

Converting a file or files to a PDF/A compliant version

Contents

PDF/A document format	1
Test your PDF file	2
Add metadata	3
Conversion tools: PDF-XChange	4
PDF to PDF/A (Method 1)	4
docx to PDF/A (Method 2, not recommended for newer versions of Word)	5
Using Print to create a PDF/A file (Method 3, not recommended for newer versions of Word)	5
Creating a PDF/A file from Word	6
PDF/A-2 compliance	6
PDF/A-1a compliance	7
LaTeX (this is not a conversion tool)	8
Combining PDF files in PDF-XChange	8
PDF/A validation	9
Troubleshooting tips	9
External links	9

PDF/A document format

PDF/A is the ISO-standardized version of the Portable Document Format (PDF) specialized 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. The current available PDF/A standards are PDF/A-1 through to PDF/A-4. The valid and acceptable versions for archiving your thesis at Aalto University are PDF/A-1a, -1b, -2a and -2b. The 'a', meaning accessible, has more stringent requirements than 'b', meaning basic. '1' refers to the older standard from 2005, and it forbids the use of transparency in images, whereas '2' refers to the next version of the standard, published in 2011, and allows transparency in images. Thus, if you have images with transparencies in your thesis, which is usually the case, use PDF/A-2. The following version, '3', published in 2012, allows attaching other types of files as part of the PDF/A file and hence is considered unsuitable for archiving theses. Published in late 2020, '4' is the latest version of the standard and is not widely used yet.

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

More details on the format and creating a PDF/A compliant file are available, for example, at https://aaltodoc.aalto.fi/doc_public/ohjeet/pdfa_thesis_guide.pdf.

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, it 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. 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 rasterised text 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 1)

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.

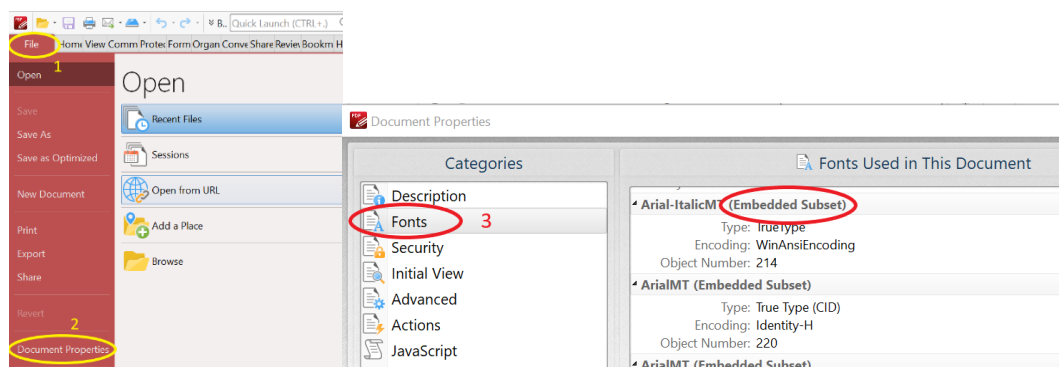


Figure 1: Using PDF-XChange to determine whether fonts are embedded in the PDF file. Embedded fonts are labelled as 'Embedded' or 'Embedded subset'.

Add metadata

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 2).

Fill in the metadata fields in the dialogue box (4 in Figure 2). Add the keywords again in 'Additional Metadata' (5 and 6 in Figure 2) 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 2). Remember to **test your file** after saving the file.

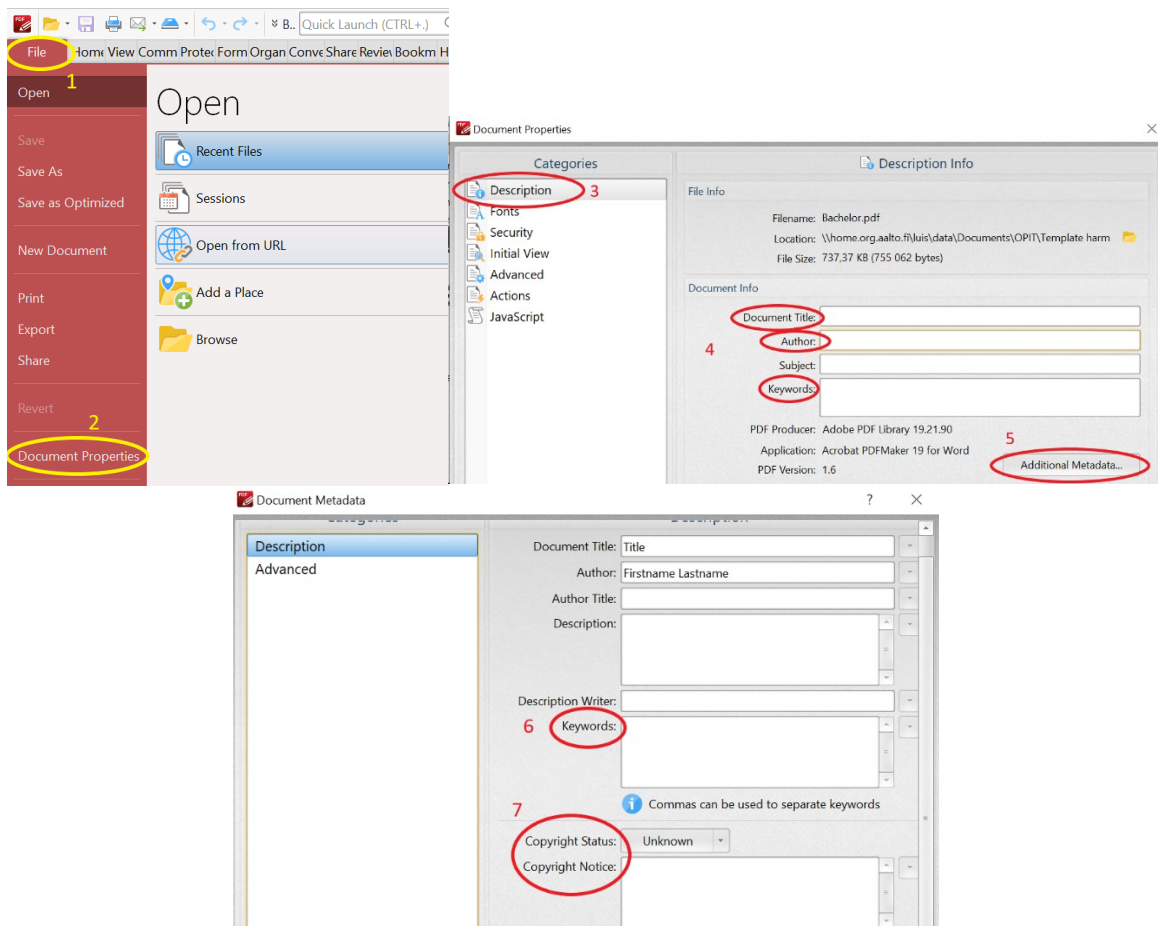


Figure 2: Adding metadata to a PDF file using PDF-XChange.

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 at <https://download.aalto.fi>.

Alternatively, use the virtual desktop 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). The PDF/A file produced by MS Word (in the Microsoft Office 365 version 2102, build 13801.21092 and the current version 2108 build 14326.20738, January 2022) when set to be PDF/A compliant claims PDF/A-3 compliance and so is unacceptable for archiving your work, say, in Aalto University's digital thesis collection. Curiously, the created file using both Word versions fail the validation at the site mentioned in *PDF/A validation* below as well as in Adobe's Acrobat Pro.

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. At the time of writing (January 2022), version 9.0 build 354.0 fails to create a proper PDF/A file using methods 2 and 3 below for Word and PowerPoint documents. The result is a valid PDF/A file but improper because it is collection of image pages. Methods 2 and 3 are presented here because they have worked with older versions of Word. Convert Word documents using the method described in *Creating a PDF/A file from Word* 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 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-2a' or 'PDF/A-2b' from the drop-down menu. Favour 'PDF/A-2a' for better accessibility, but if its creation fails, use 'PDF/A-2b'. Check the 'Embed Font Subset' box. The sequence in this process is illustrated in Figure 3.

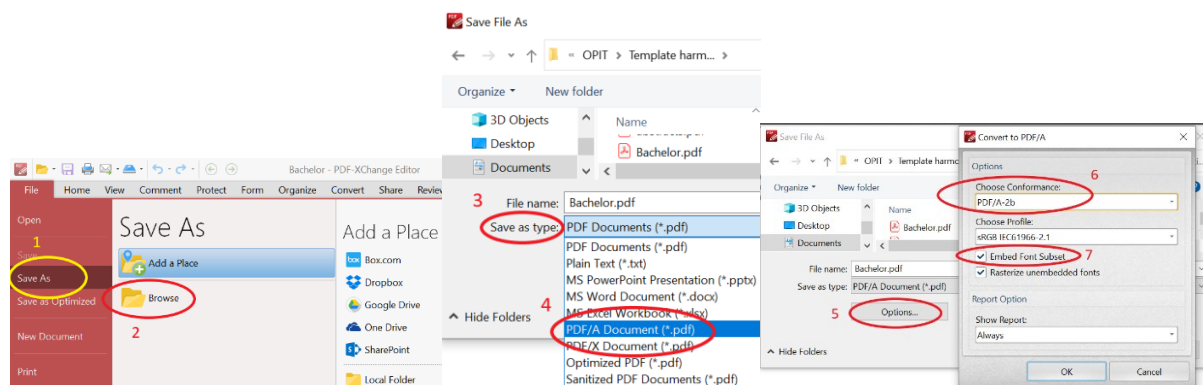


Figure 3: Saving the opened PDF file to conform with the PDF/A-2a or PDF/A-2b standard.

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, not recommended for newer versions of Word)

Open the Word (docx or dotx) document in PDF-XChange (Open → Browse) and set file type to be opened to 'MS Word Document' from the drop-down menu, as shown in Figure 4. 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.

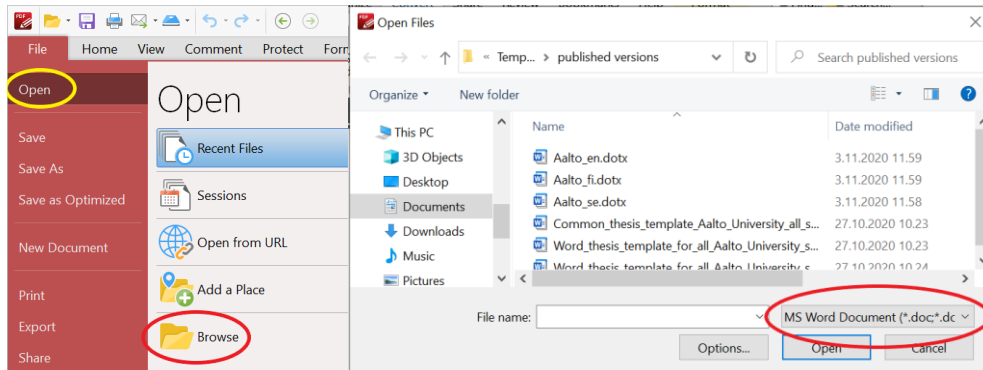


Figure 4: Opening a Word document in PDF-XChange.

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-2a 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 3 illustrates the process.

Using Print to create a PDF/A file (Method 3, not recommended for newer versions of Word)

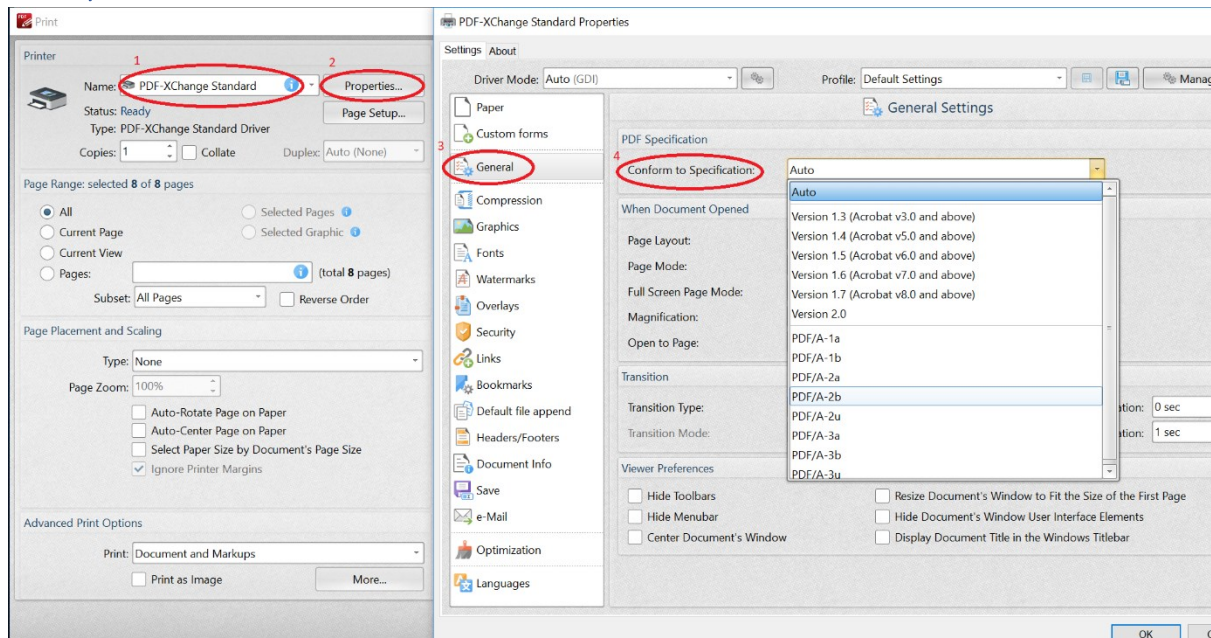


Figure 5: First set the printer name to 'PDF-XChange' (1), and then pushing the 'Properties' button (2) will open a new dialog box. Choose 'General' (3) in order to specify the desired PDF format (4).

At the time of writing (January 2022), this method tends to create an improper PDF/A file and so is not recommended. Nonetheless, to make the conversion, use the PDF-XChange printer; that is, in

the 'Print' dialog box set 'Name' to 'PDF-XChange Standard' from the drop-down menu and then push the 'Properties' button (see Figure 5):

File → Print → PDF-XChange Standard → Properties → General → choose: PDF/A-2a or PDF/A-2b

Creating a PDF/A file from Word

PDF/A-2 compliance

Creating a PDF/A-2 file from Word is a three-phase process. In the first phase, save the Word document as a PDF/A-3a¹ file as follows (see Figure 6):

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

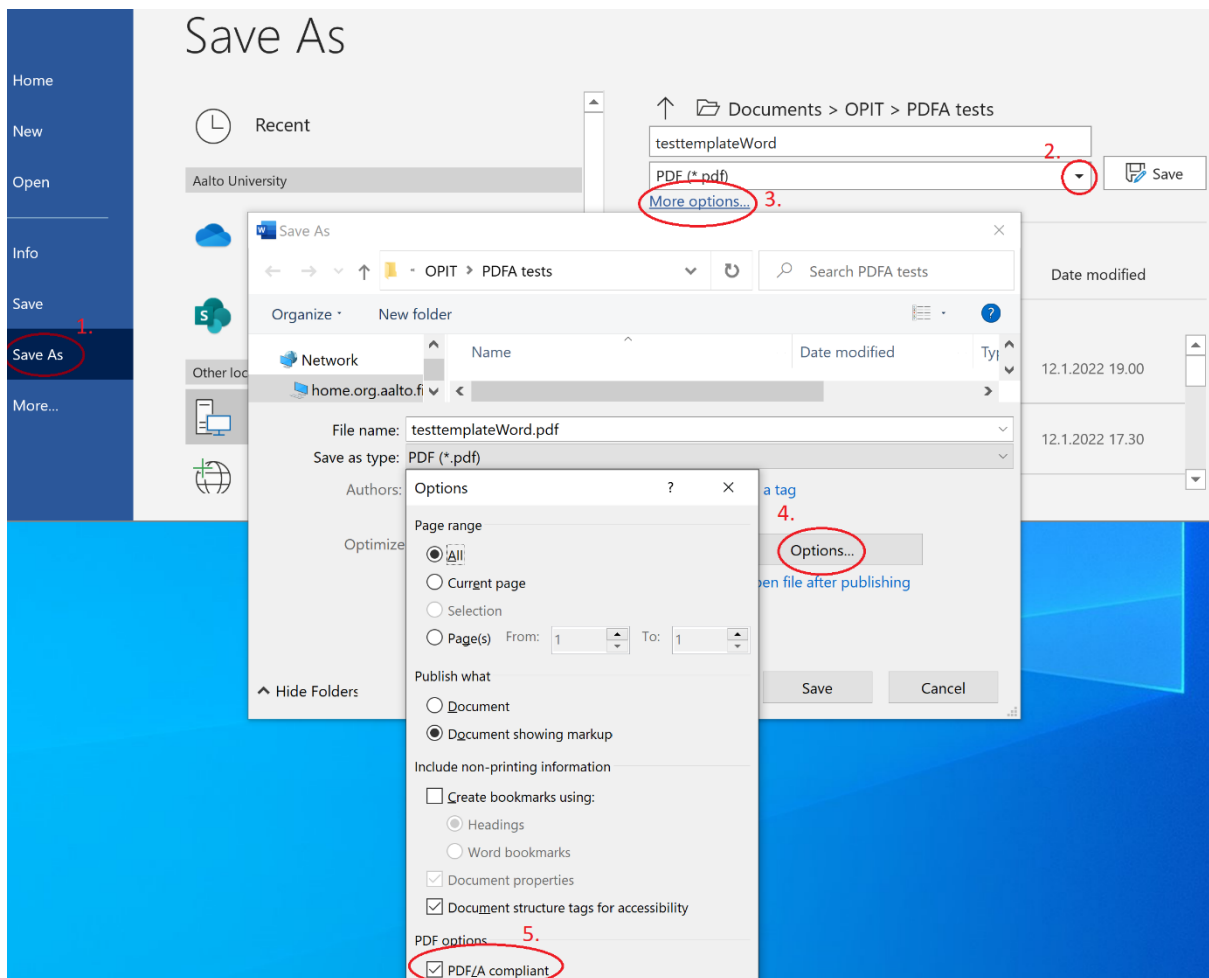


Figure 6: Saving a Word document as a PDF/A-3a file.

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. Also, as mentioned earlier, the created file fails validation elsewhere. Hence a multiphase process to create a PDF/A-2 file is necessary.

In the second phase, open the created PDF/A file in PDF-XChange and add metadata as described in *Add metadata* above. Finally, in the third and final phase, save the file as a PDF/A-2a or PDF/A-2b file, as described in *PDF to PDF/A (Method 1)* above.

PDF/A-1a compliance

If creating a PDF/A-2b compliant file fails and PDF-XChange is not available, you can try to create a file with PDF/A-1a compliance, which is acceptable for Turnitin as well as for archiving in Aalto University's digital thesis collection. Do either **File** → **Save As Adobe PDF** or **File** → **Export** → **Create Adobe PDF** → **Create Adobe PDF**, as shown in Figure 7 for the latter.

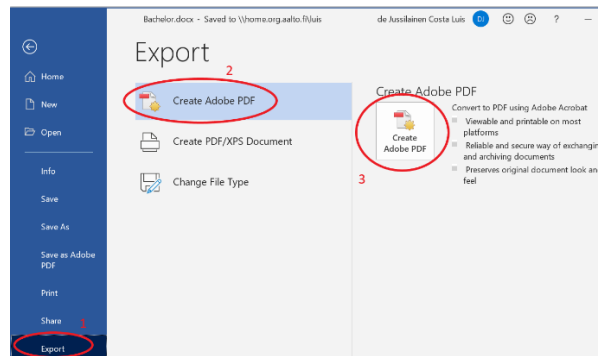


Figure 7: Exporting to create a file with PDF/A-1a compliance.

The next dialog box that appears, shown in Figure 8, 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/A1a conformance.

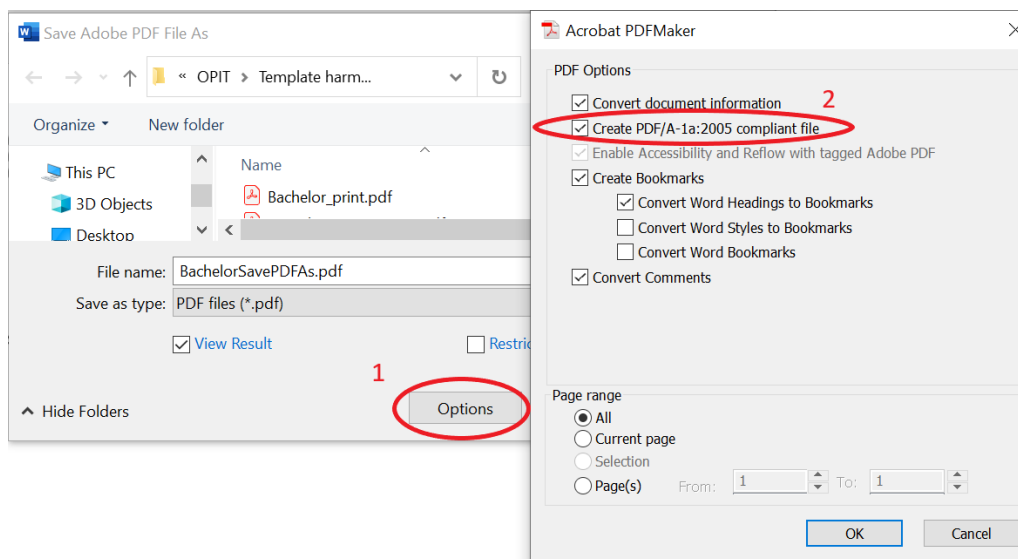


Figure 8: Settings to create a file with PDF/A-1a conformance.

LaTeX (this is not a conversion tool)

A LaTeX thesis template and class file, developed at Teacher Services, which directly produces a PDF/A-2b file is available at <https://wiki.aalto.fi/display/Aaltothesis/Aalto+Thesis+LaTeX+Template>.

The template, which supports text in Finnish, Swedish and English, is self-explanatory. Do not make any adjustments to create a PDF/A file, simply set the option in `\documentclass` to `a-1b` or `a-2b` to create a PDF/A-1b or PDF/A2-b compliant file, respectively. If the generated PDF/A turns out not to be a valid PDF/A, see the guide at https://aaltodoc.aalto.fi/doc_public/ohjeet/pdfa_thesis_guide.pdf for help. If the guide does not help you, you can try to fix the bad PDF file and convert it into a valid PDF/A format by using PDF-XChange, as described above in *PDF to PDF/A (Method 1)*.

Combining PDF files in PDF-XChange

You can combine PDF files to create one PDF/A-compliant file, as shown in Figure 9. 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.

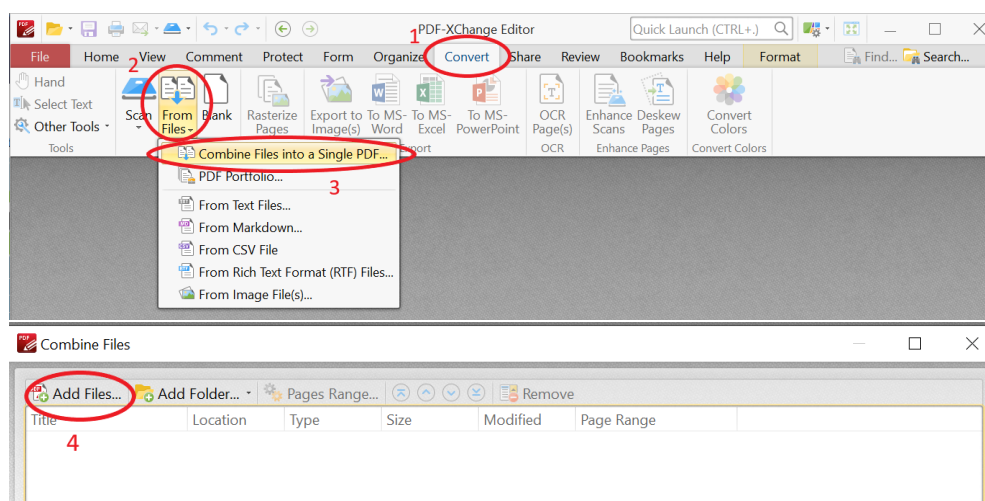


Figure 9: Combining PDF files to create one PDF/A-2a or PDF/A-2b compliant file.

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://www.pdf-online.com/osa/validate.aspx>. Drag-and-drop your file into the box there. The result of a successful validation is shown in Figure 10.

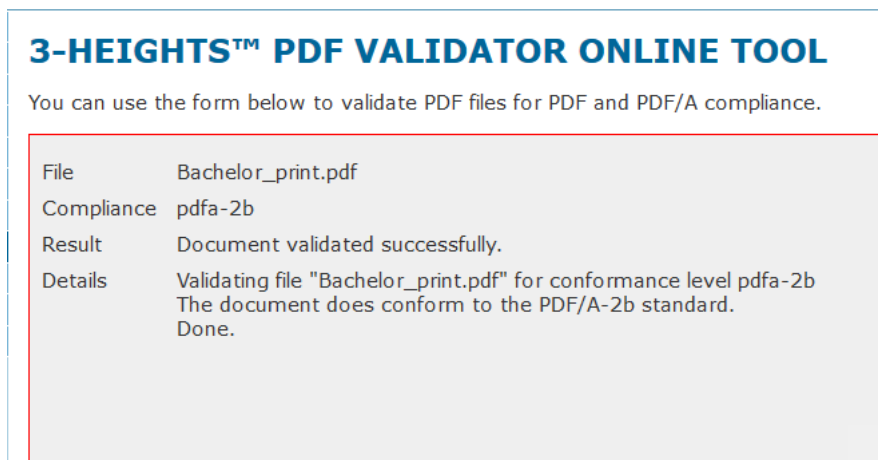


Figure 10: Result of a successful validation of a PDF/A-2b compliant file.

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.

External links

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

[Kansalliset pitkäaikaissäilytyspalvelut \(CSC\): Säilytys- ja siirtokelpoiset tiedostomuodot](#)

[Petersen-Jessen, Jari, 2009, PDF-tiedostomuodon hyödyntäminen eduskunnassa](#)

[PDF/A in a Nutshell 2.0](#)