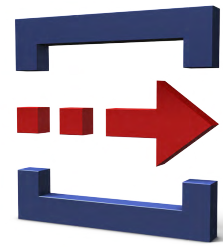


# [ myPDFconvert ]

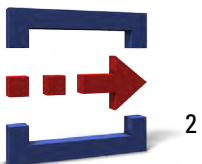


Professional PDF Converter – converting documents of various PDF and graphics formats to PDF/A-compliant documents  
An Overview



## Content

Areas of Application	3
Overview of PDF/A	3
Supported data formats	4
Optional: Compression, OCR	5
High compression	5
GUI, Engine, Plattformen, Integratibility	7
Signing PDF/A documents	7
Additional functions: splitting, merging and stamping documents	7
Reliability and performance	7
Benefit	8



## Areas of Application

The main purpose of myPDFconvert is the conversion or correction of already archived PDF or TIFF documents intended for long-term storage. Display, exact representation and reproducibility of such documents must be guaranteed for a specified period. This period can be laid down by law, e.g. 10 years for invoices in Germany.

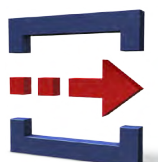
Use in document workflows: myPDFconvert not only converts various PDF documents but even graphics files to PDF/A. It uses PDF as a container for possibly compressed images. Therefore the product is also important for handling incoming mail as it allows immediate PDF/A-compliant conversion of new digital documents and of material scanned inhouse.

Splitting PDF files into pages or individual documents and merging individual documents into large PDF files are additional functions of the basic product. Even stamps or watermarks can automatically be added to a PDF/A-compliant document in order to record processing steps optically. While until now splitting into documents was only possible with PDF input files, the integration of the OCR procedure as of Release 1.1 enables this function even for scanned documents.

### Overview of PDF/A

Documents must meet these requirements among others:

- No references to external resources, no links to external documents
- No active elements that would define the appearance of the document at a specific time, e.g. Javascript, PostScript
- No actions such as starting external programs, no audio or video documents
- Used fonts and all images must be embedded in the file; the same applies to meta-data describing the file origin
- Fonts must only have some standardized encodings; apart from these, only Unicode encodings are permitted.
- Color models must be specified, i.e., device-dependent color spaces such as RGB must include an "output intent"
- No encryptions (also: no disabling of functions such as printing)
- No LZW compression (could cause problems with patent law)
- The language of the file must be specified.



**myPDFconvert: Definition and task**

With the PDF/A standard ISO 19005-1, a secure format standard for long-term archiving of electronically created documents was finally passed in September 2005. A large part of the various documents in existing document archives does not even fulfill the minimum requirements of the PDF/A standard.

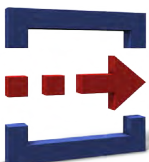
myPDFconvert is a software product for converting documents of various PDF and graphics formats to PDF/A-compliant documents or to files according to other PDF specifications.

When converting existing PDF documents to a PDF/A-compliant state, myPDFconvert checks numerous possibilities of standard violations in the document. Whether violations detected in a document are to be corrected automatically, or whether and in which cases the document is, for example, to be moved to a directory for standard-compliant or faulty documents, is defined by the user in processing profiles via the graphical user interface of the product.

**Supported data formats**

<b>Input formats</b>	<b>Output formats</b>
<b>PDF 1.3 (Acrobat 4)</b>	
<b>PDF 1.7 (Acrobat 8)</b>	<b>PDF/A-1b or</b>
<b>TIFF</b>	<b>PDF 1.3 - PDF 1.7</b>
<b>JPG</b>	
<b>BMP</b>	
<b>PNG</b>	

*Fig. 1) Supported data formats: Input formats for myPDFconvert are various PDF versions and image formats – as output formats, you can set the common PDF versions, especially PDF/A acc. to ISO 19005-1b.*



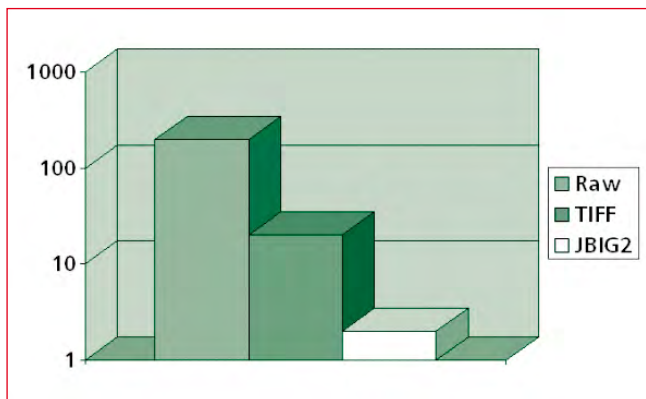


Fig. 2: General proportions: TIFF G4 compresses the scanned image by a factor of about 10. Perceptually lossless\* JBIG2 compression reduces this by a factor of 10 again.

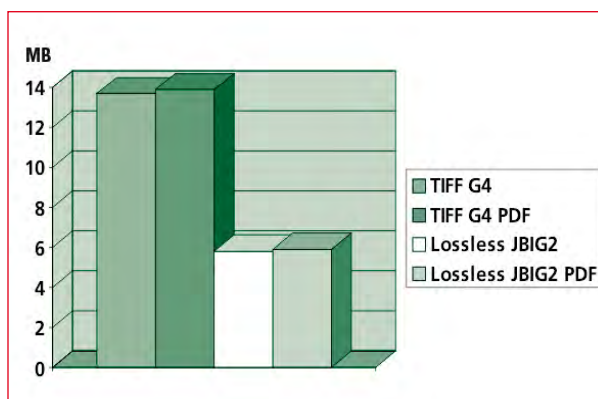


Fig. 3: If lossless compression is required: Lossless JBIG2 image data is less than half as large as the corresponding TIFF image. The additional storage requirement for the PDF container is negligible.

\* Perceptually lossless JBIG2 compression means among other things: no deterioration of OCR, i.e. no character confusion during OCR processing based on the file decompressed from JBIG2 compared to the OCR result based on raw data e.g. originating from a scan.

### Optional: Compression, OCR

Via its GUI, myPDFconvert also offers processing options for high-performing compression of documents in several PDF and graphics formats as well as an integrated OCR processing (Fig. 4).

#### High compression

New ISO-standardized compression methods such as JBIG2 (ISO/IEC 14492) and JPEG 2000 (ISO/IEC 15444) considerably reduce the storage requirement for compressed documents compared to traditional methods such as TIFF G3 and G4. Both new algorithms, JBIG2 and JPEG 2000, allow lossless as well as lossy compression (see Fig. 1 and 2) and can therefore be customized regarding quality and file size.

As of PDF Reference Version 1.4, 3rd Edition, 2001, PDF also includes a JBIG2 decoding filter. With the corresponding myPDFconvert setting, PDF documents containing JBIG2-compressed images therefore retain their PDF/A-compliance.

#### Results

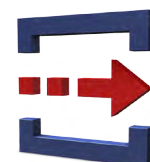
Normally, PDF/A-compliant (acc. to ISO 19005-1b) PDF files created with lossless JBIG2 compression are finally only half as large as the originally existing or scanned TIFF files (cf. Fig. 3). These documents allow full-text research by archiving systems and other auto-

matic processing, and users can immediately convert their incoming mail in compliance with the standard during scanning – and even sign it if required.

The functions used in myPDFconvert normally compress black-and-white scans by a factor 5 to 10 and color scans by a factor 10 to 100.

#### OCR (Optical Character Recognition)

If bitmap documents (e.g. TIFF, BMP) are converted to PDF documents using myPDFconvert, even an OCR processing can be performed besides compression. The text data retrieved by OCR can be stored in the created PDF as an invisible text layer, which lies above the decompressed image when viewing the document in the Acrobat viewer. In this way, users receive PDF/A-compliant, parsable PDF files as a processing result – created on the basis of raster data or of PDFs with embedded image documents.



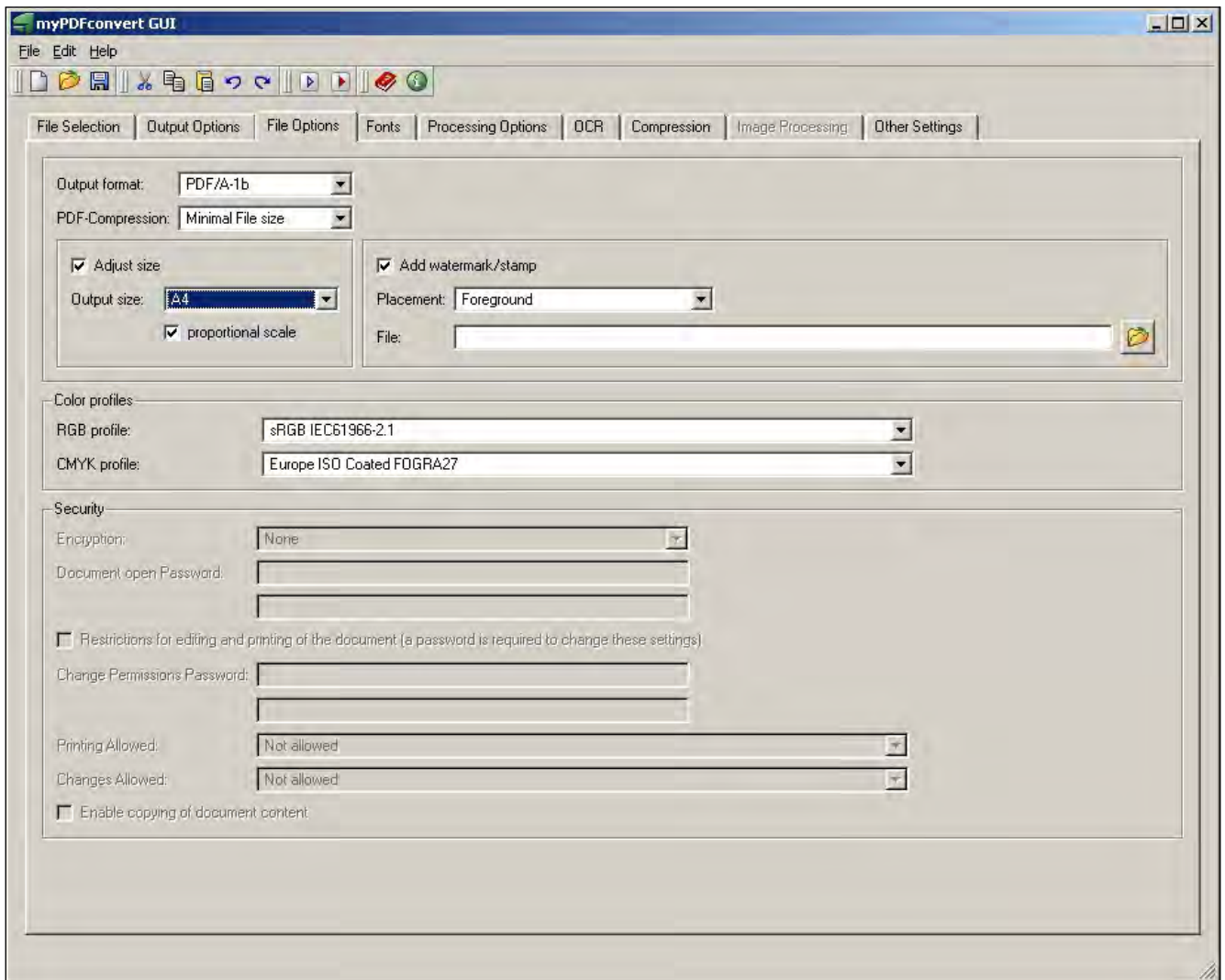
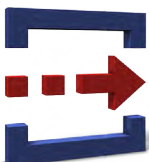


Fig. 4:  
The uniform user interface supports the user in selecting and combining all functions of a processing profile, also for compressing documents in various PDF and graphics formats and for integrated OCR processing.

### OCR: automated

Even the OCR engine in myPDFconvert is designed for industrial-strength scanning and OCR in business applications. myPDFconvert automates the OCR process by watched folders, so that users need not monitor document processing. In the corresponding mode, files undergo an OCR analysis if they are just drawn into a folder. To manage large quantities of scans, OCR batch processing runs with a high speed of approx. 1 scanned page per second.



### GUI, Engine, Plattformen, Integrability

myPDFconvert essentially consists of two elements. The first one is the graphical user interface (GUI) intended to configure procedures and processes.

The second element is the engine for the actual execution of specified tasks, which runs in the background and can be called directly from the GUI, from the command line (CLI), or from an application programming interface (API). If myPDFconvert is started as a service, it continuously checks a specified directory for documents to be processed and writes the results to a specified output directory.

### Signing PDF/A documents

myPDFconvert includes the signing of PDF documents in a configurable workflow to enable their signed storage. In this way, it can e.g. be guaranteed that TIFF documents scanned inhouse and then converted to PDF/A have not been modified since they were signed. An unnoticed modification of the PDF/A document is then excluded.

### Additional functions:

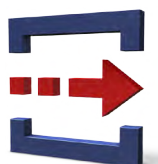
#### splitting, merging and stamping documents

myPDFconvert offers additional functions such as splitting PDF files into pages or individual documents and merging individual documents into large PDF files. Even stamps or watermarks can automatically be added to a PDF/A-compliant document in order to record processing steps optically. Sample application: Com-

plete multipage PDF files are split into individual pages, marked for enveloping and remerged into one file.

### Reliability and performance

myPDFconvert has been developed for professional use. It has been tested with thousands of various documents from various sources. The product is intended, inter alia, for converting the documents of complete archives, for which reason its development was geared to high document throughput.



## Benefit

Up to now, documents have been archived in various PDF or graphics formats – there was no standard.

- myPDFconvert offers the chance of correction: Documents archived in PDF, TIFF and other graphics formats can now automatically be converted to the secure PDF/A format and archived – optionally even including a digital signature of the PDF/A-compliant document.
- Used in the processes of incoming mail, myPDFconvert additionally guarantees that documents scanned inhouse or digitally received as PDF or graphics files will comply with the ISO standard and can therefore be archived.
- Cost reduction, because improved compression quality compared to traditional image compression methods considerably reduces the data transfer load (Intranet/Internet).

- Additional cost reduction due to correspondingly minimized storage requirements, e.g. in document archives.
- Improved document handling by standardized format (PDF) with high additional functionality compared to native image formats, e.g. performing a full-text search in the previously “pure” image or finding a specific page of a PDF document in the HTML browser.
- Revision security by compliance with the standards PDF/A (ISO/IEC-19005), JBIG2 (ISO/IEC-Standard 14492) and JPEG 2000 (ISO/IEC 15444).

As a company with main focus on output management and with our extensive experience as software manufacturer and professional service provider, we are a competent and reliable partner to many companies in resolving the above problem.