# Processing PDF/A Documents

A thorough understanding of the PDF/A standard is a prerequisite for the creation of PDF/A-compliant documents, and also for displaying them in accordance with the standard. This knowledge alone is however insufficient for optimally configuring PDF/A-related processes. The following article describes how some typical processes can be designed, in order that components that are available on the market today can be used to achieve an optimal cost / benefit relationship.

PDF/A Creation
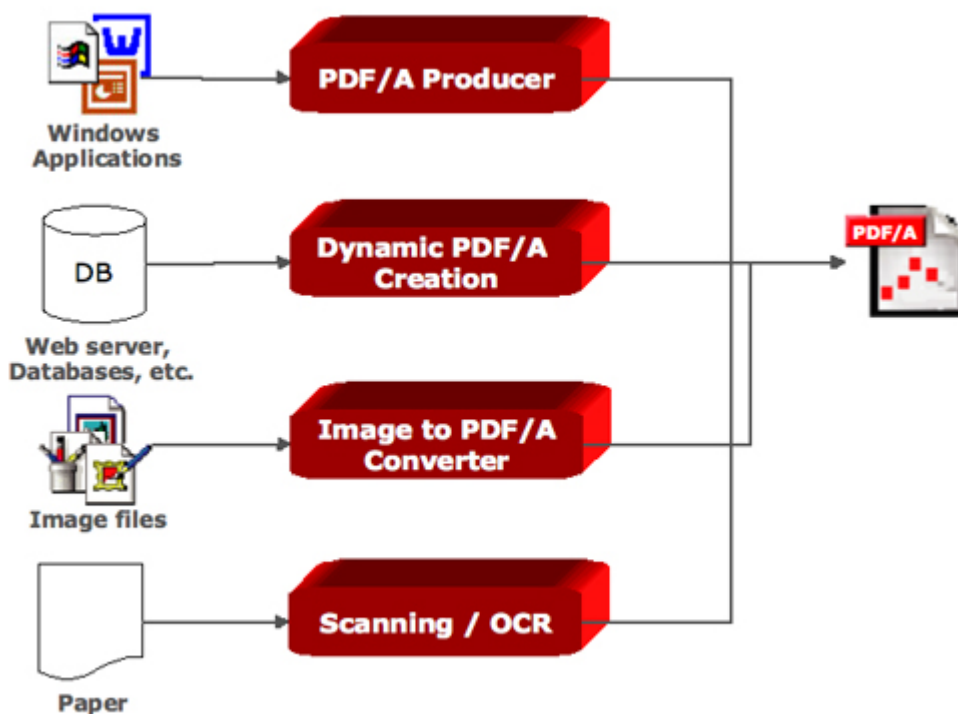
PDF/A Documents can stem from different sources.

The creation software must adhere to the do's and don'ts of PDF/A (including the corrigendum):

 ⊳ Fonts used for the text must be embedded
 ⊳ Colour profiles must be included for image sources (scanned, converted)
 ⊳ Meaningful metadata must be included and embedded as XMP

The most important creation methods:

PDF/A Producer
(e.g. Acrobat Distiller, 3-Heights PDF Producer, LuraDocument PDF Printer).

A PDF Producer (also known as PDF Creator, PDF Converter etc.) is used to create PDF documents from any Windows application with help of a printer function. MS-Office documents are most commonly converted this way. The conversion of E-Mails, including attachments, is more complex. Alternatively, PDFs can be created directly by means of a "Save to PDF" function, as is found for example in Microsoft Office 12 products.

Dynamic PDF/A creation and personalization (e.g. PDFlib)
PDFs are programmatically created directly out of an application (e.g. web server). Simple dynamic content from a database, in addition to static content, can be introduced into the document using this means.

Image to PDF/A Converter
The conversion of image files into PDF/A-conforming files requires in most cases a simple file transformation. The conversion can however become quite complex if more demanding requirements are present (for example colour management).

Scanning / OCR
Scanning, recognition of characters, and conversion into PDF/A-conforming documents is a special field that requires a high level of expertise (mixed raster content, compression methods).

PDF Processing and Conversion
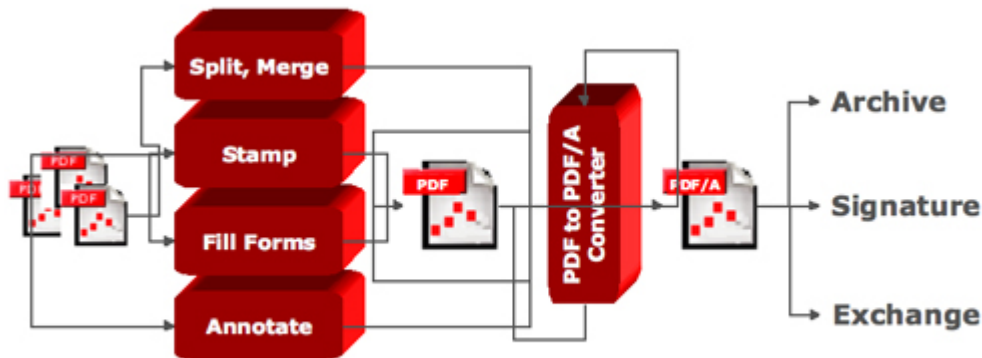
The conversion to PDF/A occurs after all other processing steps are completed, but before the documents are digitally signed.

Possible PDF processing functions:

▹ **Split:** separating a document into individual pages
▹ **Merge:** combining several separate files into one complete document
▹ **Stamp:** applying watermarks, stamping marks, headers and footers, page numbers etc.
▹ **Fill Forms:** programmatically filling in form fields with dynamic content.
▹ **Annotate:** applying interactive elements like comments, bookmarks, text marks

Most PDF processing functions do not guarantee PDF/A compliance for output documents, even when the input documents themselves were PDF/A compliant.

PDF to PDF/A Converter
The conversion to PDF/A serves in particular the following purposes:

▹ Preparation for archiving
▹ Preparation for digital signatures
▹ Preparation for document exchange (internal / external)

The conversion from PDF to PDF/A is not trivial. It entails, amongst others, the following tasks:

- Independency from output devices
- Font embedding
- Creating a static appearance for interactive content
- Eliminating transparency (transparency flattening)
- Removing prohibited content like JavaScript

Signing PDF/A
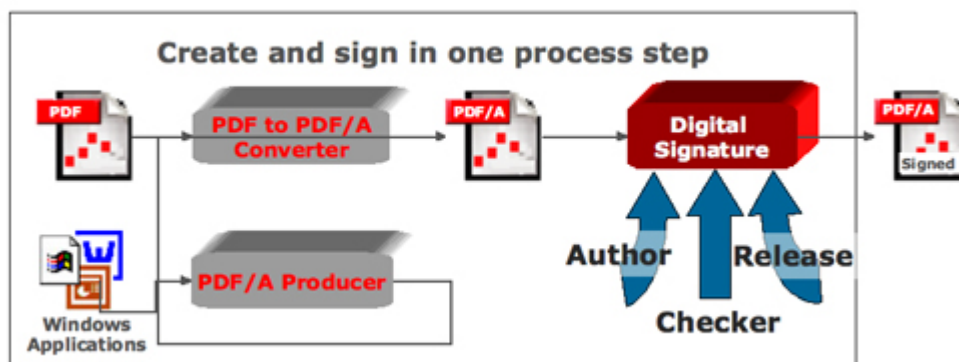
Signing takes place after the conversion to PDF/A.

The application of a digital signature to a PDF/A document is the equivalent of an incremental update to the document. The document has to be PDF/A-compliant before it is signed. The original content of the document remains unchanged and the data structure of the digital signature is added at the end of the file. The digital signature itself must also be PDF/A compliant. Several digital signatures can be applied to the document (e.g. author's signature, checker's signature, release signature).

Change to a document after a digital signature has been applied

All changes made to a PDF/A document after it has been digitally signed must also be incremental and PDF/A compliant. Common changes include redaction (removal, changing, and amendment of text, annotations etc.) as well as updating content. Currently there are no PDF/A compliant modification tools that can work with documents that have been digitally signed.

PDF/A Validation
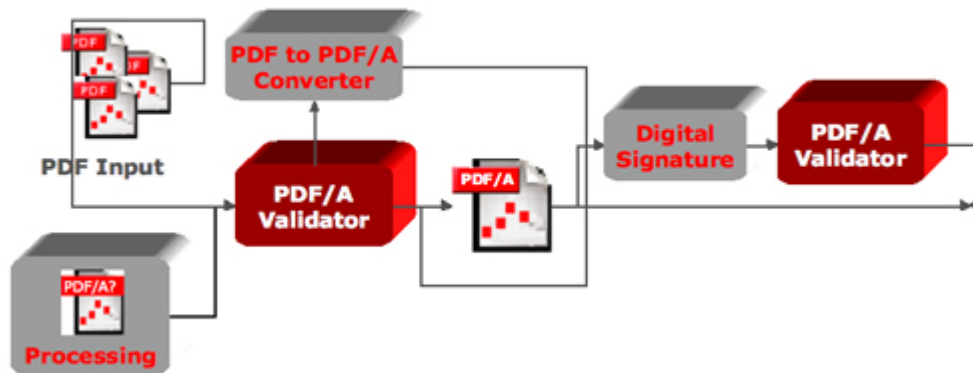
PDF/A compliance must sometimes be confirmed at different stages by means of validation.

Purpose of PDF/A Validation

The purpose of validation is to confirm whether or not a PDF document conforms to the ISO standard for PDF/A.

Areas of application

▹ Input and output verification
▹ Verification before and after specific process steps
▹ Process control (accept / reject)
▹ Generating a "Compliance Report"

Challenge

Validators must be tested and approved by an independent body using a generally accepted test suite.

Validation products:

▹ Acrobat 8 Preflight (developed by callas software)
▹ callas software: pdfInspector
▹ PDF Tools AG: 3-Heights PDF Validator
▹ Luratech: LuraDocument PDF Validator
▹ Seal Systems: PDF Checker
▹ Apago: PDF Appraiser
▹ Intarsys: PDF/A Live!

PDF/A Display
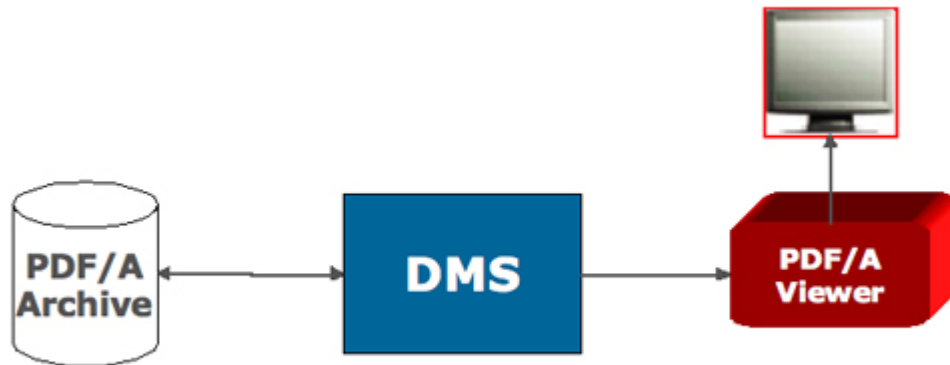
Displaying a PDF/A-compliant document is not the same as displaying a PDF/A document compliantly.

Conventional PDF Viewers

Most viewing programs are not PDF/A compliant, meaning they do not take into account all of the requirements in the ISO standard with respect to displaying PDF/A files. A PDF/A viewing component should offer the following functionality:

▹ Give a warning when a file contains elements that are not PDF/A compliant
▹ Use the embedded fonts, and not the preinstalled fonts of the same name
▹ Use the embedded colour profiles. and not the alternate colour spaces
▹ Provide a consistent display for the appearance of interactive elements contained in the file, and do not create them dynamically
▹ Provide an option for deactivating hyperlinks

Products for displaying PDF/A

▹ Acrobat 8 in PDF/A-Modus
▹ PDF Tools AG: 3-Heights PDF Viewer

PDF/A Mass Archiving
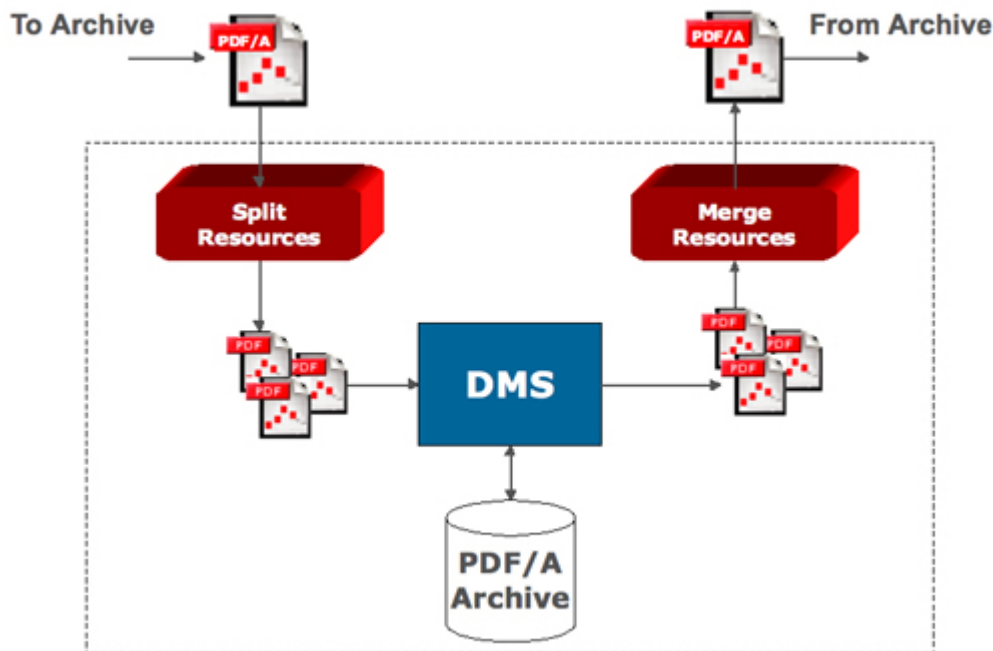
PDF/A archiving with an advanced archive system that can merge commonly used resources

Single, individual PDF/A documents can be archived as they are. When archiving large numbers of similar PDF/A documents (e.g. telecommunications invoices etc.), the situation quite often arises that each document contains the same fonts, logos, or other corporate identification elements, each of which has to be separately saved with every document. The repeated saving of the same common resources (fonts, images) is undesirable and reduces the acceptance of PDF/A.

An enhanced archiving system can resolve this problem by separating the common resources when the PDF/A files are mass archived, and storing just one instance of the resources for all of the files. When a document is retrieved, the common resources are combined with the file again to "re-"create the PDF/A compliant document.

This process can also be used for digitally signed documents, however the files must already be prepared for the separation of the resources before they are signed.

Diagram: Overview of the PDF/A Processes
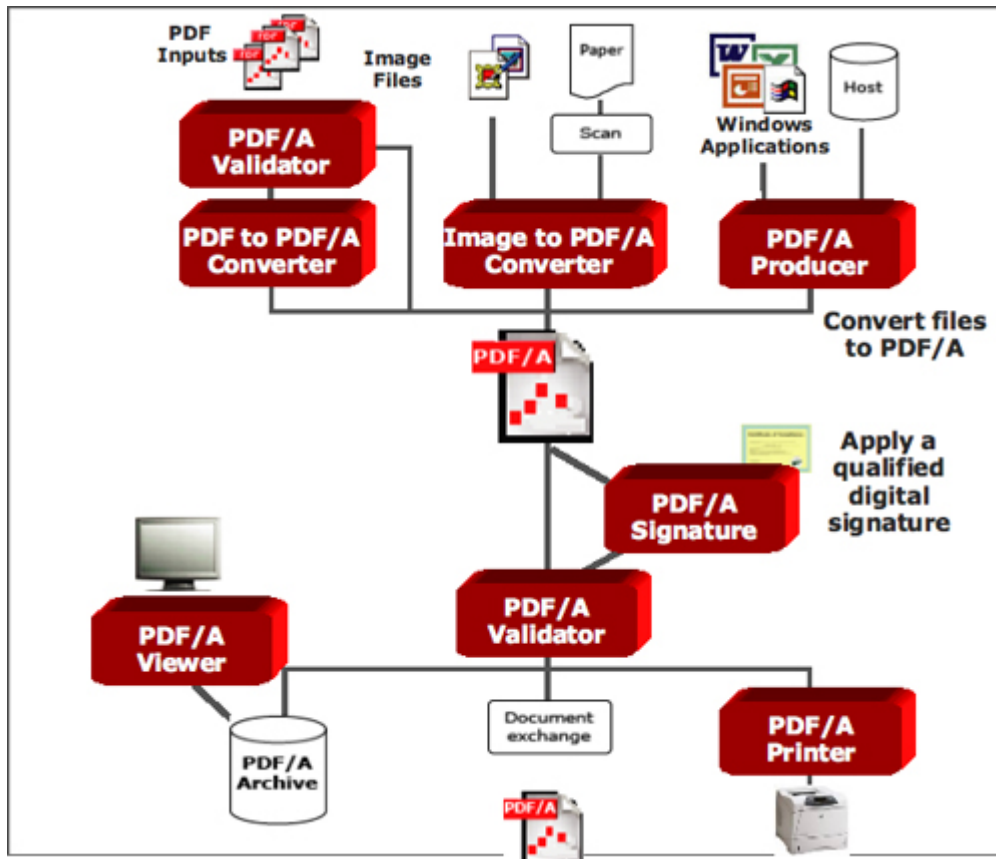
The previously described steps combined into an overall process.

PDF Tools AG

PDF Tools AG was founded as a spin-off from GLANCE (in the market for PDF tools since 1993). The company offers server-sided and developer tools for creating, processing, converting, and enhancing PDF and PDF/A documents. PDF Tools AG is active internationally with customers in over 60 countries. PDF Tools AG represents Switzerland as a voting member of the ISO Working Group 171 (PDF/A).

PDF/A Products:

Numerous solutions and applications available from PDF Tools AG have been developed to support PDF/A, including: PDF Producer, PDF to PDF/A Converter (including digital signatures), Image to PDF Converter, PDF to Image Converter, PDF Validator, PDF Viewer and PDF Printer.

Dr. Hans Bärfuss, PDF Tools AG, aoe
Translation by Roger Reeves, Reeves & Partner GmbH