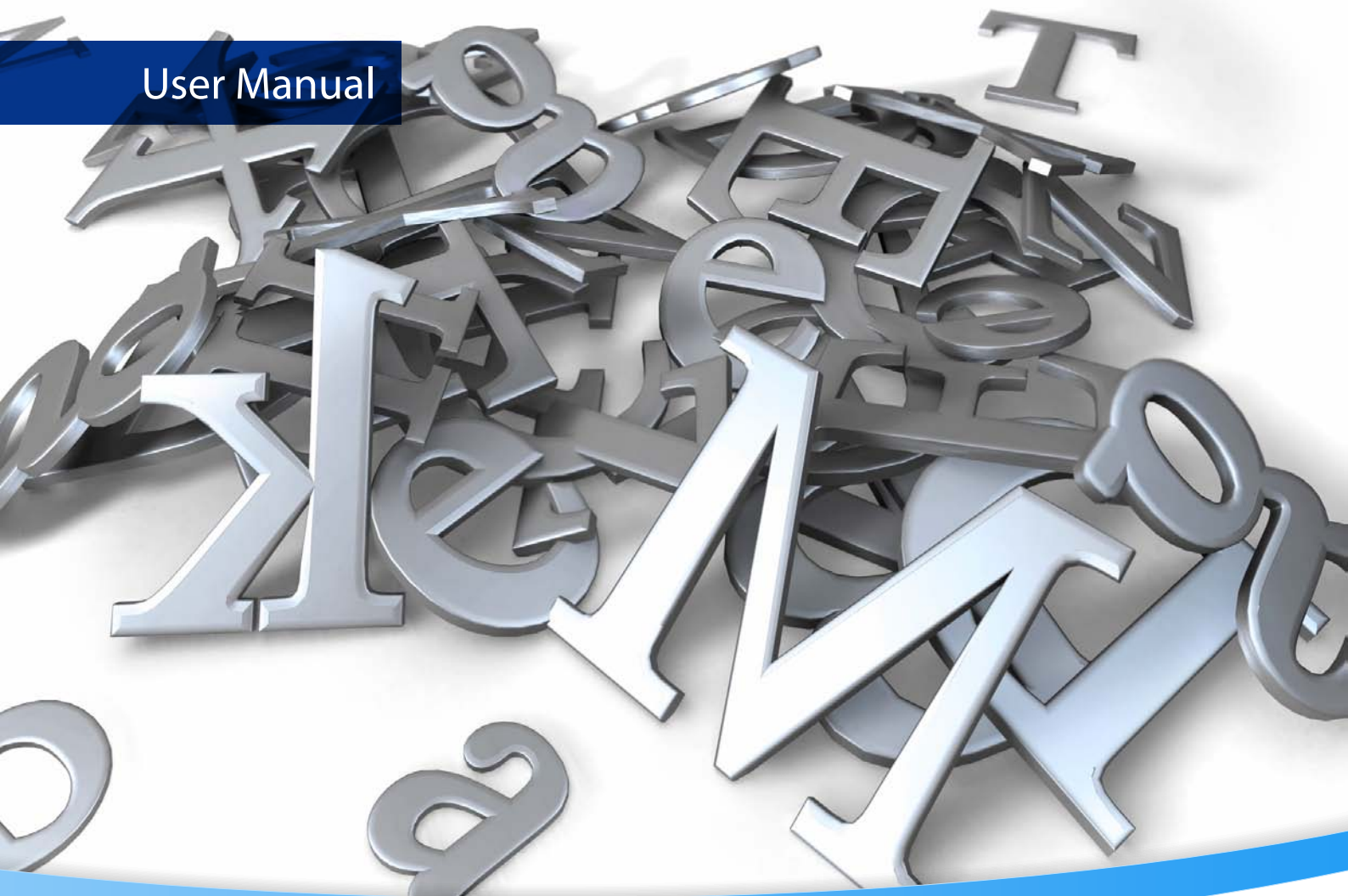


User Manual



Conversion Service Docker

Version 4.0.0



Contents

1	Introduction	3
1.1	Description	3
1.2	How the Conversion Service Docker works	3
1.3	Processing documents	4
1.4	Integration options	4
1.5	Features	5
1.6	Operating systems	5
2	Quick start guide	6
2.1	Running a container	6
2.2	Convert your first file using the preinstalled configuration	6
2.3	Configure for your scenario	7
2.4	Integrate into your system	7
3	Workflows	8
3.1	Archive PDF/A-2	8
3.1.1	Supported file formats	8
3.1.2	Process	8
3.1.3	Configuration	9
3.2	Archive PDF/A-3	11
3.2.1	Supported file formats	11
3.2.2	Features	11
3.3	Archive PDF/A-1	11
3.3.1	Supported file formats	12
3.3.2	Features and limitations	12
3.4	Conversion	12
3.4.1	Supported file formats	12
3.4.2	Features	13
3.5	Invoice	13
3.5.1	Supported file formats	13
3.5.2	Features	13
3.6	Dossier	14
3.6.1	Supported file formats	14
3.6.2	Features	14
	Title page	14
	Structuring the dossier	15
	Document outline	15
	Table of contents	16
	Stamps	17
3.7	Processing steps	17
3.7.1	Stamping	17
	Text stamp	17
	Image stamp	18
	Barcode stamp	18
	Placeholders	18
3.7.2	Metadata	18
	Standard properties	18
	Extended properties	18
	Dublin Core schema	19

	XMP Basic schema	19
	XMP Rights Management schema	19
	Custom extension schemas	20
	Advanced usage	20
3.8	Supported file formats	21
4	User guide	22
4.1	Service configuration	22
4.1.1	Service host address	22
	HTTPS	22
4.1.2	License key	22
4.1.3	Converting Office documents	23
	Office for the web service	23
	Windows Service instance	24
4.1.4	Proxy	24
4.1.5	Cross-origin requests (CORS)	24
4.1.6	Advanced service configuration	24
	Use forwarded HTTP headers from WAF	24
	Load balancer	24
	Load balancer configuration	25
	Requirements on clients	25
4.2	Profile configuration	25
4.2.1	Create and export a profile configuration	25
	Windows installation	25
	Create a profile configuration	26
	Export the profile configuration	26
4.2.2	Set a profile configuration when starting the Docker container	27
	Import profiles only	27
	Import profiles and further configuration files	28
4.3	Shell client pdfclient	28
4.4	REST API	29
4.4.1	API usage	30
4.5	Plugins	32
5	Troubleshooting	33
5.1	Error codes	33
5.2	Service log	34
5.2.1	Log file	34
5.2.2	Custom logs	35
5.3	Submitting a support request	35
6	Version history	36
6.1	Version 4	36
6.2	Version 3	36
6.3	Version 2	39
6.4	Version 1	41
7	Licensing, copyright, and contact	43

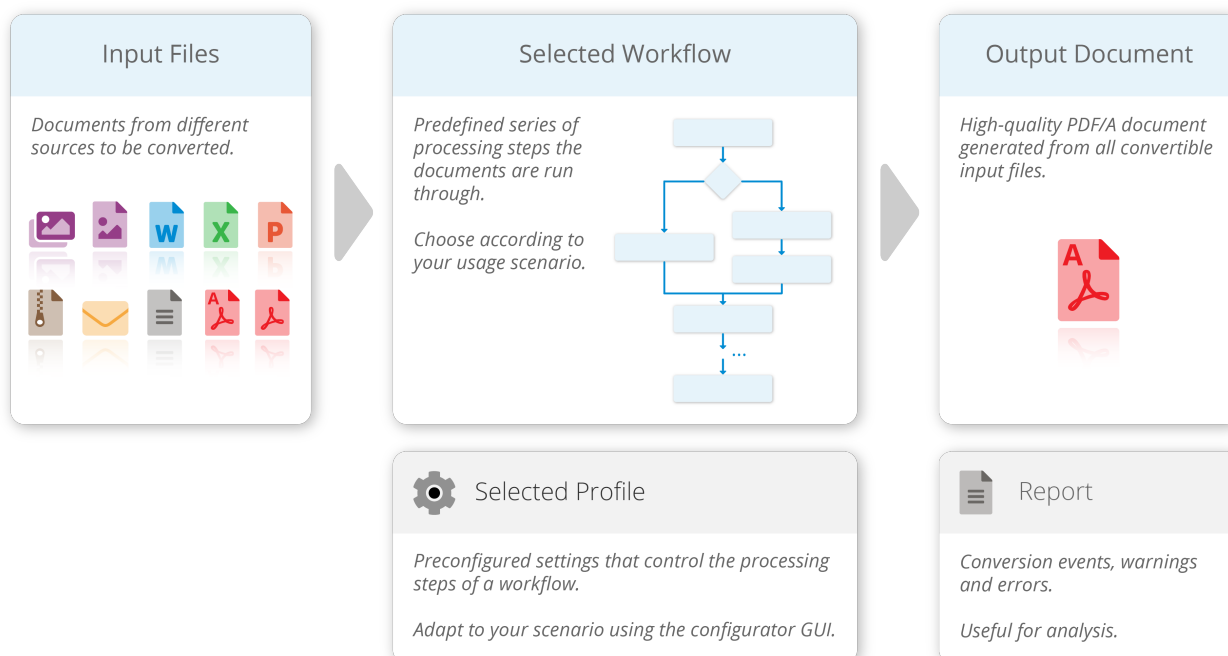
1 Introduction

1.1 Description



The Conversion Service Docker is an all-round carefree package for automating your document processes. It takes input documents of different formats from various sources and processes them according to your use case to produce high-quality PDF/A documents.

1.2 How the Conversion Service Docker works



The Conversion Service Docker can be thought of as a factory with a production line (**workflow**) for each product it offers. The products it manufactures are documents prepared for a specific use case. For example, the workflow [Archive PDF/A-2](#) is engineered specifically for preparing documents for archiving.

A production line consists of a series of **processing steps** (e.g. validate, OCR, convert, merge, sign, ...). Each step moves the raw materials (**input documents**) closer to being a finished product (**output document**).

Some steps are optional and most of them have options (e.g. paper size, image resolution, ...). These can be adjusted to your individual needs by configuring one or more setups (**profiles**) for each production line. A profile defines which processing steps to activate and what options to use.

1.3 Processing documents

If you want the factory to manufacture a product (**output document**), you send them an order (**job**) consisting of the raw materials (**input documents**), the production line (**workflow**), and setup (**profile**) to use and where to deliver the finished product (**output path**).

Input

Similarly, to prepare documents for a specific use case, you send a **job** consisting of the following input to the Conversion Service Docker:

Input documents The documents to be processed.

Workflow name Which workflow to use, i.e. the use case to prepare the documents for.

Profile name Which profile to use, i.e. how the documents should be processed.

Output path Where to store the resulting document.

Output

Once conversion is completed, the Conversion Service Docker returns:

Output document If the conversion was successful, the resulting document can be found in the specified output directory.

Report Conversion events, warnings, and errors. Useful for analysis.

1.4 Integration options

The communication with the factory can happen through different channels. You can send the order by mail, or you bring everything to the factory in person. Similarly, the Conversion Service Docker offers different **interfaces** to communicate through.

The Conversion Service Docker is a service designed to be integrated into your system using one of the following interfaces:

Shell client Easy to use shell tool. Recommended for integrating using scripts. Suited for automated and manually triggered processing.

REST API Easy to use webservice interface. Recommended for integrating into your existing application (Enterprise Application Integration). Suited for automated and manually triggered processing from within your application.

1.5 Features

- Robust client/server architecture
- Powerful REST API (HTTP or HTTPS)
- High throughput job processing
- Specialized workflows
 - Archive PDF/A-2
 - Archive PDF/A-3
 - Archive PDF/A-1
 - Conversion
 - Invoice
 - Dossier
- Multiple configuration profiles for each workflow
- High availability and stability
- Logging and monitoring

1.6 Operating systems

The Conversion Service Docker is available for the following operating systems:

- Linux:
 - Red Hat, CentOS, Oracle Linux 8+ | x64
 - Fedora 29+ | x64
 - Debian 10+ | x64
 - Other: Linux kernel 2.6+, GCC toolset 4.8+, glibc 2.27+ | x64

2 Quick start guide

The Conversion Service Docker is a complete image containing everything required to run.

2.1 Running a container

The pdftoolsag/conversion-service image is available from the [public repository on Docker Hub](#).

1. Ensure the container can communicate with any services configured in its profiles, e.g. a Conversion Service instance running on Windows, a 3-Heights® OCR Service or a cloud-based digital signature service. The examples in this manual use a simple bridge network called conversion-service, which can be created using:

```
docker network create conversion-service
```

2. Run the container using the preinstalled configuration.

```
docker run -dp 13033:13033 --network conversion-service \
-e LICENSEKEY=4H-V4-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX \
pdftoolsag/conversion-service
```

The container can be configured using environment variables. The only required configuration is the license key, which can be passed using either LICENSEKEY or LICENSEKEY_FILE.

Example: Typical container configuration with custom profile configuration.

```
docker run -dp 13033:13033 --network conversion-service \
--mount "type=bind,src=C:\path\to\ProfileExport-4.0.0.export,\
dst=/etc/convsrv/ProfileExport.export,readonly" \
-e IMPORT_PROFILES=/etc/convsrv/ProfileExport.export \
-e LICENSEKEY=4H-V4-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX \
-e WINDOWS_SERVICE_ENDPOINT=http://server:13033/conversion/v1.0/rest \
pdftoolsag/conversion-service:4.0.0
```

2.2 Convert your first file using the preinstalled configuration

The shell client pdfclient from the image is suitable to test your configuration:

```
docker run --rm --network conversion-service \
--mount "type=bind,src=c:\path\to\input,dst=/input" \
pdftoolsag/conversion-service \
bash -c "bin/pdfclient -v -url http://server:13033/conversion/v1.0/rest \
/input/input.pdf /tmp"
```

If the service uses HTTPS, its host certificate must be trusted by the Docker container running pdfclient. Otherwise no connection can be established. By default, no trusted certificates are installed.

2.3 Configure for your scenario

- Configure your service as described in [Service configuration](#).
- Adapt the default profile or create new ones as described in [Profile configuration](#).

2.4 Integrate into your system

Choose your integration:

- [Shell client pdfclient](#)
- [REST API](#)

3 Workflows

3.1 Archive PDF/A-2

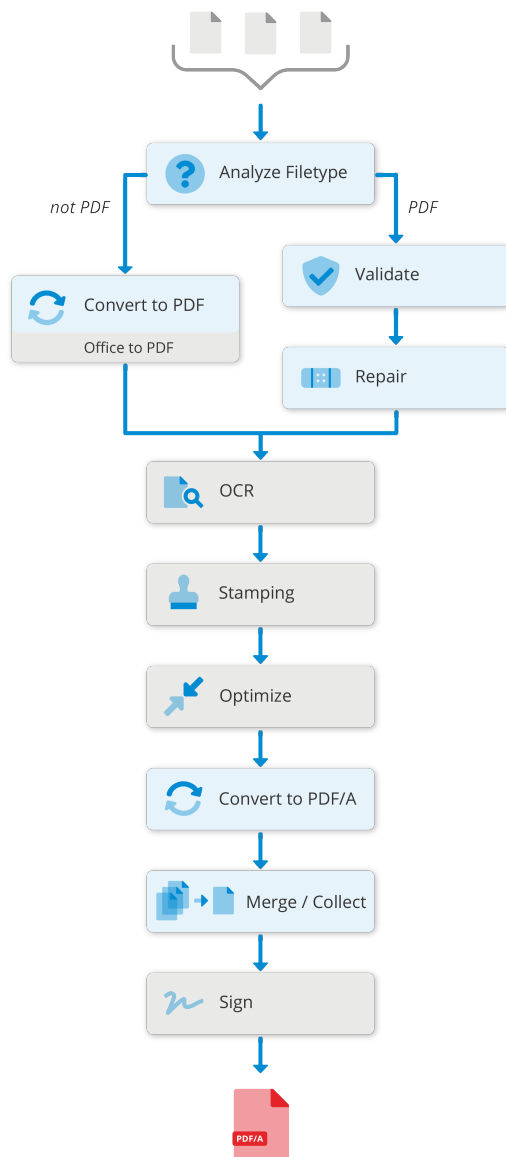
This workflow is engineered specifically for preparing documents for archiving.

3.1.1 Supported file formats

This workflow supports all file formats listed in [Supported file formats](#). The conversion of most file formats is enabled by default in the [Convert mode configuration for child documents \(Attachments\)](#).

3.1.2 Process

All input documents of a job are processed as follows:



Analyze file type In a first step, the documents are analyzed. If the file type of any document is not supported, conversion is aborted. Otherwise, they are sent to the next processing step depending on their file type.

Validate & repair (Quality Assurance) To ensure document quality, PDF and PDF/A documents are validated. If a corruption is detected, the Conversion Service Docker attempts to repair it.

Convert to PDF Non-PDF documents (e.g images, Office documents, etc.) are converted to PDF if their format is supported by the Conversion Service Docker.

Note: *the conversion of Office documents requires an additional step which can only be enabled with an appropriate license.*

OCR To make the resulting document searchable, OCR is performed on documents that require it. The recognized text is stored directly in the PDF.

Note: *this is an optional step and can only be enabled with an appropriate license.*

Stamping See [Stamping](#).

Convert to PDF/A PDF documents that are not already PDF/A-2 conforming are converted to a high-quality PDF/A-2.

Merge / Collect The converted documents of a job are merged or collected into one document, depending on the profile setting.

Optimize The resulting document is optimized for archiving. This includes several optimizations: redundant and unnecessary data for archiving is removed, images are compressed intelligently, and fonts are merged and subset.

Note: *this is an optional step and can only be enabled with an appropriate license.*

Sign The resulting document is digitally signed using the signature settings in the selected profile.

Note: *this is an optional step and can only be enabled with an appropriate license.*

3.1.3 Configuration

The workflow's profile offers a fine-grained configuration of how files are converted. All of the processing steps described can be enabled and configured in the profile configuration. Furthermore, the following configuration options are available:

Convert mode configuration for child documents (Attachments) The convert mode defines the documents to be converted (configuration "Convert") and those to be skipped (removed) from the result. When removing documents, a (configuration "Skip with warning") warning or a (configuration "Skip") informational message is generated.

The convert mode can be specified based on the type of the child document, its file name, or the type of the parent document.

For example, by default Office files are converted to PDF/A, executables are removed, and other non-convertible documents are removed with a warning.

Collect mode configuration The collect mode configuration defines how a converted document and child documents are combined.

The collect mode can be specified for each document type individually. See the documentation panel of the Conversion Service Docker Configurator for a detailed description of the available collect modes.

For example, emails can be converted by creating a PDF collection (Portfolio) of its body and attachments. When converting Word documents, their embedded files can be attached to the converted PDF.

Merge vs. attach There are two categories of collect modes: either the pages of multiple PDF documents can be merged into a single document, or child documents can be attached (embedded) into a PDF document. A PDF collection (Portfolio) is a special case of embedding child documents, where the parent document contains no pages, but shows a convenient table of the attached documents for easy navigation.

The advantage of the **Merge** collect modes is that it creates simple files that can be processed and viewed by all PDF applications. The disadvantage is that only PDF files can be merged. Furthermore, not all the information can be preserved when merging PDF files. For example, document metadata, signatures, and certain interactive form fields cannot be merged and must be removed. Also, logical structure (tagging) information may be less meaningful after merging.

Recommended collect mode configuration for the “Merge” use case:

Type	Collect mode
Job	Merge
Word	Merge or attach
Excel	Merge or attach
PowerPoint	Merge or attach
Email	Merge or attach
Archive (ZIP)	Merge or attach
PDF	Preserve structure

The advantages of the **Attach** collect modes is that all the information of the input files and the structure of the files can be preserved. PDF collections (Portfolios) provide a convenient way to show and navigate the attached documents.

Recommended collect mode configuration for the “Attach” use case:

Type	Collect mode
Job	Collection or single document
Word	Attach
Excel	Attach
PowerPoint	Attach
Email	Attach
Archive (ZIP)	Collection or single document
PDF	Preserve structure ¹

¹ If necessary, the collect mode “Flatten” can be used to flatten the structure of PDF documents. Up to version 3.1 of the Conversion Service Docker, this has been the default behavior.

The **Child error handling** configuration defines how errors during the conversion of child documents are handled. In case of an error, the child document can either be skipped (removed) from the result and a warning generated ("Skip with warning" configuration). Alternatively, the conversion of the parent document can be aborted with an error ("Strict" configuration).

3.2 Archive PDF/A-3

This workflow is engineered specifically for preparing documents for archiving. All features and processing steps of the workflow [Archive PDF/A-2](#) are also available in [Archive PDF/A-3](#). However, the PDF/A-3 format allows [additional features regarding embedded files](#).

3.2.1 Supported file formats

This workflow supports all file formats listed in [Supported file formats](#). The conversion of most file formats is enabled by default in the [Convert mode configuration for child documents \(Attachments\)](#).

3.2.2 Features

In addition to the features of the [Archive PDF/A-2](#) workflow, this workflow also supports the following features:

Convert to PDF/A-3 The workflow's profile offers a fine-grained configuration of how files are converted.

Convert mode configuration for child documents (Attachments) This extends the [Convert mode configuration for child documents \(Attachments\)](#) of the [Archive PDF/A-2](#) workflow with PDF/A-3 features. More specifically, the PDF/A-3 standard allows you to embed child documents "As is", i.e. without converting them to PDF/A.

For example, by default Office files are converted to PDF/A-3, images are used as-is, and executables are removed.

Collect mode configuration This is the same as the [Collect mode configuration](#) of the [Archive PDF/A-2](#) workflow. However, in the [Archive PDF/A-3](#) workflow, it is common to choose the "As is" convert mode for some child document types. Therefore, using "Merge" collect modes is strongly discouraged, because they are limited to PDF documents (see [Merge vs. attach](#)). Instead, the "Collection or single document" and "Attach" collect modes are recommended. Please see the documentation panel of the Conversion Service Docker Configurator for a detailed description of the available collect modes.

Attach source document The source document (original document) can be attached. The configuration allows you to decide for each file type if the source document should be attached or not. By default, the source documents for Office files are attached. Note that this may increase the file size of the result substantially.

Attach conversion report All events of a conversion can be written to a report file and attached to the result document.

3.3 Archive PDF/A-1

This workflow is engineered specifically for preparing documents for archiving. More specifically, where the conversions PDF/A-2 or PDF/A-3 are not allowed. The features and processing steps of this workflow are similar to the [Archive PDF/A-2](#) workflow.

This workflow is disabled by default because it is generally recommended you use [Archive PDF/A-2](#) for archiving. Since PDF/A-2 is based on a newer version of the PDF standard, more PDF features are allowed. These features are

commonly used and include transparency, layers, embedded files, and a less restrictive internal file structure. For these reasons, you can achieve fewer conversion errors and a better conversion quality, especially in the Docker image where [document content rasterization](#) is not available. For more information on PDF/A-2, see <https://www.pdf-tools.com/blog/pdf-tools-knowledge/pdfa-2-overview/>.

3.3.1 Supported file formats

This workflow supports all file formats listed in [Supported file formats](#). The conversion of most file formats is enabled by default in the profile's "Convert mode configuration for child documents (Attachments)".

3.3.2 Features and limitations

Compared to the [Archive PDF/A-2](#) workflow, this workflow has the following features and limitations:

Convert to PDF/A-1

Collect mode configuration The PDF/A-1 standard does not allow embedded files. Therefore, the only collect mode configuration available is merge (see [Merge vs. attach](#)).

Document content rasterization Certain graphical features of PDF documents are not allowed in PDF/A-1. For example, transparency. The removal of these features can lead to severe visual differences that may render the page's content unreadable.

The product can preserve the visual appearance of such documents by converting the page content to images (content rasterization). When doing so, a "Content rasterized" warning is generated. If possible, content rasterization preserves the page's extractable text, links, outlines, and viewer preferences.

The Docker version cannot rasterize pages. Therefore, the conversion of such files leads to "Transparency removed" or "Visual differences" warnings, depending on whether the visual differences are caused by the removal of transparency.

PDF/A-1 Specific conversion warnings The following warnings occur more frequently when converting to PDF/A-1:

"Annotation removed", "Content rasterized"², "Layers removed", "Transparency removed"³, "Visual differences". The profile's conversion settings for these warnings is particularly important.

Sign Document signing is currently not implemented.

3.4 Conversion

This workflow is engineered specifically for the conversion of documents to PDF. In contrast to the [Archive PDF/A-2](#) and [Archive PDF/A-3](#) workflows, files are only converted to PDF (and not PDF/A). The file format is not validated and the output documents cannot be signed. OCR is available as an optional processing step.

3.4.1 Supported file formats

This workflow supports all file formats listed in [Supported file formats](#).

² Only in Windows, where document content rasterization is available.

³ Only in Docker image, where document content rasterization is not available.

3.4.2 Features

The Conversion workflow offers the following additional features compared to the [Archive PDF/A-2](#) and [Archive PDF/A-3](#) workflows:

Optimize for speed or size The workflow's profile offers an option to optimize for processing time (speed) or for minimal document file size.

Convert mode configuration for child documents (Attachments) Certain child documents can be skipped (removed) during conversion to PDF, such as attachments of emails or PDF documents. The convert mode can be specified based on the type of the child document, its file name, or the type of the parent document.

For example, by default executables attached to an email are removed. If desired, rules can be added to attach files that can't be converted (e.g. PDF documents containing unrendered XFA, HTML documents) in their original source format to the resulting output document.

Collect mode configuration The collect mode configuration defines how a converted document and the child documents are combined. The collect mode can be configured for each document type and also defines how errors are handled.

For example, emails can be converted by creating a PDF collection (Portfolio) of its body and attachments. Or when converting Word documents, all embedded files can be merged to the converted PDF.

3.5 Invoice

This workflow is engineered specifically for preparing invoices. This workflow converts a main document to PDF/A-3 and attaches one or more files to the result. For example, an Excel table containing additional data can be attached to a PDF invoice.

3.5.1 Supported file formats

This workflow only supports PDF and MS Word as the main document input format. (See [Supported file formats](#))

3.5.2 Features

Convert to PDF/A-3 This workflow has been optimized for the conversion of transactional documents. Supported are conversions of Word and PDF documents to PDF/A-3.

Attach documents To attach additional documents to a main document, all must be added to a single job. The documents to attach and additional properties can be set using the following document options. These values are written into the result document and are visible to a user in the list of attached documents.

DOC.ROLE (required) Defines the document's role. Supported values are:

- **main** (default): The main document. In each job, there must be one main document.
- **attached**: A document that is attached to the main document. There may be multiple attached documents in a job.

AF.RELATIONSHIP (optional) Defines the relationship of the attached document to the main document. Supported values are:

- **Unspecified** (default): Used when the relationship is not known or cannot be described using one of the other values.
- **Source**: Used if the attached document is the original source material for the main document.

- **Data:** Used if the attached document represents information used to derive a visual presentation. For example, a table or a graph.
- **Alternative:** Used if the attached document is an alternative representation of content. For example, audio.
- **Supplement:** Used if the attached document represents a supplemental representation of the original source or data that may be more easily consumable (e.g., A MathML version of an equation).

AF.MODDATE (optional) Defines the date and time when the attached document was last modified.

Example: "2009-05-19T09:15:22.0000000+02:00"

AF.DESCRPTION (optional) Descriptive text associated with the attached document.

Example:

Example using the [Shell client pdfclient](#) to attach `supplement.xls` to the main document `invoice.docx`. Note that the `-do <name> <value>` option applies a document option to the subsequent input file.

```
C:\Temp> pdfclient -v -w Invoice invoice.docx ^
-do DOC.ROLE attached -do AF.RELATIONSHIP Supplement supplement.xls ^
out.pdf
Creating job (id job123_4q5mprmmrz0)
Adding file "invoice.docx" (id 0kk1ddv0uzg)
Adding file "supplement.xls" (id 2fg4ufg3qmz)
Output: "invoice.pdf"
- Info: Converted Word document 'invoice.docx' to PDF.
- Info: Converted 'invoice.pdf' to PDF/A-3u.
- Info: Attached file 'supplement.xls' to 'invoice.pdf'.
```

3.6 Dossier

This workflow is specifically designed to compile multiple PDF documents into a single dossier.

All PDF document supplied to the job are merged into a single PDF in the order that they were added to the job.

The workflow supports the following optional features:

- Title page with title, subtitle, and a custom background.
- Extensively customizable table of contents.
- Stamps to create e.g. watermarks, header, or footer lines.
- Bookmarks to structure the dossier.

3.6.1 Supported file formats

This workflow only supports PDF files. See [Supported file formats](#)

3.6.2 Features

Title page

This workflow can be configured to create a title page for the dossier.

The title page contains the title and subtitle of the dossier. The font, size, placement, and alignment of the text is fully configurable.

Additionally, a PDF document can be configured to be used as a custom background of the title page.

The title page can be customized for each job using the following options:

DOSSIER-TITLE (optional) The title of the dossier used on the title page.

This option overrides the value specified in the profile configuration. The option should only be set at job level and is ignored at document level.

DOSSIER-SUBTITLE (optional) The subtitle of the dossier used on the title page.

This option overrides the value specified in the profile configuration. The option should only be set at job level and is ignored at document level.

Structuring the dossier

The dossier can be structured by creating a table of contents and a document outline (bookmarks).

By default, the title of each document is taken from its metadata. If the title from the metadata is not suitable, it can be overridden by using the following option:

DOCUMENT-TITLE (optional) The title of the document used for the table of contents, document outline, and stamp placeholder.

This option can be overridden by the more specific options **DOCUMENT-OUTLINE-TITLE**, **DOCUMENT-TOC-TITLE**, and **DOCUMENT-STAMP-TITLE**.

If this option is omitted and none of the more specific options is set, one of the following values is used instead (in order of precedence):

- The document title from the document metadata.
- The filename of the input document (without the extension).

This option should only be set at document level and cannot be configured in the profile configuration.

Document outline

The document outline, also called bookmarks, provides a way to quickly jump to a specific part of the dossier.

The Dossier workflow uses the existing outline tree of the input documents. Optionally, a parent bookmark can be created for each individual document.

The document outline can be customized for each job using the following options:

TOC-OUTLINE-TITLE (optional) The name of bookmark created for the table of contents (if enabled).

If this option is omitted, one of the following values is used instead (in order of precedence):

- The value of the **TOC-TITLE** option.
- The title of the table of contents configured in the profile.

This option should only be set at job level and cannot be configured in the profile configuration.

The document outline can be further configured for each individual input document using the following options:

ADD-DOCUMENT-OUTLINE (optional) Defines whether a bookmark is added to the document outline for the input document. This option can be set at the job level where it applies to all documents, or at the document level where it only applies to the specific document.

true (default) A new bookmark is added to the document outline for the input document. Existing bookmarks of the input document are preserved and added as children to the newly created bookmark.

false The existing bookmarks of the input document are added directly to the top level of the document outline.

This is useful, for example, for the following use cases:

- Special input documents that don't represent a chapter in the output document, e.g. a cover sheet or table of contents.
- Documents that already contain all necessary bookmarks

DOCUMENT-OUTLINE-TITLE (optional) The name of the newly created bookmark (if enabled).

If this option is omitted, one of the following values is used instead (in order of precedence):

- The value of the **DOCUMENT-TITLE** option.
- The document title from the document metadata.
- The filename of the input document (without the extension).

This option should only be set at document level and cannot be configured in the profile configuration.

Table of contents

The automatically generated table of contents provides an overview of the dossier.

The table of contents supports multiple individually configurable nesting levels by using the existing outline tree of the input documents. Optionally, an parent entry can be generated for each input document.

The table of contents can be customized for each job using the following options:

TOC-TITLE (optional) The title of the table of contents.

This option overrides the value specified in the profile configuration. The option should only be set at job level and is ignored at document level.

The table of contents can be further configured for each individual input document using the following options:

ADD-DOCUMENT-TOC-ITEM (optional) Defines whether an entry in the table of content is added for the input document. This option can be set at the job level where it applies to all documents, or at the document level where it only applies to the specific document.

true (default) An entry in the table of contents is added for the input document. Existing bookmarks of the input document are added as sub-entries.

false The existing bookmarks of the input document are added directly to the top level of the table of contents.

This is useful, for example, for the following use cases:

- Special input documents that don't represent a chapter in the output document, e.g. a cover sheet or table of contents.
- Documents that already contain all necessary bookmarks

DOCUMENT-TOC-TITLE (optional) The name of the entry for the document in the table of content (if enabled).

If this option is omitted, one of the following values is used instead (in order of precedence):

- The value of the **DOCUMENT-TITLE** option.
- The document title from the document metadata.
- The filename of the input document (without the extension).

This option should only be set at document level and cannot be configured in the profile configuration.

Stamps

A general overview about stamps can be found in [Stamping](#).

Stamps can be further customized for the Dossier workflow by using the following document options:

DOCUMENT-STAMP-TITLE (optional) The title of the document used for stamping, i.e. for the [input:DOCUMENT.TITLE] placeholder (if used).

If this option is omitted, one of the following values is used as placeholder value instead (in order of precedence):

- The value of the DOCUMENT-TITLE option.
- The document title from the document metadata.
- The filename of the input document (without the extension).

This option should only be set at document level and cannot be configured in the profile configuration.

CUSTOM.<OPTION-NAME> Options whose name starts with "CUSTOM." define custom placeholder variables that can be used with the placeholder format [custom:<OPTION-NAME>].

Example:

- The stamp text in the profile is configured as: "Reviewed by [custom:REVIEWER]".
- The document option CUSTOM.REVIEWER is set to "John Doe".
- The resulting stamp text is: "Reviewed by John Doe".

Example:

Example using the [Shell client pdfclient](#) to create a dossier from a cover page and two chapters. For the second chapter, the name of the bookmark is explicitly overridden.

```
C:\Temp>pdfclient -v -w Dossier -do ADD-DOCUMENT-OUTLINE false ^
chapter-1.pdf -do DOCUMENT-OUTLINE-TITLE "Chapter 2: XYZ" chapter-2.pdf out.pdf
Creating job (id job1_5g5fkycbhhe)
Adding file ".\chapter-1.pdf" (id jgvmhpyquh0)
Adding file ".\chapter-2.pdf" (id 4ogify1xd31)
Output: "out.pdf"
- Info: Assembled PDF files 'chapter-1.pdf' and 'chapter-2.pdf' into dossier 'out.pdf'.
```

3.7 Processing steps

3.7.1 Stamping

Stamping adds small bits of content to the existing content of the pages of a PDF. You can find more details regarding configuration in the documentation panel of the Conversion Service Docker Configurator.

Text stamp

Text stamps can be used to apply a single line of arbitrary text, e.g. to create header or footer lines or to apply a transparent watermark.

The stamps provide many configuration options such as font, size, color, opacity, and positioning.

Image stamp

Image stamps can be used to apply an image. For example, to apply a company logo.

Barcode stamp

Barcode stamps can be used to apply a 1D or 2D barcode. For example, to apply a stamp with the document ID.

Placeholders

The text of a text stamp or the value of a barcode stamp can be customized using placeholder variables.

The supported placeholder names are documented separately in the configurator for each workflow.

The following placeholders are supported by all workflows:

[custom: <OPTION-NAME>] Placeholder variables whose name start with "custom:" are replaced with the value of the corresponding document option CUSTOM. <OPTION-NAME>.

Example:

- The stamp text in the profile is configured as: "Reviewed by [custom: REVIEWER]".
- The document option CUSTOM. REVIEWER is set to "John Doe".
- The resulting stamp text is: "Reviewed by John Doe".

3.7.2 Metadata

The metadata of the resulting PDF can be customized in all workflows.

Note: All metadata is applied to the main result document only, and not to embedded files or any other document contained therein.
All runtime options to further customize the metadata must be set at job level, not at document level.

Standard properties

The four standard PDF metadata properties that can be customized are "Author", "Title", "Subject", and "Keywords".

A fixed value can be configured in the profile configuration for each of the properties. If no value is configured, the original value is preserved (if available).

The values can also be provided dynamically using the following job options:

META.AUTHOR The author of the document

META.TITLE The title of the document

META.SUBJECT The subject of the document

META.KEYWORDS Keywords that apply to the document

Extended properties

Extended metadata is defined by the XMP standard and the properties are grouped into schemas.

The Conversion Service Docker provides the possibility to set XMP property values from the Dublin Core schema, XMP Basic schema, XMP Rights Management schema, and custom extension schemas.

Dublin Core schema

The Dublin Core schema provides a set of commonly used properties.

The values can be configured statically for each profile in the configurator or dynamically using the following job options:

META.EXT.DC.CONTRIBUTOR Contributors to the resource (other than the authors)

Note: Although the schema technically supports multiple values, only a single value can be set here.

META.EXT.DC.COVERAGE The extent or scope of the resource.

META.EXT.DC.IDENTIFIER Unique identifier of the resource.

META.EXT.DC.RIGHTS Informal rights statement.

META.EXT.DC.SOURCE Unique identifier of the work from which this resource was derived.

META.EXT.DC.TYPE A document type. For example, "novel", "poem", or "working paper".

Note: Although the schema technically supports multiple values, only a single value can be set here.

XMP Basic schema

The XMP Basic schema contains properties that provide basic descriptive information.

The values can be configured statically for each profile in the configurator or dynamically using the following job options:

META.EXT.XMP.NICKNAME A short informal name for the resource.

META.EXT.XMP.LABEL A word or short phrase that identifies a document as a member of a user-defined collection. Used to organize documents in a file browser.

Note: This property is not available in PDF/A-1

META.EXT.XMP.RATING A number that indicates a document's status relative to other documents. Used to organize documents in a file browser. Values are user-defined within an application-defined range.

Note: This property is not available in PDF/A-1

XMP Rights Management schema

This schema includes properties related to rights management.

The values can be configured statically for each profile in the configurator or dynamically using the following job options:

META.EXT.XMP-RIGHTS.CERTIFICATE URL of an online rights management certificate.

META.EXT.XMP-RIGHTS.MARKED Indicates that this is a rights-managed resource.

META.EXT.XMP-RIGHTS.OWNER The legal owner of a resource.

Note: Although the schema technically supports multiple values, only a single value can be set here.

META.EXT.XMP-RIGHTS.USAGE-TERMS Text instructions on how a resource can be legally used.

META.EXT.XMP-RIGHTS.WEB-STATEMENT The location (URL) of a web page describing the owner and/or rights statement for this resource.

Custom extension schemas

For metadata properties that are not covered by the predefined schemas, a custom schema can be defined.

The schema definition must be provided statically in the profile configuration.

Note: XMP extension schemas are expected to be stable over time. Changes to a schema definition should only add new properties and never change the meaning or type of existing properties. If incompatible changes are necessary, a new schema should be created instead.

The actual property values can be configured statically or provided dynamically using a placeholder and the following custom job option:

CUSTOM. <OPTION-NAME> The value for the placeholder [`custom: <OPTION-NAME>`, 'Default value'].

- If the default value is omitted ([`custom: <OPTION-NAME>`]), the job option is required and an error is signaled if it is missing.
- If the default value is left empty ([`custom: <OPTION-NAME>`, ' ']), no value is set if the job option is not provided.

The single quotes (') around the default value are part of the syntax and must not be omitted.

Advanced usage

For advanced users, the metadata can be customized further using the following job options:

META.XMP A complete XMP packet that replaces the metadata of the input document.

This option can be used to set all kind of metadata, including [Standard properties](#), [Extended properties](#) and [Custom extension schemas](#).

Any properties set using the standard mechanisms above are applied to this packet and any values set using the standard mechanisms take precedence over the values provided by this XMP packet.

Note: When creating a PDF/A document, the XMP metadata must contain a full schema definition for all custom extension schemas. Such a definition can be provided

- directly in the XMP packet.
- using [Custom extension schemas](#) configuration. In this case the "Force Definition" setting in the schema should be enabled.

3.8 Supported file formats

Each workflow supports a specific subset of the following file formats, which is documented in the description of the workflow itself.

- Document formats
 - PDF 1.x, 2.0
 - PDF/A-1, PDF/A-2, PDF/A-3
- Other formats
 - Email: EML, MSG (without encryption)
 - Word⁴: DOC, DOT, DOCX, DOCM, DOTX, DOTM, RTF, XML (WordprocessingML 2003)
 - Excel⁴: XLS, XLT, XLSX, XLSM, XLTX, XLTM, XML (SpreadsheetML 2003)
 - PowerPoint⁴: PPT, PPS, PPTX, PPTM, PPSX, PPSM
 - OpenOffice^{4,5}: ODT, ODS, ODP
 - CSV
 - HTML, HTM (prepared for archiving⁶)
 - TXT
 - XML⁷
 - ZIP (without password protection)
- Image formats
 - JPEG, JPEG2000
 - TIFF
 - BMP
 - GIF
 - JBIG2
 - PNG
 - HEIC
 - HEIF

Where applicable, the Conversion Service Docker adheres to the following standards:

- ISO 32000-1 (PDF 1.7)
- ISO 32000-2 (PDF 2.0)
- ISO 19005-1 (PDF/A-1)
- ISO 19005-2 (PDF/A-2)
- ISO 19005-3 (PDF/A-3)

⁴ The conversion of Office documents requires additional configuration, see [Converting Office documents](#).

⁵ PDF conversion of OpenDocument Format depends on the rendering in Microsoft Word, Excel or PowerPoint, in particular visual differences may occur with tables and tabs. The visual differences caused by the rendering of shapes are usually not acceptable.

⁶ HTML documents need to be self-contained (layout information and images are either inline or available on the web) and suited for portrait page layout. Javascript content is disabled during processing.

⁷ Layout information and images need to be available on the web.

4 User guide

This section gives an overview of the usage and configuration of the Conversion Service Docker features and interfaces.

4.1 Service configuration

The container can be configured using environment variables passed at startup.

4.1.1 Service host address

The port exposed by the container is 13033. When running the container, the port must be published. The port defines the address of the container's [REST service](#) as:

`http[s]://<hostname>:<port>/conversion/v1.0/rest.`

This is the endpoint URL used by clients, such as the [Shell client pdfcliënt](#).

It is recommended to map the exposed port to the same port of the host machine:

`http[s]://localhost:13033/conversion/v1.0/rest`

HTTPS

By default, the service endpoint uses HTTP. Activating HTTPS disables support for HTTP. This is to prevent clients from accidentally sending sensitive information over HTTP.

service__serviceEndpoint Set this to an URL of the following form to activate HTTPS:

`https://localhost:13033/conversion/v1.0/rest`

service__certificate__path When activating HTTPS, a valid host certificate is required. The certificate must be provided as PKCS#12 file (.pfx or .p12), which includes the certificate's private key and issuer certificates.

If the private key is password-protected, the password can be configured using `service__certificate__password`.

```
docker run -dp 13033:13033 \
  -e service__serviceEndpoint=\
    https://localhost:13033/conversion/v1.0/rest \
  --secret source=service_certificate,target=service_certificate \
  -e service__certificate__path=\
    /run/secrets/service_certificate \
  pdftoolsag/conversion-service
```

4.1.2 License key

The license key must be passed using one of the following environment variables:

LICENSEKEY The value is the license key.

```
docker run -dp 13033:13033 \
```

```
-e LICENSEKEY=4H-V4-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX \
pdftoolsag/conversion-service
```

LICENSEKEY_FILE The value is the path to a text file that contains the license key.

```
docker run -dp 13033:13033 \
--secret source=service_licensekey,target=service_licensekey \
-e LICENSEKEY_FILE=/run/secrets/service_licensekey \
pdftoolsag/conversion-service
```

4.1.3 Converting Office documents

There are two options to convert Word, Excel, and PowerPoint documents to PDF with the Conversion Service Docker. The [Office for the web service](#) option converts the documents using a Microsoft Azure Cloud service. Technologies used with this option include the Microsoft Authentication Library (MSAL), the Microsoft Graph API, and the Microsoft Sharepoint API.

The [Windows Service instance](#) option requires an additional Conversion Service installation on a Windows server.

Office for the web service

A Microsoft 365 subscription (Azure cloud tenant) and a user account with a Office 365 license are required for this option. It is recommended to create a dedicated user account, as full control over this user's OneDrive is requested. This user is assumed as `myuser@myorganization.onmicrosoft.com`, hereafter.

1. Enable Office Conversion in the Processing Steps section of your profile. Note that this is active by default.
2. The `csconfig` command from the Docker image can be used to consent to the required application permissions and to initialize the token cache. It is recommended to create a volume (e.g. `convsrv-etc`) for the storage of this file based cache. This volume `convsrv-etc` is mounted to `/etc/convsrv` to provide the tokens to the Conversion Service Docker. The token cache is stored in `/etc/convsrv/TokenCaches`. The user `convsrv` requires read and write permissions to this folder. Please be aware that the tokens and the binary file containing the tokens are not encrypted. The username `myuser@myorganization.onmicrosoft.com` and Office type are provided as environment variables `SERVICE__OFFICE__TYPE` and `SERVICE__OFFICE__USERNAME`.

```
docker volume create --name convsrv-etc
docker run --rm --mount "type=volume,src=convsrv-etc,dst=/etc/convsrv" \
--user root pdftoolsag/conversion-service \
bash -c "mkdir /etc/convsrv/TokenCaches && chown convsrv:convsrv /etc/convsrv/TokenCaches"
docker run --mount "type=volume,src=convsrv-etc,dst=/etc/convsrv" \
-e SERVICE__OFFICE__TYPE=OfficeWebService \
-e SERVICE__OFFICE__USERNAME=myuser@myorganization.onmicrosoft.com \
pdftoolsag/conversion-service \
bash -c "bin/csconfig office webinit"
```

3. When starting the Docker container, mount the volume `convsrv-etc` inside the container at `/tmp/convsrv` to provide access to the token cache. Set environment variables `SERVICE__OFFICE__TYPE` and `SERVICE__OFFICE__USERNAME` according to the configured user and office type.

```
docker run -dp 13033:13033 ... \
--mount "type=volume,src=convsrv-etc,dst=/etc/convsrv" \
-e SERVICE__OFFICE__TYPE=OfficeWebService \
-e SERVICE__OFFICE__USERNAME=myuser@myorganization.onmicrosoft.com \
```


Windows Service instance

Office documents can be converted by sending them to another instance of the Conversion Service running on Windows. Office document conversion can be activated as follows:

1. Set up and run a Conversion Service on Windows.
2. Enable Office conversion in the processing steps of your profile. Note that this is active by default.
3. When starting the Docker container, set the environment variable `WINDOWS_SERVICE_ENDPOINT`:

```
docker run -dp 13033:13033 ... \
-e WINDOWS_SERVICE_ENDPOINT=http://server:13033/conversion/v1.0/rest \
pdftoolsag/conversion-service
```

Note that if the Windows service uses HTTPS, its host certificate must be trusted by the Docker container. Otherwise, no connection can be established. By default, no trusted certificates are installed. Copy trusted certificates to `/usr/local/share/ca-certificates/` and run the command `update-ca-certificates`.

4.1.4 Proxy

HTTP_PROXY Environment variable to set the proxy. The proxy is used for both http and https.

4.1.5 Cross-origin requests (CORS)

CORS_ORIGINS Environment variable to restrict cross-origin requests to a set of allowed origins. The value is a comma-separated list of URLs. A wildcard `*` can be used to allow all origins or all subdomains of a specific domain.

All parts of the URL must match, i.e. the scheme, host, and port. The URLs must not specify a path, i.e. the following URL is invalid: `https://www.example.com/`.

- Allow all origins (default): `CORS_ORIGINS=*`
- Allow single origin: `CORS_ORIGINS=https://www.example.com`
- Allow multiple domains: `CORS_ORIGINS=https://www.example.com, https://www.domain.com`
- Allow all subdomains: `CORS_ORIGINS=https://*.example.com`
- Allow single domain and port: `CORS_ORIGINS=https://www.example.com:5000`

4.1.6 Advanced service configuration

Use forwarded HTTP headers from WAF

USE_FORWARDED_HEADERS Set the environment variable to `True` to activate the use of forwarded HTTP headers:

- Header `X-Forwarded-For`: Contains the IP address of the client that initiated the request.

Load balancer

Load balancing is supported. In addition to configuring the load balancer, there are also requirements on the clients to ensure optimal operation.

The Conversion Service Docker uses no shared resources among the backend servers. So each job is processed exclusively by the backend server where it has been created. Therefore, it is important to use sticky sessions in the load balancer such that all requests for a job are forwarded to the correct backend server.

Load balancer configuration

The load balancer must be configured to use sticky sessions. For this, it is recommended to use a cookie that is set upon the first request.

Kubernetes example 1: Annotations for NGINX ingress controller

```
nginx.ingress.kubernetes.io/affinity: "cookie"  
nginx.ingress.kubernetes.io/affinity-mode: "persistent"  
nginx.ingress.kubernetes.io/session-cookie-name: "JOBSESSION"  
nginx.ingress.kubernetes.io/session-cookie-max-age: 3600
```

Kubernetes example 2: Annotations for Traefik ingress controller

```
traefik.ingress.kubernetes.io/affinity: "true"  
traefik.ingress.kubernetes.io/session-cookie-name: "JOBSESSION"
```

While the healthcheck could be implemented using the HTTP status codes of the responses, it is recommended to use the service status request of the REST API. This allows you to detect issues quicker and more reliably.

Requirements on clients

It is important that the clients support the load balancer's job session cookie.

After creating a job, the cookie must be stored and sent with subsequent requests.

It is recommended to use a dedicated cookie store for each job. This enables the load balancer to distribute the processing of multiple jobs to multiple backend servers.

The shell and GUI clients distributed with the Windows version of the Conversion Service Docker adhere to these rules. Therefore, they are suitable to test the load balancer configurations.

4.2 Profile configuration

Profiles can be created and exported using the Configurator GUI of a Windows installation of the Conversion Service Docker. In a second step, the configuration can be imported into the Docker container.

4.2.1 Create and export a profile configuration

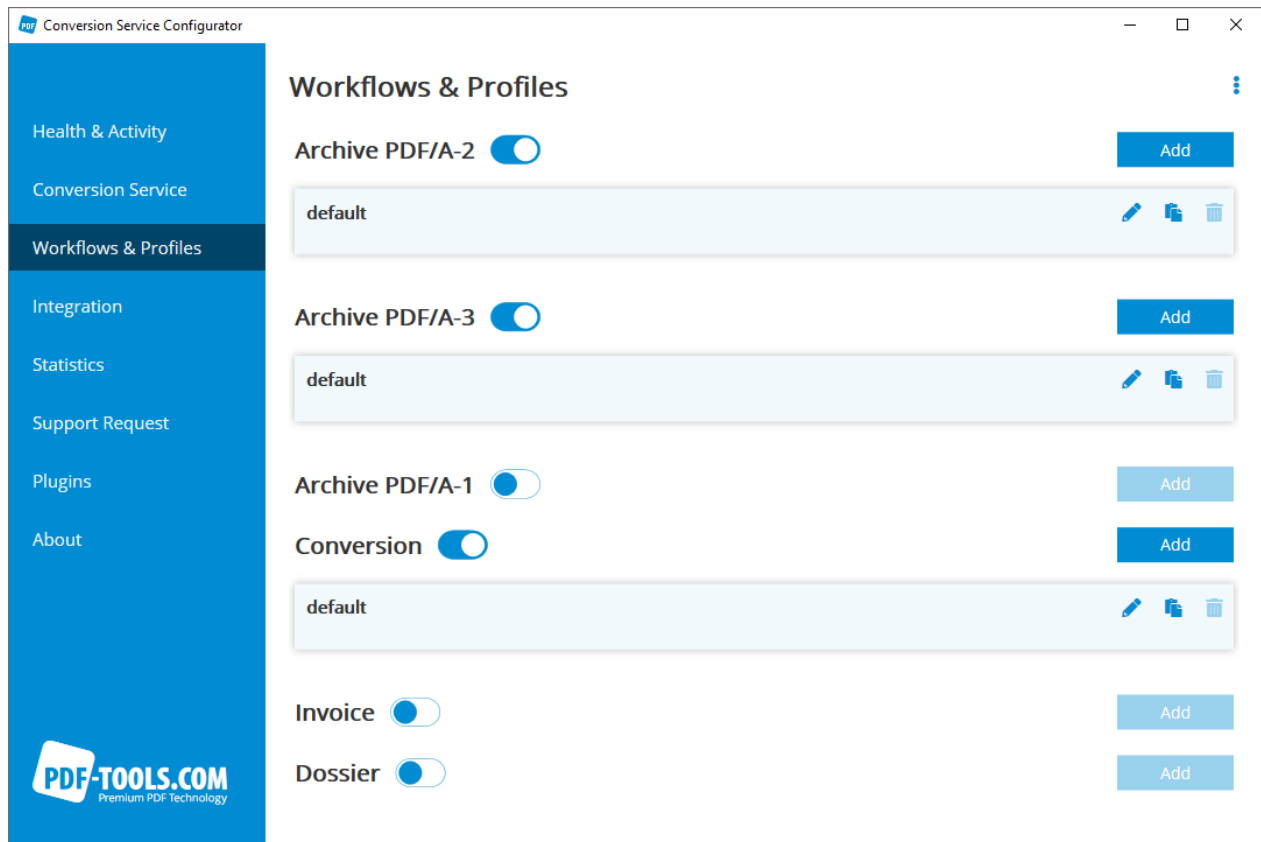
Windows installation

The version of the Conversion Service installed on Windows and the version of the container image must match. When updating to a new version, update the Windows service first. This automatically updates the configured profiles if necessary.

Create a profile configuration

Use the Conversion Service Configurator GUI, which is added to the Windows start menu during installation.

Profiles tab of the Conversion Service Configurator

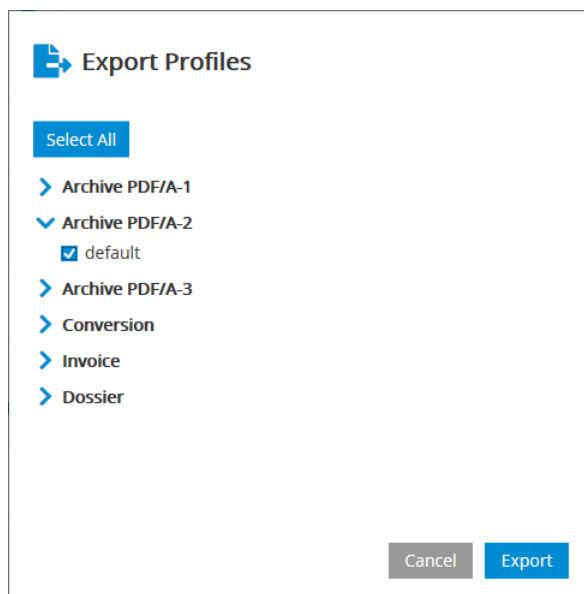


Make sure that all configuration values are valid in your Docker environment, i.e. URLs to services and paths to configuration files. Note that, as a result, the profile configuration might not be valid in your Windows environment and therefore your Windows service might not be operational.

Export the profile configuration

Click **Export Profiles** in the top-right menu to export one or more profiles to a profile configuration file.

Export window of Conversion Service Configurator



It is recommended to export the profiles to a file named `ProfileExport-<x>.<y>.<z>.export`, where `<x>.<y>.<z>` is the version number of the service, e.g. `ProfileExport-4.0.0.export`.

4.2.2 Set a profile configuration when starting the Docker container

The `ProfileExport-4.0.0.export` file containing the exported profiles must be made available in the container. To import the profiles on service startup, the environment variable `IMPORT_PROFILES` must be set.

Import profiles only

This simple profile configuration imports the file `ProfileExport-4.0.0.export` only, which is typically sufficient.

Bind `ProfileExport-4.0.0.export` to the container and set the environment variable `IMPORT_PROFILES` to activate the profile import:

```
docker run -dp 13033:13033 --network conversion-service \
  --mount "type=bind,src=C:\path\to\ProfileExport-4.0.0.export,\
    dst=/etc/convsrv/ProfileExport.export,readonly" \
  -e IMPORT_PROFILES=/etc/convsrv/ProfileExport.export \
  -e LICENSEKEY=4H-V4-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX \
  pdftoolsag/conversion-service:4.0.0
```

When importing profiles of a workflow, the following workflow properties are also imported:

- **Workflow state:** Whether the workflow is activated or deactivated.

Import profiles and further configuration files

This more advanced profile configuration is required for profiles that contain references to files. For example, SSL client certificates for connections to external servers.

1. When configuring the profiles, use the `/etc/convsrv` base path for all referenced files.
2. Create a directory `configuration` where all configuration files can be written to. Later these files are copied to a volume and mapped to `/etc/convsrv` in the container.
3. Copy `ProfileExport-4.0.0.export` and all references files to the directory `configuration`.
4. Create a volume `convsrv-etc-4.0.0` containing all configuration files from `configuration`:

```
docker volume create --name convsrv-etc-4.0.0
docker run --rm \
  --mount "type=bind,src=c:\path\to\configuration,dst=/source" \
  --mount "type=volume,src=convsrv-etc-4.0.0,dst=/etc/convsrv" \
  alpine \
  ash -c "cp /source/* /etc/convsrv/"
```

5. The folder `configuration` is not required anymore and can be deleted.
6. Run the image. Mount the volume `convsrv-etc-4.0.0` and set the environment variable `IMPORT_PROFILES` to activate the profile import:

```
docker run -dp 13033:13033 --network conversion-service \
  --mount "type=volume,src=convsrv-etc-4.0.0,dst=/etc/convsrv,readonly" \
  -e IMPORT_PROFILES=/etc/convsrv/ProfileExport-4.0.0.export \
  -e LICENSEKEY=4H-V4-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX \
  pdftoolsag/conversion-service:4.0.0
```

4.3 Shell client `pdfclient`

The shell client application, `pdfclient`, interacts with the Conversion Service Docker via the [REST API](#). The application creates jobs based on its input parameters. The jobs are then uploaded in parallel to the Conversion Service Docker and the shell client monitors the completion, before the output files are downloaded. In verbose mode (option `-v`), a detailed report including all performed actions on the documents is written to the console. If warnings or errors occur during processing, they are reported.

A selection of examples is explained in detail below.

Examples

Show usage

```
pdfclient
```

By typing `pdfclient` without parameters on the command line, the usage and examples are shown.

Basic

```
pdfclient -v -url http://<hostname>:<port>/conversion/v1.0/rest file.pdf ^
```

```
-pw <password> encrypted.pdf output.pdf
```

Sends the files `file.pdf` and `encrypted.pdf` to the Conversion Service Docker through the REST API endpoint URL `http://<hostname>:<port>/conversion/v1.0/rest` (see [Service host address](#)) and saves the result as `output.pdf`. Option `-v` turns on verbose mode, option `-url` sets the service endpoint URL, and option `-pw` supplies the password for `encrypted.pdf`. If the service endpoint URL (`-url`) is not set, `"localhost"` and the default port are used.

Workflow and profile

```
pdfclient -w "Archive PDF/A-2" -p "myCustomProfile" input.pdf output.pdf
```

Use the `-w` and `-p` options to select a specific workflow and profile for the job. In this example, the workflow is set to `"Archive PDF/A-2"` and the profile is set to `"myCustomProfile"`. If the options are not specified, the service's configured default workflow and profile are used.

Wildcards

```
pdfclient -r *.ext C:\path\to\output.pdf
```

Sends all files with extension `.ext` from the current working directory and all subdirectories. to the Conversion Service Docker, and saves the result as a single output document `C:\path\to\output.pdf`.

```
pdfclient -s *.ext C:\path\to\output
```

Process the files with extension `.ext` as separate jobs and save the results in `C:\path\to\output`. The output directory must already exist. Existing output files are overwritten by default. Output options such as creating unique filenames instead of overwriting existing files are explained in the usage of the tool.

Batch processing

```
pdfclient -r -s C:\path\to\input C:\path\to\output
```

Sends the files inside `C:\path\to\input`, including all subdirectories, to the Conversion Service Docker. It saves the results in `C:\path\to\output` with the same file structure as in the input directory. Existing output files are overwritten. Output options such as creating unique filenames instead of overwriting existing files are explained in the usage of the tool.

4.4 REST API

The Conversion Service Docker offers a REST API that allows you to schedule jobs and get service status information.

The REST API is also used by the other clients, e.g. the [Shell client pdfclient](#).

The service endpoint URL of the REST API is defined by the [Service host address](#) of the service configuration. The default value is `http://localhost:13033/conversion/v1.0/rest`.

Security considerations

The API is available by default on the local machine only. If it should be accessible remotely, the computer's firewall must be configured accordingly, e.g. using the Configurator GUI. When opening the port in the firewall it is recommended to add a rule that is as strict as possible, i.e. to not allow connections from untrusted computers.

Note that the REST API is designed for use in a protected intranet only. It offers no user authentication nor other security measures, e.g. against denial-of-service attacks. If this is required, a web application firewall (WAF) is recommended.

The Conversion Service Docker does support cross-origin requests (CORS). This is required when sending requests using JavaScript from a browser. By default, requests from all origins are allowed. If necessary, cross-origin requests can be restricted to certain origins using the service's configuration.

4.4.1 API usage

The API is described in detail by the OpenAPI document `doc/openapi.yaml` inside the installation directory. The YAML document can be viewed in an OpenAPI editor or used to generate client stub code of any programming language.

The API supports XML and JSON in the bodies of requests and responses. So it is recommended to set the headers `Content-Type` and `Accept` to the preferred type.

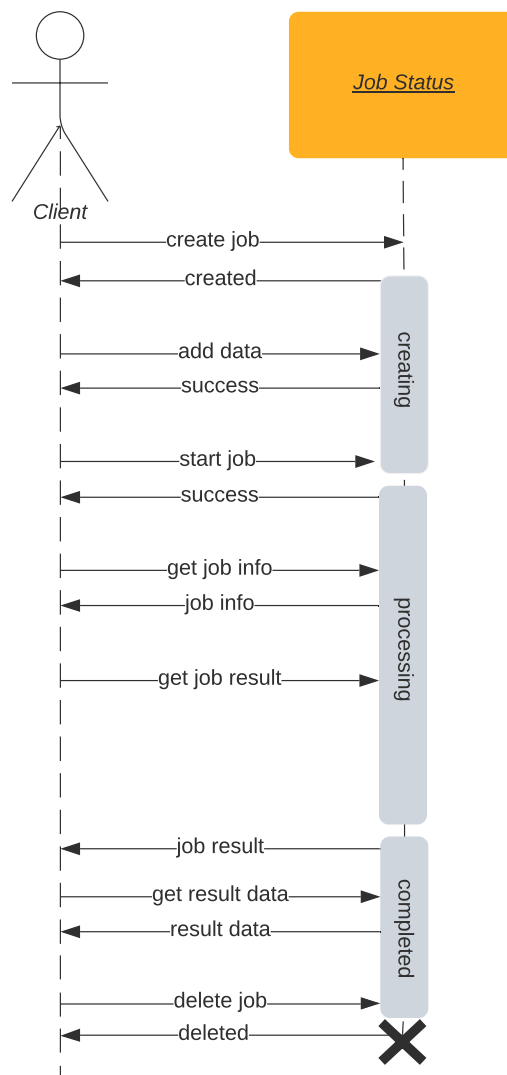
In case of an error, the API returns a suitable HTTP status code as well as a problem details object ([RFC 7807](#)). This object contains more information on the type and cause of the error. Notably, the object's property `detail` contains a human-readable explanation that is helpful to troubleshoot the issue. Therefore, it is recommended to parse and use the problem details object (`Content-Type application/problem+json` and `application/problem+xml` respectively) whenever the returned HTTP status code indicates an error.

Job processing

For a general overview of how jobs are processed by the Conversion Service Docker, see [Processing documents](#) or the documentation of a specific workflow, such as the [Archive PDF/A-2](#).

To schedule a job and retrieve its result, the following simplified sequence described in [Job processing sequence](#) can be used.

Job processing sequence



There is no limit of the number of jobs that can be started concurrently. The service will process the jobs in the order they were created, using the highest concurrency allowed by the system's CPU and the license. Nonetheless, it is recommended to not start much more jobs than the service can process. For example, on a machine with 8 CPU cores and a license for 8 cores, not much more than 8 jobs should be started⁸.

Service information

These methods return service status information.

The `getServiceStatus` method can be used to retrieve general status information. This is suitable to verify if the service is running, e.g. for health check monitoring. In addition, information on the service's load and general job count information is returned.

The `listJobs` method returns a list of all jobs and their status. This is useful to see what tasks are executing.

The `listWorkflows` method returns a list of all workflows and their profiles.

⁸ Note that this is a very simplified example. To determine the maximum concurrency, the whole system and configuration must be taken into consideration. Dependent systems, such as an OCR service or Office conversion might further limit the maximum concurrency.

4.5 Plugins

Plugins are non-standard components used for extending the Conversion Service Docker with custom workflows.

To run a Conversion Service Docker with a plugin, a custom image must be created. The `csconfig` command can be used in the `Dockerfile` to install the plugin:

Example Dockerfile: Install `Plugin-x.y.0.zip` for Conversion Service Docker version 4.0.0.

```
FROM pdftoolsag/conversion-service:4.0.0

USER root

COPY Plugin-x.y.0.zip .
RUN bin/csconfig plugins install Plugin-x.y.0.zip \
    && rm Plugin-x.y.0.zip

COPY ProfileExport-4.0.0.export .
RUN bin/csconfig profiles import -r ProfileExport-4.0.0.export \
    && rm ProfileExport-4.0.0.export

USER convsrv
```

Each plugin version is designed for a specific version of the Conversion Service Docker. So when upgrading the Conversion Service Docker, all plugins must be upgraded as well. Unless the plugins are updated, the service is not operational.

5 Troubleshooting

5.1 Error codes

If an error occurs during processing in a workflow, an error code with an explanatory message is returned.

The following are common error codes:

Internal The Conversion Service Docker is not operational because of an internal issue, e.g. an incomplete installation.

A detailed description of the problem is written to the service log file with severity [Error](#). The service administrator should be notified of the problem and submit a support request (see [Submitting a support request](#)).

No specific error message is returned because the issue is not related to the client or the request. This behavior can be changed in the service configuration. However, since the error is related to the service's configuration and the detailed description is designed to help the administrator resolve the problem quickly, the message might reveal internal configuration settings that you may not want to disclose to clients. Therefore, this is generally only recommended during installation and testing of the Conversion Service Docker.

Configuration The Conversion Service Docker is not operational because of a configuration issue, e.g. an invalid or an incompatible setting.

A detailed description of the problem is written to the service log file with severity [Error](#). The service administrator should resolve the problem with the help of the Configurator.

Generic A generic error occurred.

UnsupportedFormat The format of the input data is not supported.

UnsupportedFeature An unsupported feature was requested. This may be a feature of the input data or one requested using options.

Option An error occurred that is related to job or document options passed by the client. For example:

- A required option is missing.
- An option has an invalid or unsupported value.

Canceled The job has been canceled by the user.

Timeout The job has been canceled because of a processing timeout.

Password The data cannot be processed because of a missing or invalid password. Retry the conversion and specify the missing password using the document option `DOC . PASSWORD`. This option can be added multiple times to try several passwords for a document or to specify passwords for multiple files, e.g. attachments or embedded files.

Conformance There is a problem with the conformance of the input data. For example:

- The input data's conformance is not supported.
- The input data cannot be converted to meet the conformance required by the workflow and profile.

Corrupt Data cannot be processed because it is corrupt.

5.2 Service log

The main service process writes log messages to the standard output. The service log contains important messages and suffices to monitor the service's status and troubleshoot configuration issues.

Log entry properties

In addition to standard properties, log entries contain the following:

Level The severity of the message. The following severities are common:

Fatal Severe error. See Error below.

Error Error that prevents the service from operating, e.g. because of an incomplete installation or invalid configuration. Clients sending processing requests to the service receive an error code [Internal](#) or [Configuration](#).

It is recommended that the service administrator be notified whenever a message of severity Error or higher occurs. This can be achieved by monitoring the standard service log file or creating a [Custom logs](#).

Warn Errors that are not critical and do not prevent the service from operating. Even though no immediate action by the service administrator is required, it is advisable to review warnings periodically and decide if an action is required.

Info Informational events are useful to monitor the service's operation.

Debug Debug and tracing messages are strictly for development purposes and analysis by the Pdftools Support team. During productive use of the service, messages of this level should be disabled for performance reasons.

Trace See Debug above.

Process name The name or type of the process in which the log event occurred.

Message The log message.

Exception Critical log messages often have an exception associated that contains more detailed information on the message's cause.

Job ID Using the job id all messages of a particular processing job can be filtered for analysis.

Task ID This is an internal ID used by the service, which is useful for analysis by Pdftools Support team.

Remote IP The IP address of the remote host (client). This is meaningful only for messages that are associated with a request at the REST API.

5.2.1 Log file

The file containing all log entries is `/var/log/convsrv/ConversionService-Service.log`. This log file is primarily designed for support issues, e.g. when creating a support request file `SupportRequestFiles.zip`.

Example: See all critical errors that occurred in container <CONTAINER>.

```
docker exec <CONTAINER> grep -e ",ERROR\|FATAL," /var/log/convsrv/ConversionService-Service.log
```

5.2.2 Custom logs

For logging, the Conversion Service Docker uses [NLog](#), a very flexible logging platform. This allows to create additional log outputs, for example, to Amazon CloudWatch or Google Cloud Platform Logs. Contact pdfsupport@pdf-tools.com for more information.

5.3 Submitting a support request

To submit a support request, use the [support request form on www.pdf-tools.com](#). Depending on the type of issue, different data should be provided in addition to the problem description.

General issues

Create a `SupportRequestFiles.zip` in container <CONTAINER>:

```
docker exec <CONTAINER> bin/csconfig support /tmp/SupportRequestFiles.zip
docker cp <CONTAINER>:/tmp/SupportRequestFiles.zip .
```

The resulting file contains configuration and log files. No sensitive information is included: no input file or passwords from the configuration or client options are included.

Issue related to a particular job

Please submit the following data:

- Data described above in [General issues](#).
- Input files and options used for the conversion.
- The `jobId` of the conversion.

6 Version history

6.1 Version 4

Changes in Version 4

Installation and Deployment

- **Changed** default installation directory on Windows to `C:\Program Files\Pdftools\Conversion Service`. [v4.0]
- **Changed** default program data directory on Windows to `C:\ProgramData\Pdftools\Conversion Service`. [v4.0]

Common functionality

- **New** settings for email, HTML, CSV and XML conversion to control resizing and optimal fitting of content in case of content overflow. [v4.0]

Office Documents

- **New** settings for Excel conversion to control resizing and optimal fitting of content in case of content overflow. [v4.0]

6.2 Version 3

Changes in Version 3

REST interface

- **New** query parameter "url" on "addData" POST request to load a file directly from a webserver. [v3.4]
- **New** method "storeJobResultData" to send the result file directly to a webserver via HTTP PUT/POST request. [v3.4]
- **New** property "Status" in result of method "getJobResult" to get Success/Warning/Error status of job result instead of using legacy "Success" property. [v3.9]

Installation and deployment

- **New** Windows Server 2022 is supported. [v3.3]
- **New** Office 2021 (64 Bit) is supported. [v3.3]
- **New** dependency on Microsoft .NET Desktop Runtime 6.0 (Windows, x64) and Microsoft ASP.NET Core Runtime 6.0 (Windows, x64). [v3.7]

Workflows

- **New** workflow "Archive PDF/A-1". [v3.3]

"Archive PDF/A-2" workflow

- **New** convert mode configuration for child documents (attachments) for detailed configuration of which documents to convert. [v3.2]
- **New** support to convert HTML, CSV and XML child documents (attachments). Note that this is disabled by default. [v3.2]
- **New** collect mode configuration. This offers a detailed configuration of how to convert documents that have child documents (attachments), specifically how to combine the converted documents and how to handle errors when processing child documents. This supersedes the configuration options "Document Collection Mode" and "Child Error Handling". [v3.2]
- **New** stamping functionality to add single-line text content to the converted documents. [v3.4]
- **New** stamping functionality to add an image or 1D/2D barcode to the converted documents. [v3.11]

Archive PDF/A-3 workflow

- **New** convert mode settings "Skip" and "Skip with Warning", which allows to choose whether an informational event or a warning is generated when skipping (removing) a child document. The new default is "Skip with Warning". When updating an installation, the profiles' behavior is not changed. [v3.2]
- **New** stamping functionality to add single-line text content to the converted documents. [v3.4]
- **New** stamping functionality to add an image or 1D/2D barcode to the converted documents. [v3.11]

Conversion workflow

- **New** convert mode settings "Skip" and "Skip with Warning", which allows to choose whether an informational event or a warning is generated when skipping (removing) a child document. The new default is "Skip with Warning". When updating an installation, the profiles' behavior is not changed. [v3.2]
- **New** stamping functionality to add single-line text content to the converted documents. [v3.4]
- **New** stamping functionality to add an image or 1D/2D barcode to the converted documents. [v3.11]

Dossier workflow

- **New** profile setting to control the horizontal alignment (Left, Center, Right) of the title of the table of contents. [v3.2]
- **Changed** the document outline to contain a bookmark for the table of contents. [v3.2]
- **New** job option **TOC-OUTLINE-TITLE** to set the title of the bookmark for the table of content. [v3.2]
- **New** profile setting for automatic numbering of the entries in the table of contents. [v3.2]
- **New** profile setting to apply stamps to the pages of the table of contents. [v3.2]
- **New** optional preprocessing step to convert input documents to PDF. [v3.2]
- **New** support for custom placeholder variables in Text Stamps. [v3.4]
- **New** stamping functionality to add an image or 1D/2D barcode to the converted documents. [v3.11]

Common functionality

- **New** conversion of CSV, HTML (prepared for archiving) and XML documents. [v3.1]
- **New** conversion of WordprocessingML 2003 (.xml) and SpreadsheetML 2003 (.xml) documents via Microsoft Word and Microsoft Excel, respectively. [v3.3]

- **Changed** newly created bookmarks to be closed instead of open by default. [v3.3]
- **New** conversion of signed Emails (s/mime). [v3.4]
- **New** conversion of OpenDocument Text (.odt), Spreadsheet (.ods) and Presentation (.odp) via Microsoft Word, Excel and PowerPoint, respectively. [v3.4]
- **New** profile settings and options in all workflows to add metadata to the documents. [v3.5]
- **Improved** performance especially for small files. [v3.7]
- **Improved** robustness in low memory conditions on Windows. [v3.8]
- **New** option to convert Word, Excel, PowerPoint documents through the Azure cloud with the Microsoft Office for the web service. [v3.10]

Configurator

- **Changed** export and import of profiles to also include the workflow's activation state. [v3.3]
- **Changed** name of "Profiles" tab to "Workflows & Profiles". [v3.1]
- **New** possibility to activate and deactivate workflows in the "Workflows & Profiles" tab. [v3.1]

GUI client

- **Improved** UI design to provide a clear overview of the job progress and results. [v3.1]
- **New** Office add-ins for Word, Outlook, Excel and PowerPoint. [v3.3]

Integration/Connections

- **New** input connector "Watched Mailbox (IMAP)" to automatically convert emails or files sent to a configurable mailbox on an IMAP server. [v3.2]
- **New** input connector "Watched Mailbox (Exchange Online)" to automatically convert emails or files sent to a configurable mailbox on a Microsoft Exchange Online server. [v3.5]
- **New** output connector "Output Mailbox (IMAP)" to copy output files to a mailbox without actually sending an email. [v3.4]
- **New** output connector "Output Mailbox (Exchange Online)" to copy output files to a mailbox without actually sending an email. [v3.5]
- **New** output connector "Send Email (SMTP)" to send an email to a configurable or dynamic email address, containing the results or original files [v3.4] and/or additional information (error message, warnings) [v3.6].
- **New** output connector "Send Email (Exchange Online)" to send an email to a configurable or dynamic email address, containing the results or original files [v3.4] and/or additional information (error message, warnings) [v3.6].
- **New** output connector "Create Text File" to create a text file with additional information (error message, warnings) in an output folder. [v3.6]

Watched Folder connector

- **New** support for companion files to pass information as variables directly to the output connectors. [v3.10]

REST Input (JSON) connector

- **New** key "options" in JSON format to pass job-level and document-level options to the workflow. [v3.1]
- **New** key "variables" in JSON format to pass information directly to the output connectors. [v3.1]
- **Changed** default request body size limit from unlimited to 100MB. [v3.1]

REST Output connector

- **New** support for placeholder variables in URL setting. [v3.1]
- **New** support for sending form fields in addition to the files. [v3.10]
- **New** support for username/password based authentication. [v3.10]
- **New** support for setting the Accept header. [v3.10]

6.3 Version 2

Changes in Version 2

REST interface

- **New** support for Cross-Origin Requests (CORS). By default, requests from all origins are allowed. [v2.3]
- **Changed** value of property type of problem details object (RFC 7807) to `http://www.pdf-tools.com/service/rest/errors/<type>`. [v2.4]
- **New** Configuration and Timeout job result error codes. [v2.6]
- **New** support for HTTPS service endpoint. [v2.9]

Installation and deployment

- **New** Docker image "pdftoolsag/conversion-service" of the Conversion Service Docker. [v2.2]
- **Changed** license key format to 4H-VX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX. License keys with the format 1-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX-XXXXX are no longer supported. [v2.2]
- **New** dependency on Microsoft .NET Desktop Runtime 5.0 (Windows, x64) and Microsoft ASP.NET Core Runtime 5.0 (Windows, x64). [v2.9]
- **New** client installer (msi). [v2.10]

Workflows

- **New** workflow "Archive PDF/A-3" similar to "Archive PDF/A-2" but with support for non-PDF/A attachments. [v2.5]
- **New** workflow "Invoice" for creating PDF/A-3 conformant invoices with attachments. [v2.5]
- **New** workflow "Dossier" for merging PDF documents into a dossier while applying headers, footers or watermarks [v2.8] and support for a title page and table of contents. [v2.11]
- **New** workflow "Conversion" provides document conversions from multiple file formats to PDF with an emphasis on performance and document size. [v2.8]

Archive PDF/A-2 workflow

- **New** profile option "Flatten annotations" in section "Document Optimization". [v2.3]
- **New** profile option "Child error handling". [v2.7]
- **New** profile option for emails "Use subject as filename". [v2.8]

Common functionality

- **New** conversion of HEIC and HEIF images. [v2.11]
- **Improved** email conversion: Attachments of emails where parts of the attachment are removed during the conversion process (e.g. a not convertible part of a ZIP file) are now listed as "Partially removed" inside the email header. [v2.7]
- **New** profile setting "Task timeout" for all workflows, to avoid the situation where a single long-running processing step can lead to timeouts in other unrelated jobs.
- **Improved** and harmonized messages of warnings and events.

Office documents

- **New** conversion of RTF documents. [v2.2]
- **Improved** conversion of Word, Excel, and PowerPoint: These types of documents are now converted directly within the Conversion Service Docker through Microsoft Office applications. A parallel installation of the 3-Heights® Document Converter is no longer required and no longer supported for this purpose. [v2.1]
- **Improved** conversion of DOCX documents: Certain corrupt documents are now repaired automatically. [v2.6]
- **Improved** conversion of password-protected DOCX documents: Embedded files from these documents can now be processed. [v2.6]

Digital signatures

- **New** signature types "Windows - PAdES-B-B" and "PKCS#11 - PAdES-B-B" to create basic signatures that require no TSA nor revocation information. [v2.5]
- **Improved** processing of signed PDF documents: If signatures need to be removed, their visual appearance is preserved. [v2.10]

OCR

- **Improved** OCR recognition task scheduling: Each page is now processed in a separate task, which significantly reduces the amount of time a worker is blocked by OCR tasks and thus improves the latency of other high priority tasks. [v2.4]
- **Improved** OCR engine configuration: Specific configuration for each engine type. [v2.4]
- **New** support for load-balancing to multiple instances of 3-Heights® OCR Service. [v2.6]

PDF to PDF/A conversion

- **Improved** conversion event messages to be more specific. [v2.11]
- **Improved** detection of optimal target conformance level where the highest level is chosen that can be achieved with a perfect conversion. [v2.11]
- **New** warning types "Annotation removed" and "Multimedia removed" for events that were previously reported as "Action removed". [v2.11]

Configurator

- **Improved** UI design of the profiles tab to provide a clear overview of profiles and their related workflow. [v2.7]
- **New** tab "Integration" for configuring connections to allow easy integration of the Conversion Service Docker into existing systems. [v2.4]
- **New** tab "Health & Activity" for monitoring the state and recent activity of the service. [v2.10]
- **New** tab "Statistics" to report and analyze the service's conversion history. [v2.11]

- **New** possibility to copy profiles and connections. [v2.11]
- **Improved** UI validation of service, profiles and connection configuration to provide greater clarity of errors and warnings. [v2.12]

Integration/Connections

- **New** service "Conversion Service Docker Connections" that allows combining input sources with output for easy integration into existing systems. [v2.4]
- **Removed** service "Conversion Service Docker Folders" (superseded by the new service "Conversion Service Docker Connections") [v2.4]

Input connectors

- **New** REST input connector for plain HTTP requests. [v2.9]
- **New** REST input connector for structured JSON requests. [v2.10]

Output connectors

- **New** REST output connector. [v2.9]
- **New** Command Execute output connector. [v2.9]

Various

- **Changed** trademark [™] to registered trademark [®]. [v2.10]

6.4 Version 1

Changes in Version 1

- **Changed** minimum required version of 3-Heights[®] Document Converter Enterprise Edition to version 6.5.
- **Changed** port to connect to 3-Heights[®] Document Converter Enterprise Edition to 7983.
- **New** requirement of Microsoft Windows Desktop Runtime 3.1 (Windows, x64) and Microsoft ASP.NET Core Runtime 3.1 (Windows, x64). This is in addition to the .NET Framework 4.7.
- Changes to the REST API
 - **Changed** to return RFC 7807 Problem object instead of proprietary error object.
 - **Changed** JSON serialization of enum values from integer to string.
 - **Changed** XML objects to use no namespace in accordance with openapi.yaml.
 - **New** codes in the service status response.
- **New** GUI Client for manual processing.
- **New** view in the 4-Heights[®] Conversion Service Configurator GUI for submitting a support request.
- **New** profile option for job priority in workflow "Archive PDF/A2".
- **New** profile configuration (Digital Signature) available in the configurator GUI.
- **New** attachment information is shown as part of the email header.
- **New** possibility to import/export profile configurations.
- **New** license dependent view in the Conversion Service Configurator GUI for installing plugins.
- **New** document content overflow into the margin when converting emails is signaled with ContentOverflowWarning.

- **New** service setting for proxy configuration.
- **New** documentation panel in the Conversion Service Configurator GUI.

7 Licensing, copyright, and contact

Pdftools is a world leader in PDF (Portable Document Format) software, delivering reliable PDF products to international customers in all market segments.

Pdftools provides server-based software products designed specifically for developers, integrators, consultants, customizing specialists, and IT departments. Thousands of companies worldwide use our products directly and hundreds of thousands of users benefit from the technology indirectly via a global network of OEM partners. The tools can be easily embedded into application programs and are available for a multitude of operating system platforms.

Licensing and copyright The Conversion Service Docker is copyrighted. This user manual is also copyright protected; It may be copied and given away provided that it remains unchanged including the copyright notice.

Contact

PDF Tools AG
Brown-Boveri-Strasse 5
8050 Zürich
Switzerland
<https://www.pdf-tools.com>
pdfsales@pdf-tools.com