the document you are looking and it becomes very difficult to use the item with a handheld device and causes concern for accessibility. The pages should be cropped so that they are uniform as the document is scrolled through. Each page should be close to the same size so that they transition from one page to the next without a major jump in size or shape unless the document is purposely formatted that way.

Posters that are created in a slide program like PowerPoint can be converted to PDF. These pose a problem for viewing and accessibility because of the size and layout. When possible the poster should have a separate text file included that gives the information on the poster in an easy to read format. This will allow users with accessibility needs to have an easy to use format and still have the original poster for review as well.

If the document is handwritten or the type is not clear and crisp a transcription should be included to aid in searching, text mining, and accessibility. The text should be transcribed so that it is identical to the original item. If the item is in an uncommon dialect or different language than what is commonly used a translation should be provided as well. Transcriptions and translations can be provided as a PDF so that they can be easily downloaded and searched.

### **OCR:**

If the document has text in it then it should be reviewed to determine if it is worth performing optical character recognition (OCR) on it to help with searching and machine reading. If the document is not going to read well then it should be transcribed to create an easy to read version.

The OCR software in Adobe Acrobat Professional is satisfactory but it is not as good as other programs such as ABBYY Fine Reader or OmniPage. The Adobe Acrobat OCR is the quickest and easiest to use but the accuracy will need to be checked more closely depended on the type of document.

For the best quality OCR text the text should be high contrast compared to the background. Generally black text on white background works the best. Regular paragraphs work best compared to columns since the formatting is easier to follow and it is easier to set a reading order.

For the exact steps in performing OCR and Accessibility in Acrobat please review the document titled OCR Steps.

#### Accessibility:

All documents with text that is a typed format and that have been OCR'd should be made accessible as well. The accessibility steps will set the reading order of the document, language specifications, and add tags to the document. If the document is formatted in a way that could be confusing or difficult to use such as a poster or article with several columns of text the document should be considered for a reformatted basic version to help with accessibility.

After running the full check for accessibility on the document if errors are found that are not easily corrected it is best to perform a Google Search on the error type to determine the appropriate repair. Adobe also has several forums where errors are discussed and this may help with finding a solution.

For the exact steps in performing OCR and Accessibility in Acrobat please review the document titled OCR Steps.

## **OCR/Accessibility Basics Acrobat Pro X**

Rev:112712app

### **OCR:**

Open Acrobat Professional and choose Combine Files into a Single PDF.

Drag the files that you are combining into the window.

Save the file as the designated file name.

Choose Tools – Recognize text—Exact.

When the OCR is complete hit save.

Left side of the panel choose the Order panel (looks like a Z) then under Tools choose Accessibility—Touch Up Reading Order.

Start on Page 1 and work to the end of the document.

Use the mouse to outline the text, if it is in columns do it in the order of how you would read it.

Highlight the text with the mouse (small blue boxes will go around the letters) and hit Text on the Touch Up Reading Order panel.

Do this to all pages and hit save when done.

Under Tools—Accessibility—Add Tags hit save when the tags are added.

Start on Page one and go through page by page and make sure that the text is highlighted and that the reading order is how you would read it. If blank spots are highlighted or if images (like logos) are highlighted choose background in the touch up reading order panel.

If there is an image right click on the image and choose edit alt text. Write in the box what the picture is but keep it brief.

After all pages are checked and all alt text is added click save.

Under File—Properties—Advanced and choose English as the language and save.

Under File—Properties—check to make sure that Fast Web View says yes, if not Save As—Optimized PDF. It can be a good practice to save the file as file name\_opt or something similar to avoid confusion.

Choose Tools—Accessibility –Full Check and make sure that the pdf checks ok.

### **Large File Size:**

Create a new folder to put all of the reduced files in together.

If the item has pages that are color reduce the images to 300dpi and save as JPEG.

If the item has black and white photographs, charts, maps, or drawings reduce to 300dpi, convert to greyscale, and save as JPEG.

If the item has black text on white paper or simple line drawings then reduce to Bitmap with the setting of 500dpi in the bitmap pop up window. Save these as TIFF images.

Combine the new files in a pdf as listed above and then save the item as the file name.

Immediately reduce the item and save as file name\_redux.

Create a new folder in the file and call it Reimage.

Save the pdf as –Save As—Image—TIFF to the Reimage folder.

Combine the Reimage files to make a new pdf. Run the ocr and accessibility as listed above. If need be optimize the pdf at the end to reduce the file size slightly and have it be optimal for viewing on the internet.

#### Text Correction:

Go to Tools—Edit Document Text

Use the mouse to navigate to the area of text that you want to review.

Click on the area and a bounding box should appear around the text.

Click and drag inside the box and highlight all of the text.

Right click on the highlighted text and choose Properties from the list.

Choose the Text Tab and then click on the Fill with the black square next to it.

Choose a color that is sharp contrast to the background.

By doing this you have just highlighted the OCR hidden text so you can now see errors in the hidden text that will be searched and machine read.

Use the touch up text tool to adjust the text as needed.

Reset the Fill color back to black and move to the next section of text.

This will also show you if the reading order is out of place and needs to be corrected.