CAUL Webinar Notes

The focus of this session will be on scanning from print and adding accessibility features to scanned PDFs in Adobe Acrobat Pro XI. Born-digital documents (electronic journal articles, e-books, word processor documents, and PDFs created from word processor/publisher documents) are usually (though not always) more accessible than documents that are scanned from print copies.

Regarding electronic documents, we should encourage vendors / database owners to make their documents more accessible whenever possible. (OCR'd text or e-text, not flat PDF images, not behind a wall of DRM.) We should complain to vendors and publishers if accessible documents are hidden within inaccessible websites/platforms that do not cooperate with screen readers.

You can add accessibility features to Word documents using an Adobe Acrobat plugin for Word. It adds bookmarks and other basic features as the document is rendered into a PDF.

Additionally, you can enhance the accessibility of many electronic documents by following some of the procedures outlined below, where appropriate.

To test the screen readability of a document or webpage, use a screen reader testing app, such as:
https://www.naturalreaders.com/online/

Workshop component:

1. Do I have to scan this? Providing an already-existing electronic copy is often the better choice in terms of copyright and baked-in accessibility.
2. Can I scan this effectively?
   a. Is it a clean copy? (No marks/highlights that obstruct the text; gutters wide enough that the scanner can read every word. Printing quality—not a pulpy mimeograph.)
      i. Show images of what NOT to scan.
   b. If not, can the Accessibility Office at your institution provide transcription or narration service? Can you source a better copy via doc del? If this is for Reserves, does the professor have a better quality copy? Would the Library consider purchasing a new print copy?
3. Effective scanning:
   a. MSVU Library staff use a networked Konica multi-function printer/scanner/copier, not a traditional flatbed scanner attached to a workstation (better speed and size options).
   b. Choose appropriate scan dimensions for BOTH pages (right & left). May require B4, or A3 scan dimensions.
   c. 600 x 600 dpi will improve OCR (character recognition / quality) later on.
   d. With higher resolution, longer chapters / excerpts may have to be split up.
   e. May be better to send your scanned files to a network folder rather than your e-mail, in case they bounce due to size.
   f. Hold pages as flat and steady as possible, and keep them flush to the edge of the scanner glass.
   g. If you're worried you skipped a page, don't start over! It's easy to delete an extra page in Adobe Acrobat.
   h. First or last page of your scan should be the monograph or journal cover page; helpful to students and faculty alike.
   i. Consider including endnotes for book chapters, where applicable.
4. Processing in Adobe Acrobat:
   a. Adobe Acrobat Reader (free) should not be mistaken for Adobe Acrobat Pro (paid/licensed). AAP is a full-function PDF editor, which includes a Microsoft Word plugin for further improving the accessibility of Word documents converted to PDF. I use this plugin for the initial conversion/rendering of galley PDFs for the Partnership journal.

b. Open your PDF in Adobe Acrobat Pro, not Adobe Reader.
c. Add metadata to your document.
    i. File – Properties – Description: set title and author, at least.
    ii. Initial view tab: Show – Document Title
    iii. Advanced tab: set Language.
d. Duplicate and crop pages to facilitate straight, accurate OCR. (Duplication can be skipped for single-page scans.)
    i. To duplicate, open page thumbnails from the left-hand menu.
    ii. Select each page and Ctrl + click/drag.
    iii. Select the first page of the document. Use the Pages-Crop tool (from the right-hand menu) to select the left half of the page.
    iv. Press Enter; a new window will appear. Adjust the page range to include all pages, and apply to Odd Pages Only.
    v. Select the 2nd page of the document, and follow the same procedure again, but apply it to even pages only.
    vi. Examine document to ensure all pages are in order and no text has been cropped out. Delete extraneous pages.
e. Set page numbers to accurately reflect book/journal pagination.
    i. Right-click on any page and select Number Pages from the dialog box.
    ii. Use Roman numerals for non-numbered pages.
f. Add OCR (text recognition) to the document.
    i. From the right-hand menu, select Text Recognition – In This File
    ii. All pages – Advanced Options – ClearScan – 600dpi. (ClearScan replaces photoreprographic images of characters with digital characters; results in a more compact end product.
    iii. As the OCR engine runs, you'll notice the pages straightening along the horizontal. This is why two-page scans need to be duplicated and cropped. Failure to do so will inevitably result in OCR'd documents with one crooked page.
    iv. At this point, consider using a screen reader app like the one available from NaturalReaders.com to test your OCR for accuracy.
g. For general scanning requests (for Reserves or Doc Del, not coming from Accessibility Services or a patron with a disability), MSVU staff would normally stop here, in the interest of time. For increased accessibility, there are a few more steps to follow.
h. Review document and delete scanning artifacts: shadows, scratches, dust, etc.
    i. From the right-hand menu, select Content Editing – Edit Text & Images.
    ii. Select (click on) and delete extraneous artifacts. Be on the lookout for images on pages that do not actually contain pictures or illustrations.
    iii. For more detailed cleanup, you may wish to consult the Content tab in the left-hand menu. It will take you through each and every image, text, and artifact that a screen reader can "see." You can delete or modify these as needed.
i. Clean up the reading order of the page. Think about what you would say if you were reading the document aloud to someone.
    i. Go to the Order menu in the right-hand panel. From the drop-down, select Show Reading Order Panel.
    ii. Set all page numbers and running headers / running footers as "background"; you would not read these aloud, and we added pagination metadata earlier.
    iii. Distinguish headings from body text.
    iv. Distinguish images/figures and captions from body text.
j. Add alt text to images and figures.

       i. From the right-hand menu, select Accessibility – Set Alternate Text. The pop-up windows will lead you through each image in the document.

       ii. Good alternate text tells you what the image actually looks like: its style and content. The function is slightly different from a caption, but you may also want to include the text of any captions.

   k. Add bookmarks for easy navigation.

       i. From the left-hand panel, select the Bookmarks tab. It will likely be empty at this point.

       ii. Bookmarks should be used to indicate main headers in the text. Highlight a header, then type Ctrl+B to create a new bookmark.

       iii. Edit the text of the bookmark, if necessary, and drag it to nest appropriately.

   l. Run Adobe's Accessibility Checker to address any lingering issues.

       i. From the right-hand menu, go to Accessibility – Full Check.

       ii. The output will display in the left-hand menu. Right-clicking on problems or errors often gives the option of "Fix," which will open up a new tool or utility in Acrobat, or "Explain," which will open up a relevant Adobe help document. Responding to problems in the Accessibility Report and reading Help documents/tutorials is one of the best ways to improve your proficiency with Adobe.

5. MSVU procedures:

   a. Bare minimum for all requests: high quality, cleaned-up scan with OCR. This is a universal design feature, in that it makes the text searchable for all users. As well, most students with print disabilities have personal or institutional access to software that can properly parse/process a high-quality scan.

   b. We add more accessibility features (alt text, tags, etc.) when we **know** the requesting patron has a print disability (self-identified, or flagged via Accessibility Services).

   c. We have recently established a small repository of alternate format requests. This repository is a collaborative effort between our Library, Print Shop, and Accessibility Services. In time, and with sound legal advice (Marrakesh Treaty implores us to share), we would like to expand this cooperative effort to other Novanet or CAUL partner libraries.