Converting a PDF file to a PDF/A compliant version

Modified 12.8.2026

Contents

PDF/A document format

PDF/A is the ISO-standardized version of the Portable Document Format (PDF) specialised for the digital preservation of electronic documents. The extension in the file name is pdf. PDF/A differs from 'normal' PDF in that features ill-suited for long-term archiving are omitted. A PDF/A file has all the fonts used in the document embedded within the PDF file, so that the viewer of the file need not have the same fonts used to create the file installed on their computer to read it. Also, the file has a hierarchical structure, and the content is tagged for reliable text extraction for use, for example, in assistive technology. The PDF/A standard currently has four parts, PDF/A-1 through to PDF/A-4.

Part	Published	Comments
PDF/A-1	2005	Transparencies in images forbidden
PDF/A-2	2011	Transparencies in images allowed
PDF/A-3	2012	Same as PDF/A-2 but allows embedding of other file types
PDF/A-4	2020	Not allowed yet for official archiving in Finland. Not yet available in most consumer software

In addition to the parts in the standard (which in practice are newer versions of the standard), the standard specifies four levels of conformance:

Level	Conformance	Comment
b	basic	used in parts 1–3 only
a	accessible	used in parts 1–3 only; contains all requirements of b in addition to those necessary for better digital accessibility
f	embedding of arbitrary files	used only in part 4
e	engineering	used only in part 4

Metadata is an important part of the standard. It facilitates finding information about the document content and so helps search engines find your document. It can be added in the word-processor file, like Word, or it can be added to the PDF file manually before the PDF/A file is created (see *Add metadata* below).

At Aalto University, theses are stored using one of the following PDF/A formats: PDF/A-3a or PDF/A-2a, PDF/A-3b or PDF/A2b, PDF/A-1a or PDF/A1-b, in order of preference.

For example, if you have images with transparencies in your thesis, which is usually the case, use PDF/A-3 or PDF/A-2. Since Finnish law requires all documents published by all institutions—this includes theses—to be accessible, strive to create PDF/A-3a or PDF/A-2a files of your thesis.

Creating a PDF/A file from Word in Windows

The PDF/A file currently created by Word is a PDF/A-3a file. However, to properly comply with the PDF/A-3a standard, you must at least add metadata and alt text to the images in your document.

1. Add metadata to the Word document

Add the document metadata already in the Word document (this is a good idea, since you might forget to do so when doing the PDF/A conversion later) as shown here in Figure 1. *Alternatively*, add it to the PDF file later, as described in step 4 of this section **Add copyright status (and other metadata if not added earlier)** and the section **Add metadata to the PDF file** below.

File → Info → Show All Properties

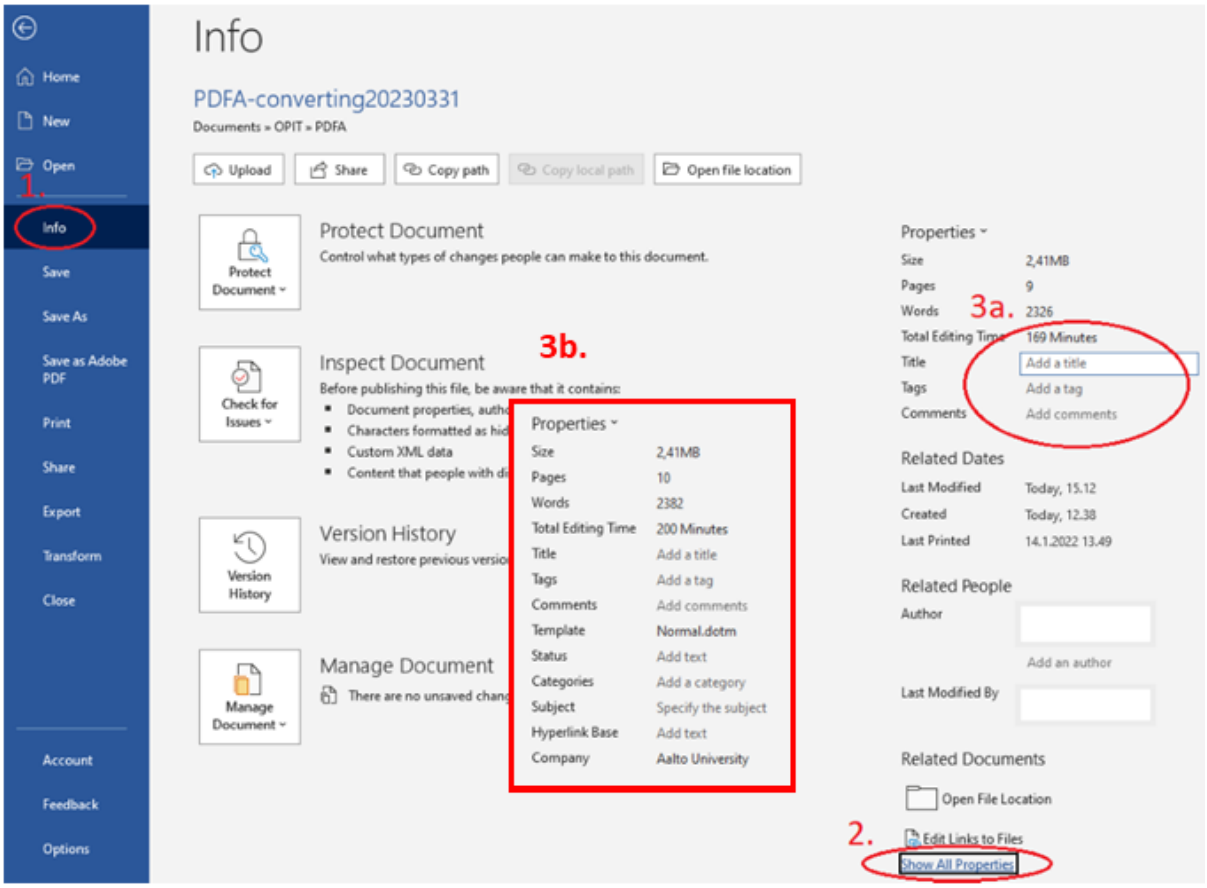

Figure 1: Adding metadata to a Word file in Windows. Clicking on 'Show All Properties' (2.) displays the list of metadata fields (3b.) you can edit. Hover the cursor on the respective field area for the text box to appear (3a).

Add at least your name in the **Author** field, the title of your thesis in **Title**, and some keywords in the **Tags** field, but we recommend you add the abstract as well in the **Subject** field. Also, add the copyright notice you want. This information can be added only on the pdf file. See section **Add metadata to the PDF file** for details on how to do this.

2. Add alt text to all the images in your document

Here are three alternative ways of doing this:

1. Right click image → **View Alt Text…** OR
2. Left-click image to select it → **Review** → **Check Accessibility** → **Alt Text** OR
3. Check the accessibility of your document: **Review** → **Check Accessibility** → **Check Accessibility**

3. Save the document as a PDF/A-3a file[1]

File → **Save as** → choose **PDF (*.pdf)** → **More options** → **Options** → **PDF/A compliant**

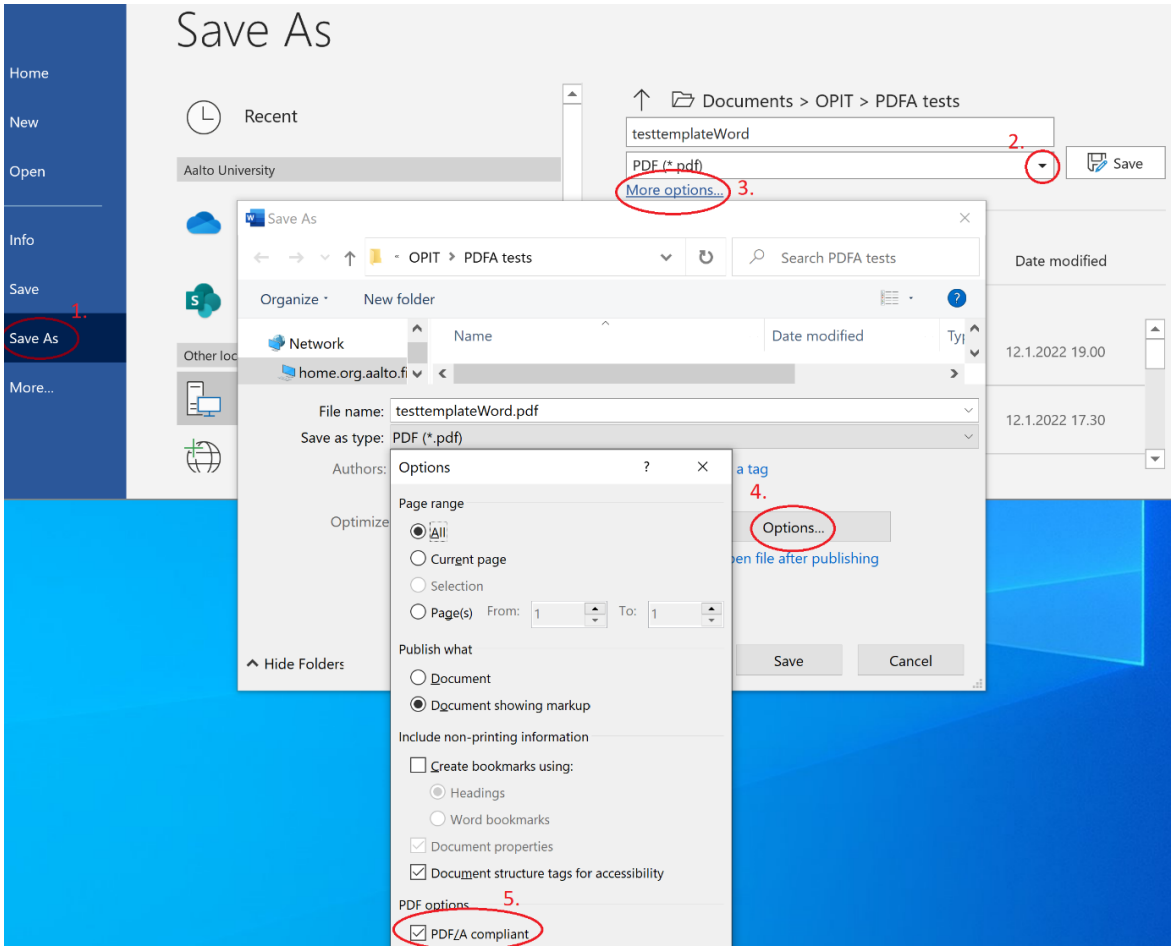

Figure 2: Saving a Word document as a PDF/A-3a file. Check also the 'Document structure tags for accessibility.

4. Add copyright status (and other metadata if not added earlier)

1. Open the created PDF/A file in PDF-XChange.
2. Click on the **Enable Editing** button
3. Do **File** → **Document Properties** → **Additional Metadata** → **Copyright Status**

[1] Word neither gives you a choice on the type of PDF/A file to be created nor says what type of PDF/A file is created. When checked elsewhere the created file claims PDF/A-3a compliance. Note that the created file may fail validation.

A typical choice is **Copyrighted** with the following copyright notice and URL:
This work is licensed under a Creative Commons "Attribution-NonCommercial-ShareAlike 4.0 International" (CC BY-NC-SA 4.0) license.
https://creativecommons.org/licenses/by-nc-sa/4.0/

For more alternatives on copyright notices, see https://creativecommons.org/cc-licenses/

If you haven't added metadata earlier (step 1), you can do so during this step in the **Document Properties** view. See section **Add Metadata to the file** below for screenshots.

4. Save the additions made by clicking on **OK**.

5. Resave the file as a PDF/A-3a file

Your document is still open in PDF-XChange. Now, finally, resave your document as a PDF/A-3a file (for details and screenshots, see section **PDF to PDF/A (Method 1)** below):

File → Save As → Browse → Save as type → PDF/A Document (*.pdf) → Options → Choose Conformance: → PDF/A-3a (ensure that the **Embed Font Subset** box is checked) → **OK**

If all has gone as it should, you should now have a proper PDF/A-3a compliant file. Ensure that the file is indeed a proper PDF file by testing it as described in section **Test your PDF file**. Then, verify that the file conforms to the PDF/A-3a standard by validating the file (see section **PDF/A validation**).

If things go wrong

First, try to fix the created PDF/A-3a file that failed validation using Aalto University's PDF conversion and validation service at https://aaltodoc.aalto.fi/sec/pdfa/. Upload the file to be fixed using the **Select file** button in the **PDF/A conversion to PDF/A format** dialog box, pick the desired PDF/A standard (begin with PDF/A-3a, and it that fails, try PDF/A-2a, PDF/A-3b and PDF/A-2b in that order). Remember to test your PDF file and validate it in the validator in the same service. Note that the validation may fail after conversion, in which case read on for further advice.

If the conversion fails, try creating a PDF/A-3b compliant file in PDF-XChange similarly as you would the PDF/A-3a, but choosing the setting PDF/A-3b in Figure 8. If this fails, try creating a PDF/A-2b file.

If creating a both a PDF/A-3b and a PDF/A-2b compliant file fails and/or PDF-XChange is not available, you can try to create a file with PDF/A-1a compliance. Do either **File → Save As Adobe PDF** or **File → Export → Create Adobe PDF → Create Adobe PDF**, as shown in Figure 3 for the latter.

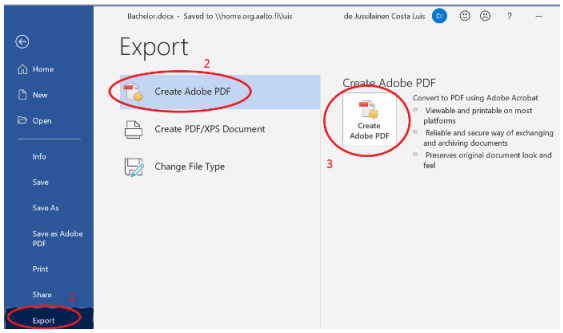

Figure 3: Exporting to create a file with PDF/A-1a compliance.

The next dialog box that appears, shown in Figure 4, is identical for both approaches. Press **Options** → check the box '**Create PDF/A-1a: 2005 compliant file**' → **OK** → name the file appropriately → **Save** to create the file with PDF/A-1a conformance.

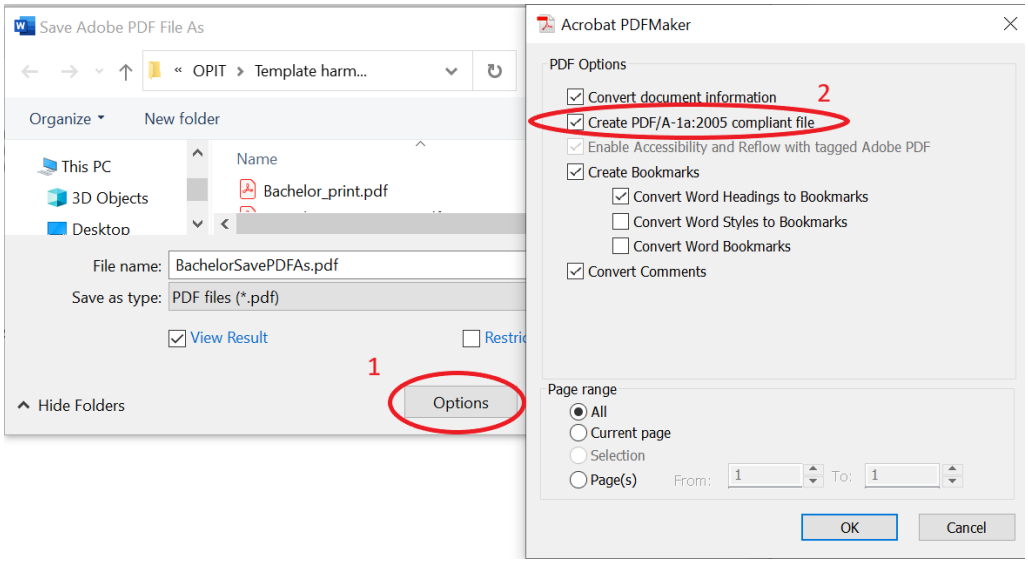

Figure 4: Settings to create a file with PDF/A-1a conformance.

If creating a PDF/A-1also fails, try using the approach described in section **docx to PDF/A (Method 2)**.

Creating a PDF/A file from Word in a Mac

Creating a PDF/A-compliant file from a Word document involves the following:

1. Create a PDF/A-3a compliant file as shown in Figure 5.
2. Add metadata to the resulting PDF/A-3a file as described in section **Add metadata**.
3. Finally, convert this file into a PDF/A-3a file. You can do this with PDF-XChange in the Windows environment (see sections **Conversion tools: PDF-XChange** and **PDF to PDF/A (Method 1)**) via the Virtual Desktop Infrastructure (VDI). Instructions for the use of VDI are available here.

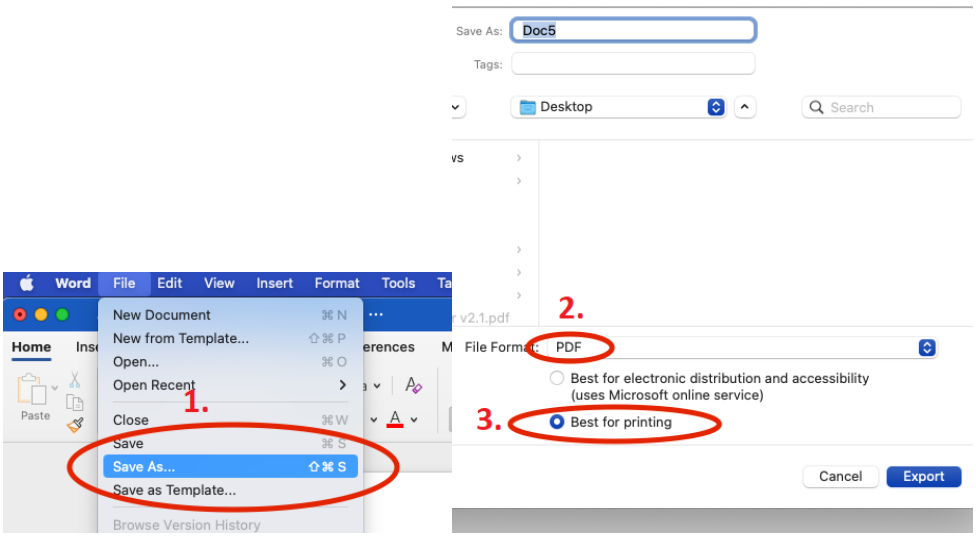

Figure 5: Creating a PDF/A-3a compliant file in a Mac.

Add metadata to the PDF file

Add at least the title of the document (thesis name), the author's name, and the relevant keywords to the PDF file's metadata. Adding your abstract is strongly recommended. Add the abstract text without paragraph breaks in the field 'Subject'. Add the metadata before saving your file as the final PDF/A file. If you add the metadata to your final PDF/A file, you will have to enable editing it and resave it with the appropriate compliance after adding the metadata.

To add the metadata, go to **File → Document Properties → Description** (1, 2 and 3 in Figure 6).

Fill in the metadata fields in the dialogue box (4 in Figure 6). Add the keywords again in 'Additional Metadata' (5 and 6 in Figure 6) because, unlike the title and author, the keywords are not transferred automatically. Also note that, if you have added the abstract text in the subject field, it will appear in the 'Description' field here. Specify the copyright status you wish to give your document (7 in Figure 6). We recommend using a Creative Commons license of your choice. Remember to **test your file** after saving the file.

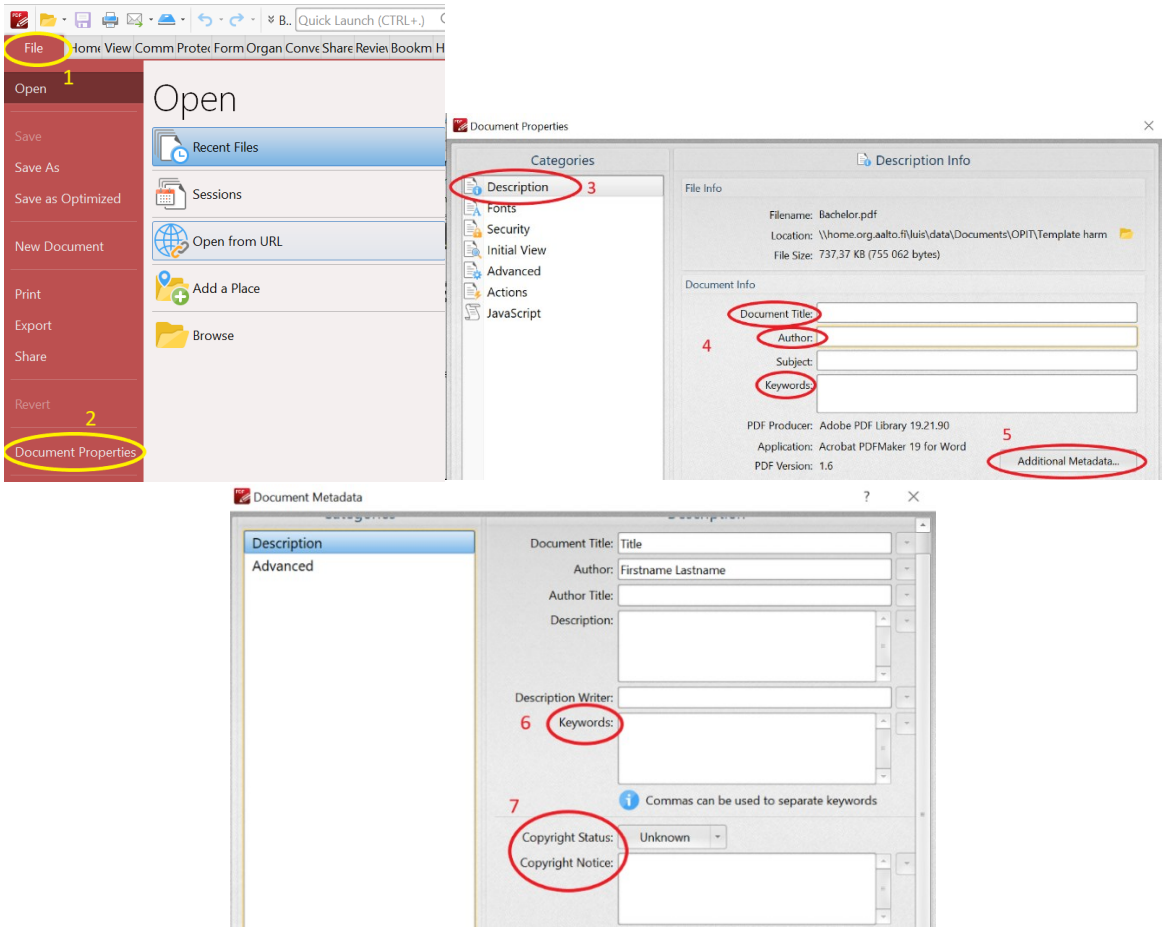

Figure 6: Adding metadata to a PDF file using PDF-XChange.

Test your PDF file

In a proper PDF file, both 'normal' and PDF/A, text is stored as text in the file. Hence, text can be highlighted with your mouse—one word and even one letter at a time—for instance to copy it. This **highlight test** is a simple and effective quality check for your file. If the text cannot be highlighted, the content has been stored as a bitmap (rasterised image) in the file. Such a file is not a 'proper' PDF and will not be accepted, for example, by Turnitin or Aalto's document archiving system. So, always check the quality of the PDF file you create.

Converting a file from some format to PDF or PDF/A or combining several PDF files can result in an improper PDF file with content as an image if the settings are incorrect or fonts are not embedded in the original PDF files. You will find below instructions on how to convert or combine your files correctly.

Having all the fonts used in your PDF file embedded in it is essential to create a PDF/A file; a PDF file without embedded fonts will result in the text being rasterised, sometimes the entire resulting PDF/A document being a collection of page images, or characters in that font are omitted or replaced with some symbol, say a square. The list of fonts—the embedded fonts have the text 'Embedded Set' or 'Embedded subset' alongside—can be viewed as follows:

Acrobat Reader: **File → Properties → Fonts** tab

PDF-XChange: **File → Document Properties → Fonts** (see Figure 7)

If no fonts are listed, the PDF file contains rasterised fonts only and so is improper. Use only proper PDF files to create PDF/A-compliant files.

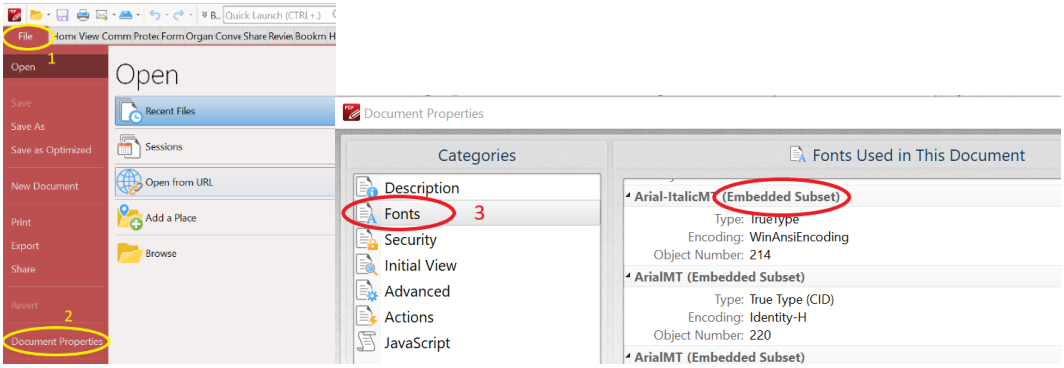

Figure 7: Using PDF-XChange to determine whether fonts are embedded in the PDF file. Embedded fonts are labelled as 'Embedded' or 'Embedded subset'.

Conversion tools: PDF-XChange

PDF-XChange pro/Editor, installed on all Windows workstations and managed by Aalto IT, is the recommended conversion tool. It is also available for home use by downloading it from https://download.aalto.fi. Alternatively, use the virtual desktop infrastructure environment, https://www.aalto.fi/en/services/vdiaaltofi-how-to-use-aalto-virtual-desktop-infrastructure, to use PDF-XChange.

Word-processor applications, like MS Word or LibreOffice, are also able to save documents as PDF/A -compliant files. However, the properness of the created PDF/A file may be compromised (see **Test your PDF file** above).

There are few ways to convert your file to PDF/A with PDF-XChange, some of which are successful in some situations and not so in others. The most important requirement dictating the success of the conversion is the availability of the fonts and glyphs used in the document. If a glyph in a font is unavailable, the conversion will fail. Also, the PDF-XChange version may affect the result. Convert Word documents using the method described in **Creating a PDF/A file from Word in Windows** or **Creating a PDF/A file from Word in a Mac** below. Regardless of how you create the PDF/A file, **always test the result** as described in **Test your PDF file** above and validate it (see *PDF/A validation* below).

PDF to PDF/A (Method 1)

Open the PDF file with PDF-XChange. Ensure that all the fonts used are embedded (see **Test your PDF file** above). Then do

- **File → Save As → Browse**;
- go to the folder where you want to save your file,
- set **Save as type → PDF/A (*.pdf)** from the drop-down menu,
- press '**Options**' and set '**Choose Conformance**' to '**PDF/A-3a**', '**PDF/A-2a**', '**PDF/A-3b**' or '**PDF/A-2b**' from the drop-down menu. Favour 'PDF/A-3a' or 'PDF/A-2a' for better accessibility, but if its creation fails, try 'PDF/A-3b' or 'PDF/A-2b'. Check the '**Embed Font Subset**' box. The sequence in this process is illustrated in Figure 8.

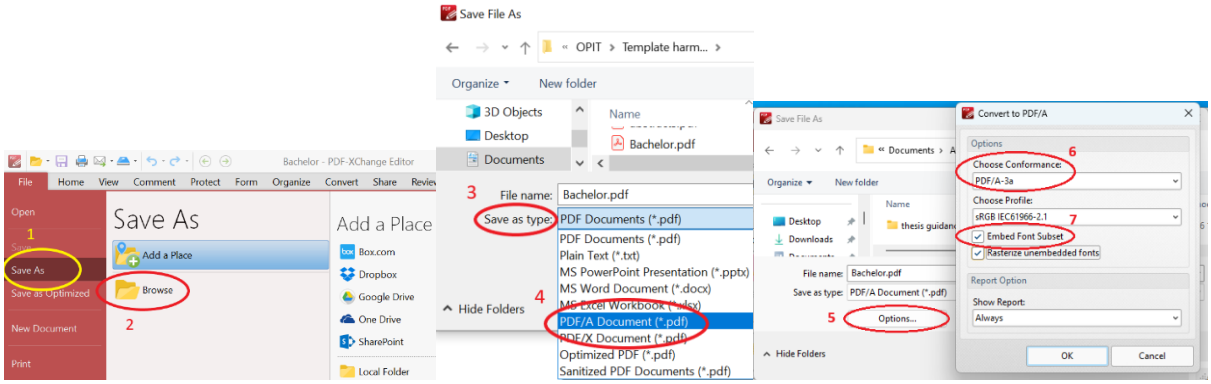

Figure 8: Saving the opened PDF file to conform with the PDF/A-2a or PDF/A-2b standard.

You may also check the 'Rasterize unembedded fonts' box, but this should not be necessary if all the fonts used in the original PDF are embedded in it.

docx to PDF/A (Method 2)

Open the Word (docx or dotx) document in PDF-XChange (**File → Open → Browse**) and the set file type to be opened to '**MS Word Document**' from the drop-down menu, as shown in Figure 9. Navigate to the folder where the file you want to open is located, choose or name the file to be opened, and press '**Open**'. PDF-XChange converts the file on-the-fly to PDF, which can take a while.

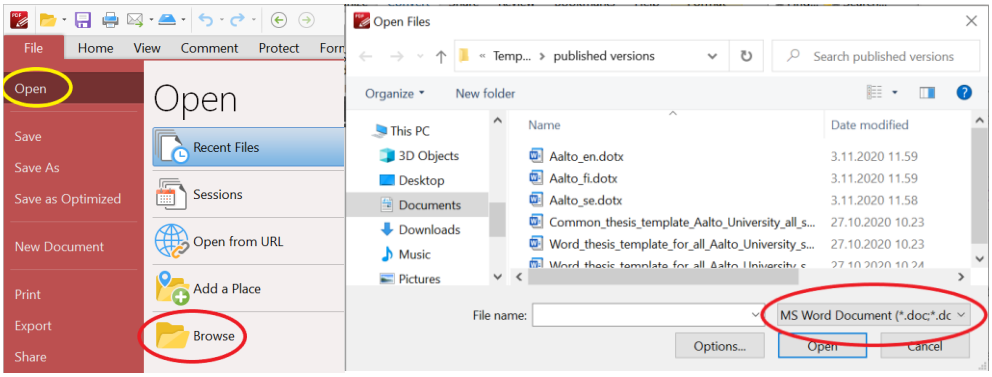

Figure 9: Opening a Word document in PDF-XChange.

The next step, saving the file with the desired PDF/A conformance, is identical to Method 1 above. That is, do **Save As → Browse**, go to the desired folder, set **Save as type → PDF/A Document (*.pdf) → Options → PDF/A-3a**, **PDF/A-2a**, **PDF/A-3b**, or **PDF/A-2b**. Check the box '**Embed font subset**'. Check also '**Rasterize unembedded fonts**', but this should be unnecessary because all fonts used in the Word document should be embedded in the converted PDF file. Figure 8 illustrates the process.

Combining PDF files in PDF-XChange

You can combine PDF files to create one PDF/A-compliant file, as shown in Figure 10. Add the files to be combined either by clicking '**Add files**' (4 in the figure) and picking the desired files or drag-and-drop them into the box in the order they are to be combined.

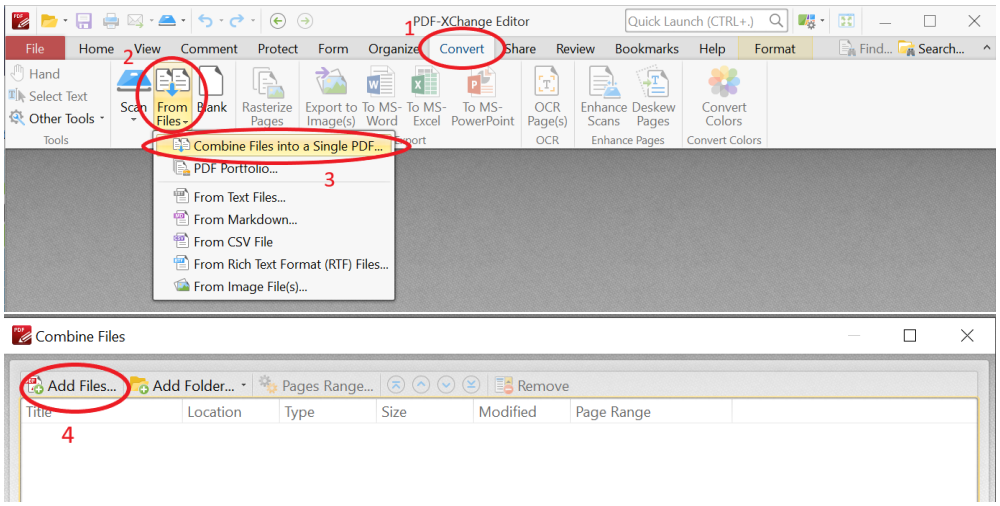

Figure 10: Combining PDF files to create one PDF/A-2a or PDF/A-2b compliant file.

Do not use docx files directly when combining files since some fonts may fail to get embedded without you knowing. Test the resulting file first (see **Test your PDF file** above) and then validate it (see **PDF/A validation** below). If the validation fails, create PDF/A-compliant files from the individual PDF ones first, combine them next, and then create the final PDF/A file.

PDF/A validation

Validate your file at https://aaltodoc.aalto.fi/sec/pdfa/. You can choose the interface language, **English** or **Finnish**. In the **PDF/A -file validation** dialog box, use the **Select file** button to choose your file, select the desired PDF/A standard (optimally PDF/A-3a), and click on the **Send file for validation** button. The result of a successful validation is shown in Figure 11.

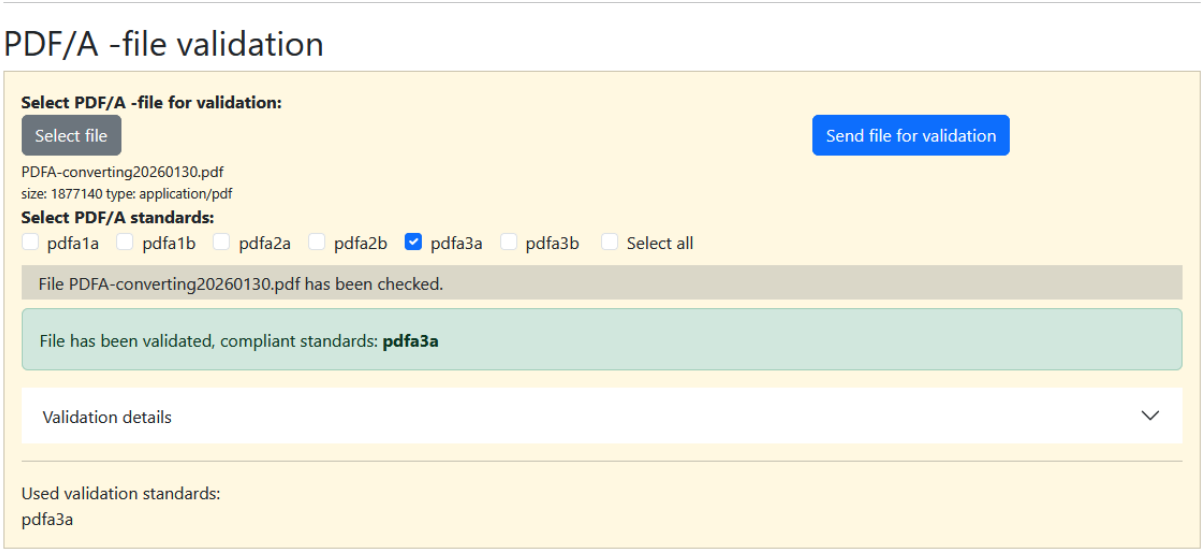

Figure 11: Result of a successful validation of a PDF/A-2b compliant file.

In the event of a failed validation, you can click on Validation details to try and figure out why the validation failed. The details might not be very informative to the unacquainted person.

Troubleshooting tips

- Favour the pdf format for images in LaTeX. Ensure that all fonts used in the image are embedded in the file.
- Use the jpg or png format for image files you add to your Word or LibreOffice publication.
- Use the Insert image function in the word-processing program:
  Microsoft Word: Insert -menu → Picture → choose file
  LibreOffice: Insert → Image → From file → choose file
- Do not use Copy–Paste or drag-and-drop functions for inserting images
- Test your pdf for its 'properness' and ensure that all fonts are embedded
- Add metadata to your PDF (File → Document Properties)

External links

https://en.wikipedia.org/wiki/PDF/A

Digital preservation: File Formats

PDF/A in a Nutshell 2.0