2020

Best practices for producing high quality PDF files

Formats Group, Deep Blue

https://hdl.handle.net/2027.42/58005

http://creativecommons.org/licenses/by/3.0/

Best practices for producing high quality PDF files

Version 4.0 2020

The Portable Document Format (PDF) file type is a **recommended and supported format** for deposit in Deep Blue.

The following guidelines will help you prepare PDF that are of **archival quality and accessible** to people with a range of disabilities and across a range of reading technologies.

PDF created according to the **PDF/A standard** are files that meet accessibility standards, can be reproduced with consistent visual presentation, and use standard Unicode characters for all text. The guidelines below will allow you to create PDF/A files.

You will need access to **Adobe Acrobat Professional**, a software program that is available for use at all U-M Campus Computing Sites. (Adobe Acrobat Professional is also known as Acrobat DC, but is not to be confused with the free software Acrobat Reader.)

This document addresses creating PDF from word-processor documents, from scans, and from webpages.

Questions?

If you have any questions, please contact us at deepblue@umich.edu and we will be happy to help you if we can.

Suggested citation for re-use

> Based on the University of Michigan Library's "Best practices for producing high quality PDF files," available at http://hdl.handle.net/2027.42/58005 , where it is available under a Creative Commons "Attribution" license: http://creativecommons.org/licenses/by/3.0/ .

Document Sections

Creating PDF from Word Processor Documents

The instructions below refer to Microsoft Word, which is available for use at all [U-M Campus Computing Sites](#). U-M users may also [download Word onto your personal computer](#).

Quick Wins: Improve your document quality and accessibility right now

These minor changes have a major impact.

1. Styles

Use the built-in "Styles" to mark up text.

Main title should be marked as "Heading 1," section titles should be marked as "Heading 2," subsections should be marked as "Heading 3," and so on. Regular text should be marked as "Normal," block quotes should be marked as "Quote."

You can change the appearance of the Styles. For example, you can modify "Heading 2" to be 16 point, bold, Helvetica, left justified. If you decide later that "Heading 2" should be 14 point, bold, Times, centered, you can modify the Style once, and all text marked Heading 2 in your document will be updated.

2. Language

Indicate the main language of your document within the document settings.

If portions of your document are in a secondary language (for example, Portuguese quotations within an English-language article) highlight those portions and indicate their language using the Tools menu.

3. Links

Text used as hyperlinks should describe or explain where the link leads.

See for example the sentence from earlier in this document:

> The instructions below refer to Microsoft Word, which is available for use at all [U-M Campus Computing Sites](#). U-M users may also [download Word onto your personal computer](#).

It is clear from the link text "U-M Campus Computing Sites" that the first link leads to a website about U-M Campus Computing Sites. It is clear from the link text "download Word onto your personal computer" that the second link leads to a website that allows U-M affiliates to download Microsoft Word.

Links can also be used to link to sections within the document itself. A linked table of contents will make a long document easier to navigate and use.

4. Images and embedded media

Include textual description of images. Effective textual description provides the equivalent information as the image, in text. It can be included in the image "alt text," or in the surrounding text.

For audio files, include captions and/or transcripts. For video files, include captions and/or transcripts.

Provide titles or captions for all images, tables, and embedded media.

5. Structuring elements

Use built-in features to structure text in lists, columns, headers/footers, tables, pages, etc.

Word has features for common text structures. These features should always be used (rather than the space bar, tab, or enter) to structure text in lists, columns, headers and footers, tables, or pages.

For more information on making data tables accessible, please consult Microsoft Office: Create Accessible Tables.

6. Checker

Use the built-in "Check Accessibility" Tool to scan your document and fix issues.

The checker will test for the items listed here, and a few more. It cannot catch everything, but it can alert you to potential problems and suggest fixes or resources to learn more.

Save as PDF

After taking steps to create an accessible document in Microsoft Word, save the file as PDF to preserve its accessibility features. To do this, choose Save As, PDF, then review additional options to ensure that structural and accessibility features are preserved and converted into "tags" in the PDF.

If you have the option, save as PDF/A rather than PDF. For more information on this process, which differs slightly by operating system and software version, see the following resources:

- Microsoft Office: Create accessible PDFs
- WebAIM PDF Accessibility: Converting Documents to PDFs

When saving your file, the file name should:

- Use only letters (Aa-Zz), numbers (0-9), dash (-) and underscore (_)
- Avoid any other characters, especially commas, spaces, forward-slash (/) or back-slash (\), as these characters can cause problems on the web and in our preservation storage systems
- Be 32 characters or less

Learn More: Further information about document quality and accessibility

If you wish to learn more, or review more detailed instructions, about any of the above information, please consult the following resources:

- [Microsoft Office: Create accessible PDFs](#)
- [WebAIM PDF Accessibility: Converting Documents to PDFs](#)
- [University of Washington: Creating Accessible PDFs from Microsoft Word](#)
- [Adobe Support: Create PDFs from Microsoft Office](#)

For direct support on campus, please consult with ScholarSpace in the Hatcher Graduate Library, Room 206 Hatcher North, ScholarSpace@umich.edu.

Creating PDF files by scanning print material

General Principles

PDF files may be created by scanning print materials.

Compared to PDF files created from word processor documents, PDF files from scanned material may have increased file size and compromised viewing, printing, or usability. Nevertheless, PDF files created from scanning are a reliable way to preserve materials when no other options are available.

When scanning material, use the following specifications:

- All materials should be scanned at 100% scale to the dimensions of the original.
- The PDF file should be optimized and should be in ("Searchable Image (Exact)"/"Image+Text") format.
    - Perform OCR (optical character recognition) and embed the generated text in the document.
- Pages containing text and/or line art should be monochrome (black and white), 600dpi, and compressed using ITU Group IV compression.
- Oversized materials (greater than 11x16 in/28x41 cm) should be 300dpi, grayscale/color.
- Pages containing photographs and/or illustrations should be 24-bit color using the sRGB color space or 8-bit grayscale, 400dpi (or 300dpi for oversized materials greater than 11x16 in/28x41 cm), and be compressed with JPEG compression using the highest quality setting.
- Missing or blank pages should be represented as blank images of the same size as the original.
- If your process creates intermediate files (TIFF, PostScript, etc.), it must not change the resolution of the original scans (down sampling) or use lossy compression such as JPEG. (These methods, although they do reduce file size, will cause irreversible quality loss.) TIFF is an excellent choice for an intermediate file format, and acceptable compression schemes (which can all be used in combination with TIFF) are ITU Group 4 (for black and white material only), LZW, and Flate (Zip).
- The PDF should have tags as described elsewhere in this document.

Using Adobe Acrobat Professional

Adobe Acrobat Professional has built-in scanning capabilities that can be used to create PDF files using printed materials. (Note that Acrobat Professional is not the same as Acrobat Reader or Acrobat Standard; only Acrobat Professional has support for scanning.)

To learn more and review detailed instructions, please consult the following resource:

[Adobe Help: Scan documents to PDF](#)

Scanning on a large scale

If you need to create a large number of PDF files, these approaches may not be efficient (although the principles outlined here still apply). Before you start your scanning project, contact us and we will be

happy to advise you, your staff, or a scanning vendor to make sure that your materials are in the best possible form for preservation.

Scanning Support

Contact [ScholarSpace](#) in the Library for guidance and getting started with scanning on any scale, scholarspace@umich.edu.

Creating PDF from Web

For content that currently exists as a webpage (e.g. a blog post, an article in an online journal, an informational webpage), you can efficiently create a high-quality PDF, that preserves the webpage's structure and styling, by using Adobe Acrobat Pro.

Note that this is more effective than selecting Print > Save to PDF from your web browser, which will not preserve structural and styling information.

To create a PDF from the web in Adobe Acrobat Pro, go to File > Create > PDF from Web Page.

To learn more and review detailed instructions, please consult the following resource:

[Adobe Help: Converting web pages to PDF](#)

Improving PDF Once Created

A PDF created by any of the above methods can be manually checked and improved in Adobe Acrobat Pro as a final step.

Basic Principles

When manually checking and improving PDF quality and accessibility, the following principles serve as a guide:

- Document metadata, including title and language, should be complete and accurate.
- Document structure (headings, paragraphs, tables, images, lists, etc.) should be accurately indicated by tags and in correct reading order.
- Non-text elements (images, diagrams) should have a text equivalent the describes the element or provides equivalent information in text.
- Links to external webpages and internal document anchors should function as expected.
- The order of elements on a page should match the intended reading order of the content.
- Text and other information should be legible, with sufficient contrast.
- The built-in Adobe Acrobat Accessibility Checker should be used to review the document and the built-in Adobe Make Accessible Action Wizard should be used to fix issues.

Learn more

To learn more and review detailed instructions, please consult the following resources:

[Adobe Help: Create and Verify PDF Accessibility](#)

[Adobe: PDF Accessibility Overview](#)

[Michigan State University Accessible PDFs Tutorial](#)