Conversion of Microsoft Word or PowerPoint Files to PDF /A-1b

The ISO standard PDF/A was established to use the popular PDF file format by Adobe Systems for long-term preservation. Word or PowerPoint files are thus often converted to PDF/A before they are delivered to digital archives. In contrast to conventional PDF formats, the PDF/A format does not allow features that are not suitable for long-term archiving. In particular, PDF/A stores all information needed for proper visual presentation within the file (such as all used fonts). We recommend using the compliance level PDF/A-1b, which is designed for accurate visual presentation and does not allow embedded files.

The current document is divided into three sections. In the first section, we explain how to convert elaborate scientific documents to PDF/A-1b. In the second section, we describe several alternative conversion methods. In the third section, we summarize the results of our tests with these methods.

1. Instructions

To convert an elaborate scientific document from MS Word or PowerPoint to PDF/A-1b, we recommend to explicitly show URLs of hyperlinks in the text and to use the following method[1]:

a) In MS Word or MS PowerPoint select the tab "File" and click "Print".

b) In the printer drop down menu select "Adobe PDF". (If the entry "Adobe PDF" is missing, only the Adobe Reader may be installed on your computer. In this case you may either use conversion method 3 or install Adobe Acrobat Standard or Professional2.)

c) Select the "Properties" box. In the tab "Adobe PDF Settings" choose in the "Default Settings" drop down menu the option "PDF/A-1b: 2005 (RGB)".

d) Click "Print" to create the PDF file.

You should carefully verify the quality of the generated PDF file by inspecting formulas, hyperlinks, special characters, special fonts, text errors, selecting and searching text, comments, tables, colours, transparent objects, vector graphics, and multiple graphic layers.

2. Description of conversion methods

We tested four methods to generate PDF/A-1b files from MS Word or MS PowerPoint files (MS Office Professional Plus 2010, version 14.0). Only method 3 uses free software. Methods 1, 2 and 4 require the paid products Adobe Acrobat Standard or Adobe Acrobat Professional[2]:

1) Adobe Acrobat Professional may be used for conversions. Start Adobe Acrobat Professional and open your Word or PowerPoint File from within Adobe Acrobat Professional. In the menu "File" you select "Save as …", then "More options". In the new menu you select "PDF/A". After clicking on "Settings" you select "PDF/A-1b" and "sRGB".[3]

2) A Word or PowerPoint file may be converted to PDF/A-1b using the Acrobat PDFMaker. Start Word or PowerPoint, open your document, in the Acrobat tab select "Conversion settings", and then select "PDF/A-1b" (once for each installation). To create the PDF file you may either select "Create PDF" in the tab "Acrobat" or "Save as Adobe PDF" in the tab "File".[4]

3) You may save your file as PDF/A-1b with Word or PowerPoint. In the tab "File", click "Save as". In the new dialog box use the drop down menu to select the format PDF/A. Click "Options" and in the new dialog box select „ISO19005-1-compatible PDF/A" (ISO19005-1 is PDF/A-1; the created file should thus be consistent with PDF/A-1b).[5]

4) You may "print" a file to an Adobe PDF printer with Word or PowerPoint. This is the recommended method described above in the section "Instructions".

3. Tests with these conversion methods

Five Word files and five PowerPoint were selected to test these four conversion methods. These files contained elaborate scientific publications with figures, equations, and special characters. The four conversion methods were evaluated as follows:

- It was visually checked whether the printable properties were mapped faithfully. In particular, figures, formulas, special fonts and text were examined.
- Endnote links and implicit hyperlinks (web address not given in text) were checked. This issue was given low priority as such links are not common in scientific publications.
- Computation time and file size
- The validity of the generated PDF / A-1b was verified with Adobe Preflight (Ctrl+Shift+X in Adobe Professional XI)

The preservation of tags or metadata was not assessed. Likewise, preservation of comments below PowerPoint slides was not checked.

Our test results show that one should use method 4 (table 1). With this method no differences to the originals were found by visual inspections, and all generated files were compliant with the PDF/A-1b standard. In addition, this method produced the smallest files and the computation time for these documents (up to 40 pages) did not exceed a few seconds.

Table 1: Results of the four conversion methods

Criteria	1) Open and convert using Adobe Acrobat	2) Acrobat PDFMaker	3) Save as PDF	4) Adobe PDF printer (Distiller)
Visual inspection	No errors, but partially unsufficient image quality. Furthermore, some text could neither be searched nor copied.	One file could not be reliably converted	In 4 out of 10 files some figures were not acceptable, as transparent objects had become non-transparent.	No errors
Links und references	No errors	No errors	Broken endnote links	Broken hyperlinks, broken endnote links
Compliance with PDF/A-1b standard	No errors	Four of the 10 files were not compliant with the PDF/A-1b standard	One of the 10 files was not compliant with the PDF/A-1b standard.	No errors
Size of the PDF /A-1b files	Up to four times larger than original	Up to five times larger than original	Up to six times larger than original	Up to twice as large as original
Computing time (Intel Core i7, 3.4 GHz)	Up to several minutes	Up to several minutes	Up to several seconds	Up to several seconds

[1] See the following website for further instructions (accessed 01/29/2015): http://blogs.adobe.com/acrolaw/2007/01/pdfa_in_action . Acrobat Distiller converts from Postscript (*.ps) to PDF. Also with the free OpenOffice 3 for Windows (Oracle) you can convert to PDF/A-1b with Adobe Distiller.

[2] Members of the ETH may order Adobe Acrobat Professional at https://ides.ethz.ch. Please contact your IT support. See also Adobe Acrobat XI : Product comparison, http://www.adobe.com/products/acrobat/product-comparison.html (accessed 01/29/2015)

[3] See the following website for further instructions (accessed 29/01/2015): http://blogs.adobe.com/acrolaw/2011/05/using-save-as-to-to-conform-to-pdfa

[4] See the following website for further instructions (accessed 29/01/2015): http://helpx.adobe.com/acrobat/using/creating-pdfs-pdfmaker-windows.html

[5] See the following website for further instructions (accessed 29/01/2015): https://support.office.com/en-us/article/Save-as-PDF-d85416c5-7d77-4fd6-a216-6f4bf7c7c110#bm11