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

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 user of the file need not have the same fonts used to create the file installed on their computer in order to read it. The valid and acceptable versions for archiving your thesis 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, and it forbids the use of transparency in images, whereas '2' refers to the newer one and allows transparency in images. Thus, if you have images with transparencies in your thesis, which is usually the case, use PDF/A-2b.

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 of paramount importance to create a PDF/A file; a PDF file without embedded fonts will result in the text being rasterised, sometimes the entire PDF 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 extracted 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.

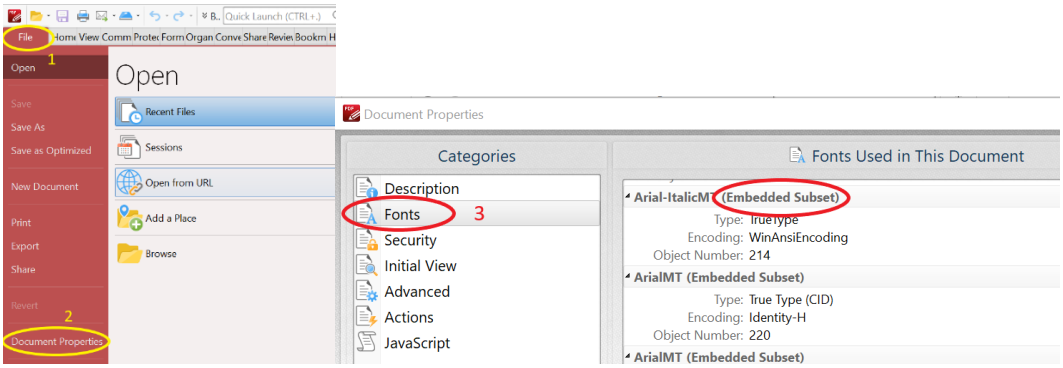


Figure 1: Using PDF-XChange tp determine whether fonts are embedded in the PDF file. These fonts are embedded.

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. If you work on your final PDF/A file, you will have to enable editing it and resave it with the appropriate compliance after adding 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. Specify the copyright status you wish to give your document (7 in Figure 2). Remember to test your file at the end after saving.

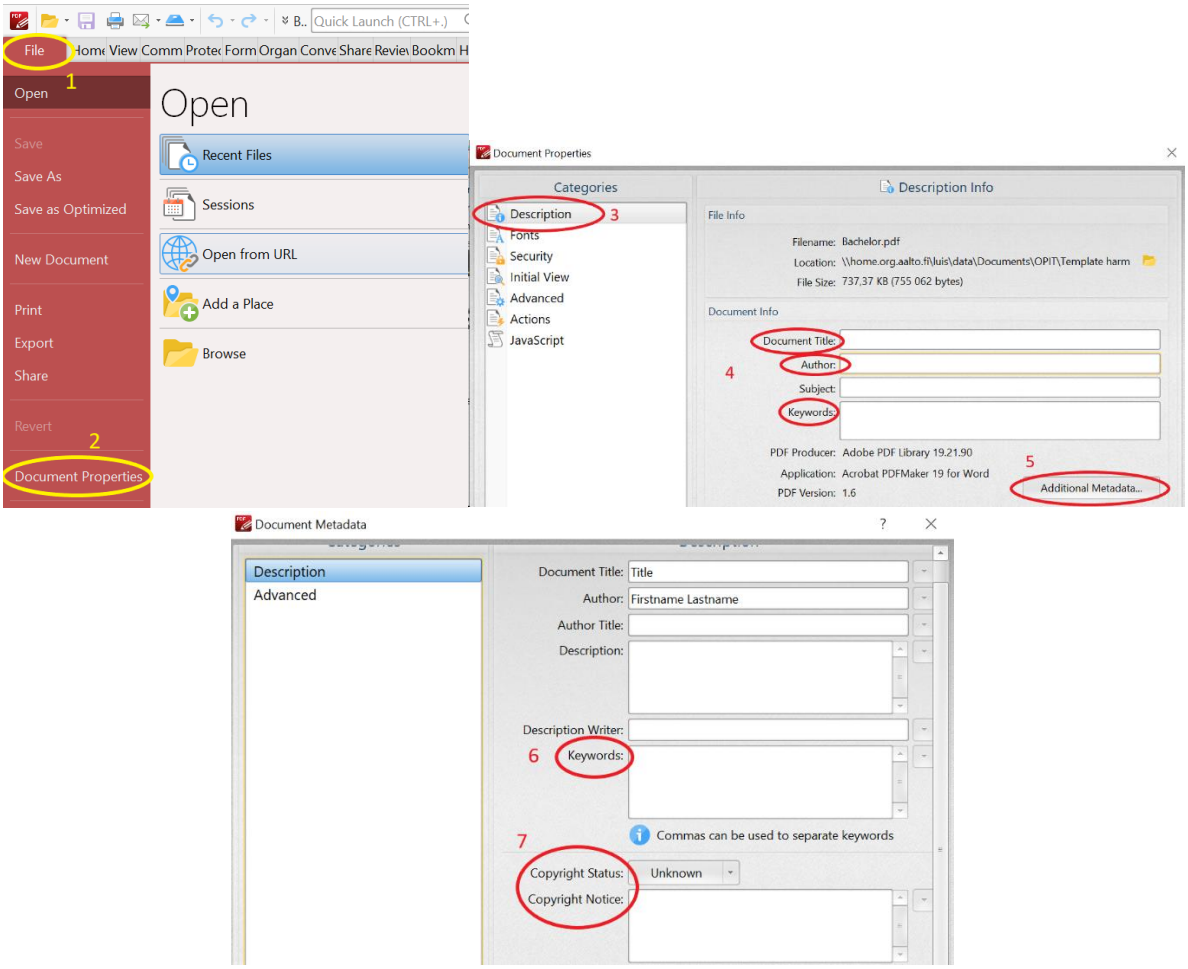


Figure 2: Adding metadata to a PDF file using PDF-XChange.

Conversion tools

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 quality or properness of the created PDF/A file may be compromised (see *Test your PDF file* above). The PDF/A files produced by MS Word (in the current Microsoft Office 365 version) are easily flawed if the settings are not adjusted.

PDF-XChange

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. Also, the PDF-XChange version may affect the end-result. At the time of writing (December 2020), version 8.0 build 339.0 fails to create a proper PDF/A file in many situations using the Print method (Method 3). This is the only the method given in the older instructions to create your PDF/A file and so is described here as well. Hence methods 1 and maybe 2 are the preferred methods of conversion. 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-2b' from the drop-down menu. Check the 'Embed Font Subset' box. The sequence in this process is illustrated in Figure 3.

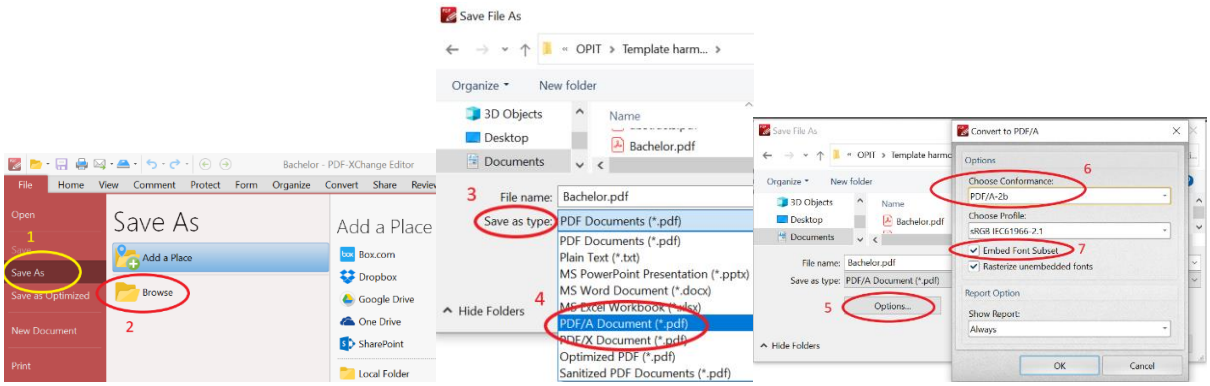


Figure 3: Saving the opened PDF file to conform with the PDF/A-2b standard.

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 (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 will convert the file on-the-fly to PDF, which can take a while.

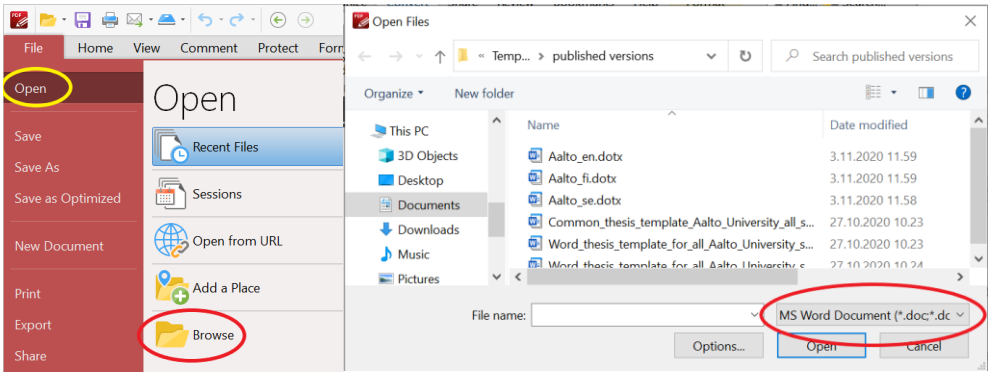


Figure 4: Opening a Word document in PDF-XChange.

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-2b. Check the box 'Embed font subset'. Check also 'Rasterize unembedded fonts', but, again, this should not be required because all fonts used in the Word document should be embedded in the converted PDF file on opening it. This process is illustrated in Figure 3.

Using Print to create a PDF/A file (Method 3)

At the time of writing (December 2020), 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:

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

Ensure first in the 'Print' dialog box that 'Name' is set to 'PDF-XChange Standard' from the drop-down menu and only then push the 'Properties' button (see Figure 5).

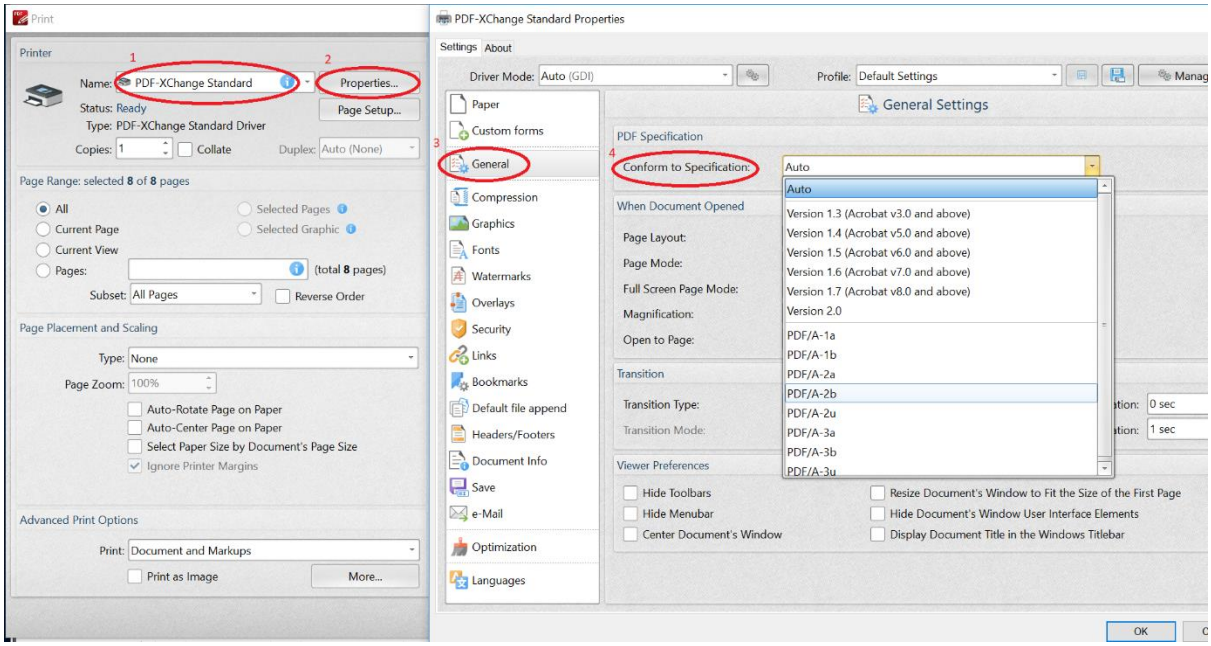


Figure 5: First set the printer name to 'PDF-XChange' (1) and then pushing the 'Properties' button (2) will open a new dialog box. Choose 'General' (3) in order to be able to specify the desired PDF format (4).

Creating a PDF/A file directly with Word

PDF/A-2b conformance

Currently, the only way to create a file with PDF/A-2b conformance is to 'Print' the file, as shown in Figure 6.

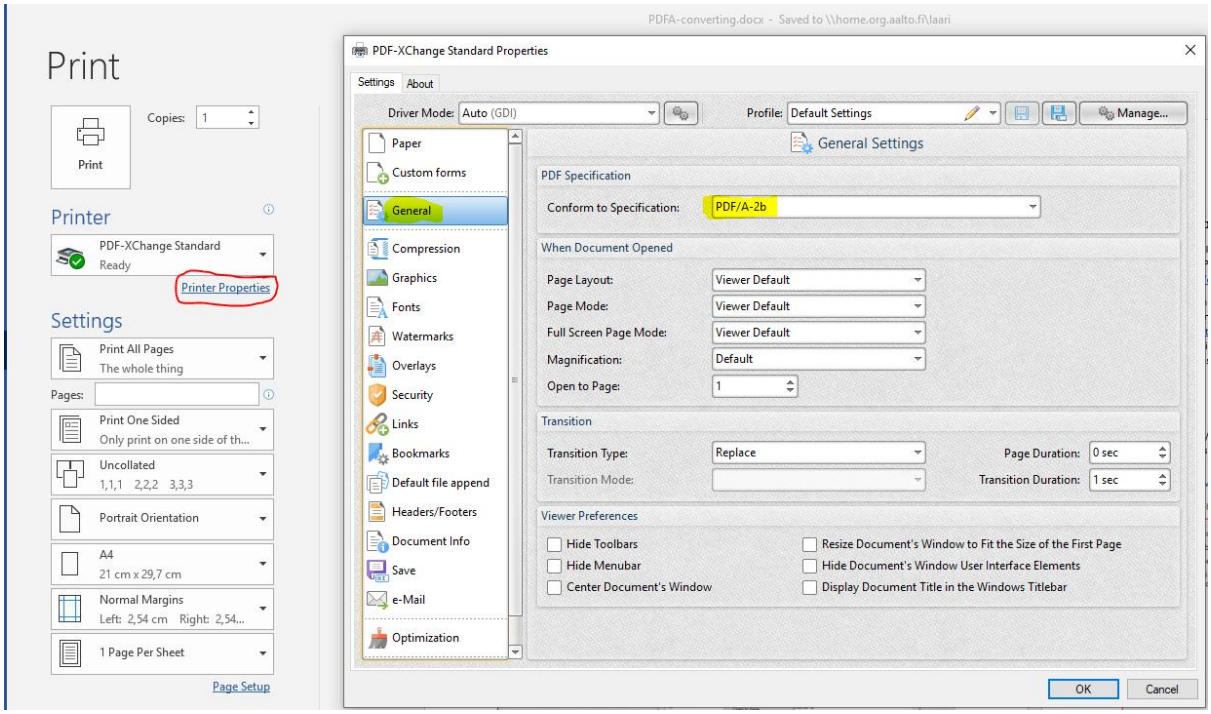


Figure 6: Creating a file with PDF/A-2b conformance directly from MS Word (Microsoft Office 365).

PDF/A-1a conformance

If creating a file with PDF/A-2b conformance fails and PDF-XChange is not available, you can try to create a file with PDF/A-1a conformance, 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 the latter.

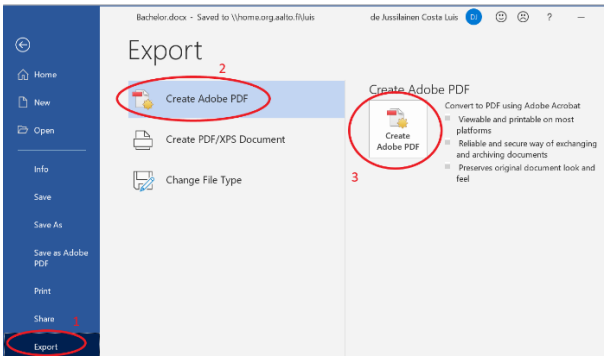


Figure 7: Exporting to create a file with PDF/A-1a conformance.

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.

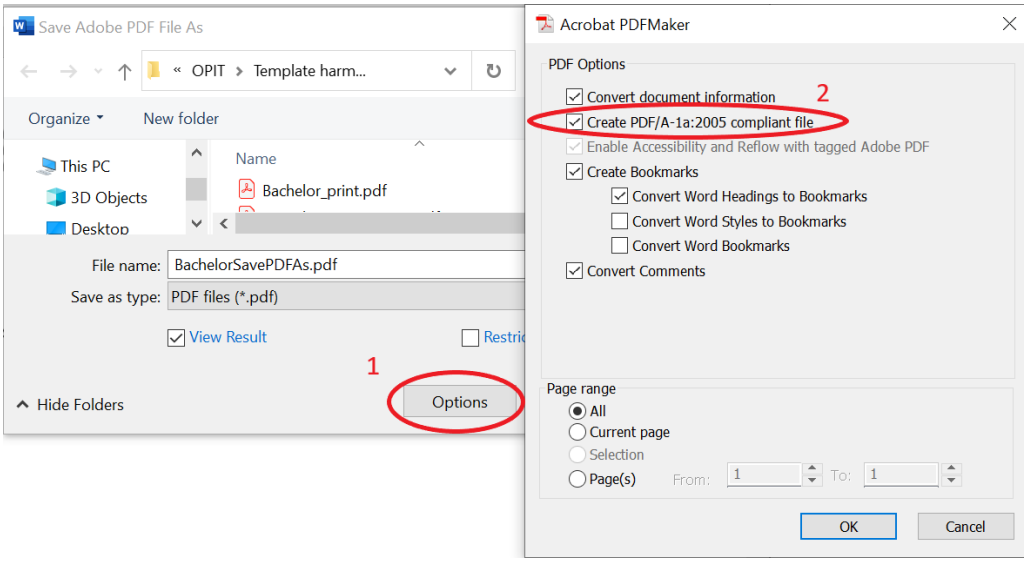


Figure 8: Settings to create a file with PDF/A-1a conformance.

LaTeX (this is not a conversion tool)

A LaTeX thesis template and class file, developed at the School of Electrical Engineering, which directly produces a PDF/A 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-documentary. If the generated PDF/A is not a valid PDF/A, see the guide at https://aaltodoc.aalto.fi/doc_public/ohjeet/pdfa_thesis_guide.pdf for help. If that does not help you, you can try to fix the PDF file and convert it into a valid PDF/A -format by using the PDF-XChange program, as described above in Method 1.

Combining 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 combine 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.

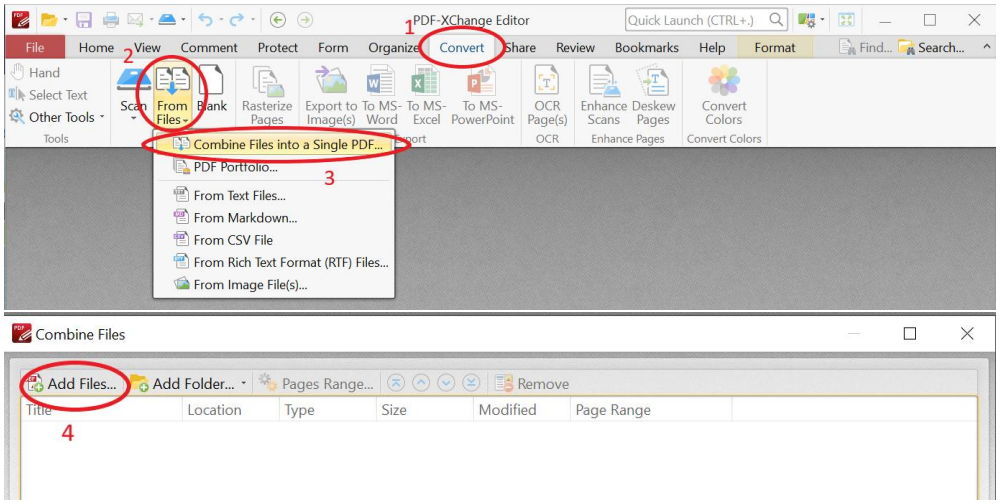
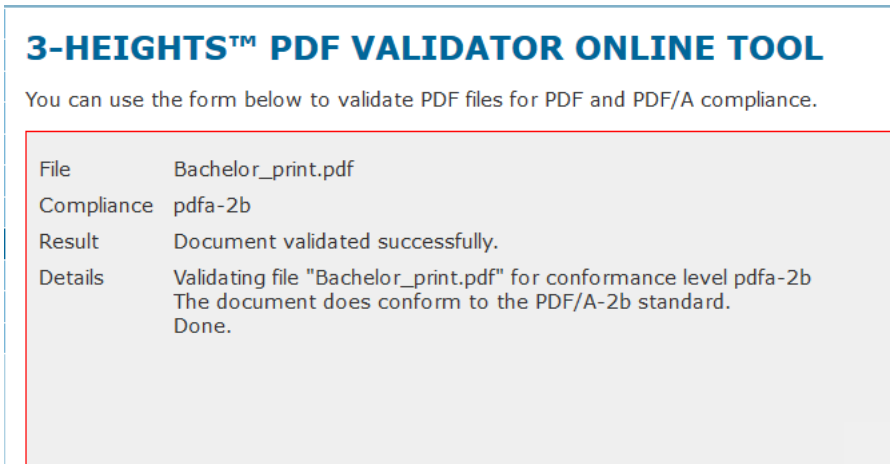


Figure 9: Combining PDF files to create one PDF/A-2b compliant file.

Do not use docx files directly when combining files since some fonts may fail to get embedded without your knowledge. Validate the resulting file (see *PDF/A Validation* below). If the validation fails, create PDF/A-compliant files from the individual PDF ones, then combine them, 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 will look like this (pdfa-1b is also an acceptable file type):



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.

If you have a problem with converting or validating file, contact: IT Service Desk: servicedesk@aalto.fi

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