



PDF/A Guidelines

*Guide to PDF/A document creation,
conversion and validation*

**Version 1.0
May 17 2022**

Change Log

Revision	Date	Description	Author
Draft	5/17/22	v1.0 release	

Background

Per [section 8A.2.2 of the PDS4 Standards Reference](#), PDS4-compliant formats for documents are “flat UTF-8” text, PDF/A-1a (which is preferred), or PDF/A-1b. The International Organization for Standardization (ISO) has created various PDF subset standards which are specialized for different uses. The PDF/A standard (ISO 19005) is designed for long-term preservation. PDS4 requires compliance specifically to the first version, PDF/A-1.

PDF/A differs from PDF by prohibiting features unsuitable for long-term archiving, such as font linking (as opposed to font embedding) and encryption. PDF/A comes in many varieties, referred to as profiles. Profiles are the specific sets of validation rules that each document conforms to. Per the profiles that the PDS uses are PDF/A 1-A and PDF/A 1-B.

The following document provides best practices for authoring documents for PDF/A generation, as well as how to create, convert, and validate PDF/A documents prior to submission to the PDS.

Creating PDF/A-compatible Documents

The following set of best practices can be used to help create documents that can help ensure valid conversion to a PDF/A-1a- or PDF/A-1b-compliant PDF. These best practices are based upon experience of PDS personnel and summarized from various sources. See the [More Details section below](#) for references and if you are interested in learning more.

No multimedia

Multimedia documents are not supported by PDF/A.

No scripts

Scripts are not supported by PDF/A.

No external references

PDF files can dynamically alter their content during visualization. Active content that imports data from an external file is not supported by PDF/A. The document must be entirely self-contained.

No transparencies

Transparencies need to be reduced down to their non-transparent equivalents, which causes artifacts, and increases file sizes. Many image formats introduce transparencies, without you knowing (e.g. IF, PNG, vectors, etc.). To avoid this, JPEG images are recommended as transparencies are *flattened* by default in this format. [See below for more details on transparencies.](#)

No encryption

Encryption is not supported by PDF/A.

Free / Open Source fonts when possible

Some non-free fonts may be allowed when working with their originating programs (such as using Microsoft fonts with Word or PowerPoint), but they will prevent a converter from converting them to PDF/A later.

Additionally, fonts protected by copyright cannot be copied into a PDF/A document and can often cause a failure in generation. Usually a program that creates PDF/A files will give a warning if a font file cannot be embedded. The problem is not very prevalent with standard western fonts, since most font developers allow their fonts to be embedded. However, some fonts (including many Japanese fonts) are copyrighted and cannot be embedded. [3]

Accessibility

- Use tools that preserve the structure of the document - Carefully authored Word documents are the most likely candidate here.
- Describe images and symbols with alt text.

PDF/A does not support 3D models

Designs created in 2D can be archived as PDF/A without any problems. This is not the case for 3D models. Three-dimensional designs have only been supported since Acrobat 7 (PDF 1.6). They are therefore not permitted in PDF/A-compliant files. [2]

Other

Scanned documents are OK, but they will not be PDF/A 1-a compliant, only PDF/A 1-b

Creating a PDF/A

To create a PDF/A from scratch there are multiple software options to do so. The process for creating a PDF/A document varies depending on which PDF/A compatible software is being used.

Note: Be sure to review the Best Practices above prior to attempting to create a PDF/A to ensure a successful conversion and validation of the document files.

Microsoft Office

The Windows version of Microsoft Office supports PDF-A 1-B export.

1. Choose **File** → **Export** → **Create PDF/XPS**
2. Check the **ISO 19005-1 Compliant (PDF/A)** checkbox
3. Click **OK**

Adobe Acrobat

Adobe Acrobat supports 1-A and 1-B export. There are multiple ways to do this in Acrobat. See [Adobe's online documentation](#) for more details.

LibreOffice

To export a PDF/A:

1. Select **File** → **Export...** → **Export as PDF...**
2. Menu of PDF export options should appear.
3. From menu, check **Archive (PDF/A, ISO 19005)**
4. Select PDF/A 1-b under **PDF/A version**
5. Click **Export**

Converting to PDF/A

Converting an existing document to a PDF/A can also be done using different methods

Adobe Acrobat

Adobe Acrobat is probably your best choice for converting PDF/A documents. In Adobe Acrobat DC, you can choose `File -> Save as Other -> Archivable PDF (PDF/A)`. Adobe supports several other programs, and there are other ways to do this that give you more fine-grained control. See (<https://helpx.adobe.com/acrobat/using/pdf-x-pdf-a-pdf.html>) for more information.

Online converters

Online converters are an easy way to convert small batches of PDF documents. One converter that seems relatively easy to use is [PDFTron](#). To convert:

1. Drag your file onto the page
2. Select the output profile that you want
3. Click "Convert"

LibreOffice

Libre Office can import PDF documents and re-export them as PDF/A, as described above. However, you'll need to spend some time cleaning up the document. You may run into font issues, or your paragraphs may be out of order.

OCRMyPDF

A great way to convert PDF files using the command-line is with OCRmyPDF (<https://ocrmypdf.readthedocs.io/en/latest/>). OCRmyPDF will automatically convert a document to PDF/A, optionally running OCR on the document during the conversion. This makes OCRMyPDF suitable for converting both scanned and native PDF documents. Here is a sample command:

```
ocrmypdf --output-type=pdfa-1 document.pdf document.pdfa.pdf
```

PDF/A Validation Tools

PDS Validate Tool (recommended)

The PDS Validate Tool wraps the [VeraPDF Validator library](#) and validates the PDF is compliant with PDF/A-1a or PDF/A-1b, depending upon which is used, as well as the PDS4 XML label associated with the PDF document. For more information on how to [install](#) and execute the software, see the [online documentation here](#).

Note: Validate tool requires a PDS4 label in order to check compliance. If one is not available for the PDF document, proceed with using other validation tools below.

VeraPDF

VeraPDF is a PDF/A validator that supports every profile level of PDF/A, and has both a usable GUI and command line interface. This is the best bet, if your system runs Java.

For help debugging the validation output reports, see [PDF/A-1 Validation Rules](#).

Apache PDFBox

PDFBox is another strong candidate, if you have a system that supports Java. However, it is a code library, and not a direct tool, so it is mostly useful if you want to include PDF/A validation in an application.

Other Validators

Other PDF/A validators are available, such as the 3-Heights validator, or the Solid Documents validator. However, they are not free so we have not evaluated them. But they do exist.

Online validators

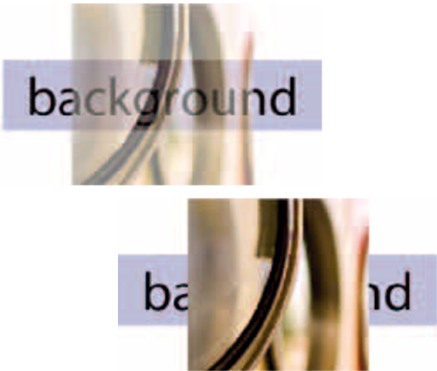
The online validator that PDS has had the most success with is the Solid Documents PDF/A validator, at (<http://www.validatepdfa.com/>). This works by emailing them your document, and

they will email you back a validation report. This takes some time, however, and it can be hard to keep track of multiple versions of the same document, so it is probably only suitable for small batches.

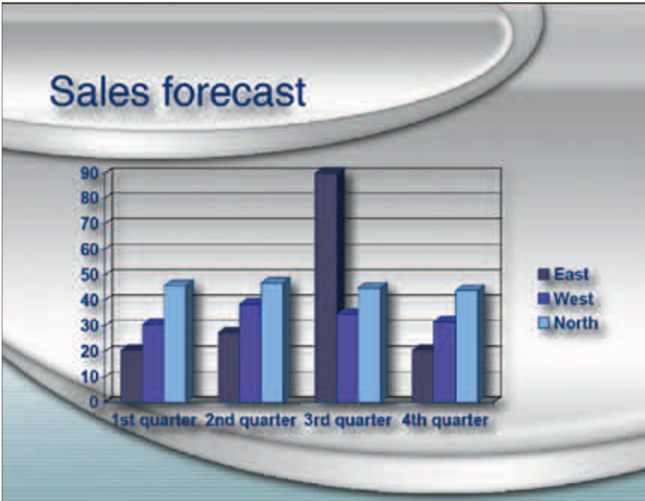
More Details

More on Transparencies

Transparent objects are not allowed in PDF/A-compliant documents. Transparent objects are not 100% opaque – instead, their background shows through, as is the case for glass or thin parchment. Transparent objects cannot always be detected with the naked eye, especially since opacity can be as high as 99% [2]. Adobe Acrobat Professional (Version 6, 7, or 8) includes a flattener module that can be used to remove transparencies. If you continue to receive errors referencing transparencies, please contact your PDS Node representative or the [PDS Help Desk](#) for assistance.



Transparency: The upper image above is transparent. The transparent object can be recognized easily, since the background is visible through the image. The lower image has an opaque foreground image. [2]



Risk of transparent objects: Drop shadows and soft object edges can cause the creation of transparent elements in a PDF file. This example is from a PowerPoint presentation. [2]

References

For more details and information about PDF/A see the following references:

- [1] [PDF/A Family, PDF for Long-term Preservation](#), Library of Congress
- [2] [PDF/A in a Nutshell](#), PDF Association Staff
- [3] [PDF/A FAQ - PDF Association](#), PDF Association Staff
- [4] [PDF/A-1 Validation Rules](#) -