QS-DocumentAssembler

Version 4.6

DocumentAssembler

by QS QualitySoft GmbH

Product Documentation

Software-Version 4.6

01/2024

Table of Contents

Note from Editors

Quality Soft strives to keep this documentation up to date. If you find any errors or incomprehensible parts in the manual, just send us an e-mail. The QS-DocumentAssembler is constantly maintained, improved and expanded. Changes to the product are possible at any time.

1   Introduction

1.1   **Task**

Documents that have been scanned in by document scanners, have been captured by digital cameras or have been received from fax servers are present as image files. Barcodes should be recognised in these image files. Likewise, Adobe PDF documents could contain barcodes that should be read. The recognised barcodes should be made available for further processing or the files should be renamed according to the barcodes recognised.

1.2   **Features of QS-DocumentAssembler**

The QS-Document Assembler is a powerful, highly configurable software for mass reading of barcodes and OCR areas from images or Adobe PDF files.

The automated recognition and sorting, indexing, structuring, mapping and renaming of scanned documents based on barcodes are the most common application areas.

Barcodes can be searched on the whole page at any angle. Also file formats with multiple pages per file are supported. The results of any number of recognised barcodes can be written to result files.

QS-DocumentAssembler can read both 1-dimensional (linear) barcodes and 2-dimensional barcode types like PDF417, Data Matrix, QR Code and Aztec Code, which can contain large amounts of data. Thus, in a specific version for the German Tax Office ("Elster-Barcode") more than 800 encrypted characters are recognized and provided for subsequent processing. The recognition time for a PDF 417 barcode – as used in the banking sector - is less than 50 ms.

The barcode recognition of the QS-DocumentAssembler is based on the "QS-Barcode Library" which is also available as a stand-alone product for software developers - "QS-Barcode SDK" (Software Development Kit).

OCR recoginition on documents can be configured in such a way that the results can be used as a barcode result for further processing of the documents.

The QS-DocumentAssembler will support the creation and administration of a number of individual configurations that will be called jobs in the following document. In each job, an individual set of processing characteristics will be collected. Among these characteristics are for example the source and target directories, the barcode type that is to be detected, the processing of the target files and many more.

The jobs of the QS-DocumentAssembler can be carried out manually or they can be configured so that they are carried out by a scheduler in defined periods of time, either once or continually.

The QS-DocumentAssembler has also been designed for batch operations, in which documents will be processed automatically in large numbers.

The QS-DocumentAssembler also offers you the opportunity to carry out the processing of documents from local drives and from network drives without user interaction also as a Windows service.

1.2.1 **Supported Barcode Types**

1-dimensional barcode types (linear barcodes)

- Code39
- Code93
- Code 32
- Code 11
- Codabar
- Code 128
- Codablock F
- EAN 8 and 13
- EAN 128 / UCC 128
- UPC A / UPC E
- Code 2/5 (Interleaved, Industrie, IATA, BCD-Matrix, Datalogic)
- GS1-Family (EAN-13 and DataBar)
- Pharmacode
- Patchcode

2-dimensional barcode types

- Aztec
- Data Matrix
- QR Code
- PDF417.

1.2.2 **Processing of Documents**

Within the framework of the processing of documents, the following use cases are differentiated between in the QS-DocumentAssembler.

A barcode is attached to each page of a source document

The source files can be both one-page documents and documents with multiple pages. If a barcode is found, then the page will be filed in the defined target directory according to the pre-set rules. Pages without barcodes will be filed in the defined error directory.

Keeping a multiple page source document together

Every page of the source document will be searched for barcodes. The naming of the entire document will be performed according to the first barcode that is found. If no barcode is found, then the document will be moved into a defined error directory.

Every barcode change defines the beginning of a new stack (target file)

The source files can be both one-page documents and documents with multiple pages. It has to be ensured that the scanned documents are numbered consecutively.

When a page with a barcode is found, then a new stack (target file) will be created. All following pages without barcodes will be attached to this stack, until a new barcode that differs from the previous one is found, which will in turn begin a new stack.

Every found barcode defines the beginning of a new stack (target file)

In comparison to the previous definition, with this constellation a new stack is defined with every found barcode. This serves the purpose to separate documents with identical barcodes within the stack mode, instead of writing them into one target file.

The barcode defines the beginning of a new stack with the removal of cover sheets/insert sheets.

You can create insert sheets with a barcode that you can then place in front of your document stack before scanning.

The source files can be both one-page documents and documents with multiple pages. It has to be ensured that the scanned documents are numbered consecutively. The stack (target file) will be created via the barcode; however, pages with a barcode will be removed and not entered into the target document.

The target document will receive the name of the barcode of the preceding cover sheet. This option will be provided if your documents do not have any barcodes.

The barcode defines the beginning of a new stack and the following pages have a predefined barcode.

There is always the danger that the barcodes will not be detected and that then two stacks are combined into one. If you attach a placeholder barcode with a fixed and constant value to every following page, then the incorrect combining together of two stacks can be prevented.

Identical barcodes define the beginning and end of the stack.

An additional method to prevent the incorrect combining of two stacks is to label the beginning and the end of the stack with an identical barcode. If the second barcode cannot be found here, then the entire region until the next barcode that is found will be saved as a document in the errata. There is also the possibility here to remove the pages with the barcodes.

Reading and recording barcodes in documents

QS-DocumentAssembler can also be used to read a number of barcodes from a source file for protocol purposes and to deposit them into a results data file. This results data file can then, after recognition by a third party program, be used again.

Examples of applications are the recording of outgoing goods, in which a large number of small barcode labels with sequential numbers are attached to large control sheets, which are then scanned in. Reading using QS-DocumentAssembler is easier and quicker using this method than using a hand-held scanner.

As part of the processing opportunities that are described here, DocumentAssembler can provide additional functions that can be used when necessary.

Fundamentally, there is the option in all use cases to adapt the filenames of the target files over the variable target filenames individually. To carry this out, there are a number of predefined variables available for the individual configuration of the target filename.

In addition, in all use cases, blank pages can be filtered out of all source documents. Here, there is the option of filtering out blank pages via the definition of the darkness level of a blank page.

Additionally, there is also the possibility of using the same set of variables for the individual application of an imprinter stamp to a freely definable position in the target document.

If a number of barcodes are located in the source document, then you can filter out the desired barcode via a filtering separator using regular expressions.

Occasionally, pages are scanned in the wrong way round. Here, there is the possibility of rotating these pages in the direction of the barcode.

When writing the target files, you can select which format the target file should have. We support the PDF and TIFF formats as one-page and multiple page formats, as well as the BMP, PNG, JPG and GIF formats as one-page formats. If, when scanning you have to produce a graphical format, but also require the target format to be a PDF, then you can change the output format without any problems.

Within the framework of the processing of documents, the following use cases are differentiated between in the QS-DocumentAssembler.

The barcode recognition is based on the QS-Barcode Library. The optimal barcode settings can be found with our Freeware bcTester: https://www.bctester.de.

For details of individual parameters, please see the corresponding QS-Barcode SDK documentation, available online at:

https://qualitysoft.de/wp-content/uploads/qsbarsdk_en.pdf

Basic information on QS-Barcode recognition (types, printing advice, etc.)

https://qualitysoft.de/wp-content/uploads/bcinf_en.pdf

QS-DocumentAssembler prices and ordering advice:

https://qualitysoft.de/wp-content/uploads/docasm_prices.pdf

Should you experience problems in recognizing barcodes please contact QualitySoft at any time.

Please send your images to info@qualitysoft.de.

1.3   **Standard and Professional Edition**

The QS-DocumentAssembler is available in a Standard and a Professional Edition. The Professional Edition differs from the Standard Edition in that the Professional Edition supports the processing of documents as a Windows service as well as it supports any number of parallel processes for the concurrent processing of several jobs.

2   **System requirements / Installation**

2.1   **System requirements**

QS-DocumentAssembler is compatible with the Microsoft Windows operating systems Server 2008, Server 2012, Server 2016, Server 2019, Server 2022, Windows 7/8/10/11. The Microsoft .NET Framework 4.8 has to be installed.

Every PC with at least 210 MB free hard drive storage is suitable for the use of QS-DocumentAssembler. The image files which are to be processed with the software often require much more space. 210 MB is just the space needed for the software.

There is a large selection of devices available for the scanning of documents with document scanners and for the automated filing of the digitalised images that are to be processed by the QS-DocumentAssembler that can be operated using various different scanning software.

QualitySoft would be very happy to advise you in the selection of appropriate components and itself possesses a corresponding range of products.

2.2 Installation

A free evaluation copy of QS-DocumentAssembler can be downloaded from https://www.qualitysoft.de/products/qs-documentassembler*.*

The QS-Document assembler only supports 64 bit operating systems.

With the QS-DocumentAssembler installer an update of already existing installations from version 4.0 onwards can be performed. A deinstallation of the older version is not necessary.

Please unzip the file before running setup. The evaluation copy runs in demo mode.

Take the following steps to install the program:

1. Install the Microsoft .NET Framework 4.8 if not installed on the target computer.

2. For the standard or professional version copy your license file (qsbc.lic) to the directory where you have unpacked the archive with the demo version and overwrite the existing Demo license.

3. Start the SETUP.EXE

4. Follow the instructions of the setup program. Change the target installation directory if desired.

5. During setup, your individual license file QSBC.LIC is automatically copied from the install directory into the target directory (not applicable in the evaluation version).

6. You can now start the QS-DocumentAssembler.

7. Configure the QS-DocumentAssembler according to your requirements.

Notes

If you want to use the QS-DocumentAssembler service, we recommend that you do not change the default "Install all users" setting. If you change this setting, the QS-DocumentAssembler service can be executed only for those users who are logged on during the installation process, as the relevant components that need to be registered will not be registered for other users.


2.3 Demo-Version

During the installation, a demo license is installed by default, if you have not replaced the demo license by your purchased license.

You can use demo version of the QS-DocumentAssembler without any time limit. The only limitation in use is the falsification of the barcode results. The following replacements are made:

3 turns to 1, A,a turns to Q,q; B,b turns to S,s. For 2D barcodes also the text "DEMO" will randomly appear in the result.

A note at the start of the QS DocumentAssembler brings this to your attention.
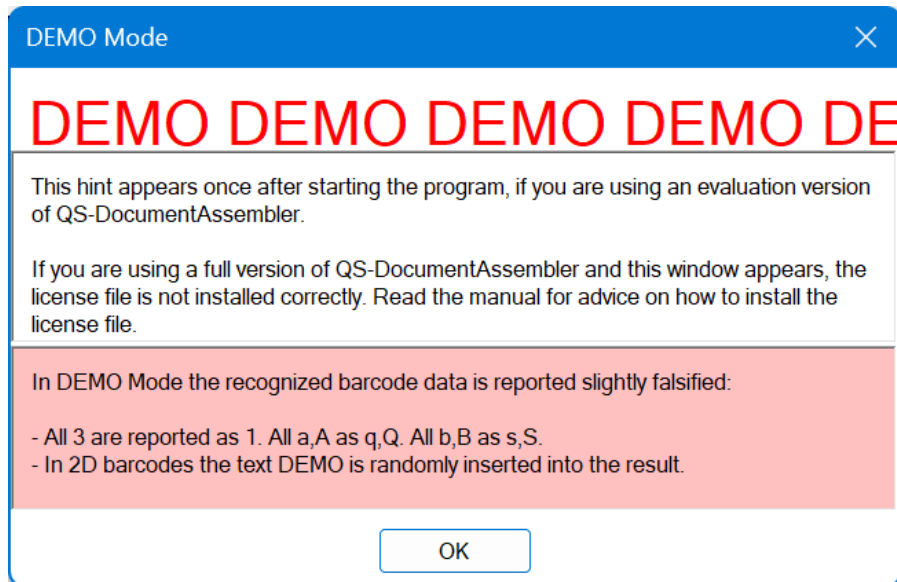
2.4   **Full Version**

If you have purchased the standard or professional version of QS-DocumentAssembler, you can replace the demo license by your full license before the installation process. If you have purchased the full version after installing the demo version, simply copy your license file (qsbc.lic) to the program directory of the QS-DocumentAssembler. You can immediately start working with the full version after a restart of the QS-DocumentAssembler.

3   Operation of QS-DocumentAssembler

3.1   Main Window

After the start, the QS-DocumentAssembler shows the following dialogue:

3.1.1   Job List

The basis of the DocumentAssembler is the job. A job corresponds to a configuration setting for the processing of documents with barcodes.

All of the jobs that are available for processing will be displaced in the job list. To allow a job to be carried out, it must have the "checked" status. This status indicates that the job contains all of the necessary information to enable it to be carried out.

When saving a job, plausibility checks will be carried out on the input data (see chapter 4, Plausibility check) . If there is still data missing, (for example, the source directory has not been entered or similar problems) or if the data are inconsistent, then the job can also be

saved as unchecked. The job will then be displayed as having an "unchecked" status. These jobs will additionally be displayed as being greyed out.

In the "Next start" column, you can see how and when a job will be carried out and/or can be carried out (see chapter 3.1.2, Execution of Jobs.

On the right hand side next to the job list is a vertical bar of buttons in which all of the functions for the processing of jobs are made available. You have the option to create a new job, process a job, copy a job or to delete a job.

3.1.2   Execution of Jobs

An active job can always be carried out in two ways. It will either be started manually on the button <Start> by the operator at a desired time, or it will be started automatically via the scheduler (see chapter 3.1.2.2, Scheduler) at a defined point in time.

The "Active" column displays whether or not a job can be carried out. Jobs to be carried out manually must also be set to be active before they can be carried out. Jobs that have not been checked cannot be set as active, as their settings will not permit successful execution.

3.1.2.1   Manual Execution

Via the <Start> button and via the <Execution, Start> menu, a processing run can be initiated manually for the selected job, if the job is intended to be executed manually.

3.1.2.2   Scheduler

The scheduler regulates the time-controlled execution of jobs, if they are intended for automatic processing.

The scheduler can either be started automatically when the program is started or it can be started manually. If you would like the scheduler to be started automatically when the DocumentAssembler is started, then you can activate the "Start the scheduler when the program starts" checkbox. Otherwise, the scheduler must be started manually via the < Start Scheduler> button.

Only active jobs will be included by the scheduling.

If a job has been scheduled for execution, then the next processing date will be displayed in the "Next Start" column and the job will be will be set to "waiting" status. "Processing" area, the time remaining until the start of the job will be displayed in the "Next Start in" field.

In addition, you can also define here the maximum number of jobs to be carried out at the same time. This makes particular sense if you are carrying out many jobs at short intervals. By setting a limit, you can regulate the maximum load on the processor of the corresponding computer.

3.1.3 **Processing**

In the "Processing" area, the progress of the processing of a selected, ongoing job in the job list will be displayed. If a number of jobs are running in parallel, you can have the progress of the processing of a different job that is running displayed by selecting it in the job list.

In the "Current Document" field, the document that is currently being processed including the number of pages that have been processed will be displayed. The progress bar shows the number of processed documents in relation to the total number that are displayed in the "Number of Files in Run" field. Successfully and incorrectly processed documents will be displayed in the "Number of Files Successfully Processed" and "Number of Files Processed with Errors" fields respectively. Documents that have not yet been processed will be displayed in the "Number of Files not yet Processed".

The display will reflect only the last and the current run. The display will provide no historical perspective about the complete number of processed files.

A processing run will always process all of the image files that are present in the source directory. Files that are added during the process into the source directory will be displayed in the "Number of New Files" field and will only be considered in the following processing run.

3.1.4 **The Menu Bar**

Using the menu bar many options are provided. Some of them can be accessed from the dialog directly, such as the start and finish of jobs as well as some advanced features.

Via menu <File> the application can be closed.

In menu <Optiones> the following settings can be made:

"*Start DocumentAssembler minimised*"
The QS-DocumentAssembler starts minimised. Once started, only the icon in the task bar is visible.
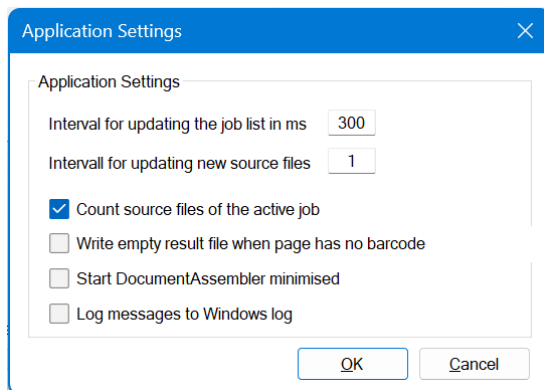
" Log Error messages to Windows Log"
> By default, the QS-DocumentAssembler logs to its own log file. You can switch on additional logging to the Windows Event Log.

"Count source files of the active job"
> If a job shall process files from subfolders and there are already many existing files, the counting of the files may lead to a slowdown of the interface reaction. If required you can therefore deactivate the counting of the source files.

"Application settings"

> Within this dialog you can change different settings that concern the application.

Application Settings

Application Settings

Interval for updating the job list in ms   300

Intervall for updating new source files   1

☑ Count source files of the active job

☐ Write empty result file when page has no barcode

☐ Start DocumentAssembler minimised

☐ Log messages to Windows log

OK   Cancel

Interval for updating the job list update in ms

The report of a job file will be updated every 300 ms. If required, you can raise this value.

Interval for updating new source files

The number of source files will be updated in every time-controlled iteration. If the source file shall only be updated in every third iteration, put the value to 3.

Count source files of the active job

See above.

Writing empty result file when page has no barcode

Normally, a result file is only written when a barcode has been found. If you choose this option, a result file with the size of 0 byte will be written if no barcode has been found.

Start DocumentAssembler minimised

See above.

Log messages to Windows log

See above.

By using the menu <service> you can install and uninstall the QS-DocumentAssembler service. Here you can also start the service and stop a running service. To perform these functions, the QS-DocumentAssembler must be started with administrator rights. If this function is performed without administrator rights, you will be asked if the DocumentAssembler shall be restarted with administrator rights.

The <View> menu displays the various log files. Thus, a search for these files in the directory structure is not necessary.

"Log File"
    Displays the corresponding log file of the the selected job.

"*Result File*"
    Displays the corresponding result file of the the selected job.

"*Error File*"
    Displays the corresponding error file of the the selected job.

"*Application Error File*"
    Displays QS-DocumentAssembler the error file

The <Language> menu allows you to switch between the available languages. At present, German and English are supported.

In menu <Help>, you can open the manual of the QS-DocumentAssembler as a PDF document and a PDF document with general information about the recognition of barcodes. In addition, the "Quick Support" of the TeamViewer can be started for support via the menu <Remote Maintenance>.

In addition you can here start an "About" dialog, which includes information on program version and manufacturer contact information.

3.1.5 The creation of a new Job

A new job can be set up at any time using the "New Job" button and can then subsequently be processed in the "Settings" dialogue box. Processing will be carried out over a number of the register cards that are located in the dialogue box.

3.1.6 The Handling of Jobs

Via the vertical button bar to the right of the list of jobs, new jobs can be created. In addition, existing jobs can be edited, copied or deleted.

New Job: A new job can be created. It will be preassigned with default settings.

Edit Job: Calls the dialogue for editing the selected job.

Copy Job: Duplicate the selected job.

Delete Job: Delete the selected job.

Depending on the status of a selected job, not all functions may be available. E.g. an active job can neiter be edited nor deleted.

The edit of an existing job can be released by a double-click on the job. The job must not be active during this procedure.

3.2 Settings

The "**Edit job**" button takes you to the job settings.

3.2.1 Tab "General"

Name

Every job must be given a unique name. Duplicate names are not possible due to technical reasons. The internal ID of the job will be shown under the name of the job.

Scheduler

The DocumentAssembler provides the option of executing jobs both manually and time-controlled in a variety of time slots.

The scheduler monitors the list of jobs to be processed and starts the jobs at the defined time. Only the maximum number of jobs to be carried out at the same time will be started. Any additional jobs that are to be started will be placed in the queue and will then be started when the number of jobs that are to be processed simultaneously permits it.

If there are no further execution times for a job, then a job will be deactivated and will then receive the status "expired".

If the scheduler is started at a time at which a job has already expired, then the job will no longer be executed and will instead immediately be given the "expired" status.

In the following section, the individually time-controlled trigger time points and their settings will be described.

- **Manually**
  No settings are possible because the job has to be started manually and can not be executed by the scheduler.

- **Once**

  If you check the radio button "Once", you must select a start date and start time at which the job should be started. The date must be in the future. After successful completion or if the start time is expired, the job is disabled and gets the status "expired".

- **Daily**

  For the daily execution of a job, you must specify the execution interval in seconds. The

execution interval defines the intervals at which a job will be started by the scheduler. Furthermore, you may also determine the start and end dates, though this is not compulsory. If you do not define any start or end dates, then the job will be scheduled both with immediate effect and permanently.

Execution can be restricted on a daily basis (every day, every second day etc.) or on a weekday, or on the weekend. The run time can additionally be restricted over the day.

- **Weekly**

For the weekly execution of a job, you must specify the execution interval in seconds. The execution interval defines the intervals at which a job will be started by the scheduler.

Furthermore, you may also determine the start and end dates, though this is not compulsory. If you do not define any start or end dates, then the job will be scheduled both with immediate effect and permanently.

Execution can be restricted on a weekly basis (every week, every second week etc.) and to defined weekdays. The run time can additionally be restricted over the day.

- **Monthly**

For the monthly execution of a job, you must specify the execution interval in se-

conds. The execution interval defines the intervals at which a job will be started by the scheduler.

Furthermore, you may also determine the start and end dates, though this is not compulsory. If you do not define any start or end dates, then the job will be scheduled both with immediate effect and permanently.

Execution can either be restricted on a weekly basis and to individual days within the week or

to individual days within the month.

The run time can additionally be restricted over the day.

Post Processing Application

QS-DocumentAssembler can start a post processing application after the job has finished (when all files in the source directory have been processed) or after the processing of each file. You can also decide whether the QS-DocumentAssembler should wait for the post processing application to end before a new job is started.

"Command Line"
> The desired program call is entered here.

"Run post processing application even if no file has been processed"
> Occasionally, it is desired that the *post processing application is started* when no files have been processed

"DocumentAssembler waits for the post processing application"
> Determines whether the QS-DocumentAssembler should wait for the end of the post processing application

"Execution"
> The post processing application can be executed after each job or after each document.

"Run post processing application hidden"
> The post processing application can be started visibly, if a user interaction is still to take place, or invisibly, so that the window of the executing program is not displayed. This may also prevent a disruptive interruption if you want to work on the computer in parallel.

Custom DLL

If you need special functions which the DocumentAssembler does not provide, please contact us. We will gladly provide an offer for the implementation of customer-specific functions.

Error Handling

Specify the behaviour of the QS-DocumentAssembler when errors occur.

"Display a message on error (processing is interrupted)"
> If errors occur during the processing of files, a message can be displayed. Thus the processing is interrupted. This option is off by default.

"Move unreadable files to the directory error"
> If files can not be read, they can be moved to the error directory. This option is enabled by default. If this option is not enabled, the entire processing is terminated. Processing can only continue if this file has been removed.

E-Mail dispatch

You can opt to receive an e-mail from the QS-DocumentAssembler whenever a processing error occurs that cannot be rectified. If you choose this option, you will be informed immediately whenever processing is stopped. You can enter multiple recipients separated by a semicolon.

"in the case of a program error"

> Send an e-mail when a program error interrupts the execution.

"in the case of a barcode error"

> Send an e-mail if a barcode could not be read based on the settings or if a document is moved to the error directory.

"after completing the job"

> Send a mail when the job has processed all files in the source directory

"in the case of a mail error "

> Send an email when you pick up documents from the Inbox and an error occurred while retrieving the mail or saving attachments.

Server type SMTP

SMTP Server		SMTP Port	25
Recipient		Subject	
Sender		Encryption	None
☐ Authentication		☐ Show Password	
User		Password	

"SMTP Server

> SMTP Server used for sending mails

"SMTP Port"

> The port to communicate with your SMTP server.

"Recipient"

> Mail address of the r*ecipient.* You can enter several recipients separated by a semicolon.

"Subject"

> If you use the QS-DocumentAssembler on several computers, you can distinguish between senders in the subject line.

"Sender"

> Mail address of the sender

"*Encryption*"

> Encryption, which the SMTP server uses to encrypt mail.*"Authentication"*
> If your SMTP server supports anonymous login, you can disable this option.

"User "

> User name for the SMTP server login.

"Password"

> Password for the SMTP server login.

Server type Outlook356

Tenant ID	
Application ID	
Sender	
Subject	
Recipient	

☐ Outlook365 mailbox is enabled   [Activate new user]

"Tenant ID"

> This is a specific instance in Azure Active Directory (AAD) that stores user accounts and groups. The ID is a unique GUID.

"Application ID "

> Application (GUID) created in the AAD for the DocumentAssembler to access a mailbox from Outlook356.

"Sender"

> Shows the mail address of the sender, after an Outlook365 user is configured to send mail. This field is read-only.

"Subject"

> If you use the QS-DocumentAssembler on several computers, you can distinguish between senders in the subject line.

"Recipient"

> Mail address of the r*ecipient.* You can enter several recipients separated by a semicolon.

"Outlook365 mailbox is enabled."

> Indicates whether a mailbox has already been configured to send mail. This field is read-only.

The configuration of Outlook365 for sending and retrieving mails is described in more detail in Chapter 3.2.2.1 Configuring Outlook365 to send and retrieve mail.

Activate new user

To set up an Outlook365 user for the DocumentAssembler, you need your tenant ID and an application ID. You can find out where to find the tenant ID and how to create an application ID in chapter 3.2.2.1 Configuring Outlook365 to send and retrieve mail.

Select "Outlook365" as the server type and enter tenant ID and application ID and click on the button "Activate new user".

Then the login dialog for Outlook365 appears. There you have to log in to Outlook 365 once with the desired user and password.

If you have already been logged in to Outlook365 on your computer, you may only have to select the user to log in.

Upon successful registration, you will receive the message:

During the sign-in, Outlook365 generates a token, which will be used for the login in the future. This token is stored in the database encrypted with AES-256. Before each access to the mailbox, the token is updated.

Acoustic signal after each job

You have the option to play a sound after each run of the job. You can choose whether you want "no signal", a "simple beep" or play sound file. With the run button, you can test a sound file.

3.2.2 Tab "File/Folder"

Directory and File Names

"Source Directory"

> The source directory is the directory in which the documents to be processed are located.

"Target Directory"

> The target directory is used to store the documents after processing.

"Error Directory"

> If you configure processing to sort out image files without barcodes, the error directory is used. Documents where errors occur during processing are also moved to this directory.

"Log File"

> The processing of each page is documented in the log file. Enter the log file name with complete path. It is possible to integrate date parameters (day, month, year) in the file name. The following parameters are valid:
>
> %YYYY%   four-digit year, e.g. 2014
> %YY%   two-digit year, e.g. 14
> %MMMM%   Month in long form,e.g. "September"
> %MMM%   Month in short form, e.g. "Sep"

%MM%	two-digit month with leading zero, e.g. "07"
%M%	one-digit month
%DDDD%	Day in long form, e.g. "Tuesday"
%DDD%	Day in short form, e.g. "Tu"
%DD%	two-digit day with leading zero, e.g."01"
%D%	one-digit day
%DOY%	Day of the year, e.g. 125

The entry could be e.g.:

c:\docasm\log\Protocol_%MM%%YYYY%.log

With this setting, a new log file would be created monthly.

"Result File"

When all barcode information has to be collected in one file, the file name can be speci-fied here. For creating the file name of the result file, the above date parameters can also be used.

"Temp. Directory"

Temporary files created during processing are stored in this directory.
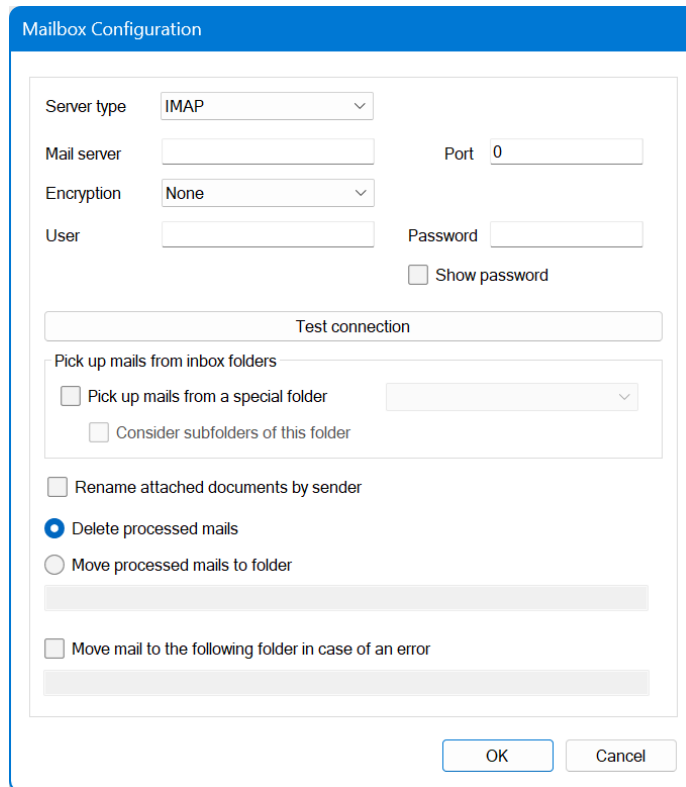
One or more environmental variables can be used for the formation of all directory names. The environmental variables should be masked with a percent sign (%) on both sides. The path specification for the source directory could for example read:

%SystemDrive%\docasm\source\

Collect documents from an e-mail inbox

In addition to reading source documents from a directory structure, the documentation also supports the collection of documents (as an attachment) from e-mail mailboxes.

"Server type"
> IMAP, POP3, Exchange-and Outlook365 are supported.

Servertyp IMAP/POP3/EXCHANGE

"Mail server"
> Name of the server from which the mails are retrieved.

"Port"
> The port to communicate with your server.

"Encryption"
> SSL/TLS, STARTTLS and no encryption are supported.

"*User*"
> User name for the mail server login.

"*Password*"
> Password for the mail server login.

Server type Outlook365

Tenant ID	
Application ID	
User	

☐ Outlook365 mailbox is enabled   Activate new user

"Tenant ID"
> This is a specific instance in Azure Active Directory (AAD) that stores user accounts and groups. The ID is a unique GUID.

"Application ID "
> Application (GUID) created in the AAD for the DocumentAssembler to access a mailbox from Outlook356.

"User "
> Shows the mail address after an Outlook365 user is configured to send mail. This field is read-only.

"Outlook365 mailbox is enabled."
> Indicates whether a mailbox has already been configured to collect mail. This field is read-only.

The configuration of Outlook365 for sending and retrieving mails is described in more detail in Chapter 3.2.2.1 Configuring Outlook365 to send and retrieve mail.

Activate new user

To set up an Outlook365 user for the DocumentAssembler, you need your tenant ID and an application ID. You can find out where to find the tenant ID and how to create an application ID in chapter 3.2.2.1 Configuring Outlook365 to send and retrieve mail.

Select "Outlook365" as the server type and enter tenant ID and application ID and click on the button "Activate new user".

Then the login dialog for Outlook365 appears. There you have to log in to Outlook 365 once with the desired user and password.

If you have already been logged in to Outlook365 on your computer, you may only have to select the user to log in.

Upon successful registration, you will receive the message:

During the sign-in, Outlook365 generates a token, which will be used for the login in the future. This token is stored in the database encrypted with AES-256. Before each access to the mailbox, the token is updated.

Pick up mails from inbox folder

Unless otherwise defined, the mails are retrieved from the inbox by default. However, it is also possible to pick up the mails from special folders in the mailbox.

The first time you open the combo box for the Inbox folders, the default inbox is selected.

"Pick up mails from a special folder "
Enter any folder of the mailbox from which the mails are to be retrieved.

"Consider subfolders of this folder"
Here you can specify whether existing subfolders of the previously defined folder should be taken into account when picking up the mails.
Please note: If you have selected this option, the attached files will be stored in the defined source directory analogously in subfolders according to the names of the special

folders. To process these files with the DocumentAssembler, the option "Include subdi-rectories of the source directory" must be selected.

"*Delete processed mails*"

After picking up the mails and saving the attachments, the respective mail is deleted.

"*Move processed mails to folder*"

If necessary, a fetched mail can be moved to a defined folder after saving the attach-ments. Moving the mails to a folder is only supported by IMAP servers. If the folder does not exist on the server, it is created by the DocumentAssembler. You may have to subscribe to this folder in your mail program to see it.

If file attachments do not match the file filter, they are ignored when saving.

E-mails without attachments are logged in the log file and then deleted or moved depending on the settings.
E-mails are always deleted on POP3 servers.

Subdirectories

"Include subdirectories of the source directory"
> Documents in subdirectories of the source directory can also be included in the processing.

"Preserve directory structure in target directory"
> The processed documents are usually moved directly to the target directory. When files from subdirectories of the source directory are processed, you can preserve the directory structure of these subdirectories in the destination directory.

"Preserve directory structure in error directory"
> The faulty documents are usually moved directly to the error directory. When files from subdirectories of the source directory are processed, you can keep the directory structure of these subdirectories in the error directory.

Source file backup *"Backup source file to"*
> By default, source files are deleted after processing. You have the possibility to copy the source files to the specified directory before processing.

"Overwrite existing files "
> When backing up the master files, it can be specified if existing files should be overwritten. Otherwise in case of overwriting an existing file,.an error message will be displayed.

File Access Safeguards

Reading, moving and renaming the source files may cause access conflicts with other programs.

The most frequent situation: A file is already created by the scanning software, but not yet completey written while the QS-DocumentAssembler is already accessing this file. To avoid such conflicts, the QS-DocumentAssembler, provides two mechanisms:

"Do not process the source file if it is younger than (in seconds)"
> Based on the date of last access, is checked whether the file has reached a minimum age. If this is not the case, the processing is stopped.
> Note: If the files are created on a different computer, the computer clock times may not be different.

"Stop processing if the lock file exists in the source directory"
> Some scanning programs offer to create a lock file during operation. A file with a special name is created at the beginning of the scanning process and deleted at the end of this process. As long as the lock file exists, the files in that directory should be not processed by other applications. QS-DocumentAssembler can check for the existence of a lock file and will stop processing. Please note that only the filename without the associated file path has to be entered.

Adobe PDF

"Import source file with n DPI"
> For barcode recognition, the pages of the PDF documents will be internally changed into image files and the resolution will be determined via the resulting stated value. For secure recognition of the barcode, you should drop below a value of 200 DPI only in

special cases. For values significantly more than 200 DPI, you might have to adjust the barcode parameters, in particular the "light margin" to obtain good recognition.

Source File Filter

If you want to limit access just to defined file types, then you can define a file filter here in the form "*.*", "* tif" or "*.pdf". You can define any number of file filters.

"Filters are regular expressions"
Alternatively, you can define more complex file filters by using regular expressions. In the case of regular expressions, you are for example not restricted to the file extensions and can define more than one filter.

"Sort source files by "
Source files can be sorted by name or by creation date (age).

"Maximum file age in days "
Restrict the source file selection to files that are not below a certain creation date..

For the processing of TIF and PDF documents at the same time you can use the following regular expression: ^.*\.tif$|^.*\.pdf$

Delete Blank Pages

Blank pages can be deleted. This could be useful if e.g. both sides of a page are scanned and the empty back is not needed in further processing. The decision whether a page is empty or not is made based on the degree of blackening of a page.

"Delete pages with a blackening less than n%"
Specify the degree of blackening in percentage terms up to which level you consider a page to be a blank page. You can set the degree of blackening up to two decimals.

"Determine blackening "
Here, you can determine the degree of blackening in percentage terms of the blank pages that are to be deleted. Select the file and page desired (drag and drop is supported) and click on the *<Determine blackening>* button. You can then transfer this value into the field given above.

Document Size Filter

This function has proven to be a very useful addition for some QS-DocumentAssembler customers: When a variety of different document formats are being processed (e.g. even large technical drawings), processing can be accelerated if these large image files are omitted from the barcode search. Instead only A4-sized divider sheets will be scanned. The page size (pixel values) and a tolerance value can be defined, as some scanners deliver image files in slightly varying sizes, due to their "Automatic paper size recognition" function.

All documents outside the specified range will be processed as if they had been scanned for barcodes and no barcodes were were found.

3.2.2.1 **Configuring Outlook365 to send and retrieve mail**

To configure Outlook365 to send and retrieve mail, you need your tenant ID, an application ID and a user with a password to sign in to Outlook365.

To use DocumentAssembler for Microsoft 365 authentication, you must create a custom application in Azure Active Directory and assign specific permissions to it.

Attention: Access to the Outlook365 mailbox via the DocumentAssembler only works with 2-factor authentication enabled.

Where to find your Tenant ID and how to create an Application ID for use with the DocumentAssembler is explained below.

Sign in to the Azure portal (https://portal.azure.com/#home) as an administrator.

There you select the "Azure Active Directory" in "Azure services":
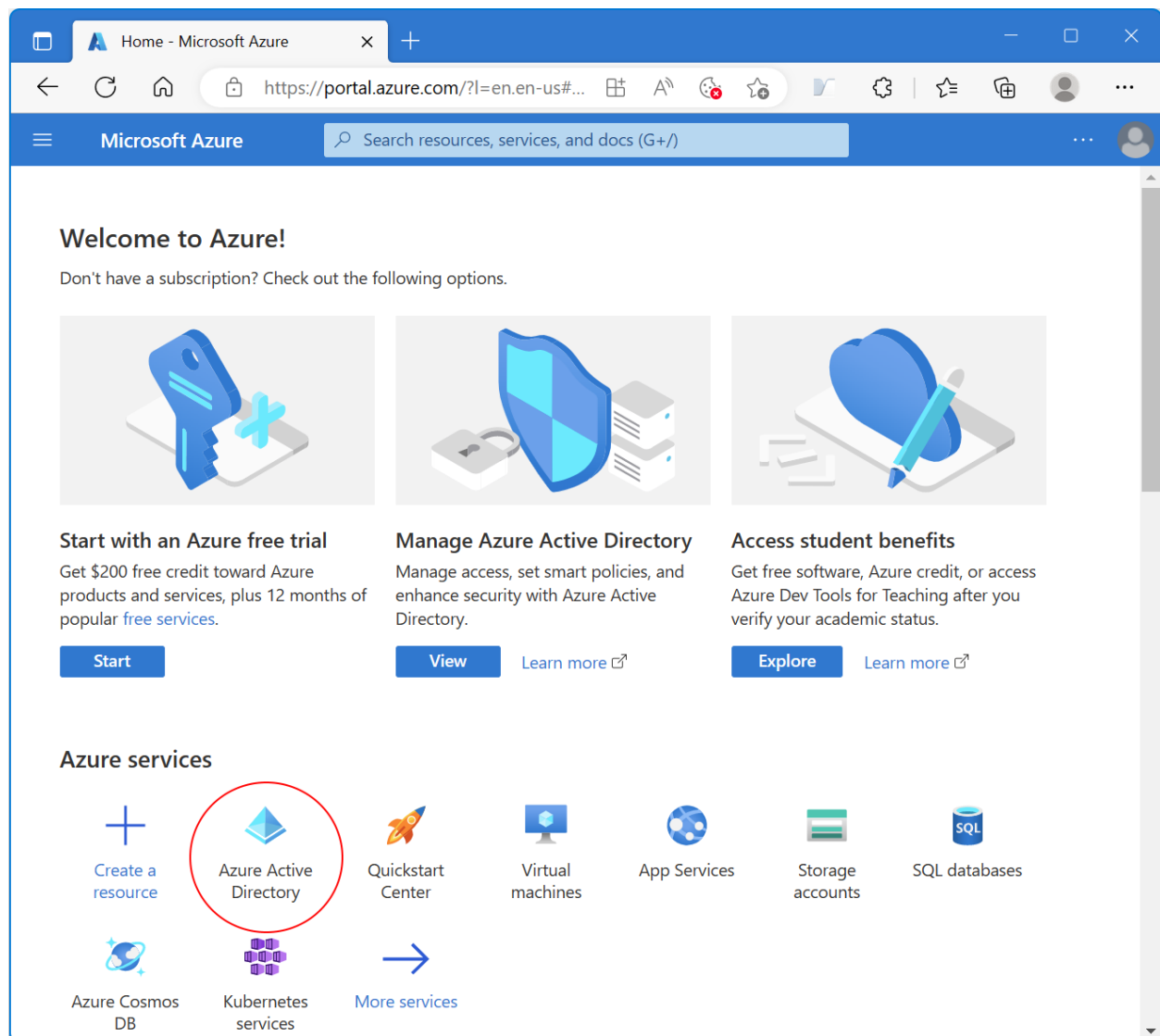
Here you will find your Tenant ID, which you will need later for configuration in DocumenAssembler.

Then go to "App registrations"

Click on "New registration".

Specify a name for your custom application, such as QS-DocumentAssembler, and then click on "Register". You can also assign an individual name here.
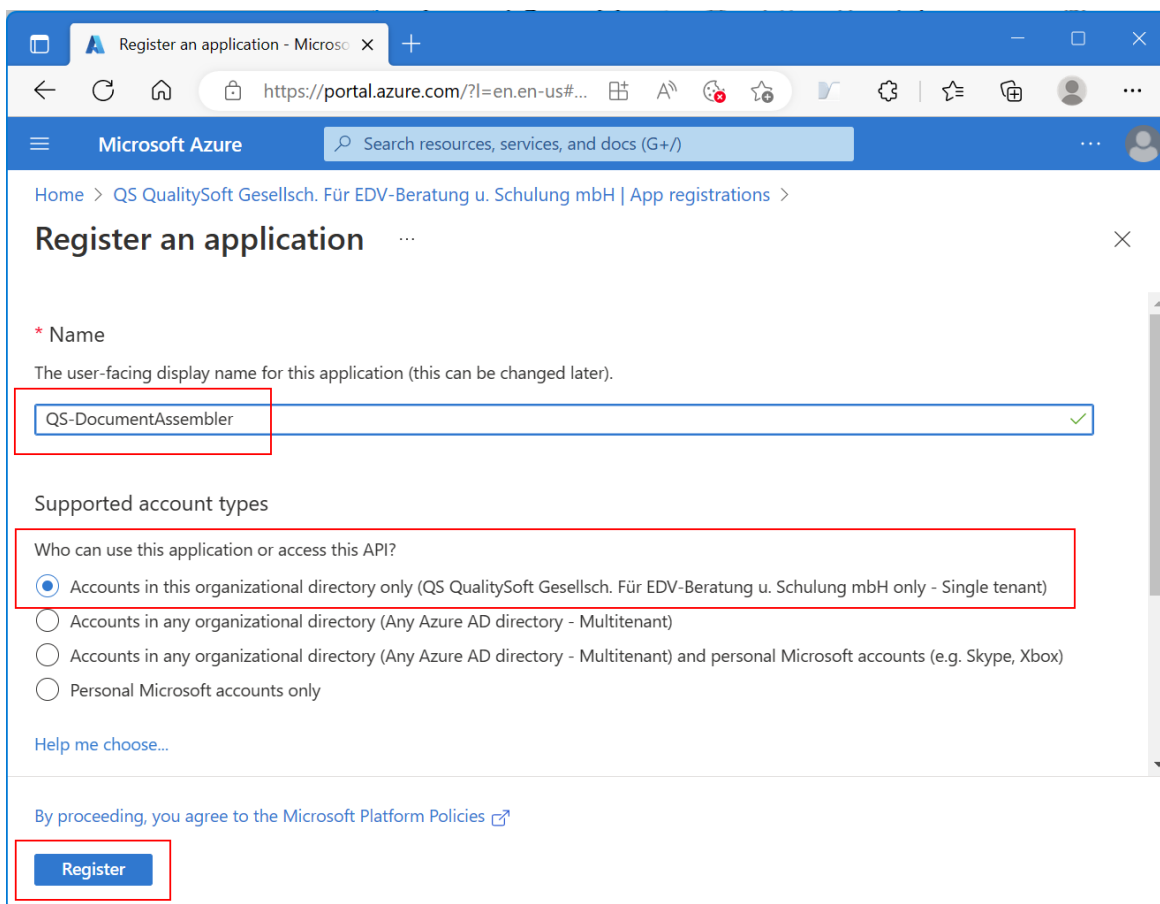
You have created an app in Azure. Here you will now find your Application ID, which you will need later for configuration in the DocumentAssembler.

Now select "Authentication".
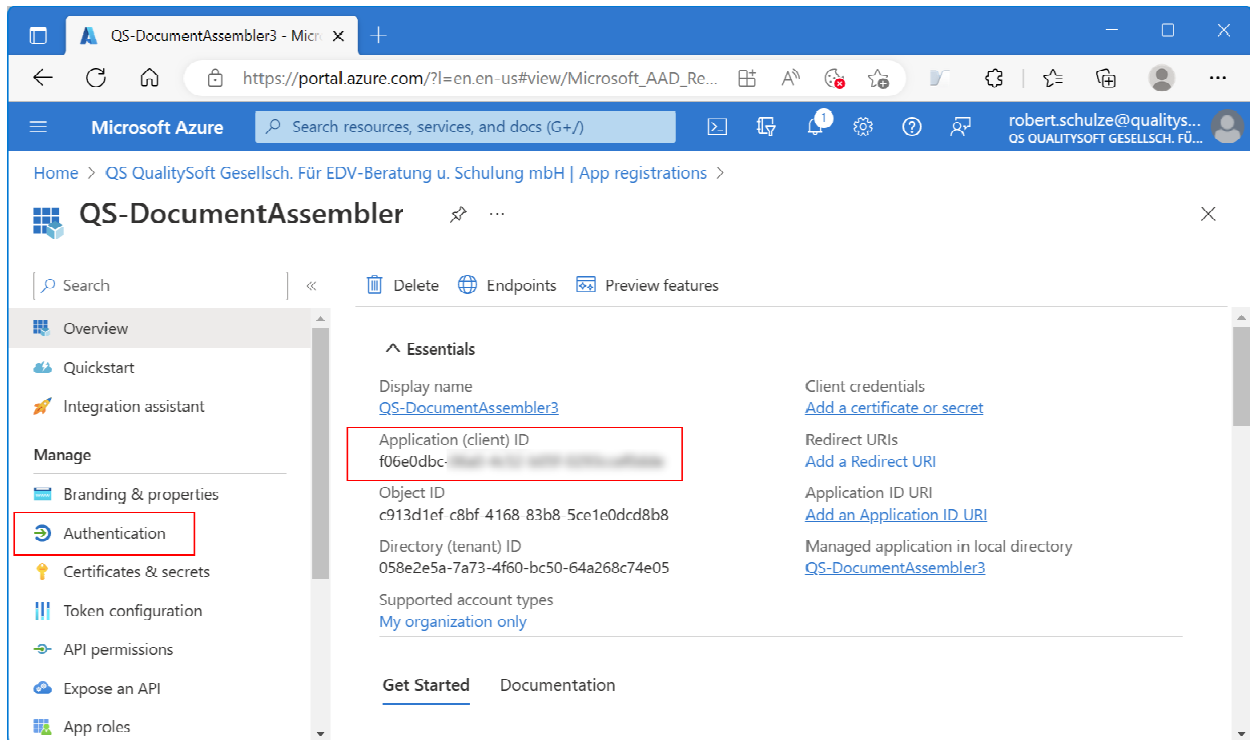
Under Platform configurations, select Add a platform.
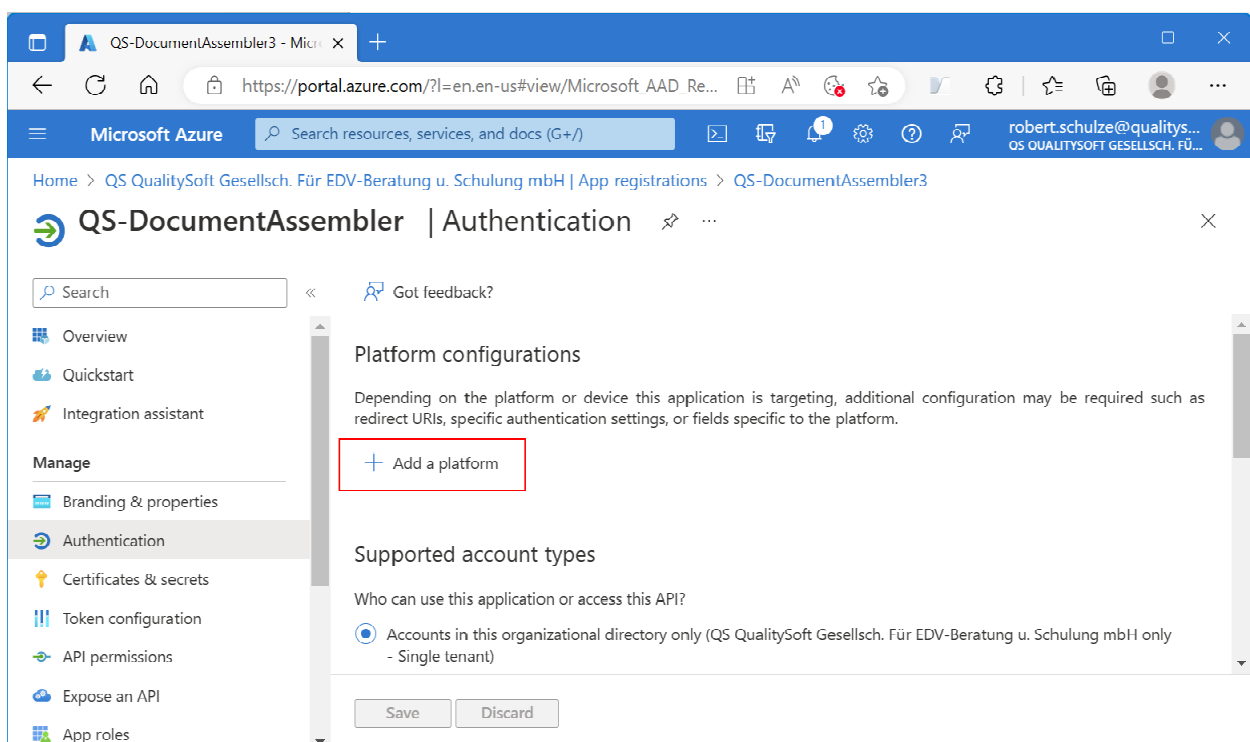
Under Configure platforms, select the tile "Mobile and desktop applications" to configure its settings.

Configure platforms

Web applications

Web

Build, host, and deploy a web server application. .NET, Java, Python

Single-page application

Configure browser client applications and progressive web applications. Javascript.

Mobile and desktop applications

iOS / macOS

Objective-C, Swift, Xamarin

Android

Java, Kotlin, Xamarin

Mobile and desktop applications

Windows, UWP, Console, IoT & Limited-entry Devices, Classic iOS + Android

Configure your app as follows and then save your settings:

Configure Desktop + devices

< All platforms   Quickstart   Docs

Redirect URIs

The URIs we will accept as destinations when returning authentication responses (tokens) after successfully authenticating users. The redirect URI you send in the request to the login server should match one listed here. Also referred to as reply URLs. Learn more about Redirect URIs and their restrictions

☑ https://login.microsoftonline.com/common/oauth2/nativeclient

☐ https://login.live.com/oauth20_desktop.srf (LiveSDK)

☐ msalf06e0dbc-06a0-4c52-b05f-0293ccef0dde://auth (MSAL only)

Custom redirect URIs

http://localhost

Configure   Cancel

Then you need to apply correct API permissions and grant the admin consent for your domain. To do this, click API Permissions, and then click "Add a Permission".

In the API permissions / Add a permission wizard, select "Microsoft Graph" and then "Delegated permissions" .

For "Delegated permissions", you must enable the following permission scopes:

- offline_access
- email
- IMAP.AccessAsUser.All
- POP.AccessAsUser.All
- SMTP.Send

To do this, enter each permission in the search field and then activate it. When you have selected all the permissions you need, you can click "Add permissions".

Finally, you must grant admin consent to the configured permissions:

Configured permissions

Applications are authorized to call APIs when they are granted permissions by users/admins as part of the consent process. The list of configured permissions should include all the permissions the application needs. Learn more about permissions and consent

+ Add a permission   ✓ Grant admin consent for QS QualitySoft Gesellsch. für EDV-Beratung u. Schulung mbH

API / Permissions name	Type	Description	Admin consent required	Status	
∨ Microsoft Graph (6)					•••
email	Delegated	View users' email address	No		•••
IMAP.AccessAsUser.All	Delegated	Read and write access to mailboxes via IMAP.	No		•••
offline_access	Delegated	Maintain access to data you have given it access to	No		•••
POP.AccessAsUser.All	Delegated	Read and write access to mailboxes via POP.	No		•••
SMTP.Send	Delegated	Send emails from mailboxes using SMTP AUTH.	No		•••
User.Read	Delegated	Sign in and read user profile	No		•••

The application configuration is now complete. You can continue configuring mail retrieval in DocumentAssembler using your Tenant ID and Application ID.

3.2.2.1.1   SMTP activation for sending mails via Outlook365

If you also want to send mails via Outlook365 from the DocumentAssembler, you must activate the SMTP authorization for the corresponding user.

Use the Microsoft 365 admin center to enable or disable SMTP authentication for specific mailboxes:

Open the Microsoft 365 admin center under https://admin.microsoft.com/Adminportal/Home and go to Users > Active users.
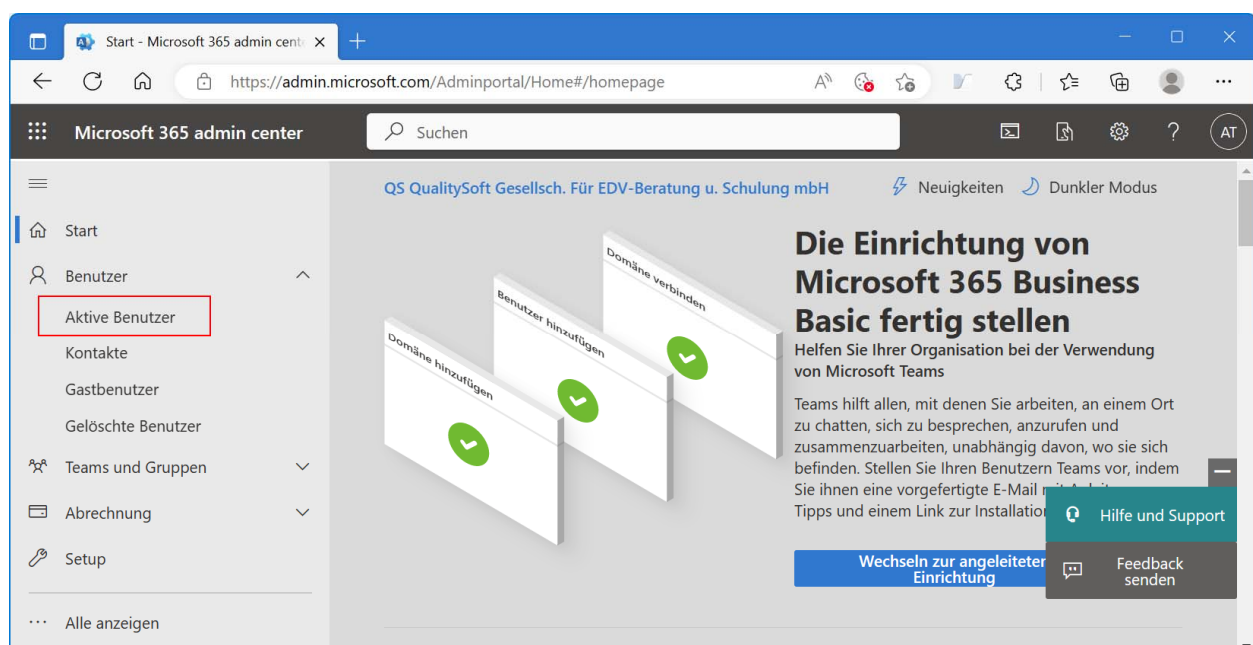
Select the user, and in the flyout that appears, click Mail.
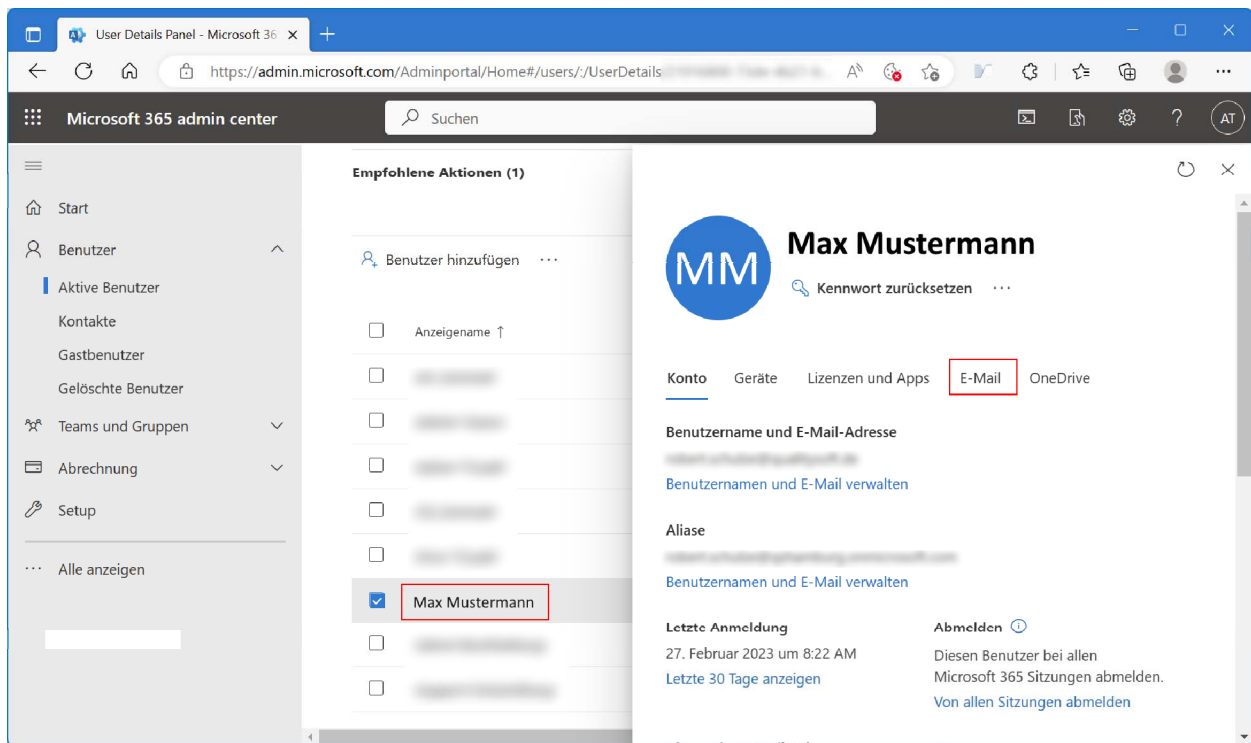
In the Email apps section, click "Manage email apps".
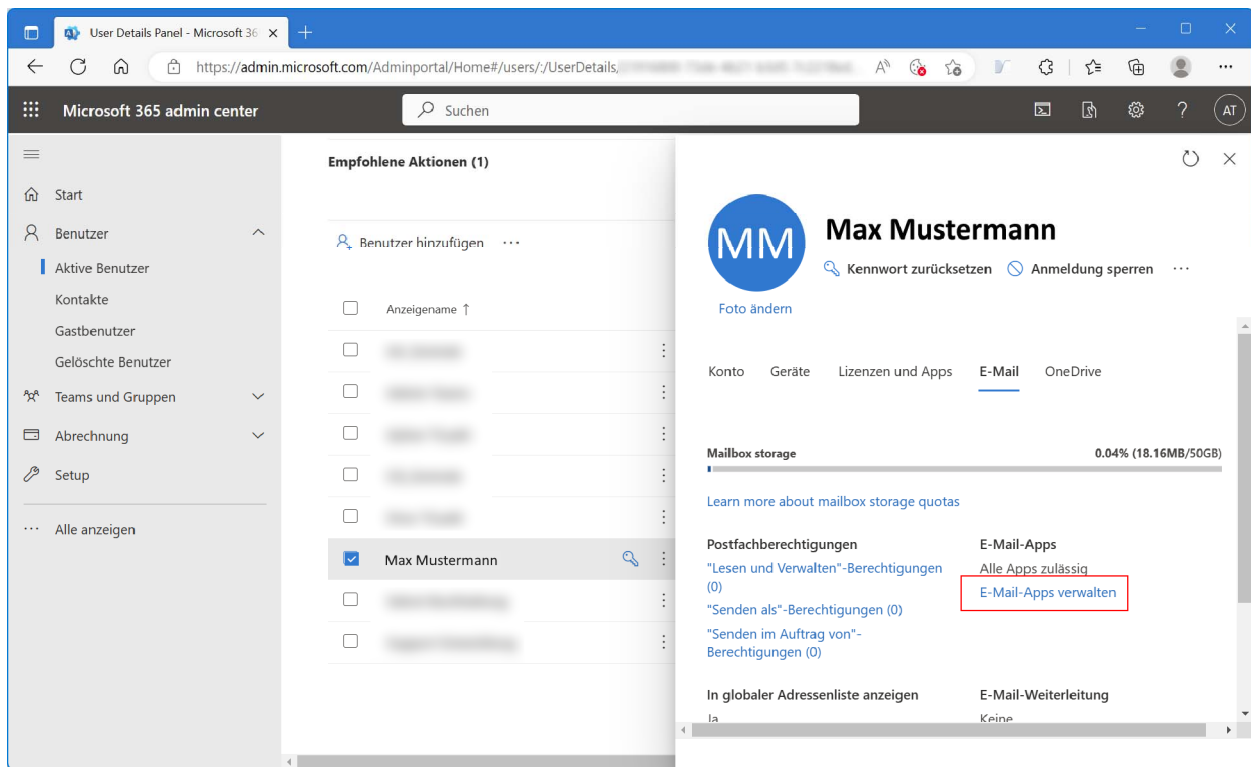
Verify the "Authenticated SMTP" setting and select it if necessary. Then save your changes.
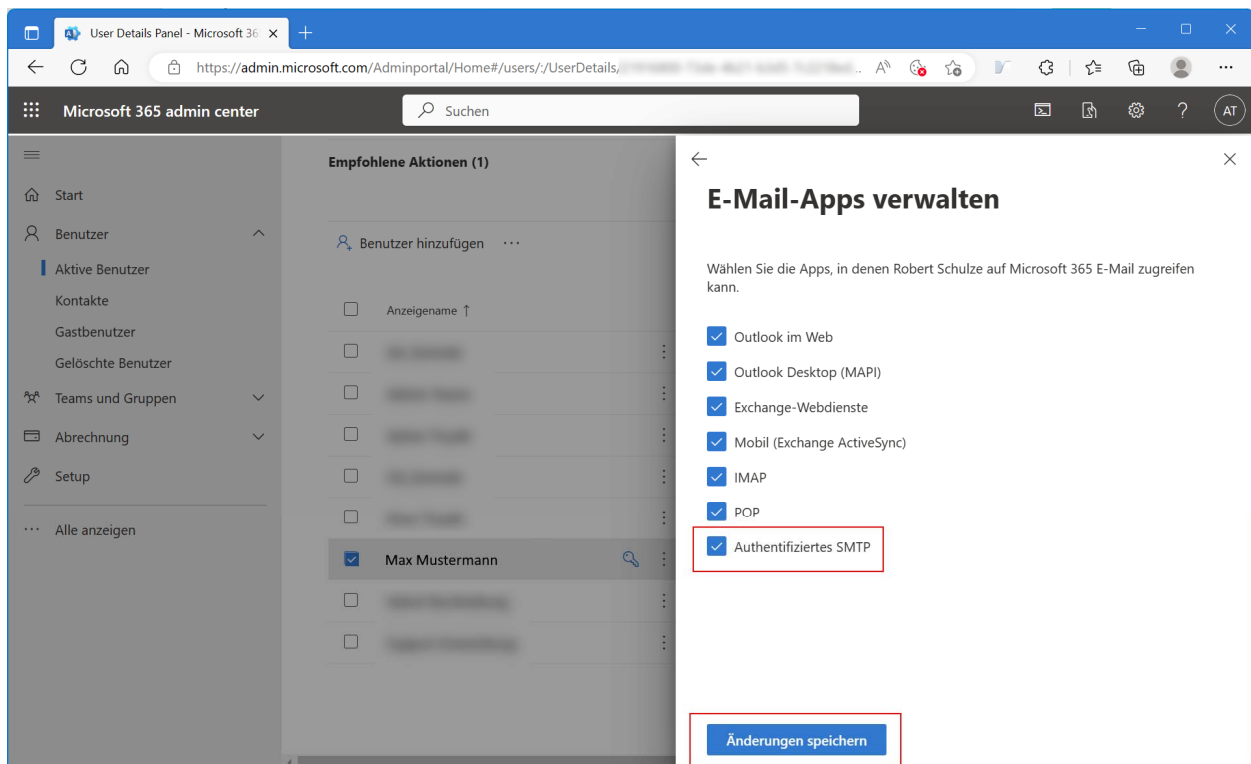
The selected user can now also send mails via DocumentAssembler.

3.2.3 **Tab "Processing"**

QS-DocumentAssembler has two main processing options available, the so called Single Mode and the Stack Mode that can be reached via the Barcode/OCR tab:

Single Mode: "Treat each source image as an individual page"
> There is no interrelation between individual source pages; each page is treated separately. If no barcode is found on a source page, the only option is to sort out this page.

Stack Mode: "Treat each source image as a part of a stack"
> Single- or multiple source documents are considered to be a stack or a part of a stack. Multiple source documents can belong together, at least the first page of a stack is provided with a barcode. The change of the barcode contents implies the beginning of a new stack. Pages without barcode can be assigned as the following pages of the stack.

Within the tab pages, these two modes of processing are distinguished by a **BLUE** and a **GREEN** area.

In addition, the DocumentAssembler supports the processing of ZUGFeRD documents or the simple conversion of source documents into another document format without barcode recognition. This can be configured via the tab "ZUGFeRD" or the tab "Convert".

3.2.3.1 **Processing Single Mode (Blue)**

Pages with Barcode

These pages are always moved to the target directory. If desired, the document can get the name of the recognised barcode.

"Move to target directory"
> The target file gets the name of the source file plus page number.

"Move to target directory and rename by barcode"
> The target file gets the name of the recognised barcode.

"Create documents with all barcodes on a page"
> If there are multiple barcodes on a page, multiple target documents with the barcodes can be created.

Prefix

The file name of the destination files may be additionally provided with a prefix. This prefix is sampled in a dialog at the start of processing. It can be used to create Department IDs or document category IDs (RE = invoice, LI = delivery note), that are not included in the bar-code.

Pages without Barcode

Here is controlled what should happen when no barcode is found.

"Move to error directory"
> The target file is written to the error directory and receives the name of the source file plus page number.

"Move to target directory"
> The target file is written to the target directory and receives the name of the source file plus page number.

"Delete
> No target file will be writen."

"Move to error directory and keep pages together"
> The faulty page is written to the error directory under the name of the source file. For multi-page source files, bad pages from the source file are appended to the file in the error directory so that a single error file is created with all the bad pages.

Please note that a barcode must not be detected although a barcode is attached on the page. This can happen if the source image is very dirty or the barcode settings do not match the barcode on the image.

Barcode Repetition or existing File Names

"Here you can decide what to do if a file name already exists in the target directory.

"Move to error directory"
> If the target file name exists, the page is written to the error directory with sequential number.

"Move to target directory with sequential number"
> The file name is always formed like <BARCODE> .NNNN. <EXTENSION>. Already the first target file has a sequential number NNNN = 0001 in its filename.

"Overwrite exissting file in target directory"
> Any existing file is overwritten.

Multi-page source files (Adobe PDF or multi-page TIFF)

If the source files to be processed are multi-page files, the processing behavior in single mode can be controlled here:

When the source file is processed, it is treated as a unit and will not be separated into individual pages. In the special case when no barcode is found on the first page, the following pages can be sought for barcodes.

"Do not split the source file"
> The source file is treated as a unit and will not be separated into individual pages regardless of whether a barcode is found or not.

"C*heck only first page for barcodes*"
> It is searched only on the first page for barcodes. If no barcode is found, the defined error handling is executed.

"Check every page for barcodes"

> If no barcode was found on the first page, the following pages are sought for a barcode until a barcode is found. If a barcode was found, the search is stopped immediately and the further processing is as if the barcode is found on the first page. If no barcode was found, the defined error handling is executed.

Result File

It can be set whether to create a result file with the recognized barcodes. If a common result file for all image files should be created, the file name results from the file name specified in the "File/Directory" tab. If a single result file should be created for each target file, then the result file will always have the same name as the corresponding target file with a different file extension.

You can choose between various export formats.

"One for all Documents"

- *Standard CSV Format:*
  "BARCODE";"FILENAME"
  "123456789";"c:\docasm\target\123456789.0001.tif"

- *All results, separated by semicolon:*
  c:\docasm\target\dok01.tif;123456789;ABCD;XYZ

- *All results, separated by semicolon, short:*
  dok01.tif;123456789;ABCD;XYZ

"One for each document"

- *Standard with RESULT header per result:*
  >>> RESULT: 1
  >>>
  123456789
  >>>

- *Standard ANSI, max. one result:*
  123456789

- *Standard UTF8, max. one result:*
  123456789

- *All results, separated by semicolon:*
  c:\docasm\target\dok01.tif;123456789;ABCD;XYZ

- *All results, separated by semicolon, short:*
  dok01.tif;123456789;ABCD;XYZ

- *All results with position information:*
  BARCODE;X;Y;DX;DY;ORIENTATION

123456789;155;288;389;123;1

Log File

The processing of each page is documented in the log file. It can also be configured whether all pages or only pages without barcode should be logged. This can be used to generate a list of errors (see Chapter 7.1, Log File).

3.2.3.2   Processing Stack Mode (Green)

Stack Type

"Delete old stack information when the job starts"
> The first page of the document after starting a job can, if it does not have a barcode, be treated as being the following page of a previous stack. In this case, the page would be attached to the last document created from the previous run. If all of the stack information is deleted when the program is started, then this will prevent new documents from being assigned to an "old" stack.

"New source file always starts new stack"
> This option ensures that, when there is a new file, a new stack will be begun in all

cases. With no barcode on the first page, this page will then be processed with an error.

"If no barcode on first page, move document to error directory"
    If no barcode is found on the first page of a source document, the whole file will be
    moved to the error directory.

Every Barcode starts a new Stack

In contrast to the previous definition, in this constellation with every found barcode, a new stack is defined. This serves the purpose of separating files with identical barcodes within the stack-mode, instead of writing them into one target file.

Every change of Barcode Content starts a new Stack

If a page with a barcode has been found, a new stack is created. All following pages with or without barcodes will be attached to this stack, until a new barcode is found that differs from the previous one, which then again starts a new stack.

Identical Barcodes mark the Beginning and End of a Stack

There is always the risk that barcodes are not recognized and consequently two stacks are assembled into one. In order to prevent this, the beginning and the end of a stack can be taged with an identical barcode. If the second barcode cannot be found, the whole area up to the next found barcode will be saved as a file in the error directory.

"Admit one-sided Files with Start-Barcode"
    If a stack consists of only one page, the identical barcode as a marker of the end of a
    stack should be avoided. Onesided files therefore only need one start-barcode.

Barcode starts a Stack and following Pages have identical Barcodes

There is always the risk that barcodes are not recognized and consequently two stacks are assembled into one. In order to prevent this, you can manually adjust the following pages with an identical barcode.

"Barcode-value"
    If on every following page a placeholder barcode with a fixed and consistent value is
    applied, the QS-DocumentAssembler is in the position to distinguish:
    - No barcode at all on the page: Error
    - Defined Barcode on the page: Continuation page
    - Another barcode on the page: Starting new stack

    The defective summarizing of two stacks can be prevented. Please give the specific
    value for the barcode.

Pages with Barcode

"Rename by barcode"
    If a barcode is found, then this barcode will be adopted for the naming of the target file.
    If a barcode occurs alternating with other barcodes multiple times in the source document, then these pages will be attached to any possible existing target documents. To
    prevent this, the target file can be provided with a sequential number. In the event of
    recurring identical barcodes in a source document, the sequential number will be increased.

Additionally, one can also create here the destination files with all of the barcodes found on a page.

"Creating files with all on the page found barcodes"

If several barcodes on one page have been found, the source for every barcode within the target directory will be multiplied.

"Move to Target Directory"
When a barcode is found, the source file name is used for the target file name.

Pages without Barcode

If no barcode was found on a page, this page is always considered following page of an existing stack. If the page is the first page of a source file and can´t be assigned to a previous stack (by clearing the stack information), this page will be moved to the error directory.

"will be appended as a new page to an existing multi-page document "
Pages without barcode are appended to an existing file. The file name results from the treatment of pages with recognized barcode.

"get file name with preceding barcode plus sequential number "
The following pages without barcode will be renamed after the last read barcode and additionally get a sequential number.

"will be moved into the error directory
Pages without barcodes will be shifted into the error directory.

"will be moved into the error directory and kept together"
The faulty page is written to the error directory under the name of the source file. For multi-page source files, bad pages from the source file are appended to the file in the error directory so that a single error file is created with all the bad pages.

Existing target file names

In contrast to former performance that attached the pages of an already existing filenames onto the originally document, you can now determine the conflict behavior of existing filenames by yourself.

"Attach to existing File"
In case of barcode repitition, pages are attached to the existing file.

"Overwriting existing File"
In case of barcode repetition, the existing file will be overwritten.

"Creating Target File with serial Number"
In case of barcode repetition, a target file with ongoing number is created. All following pages will be attached to this file until the stack changes.

'Shifting into Error Directory"
In case of barcode repetition, the new file will be moved to the error directory

Result File

It can be set whether to create a result file with the recognized barcodes. If a common result file for all image files should be created, the file name results from the file name specified in

the "File/Directory" tab. If a single result file should be created for each target file, then the result file will always have the same name as the corresponding target file with a different file extension.

You can choose between various export formats.

"One for all Documents"

- *Standard CSV format:*
  "BARCODE";"FILENAME"
  "123456789";"c:\docasm\target\123456789.0001.tif"

- *All results, separated by semicolon:*
  c:\docasm\target\dok01.tif;123456789;ABCD;XYZ

- *All results, separated by semicolon, short:*
  dok01.tif;123456789;ABCD;XYZ

"One for each document"

- *Standard with RESULT header per result:*
  >>> RESULT: 1
  >>>
  123456789
  >>>

- *Standard ANSI, max. one result:*
  123456789

- *Standard UTF8, max. one result:*
  123456789

- *All results, separated by semicolon:*
  c:\docasm\target\dok01.tif;123456789;ABCD;XYZ

- *All results, separated by semicolon, short:*
  dok01.tif;123456789;ABCD;XYZ

- *All results with position information:*
  BARCODE;X;Y;DX;DY;ORIENTATION
  123456789;155;288;389;123;1

"Log File"
     The processing of each page is documented in the log file

Cover Sheet

If you have a document with no barcode and the stack separation is carried out by using retroactively generated barcodes, then these barcodes could be removed as part of the processing.

"Remove cover sheet from stack"
>   When a barcode is found on a page, this barcode is considered stack name and the page is excluded from processing.

3.2.3.3 **The naming of the target files in the processing of Single and Stack Mode**

The target files are named differently depending on the processing mode and selected processing type. In single page mode, it may come down to a two-level numbering of the target file, in batch mode to no more than a one-stage designation. Please find the format of the target file in the following tables.

Single Mode

Pages with barcode	Format target directory	Format error directory
Move, new to error directory	SourceFileName.tif	SourceFileName.[Page].[seqno].tif
Move, sequential number	SourceFileName.[Page].[seqno].tif	-
Move, overwrite first	SourceFileName.tif	-
Rename, new to error directory	Barcode.tif	Barcode.[lfd.Nr].tif
Rename, sequential number	Barcode.[seqno].tif	-
Rename, overwrite first	Barcode.tif	-
Pages without barcode		
Target directory	SourceFileName.[Seite].[seqno].tif	
Error Directory	-	SourceFileName.[Seite].[seqno].tif
Delete	-	-

The numbering [seqno] indicates the number of occurrences in the target directory, the page reference [Page] refers to the page number in the source document.

Stack Mode

Pages with barcode	Format Zielverzeichnis	Format Fehlerverzeichnis
Move to target directory	SourceFileName.tif	-
Rename to target directory	Barcode.tif	-
Rename to target directory with sequential number	Barcode.[seqno].tif	-
Move to target directory with sequential number	SourceFileName.[seqno].tif	

Pages without barcode		-
Append to multi-page file	Last TargetFileName	
Previous barcode with sequential njmber	Barcode.[seqno].tif	-
First page of a source file and no stack		SourceFileName.[seqno].tif
No barcode on first page move document to error directory		SourceFileName.tif

In stack mode, if there exists a document with the same name then it will always be attached to this document, unless numbering has been explicitly selected for the target directory.

If a numbering is selected, the sequential number is incremented continuously.

3.2.3.4 **ZUGFeRD Processing**

In addition to Barcode and OCR processing, the DocumentAssembler can also recognize and process ZUGFeRD documents. It is possible either to recognize and move ZUGFeRD documents or to additionally extract the XML invoices from the documents.

Basically, it is possible to execute ZUGFeRD processing separately or mixed with Barcode/OCR processing. In the mixed execution, all non-ZUGFeRD documents are then transferred to barcode processing
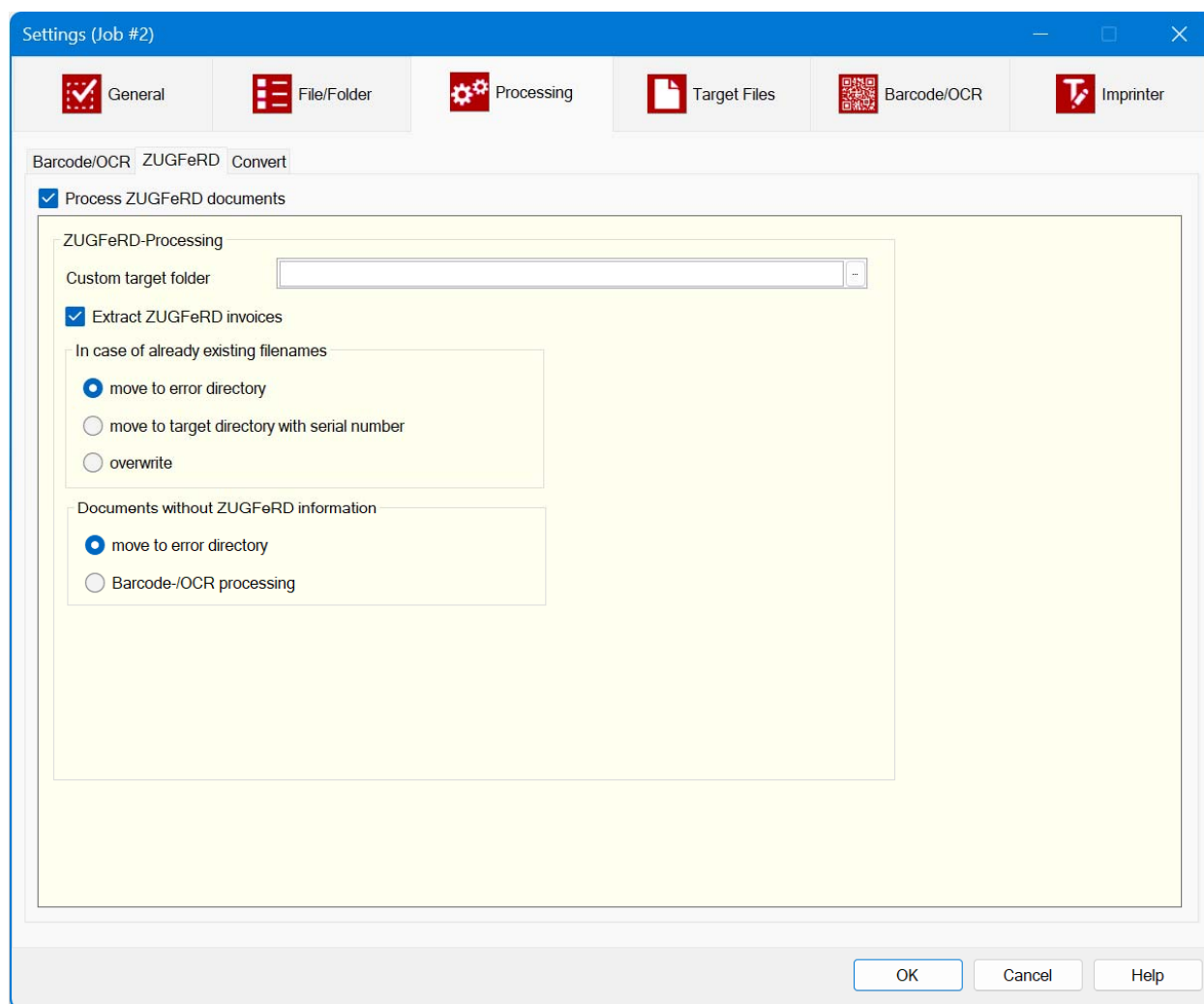
"Process *ZUGFeRD documents"*
> Activate the processing of ZUGFeRD documents.

ZUGFeRD-Processing

"Custom target folder*"*
> A target directory that differs from the default target directory can be specified for ZUGFeRD processing. All ZUGFeRD documents and invoices are written to this directory.

"Extract ZUGFeRD invoices "
> During ZUGFeRD processing, you can either move only the recognized documents to the target directory or use this option to extract the embedded invoice in XML format. The invoice is always named after the document name.

In case of already existing file names

"Here you can decide what to do if a document name or XML invoce already exists in the target directory.

"Move to error directory"
>   If the target file name exists, the page is written to the error directory.

"Move to target directory with sequential number"
>   The file name is always formed like <FILENAME> .NNNN. <EXTENSION>. Already the first target file has a sequential number NNNN = 0001 in its filename.

"Overwrite exissting file in target directory"
>   Any existing file is overwritten.

Documents without ZUGFeRD information

"Here you can decide what to do if no ZUGFeRD invoice was found in the document

"move to error directory "
>   The source file will be moved to the error directory.

"Barcode-/OCR processing"
>   If no ZUGFeRD data is found, the file is left in the source directory for Barcode/OCR processing.

3.2.3.5 **File Format Conversion**

The DocumentAssembler can also be used to convert files from one format to another. Barcode recognition is not performed in this case. It is possible to split multi-page documents into individual pages or to convert the entire document.
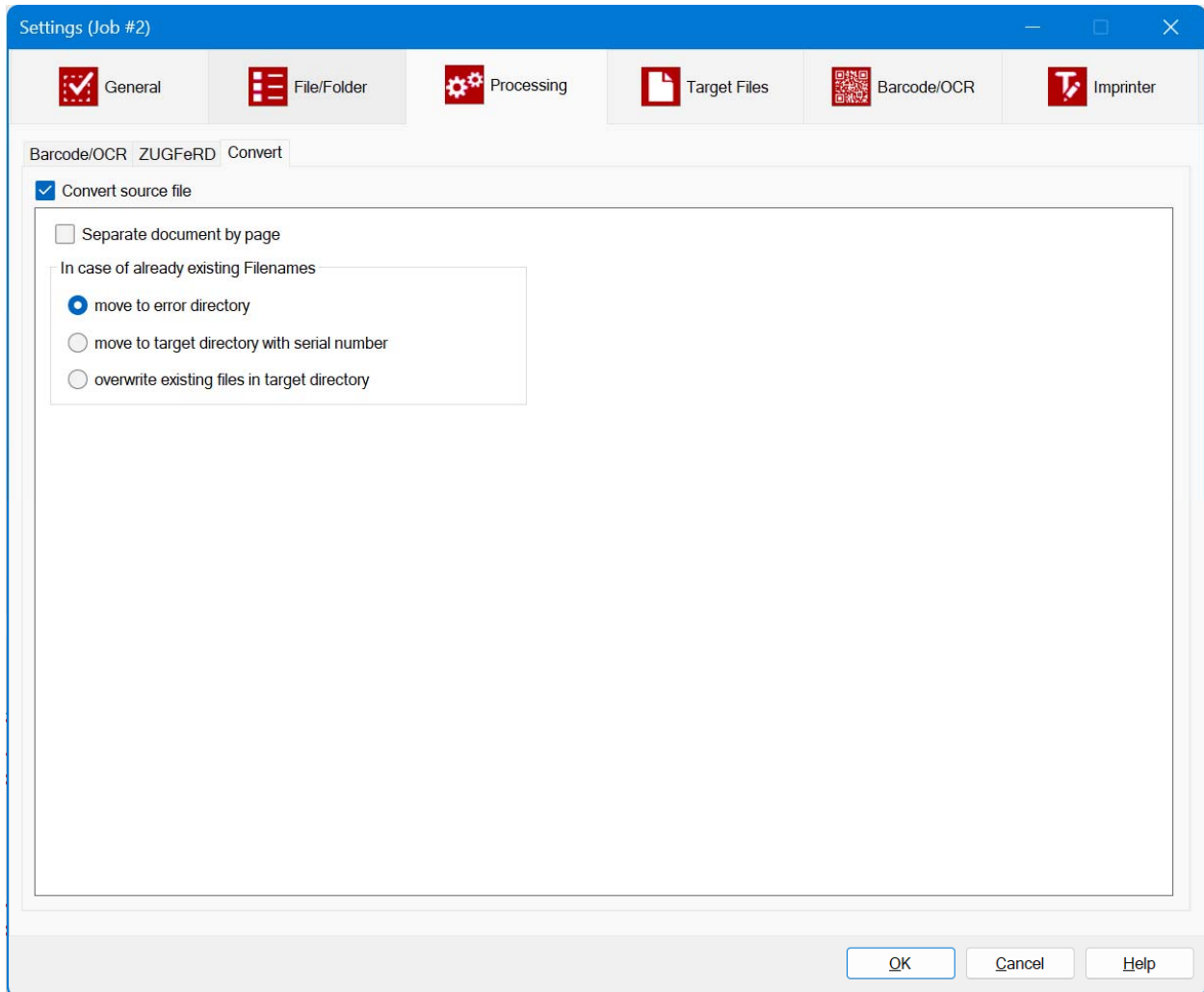
" Convert source file "
> The source file is only converted to the desired format, bypassing barcode recognition.

" Separate document by pages "
> A multi-page source document is separated into separate pages. The target files get the name of the source file plus sequential number.

In case of already existing file names

"Here you can decide what to do if a document name or XML invoce already exists in the target directory.

"Move to error directory"
> If the target file name exists, the page is written to the error directory.

"Move to target directory with sequential number"
> The file name is always formed like <FILENAME> .NNNN. <EXTENSION>. Already the first target file has a sequential number NNNN = 0001 in its filename.

"Overwrite exissting file in target directory"
> Any existing file is overwritten.

3.2.4 Tab "Target Files"

In this register card, it can be defined whether and how the target files should be converted into another file format after the recognition of the barcode. The standard setting is that the file format shall remain the same as the source files.

The QS-DocumentAssembler supports BMP, JPG, GIF, PNG and TIF formats as a target format for raster graphics plus the Adobe PDF format

When might the conversion of format be of use?

- If your source files are of differing file types, for example, TIFs captured using a document scanner and JPGs captured by digital cameras, then converting them can ensure that the target files are consistent. This is often an advantage for any subsequent applications, for example for archive systems.

- If you scan with a fairly high resolution to guarantee reliable recognition of the barcode, then by converting the files, you cannot only change the file format, you can also reduce the resolution of the image.

- You can save the images initially after they have been scanned with a low compression rate. As a result, the loss of information loss will be kept low and barcode recognition can be carried out reliably. After they have been processed, you can then save the images in a highly compressed form, without any concern for the effects that this will have on the already completed barcode recognition.

Format settings for target files.

Depending on the target format some options available on the dialog are not available.

"Convert to the following format"
Selection of the desired target format, either a named raster format or Adobe PDF. Depending on the format selected, the following options might not be available.

"Write faulty pages also in target format"
By default, faulty pages are written to the error directory in source format. With this option you have the possibility to convert these pages into the defined target format.

Target format raster graphics (BMP, JPG, GIF, PNG, TIF)

The QS-DocumentAssembler supports the raster formats BMP, JPG, GIF, PNG and TIF as a target format for bitmaps, where only the TIF format supports multi-page documents. Therefore, in Stack Mode, only the TIF format can be selected as a raster image target format.

"Bits Per Pixel"

> The colour depth for the target files is selected here. If "unchanged" is selected, the colour depth from the source file will be adopted, unless this colour depth is not supported by the selected format. Then, it will be chosen automatically: If possible, the next highest colour depth that is supported. Otherwise, the next lowest.

"Compression"

> If available, the compression format of the file type can be set.

"JPEG-Quality"

> For JPEG compression, for JPG formats, values of between 0 (high compression) and 100 (low compression) can be selected.

Target format PDF / PDF/A / PDF/A - OCR

During the explicit transformation into a PDF document also PDF/A documents can be generated. During the transformation an OCR process on the whole page can take place, which automatically stores the recognized text in the PDF document. Writing PDF/A documents without OCR is also possible.

Please note that the OCR process is very time-consuming and the processing duration can increase thereby substantially.

"Graphics Fomat in Adobe PDF documents"

> You can select from a list of common raster file formats which format is to be embed-ded in a PDF document. Depending on the format, the following parameters are availa-ble.

"Bits Per Pixel"

> The colour depth for the target files is selected here. If "unchanged" is selected, the colour depth from the source file will be adopted, unless this colour depth is not sup-ported by the selected format. Then, it will be chosen automatically: If possible, the next highest colour depth that is supported. Otherwise, the next lowest.

"Compression"

> If present, the compression format of the file type can be set.

"JPEG-Quality"

> For JPEG compression, for JPG formats, values of between 0 (high compression) and 100 (low compression) can be selected.

"Resolution of target file"

> Enter with which resolution the target file should be written. In the case of a value that is different to the source DPI, the graphic will be scaled appropriately.

Signing PDF-Files digitally

You can sign your PDF-files during processing with your own digital certificates. Your files can still be changed afterwards. As a consequence, the signature will be invalid. Files can be signed more than once. The changes to the files can thus be secured and recognized.

A test-certificate is attached to the DocumentAssembler, which you can use for your test. You can find this test-certificate in the "\certificate" folder. The password for this certificate is "demodemo".

"Certificate"
> Choose your certificate.

"Password"
> Enter the password for your certificate here. The password will be saved encrypted.

OCR

When creating searchable PDF documents, you can choose between the Tesseract engine and our new standard engine.

"OCR-Engine"
> You can choose between the faster Tesseract engine and the much more advanced default engine.

"Language file PDF/A – OCR"
> For the text recognition of scanned documents a language file ensures that the country-specific features of the text are preserved.

"Rotate pages in text direction "
> If pages are e.g. scanned upside down, the documents are rotated in text direction.

Image Modification

"Zoom Factor(Percent)"

> For all values other than 100, the output file is resized using the specified Zoom Factor (e.g. 200 -> double x and double y size, 50 -> reduce to half).

"Invert Destination File"

> The target file will be inverted.

Variable Target File Name

If you want to include additional information in the file name that deviates from the standard, you have a number of parameters to choose from:

%SOURCEFILE%	Original file name
%BARCODE	The first recognized barcode
%BARCODEn	The following recognized barcodes
%BC_REGEX%	Result of a regular expression applied to the first barcode.
%BC_REGEX_REPLACE%	Result of a regular expression replaced in the first barcode
%PAGE_NO%	The current page number of the source document
%PAGE_NOn%	The current page number of the source document, formatted to n digits, e.g. %PAGE_NO3% corresponds to "001"
%PAGE_COUNT%	The number of pages of the target document
%PAGE_COUNT n%	The number of pages of the target document,formatted to n digits, e.g. %PAGE_COUNT3% corresponds to "001"
%SEQNO%	Number of occurrences of a file name in the target directory (sequential number)
%SEQNOn%	The sequential number formatted to n digits, e.g. %SEQNO4% corresponds to "0001"
%YYYY%	four-digit year, e.g. 2014
%YY%	two-digit year, e.g. 14
%MMMM%	Month in long form,e.g. "September"
%MMM%	Month in short form, e.g. "Sep"
%MM%	two-digit month with leading zero, e.g. "07"
%M%	one-digit month
%DDDD%	Day in long form, e.g. "Tuesday"
%DDD%	Day in short form, e.g. "Tu"
%DD%	two-digit day with leading zero, e.g."01"
%D%	one-digit day
%DOY%	Day of the year, e.g. 125
%HH%	two-digit hour
%H%	one-digit hour
%mm%	two-digit minute
%m%	one-digit minute
%SS%	two-digit second
%S%	one-digit second
%AM/PM	AM/PM
%DATE_SHORT%	The current date formatted by the country settings of the OS
%DATE_LONG%	The current date formatted by the country settings of the OS
%TIME	The current time formatted by the country settings of the OS

The target file name must contain at least one of the variables "%SOURCEFILE" or "%BAR-CODE". All variables must be defined in upper case letters.

All barcodes found on a page can be used for creating the variable destination file name using the variable %BARCODEn%, e.g. %BARCODE2%, %BARCODE3%.
The barcodes are internally listed in the order in which they are found in the document. This order does not necessarily match the real order of the barcodes in the document.
Therefore, the barcodes should be sorted according to their position (see chapter 3.2.4.1, Barcode Settings).

You can add any additional text to the file name, provided the characters used are permitted for file name creation.

Note: Not all of the parameters may be useful for all operation mode options (especially time based parameters like seconds in Stack Mode).

"Ignore less number of barcode parameters than barcodes"
> If fewer barcodes are found than barcode parameters are specified in the variable target file name, the file is moved to the error directory because the target file name cannot be completely built. You can use this option to generate the file name even if fewer barcodes are found than there are barcode parameters. The excess barcode parameters are then removed from the target file name.

Variable Target Directory

The target directory can variably be defined with all above-mentioned parameters. The stated variable part will be attached to the defined target directory which you can find under "Path- and Filename".

Variable Barcode

From the barcode, through different parameters, a specific area for the variable target file-name, the vriable target directory and the imprinter can be cut out.

BC_SUBSTR	Substring from a specified position with a certain length.
BC_STARTSWITH	Substring from a given string with a given length
BC_ENDSWITH	Substring before a given string with a given length
BC_BETWEEN	Substring between two defined strings
BC_REPLACE	Replace substring in barcode.
BC_REGEX	Result of a regular expression applied to the first barcode.

Examples:

%BC_SUBSTR(5)%	All charcters from position 5 of the barcode to the end
01234==56789012==	
%BC_SUBSTR(-5)%	All characters starting at the fifth last character to the start

<mark>01234567</mark>89012	
%BC_SUBSTR(0,5)%	5 characters from position 0 (start) of the barcode.
<mark>01234</mark>56789012	
%BC_SUBSTR(10,-5)%	5 characters backwards from position 10
01234<mark>56789</mark>012	
%BC_SUBSTR(-1,-5)%	5 characters backwards from the last character (-1)
0123456<mark>89012</mark>	
%BC_SUBSTR(-5,4)%	4 characters from fifth last position
0123456<mark>8901</mark>2	
%BC_STARTSWITH(XXX)%	All characters from "XXX" to the end
012XXX<mark>3456789012</mark>	
%BC_STARTSWITH(XXX,5)%	5 characters from "XXX"
012XXX<mark>34567</mark>89012	
%BC_ENDSWITH(XXX)%	All characters before "XXX"
<mark>012345678</mark>XXX9012	
%BC_ENDSWITH(XXX,5)%	5 characters from XXX backwards
0123<mark>45678</mark>XXX9012	
%BC_BETWEEN(XXX,YYY)%	All charcters between "XXX" und "YYY"
012XXX<mark>34567890</mark>YYY12	
%BC_REPLACE(XXX,YYY)%	Replace "XXX" with "YYY"
012<mark>XXX</mark>3456789012 => 012<mark>YYY</mark>3456789012	
%BC_REPLACE(3,5,YYY)%	Replace from position 3 in length 5 with "YYY".
012<mark>XXX34</mark>56789012 => 012<mark>YYY</mark>56789012	
%BC_REGEX(ABCDd*(?=_*))%: Processing of structurally similar barcodes with different lengths. All barcodes that start with "ABCD" and have any number of digits up to the underscore. If there is no underscore, it reads to the end.	
<mark>ABCD12345678_</mark>99 <mark>ABCD123456789</mark>	
%BC_REGEX_REPLACE(AA[3,4],XX1)% Replace strings found via a regular expression with your own string. We are looking for the strings AA3/AA4 which are replaced by XX1.	
<mark>AA3</mark>BB6CC7 => <mark>XX1</mark>BB6CC7 <mark>AA4</mark>BB5CC6 => <mark>XX1</mark>BB5CC6	

Variable source file name

The source filename can also be treated with SF_SUBSTR(). The procedure is analogous to the above illustrated approach.

When using the parameter, it should be noted that the barcode, respectively the source file-name features the same lengths that are used in the parameter. If this is not the case, an error will be logged and the processing will be stopped.

Special character substitution in barcodes

If the recognised barcode is to be used as file name, then the barcode must not contain any special characters, as this would prevent a renaming of the file. To avoid this problem in file renaming, any special characters that might appear in a barcode can be replaced by user-defined characters.

"Replace non-printable characters by"

    For the renaming by barcode the ASCII values 00 to 31 are replaced by a freely defina-ble replacement character.

3.2.5 **Tab "Barcode/OCR"**

The title tells you which license is present on the system. The demo mode is additionally marked by a red "DEMO" sign in the lower left part of the dialog only when a DEMO license was found.

In addition to the recognition of barcodes on documents, the DocumentAssembler also supports recognition runs for the recognition of text content (OCR) on documents whose results, such as the contents of barcodes, can be used for further processing of documents .

The configuration is carried out in separate barcode and OCR recognition runs.

3.2.5.1 **Barcode Recognition**

In this settings dialogue, you configure the barcode recognition via the recognition runs. A recognition run contains the parameters that are relevant for the barcode search with which the barcode search will be carried out.

In most cases, a single recognition run will be sufficient. Occasionally, it can happen that the quality of the source files is very variable, so that recognition has to be carried out using differing parameters. If you are not able to model these different parameters in one recognition run, then simply create an additional recognition run.

If you want to search for both linear and 2D barcodes in the documents, or if you want to recognise two different 2D barcode types then you will have to configure these via two recognition runs.

Different linear barcode types can be read in without problems in one single recognition run. If you have created more recognition runs for certain search processes but do not require all of the runs, then you can deactivate individual recognition runs via the checkbox before the name of the run.

The editing of recognition runs can be performed via the vertical button bar to the right of the list of recognition runs.

A new recognition run is created. The recognition run is preset with default values and the dialog for the barcode settings is opened.

Opens the dialog for editing the selected recognition run.

Duplicate the currently selected recognition run.

Click this button to delete the selected recognition run.

Use these arrow keys to adjust the execution order of recognition runs.
Select an entry in the list to and move it by selecting the arrow keys.

"Execute further runs only when no barcode was found"
   If this option is checked, the recognition process for the current page in progress is stopped right after the first recognition run has delivered at least one barcode result. The remaining recognition runs will then not be executed for this page.
   If this option is not checked, all recognition runs will always be executed.

"Sort barcode results based on the position on the document (uppermost first)"
   The barcodes are internally listed in the order in which they are found in the document. This order does not necessarily match the real order of the barcodes in the document. Without this option, the recognized barcodes are reported in the same order in which they were found. This order is defined by the recognition run order (all results of the 1st recognition run followed by all results of the 2nd recognition run and so on). The order of the results within one recognition run is undefined (in most cases it will be from top to bottom, but sometimes the sequence may be different).
   If you choose this option, the barcode results will be sorted after the last recognition job has finished. The topmost barcode will be reported first. There is no analysis of whether the page is possibly rotated.

"Rotate documents in the direction of the barcode"
   If documents have accidentally been scanned upside down, the document can be rotated in the direction of the barcode.

"Only rotate pages with barcode"
   Using this option, you can restrict that only pages with barcodes are to be rotated. Pages without barcode are not affected.

3.2.5.2 **OCR Recognition**

On the "OCR" tab, you configure OCR recognition using recognition runs. A recognition run contains the parameters relevant for the OCR search.

The recognition runs are processed via the vertical button bar on the right next to the list of recognition runs.

A new recognition run is created. The recognition run is preset with standard parameters and the dialog for the OCR settings is opened.

Opens the dialog for editing the selected recognition run.

Duplicate the currently selected recognition run.

With this button you delete the selected recognition run.

You can use these arrow buttons to adjust the execution order of the recognition runs. Mark an entry in the list and then move it by pressing the arrow keys.

" Execute further runs only when no text was found"
> If this option is checked, the recognition process for the current page in progress is stopped right after the first recognition run has delivered at least one result. The remaining recognition runs will then not be executed for this page.
> If this option is not checked, all recognition runs will always be executed.

" Rotate documents in the direction of the text"
> If documents have accidentally been scanned upside down, the document can be rotated in the direction of the text.

" Only rotate pages with found text"
> Using this option, you can restrict that only pages with found text are to be rotated. Pages without found texta are not affected.

3.2.5.3 **Barcode Filter**

Documents may contain barcodes from different sources. By using a barcode filter, the barcodes which are not of interest can be suppressed. If a barcode does not match the filter criteria, it will be filtered and ignored in further processing.

If a barcode corresponds to just one of several possible defined filters, then it will be added to the list of recognised barcodes. Filters will therefore be connected internally with "or". Any other barcodes will not be considered in further processing.

Tab "Create Filter"
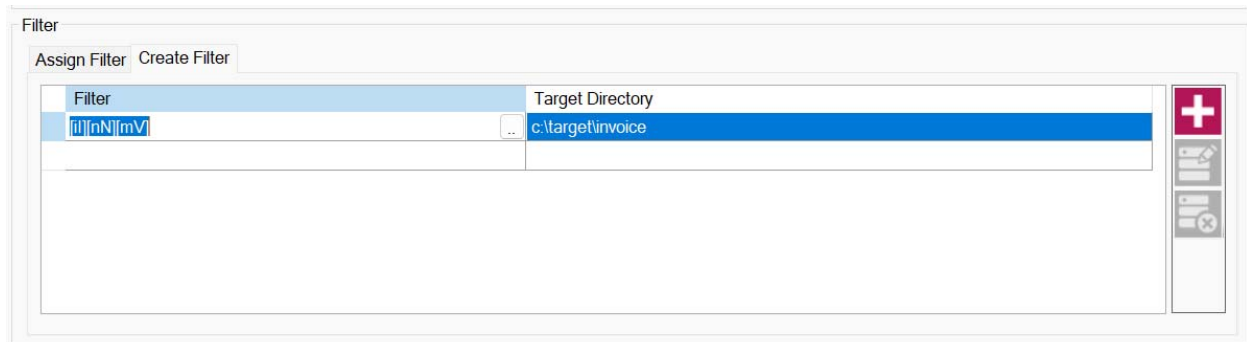
One or more barcode filters can be added by using the "Create filter" tab. In this way, every filter can be assigned to an individual target directory for the creation of the target document. This individual target directory then overwrites the target directory that has been defined on the "File/Directory" register card.

Inputting of the filter and of the target directory is performed by direct data entry into the list of the barcode filter. In addition, a button on the right edge of the respective input field can be used to open a window for editing a longer regular expression for the filter or a selection dialog for the directory.

The selected filter can be deleted or edited via the button bar at the right-hand side of the filter list.

It is recommended to reduce the number of recognisable barcodes by selecting barcode parameters first (typ, length and so on) and only then use barcode filters to limit the number further. The filter have to be assigned to individual recognition runs by using the "Assign Filter" tab.
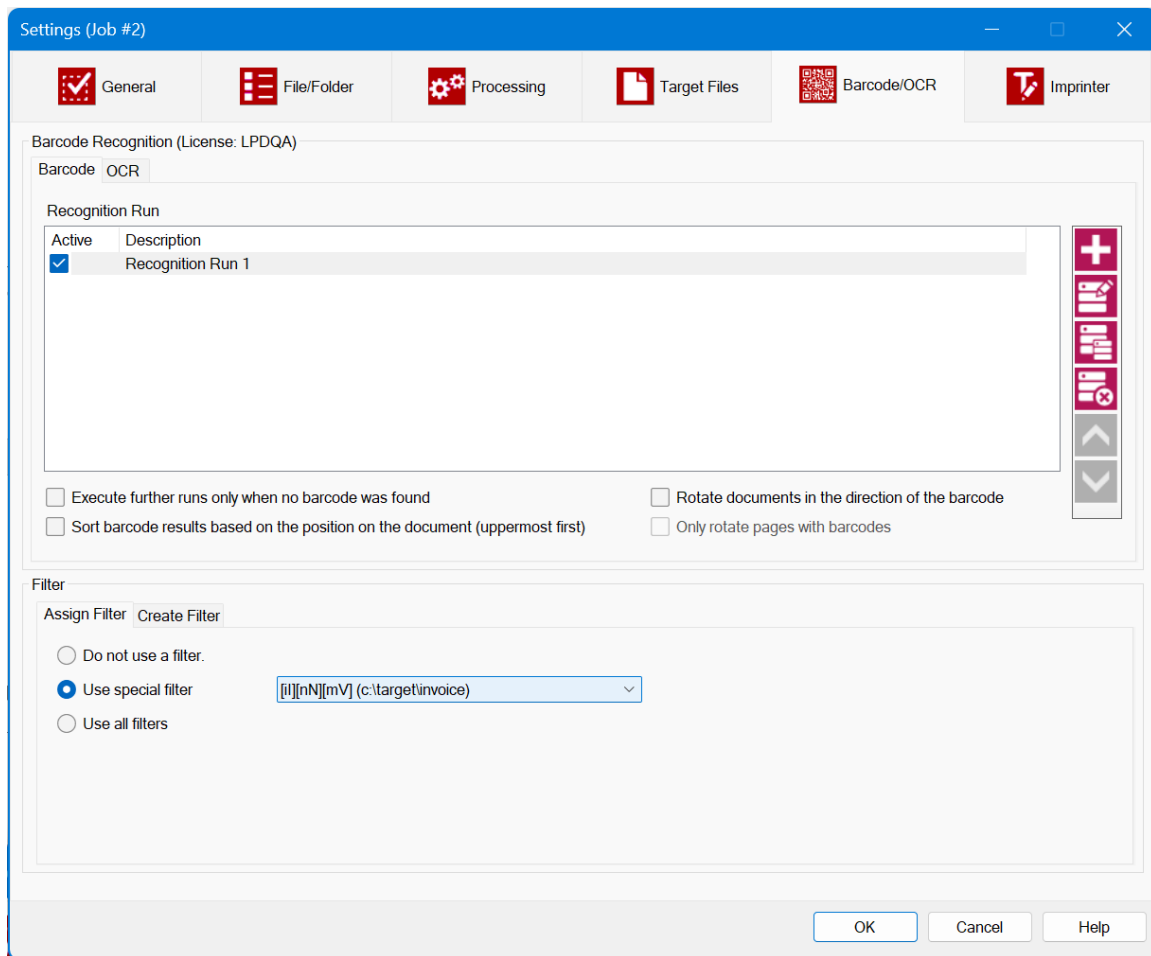
Assign Filter

The mere creation of barcode filters is not sufficient. To use the filters, they must be assigned to a recognition run. To perform this, please select the desired run in the list of recognition runs to which the filter should be assigned. You have here the opportunity to assign a special filter that you can select from the corresponding combo box to the recognition run, or to assign all existing filters to the recognition run.

"Do not use a barcode filter"
    No filters are used.

"Use special filter"
    If you have created barcode filters, you can assign the desired filter from the combo box.

"Use all barcode filters"
    All defined filters are used

Using Regular Expressions

The barcode filters in QS-DocumentAssembler have to be created using regular expressions, based on the rules of the Microsoft .NET Framework. For further information about the regular expression in the Microsoft.NET Framework please see https://docs.microsoft.com/de-de/dotnet/standard/base-types/regular-expression-language-quick-reference

The defined filters support a pattern recognition, whereby placeholder characters are permitted. This pattern recognition provides a very wide range of opportunities.

The following table shows some selected placeholder characters that you can use, as well as the number of digits or character strings for which an equivalent will be recognised.

All characters are case sensitiv.

Sign in Pattern	Meaning
.	Any single character.
*	Matches the previous element zero or more times.
+	Matches the previous element one or more times.
\d	Any decimal digit. (0-9)
\D	Any character other than a decimal digit.
[character group]	Matches any single character in character_group.
[^character group]	Matches any single character that is not in character_group.
?	Matches the previous element zero or one time.
^	The match must start at the beginning of the string.
$	The match must occur at the end of the string.
(?i)	Ignore upper and lower case for all subsequent characters.
(?-i)	Case-insensitive notation is no longer ignored.
|	Matches any one element separated by the vertical bar (|) character.

A group of one of more characters (list of characters) in square brackets ([aei]) can be used to detect an equivalent with every individual character in the expression, whereby only one character from the list of characters has to match. If you want to search for a fixed character string, then enter the characters without the square brackets.

Alongside a simple list of characters that are contained within square brackets, the list of characters can specify a range of characters with a hyphen ([a-i]) that separates the upper and the lower values of the range from each other. If [A-Z] is used in the pattern, then for example, an equivalent will be detected if the corresponding character position in the expression contains an upper case letter in the range from A to Z. A number of ranges can be specified within the brackets without separator characters. For [a-zA-Z0-9] for example, equivalents (for one character) with all alphanumeric characters will be detected.

There are some additional important rules for the pattern recognition;

- A negation symbol (^) at the beginning of the character list ([^A-Z]) means that an equivalent will be detected if a character in the expressions is found that is not contained in the character list, here for example no upper case letters.

- The hyphen (-) is used to label a range of characters.

- If a range of characters is specified, then the characters must be specified and sorted into ascending order (A-Z or 0-100). [A-Z] is a valid pattern, but [Z-A] is not valid.

If you would like use characters with special importance, so-called meta-characters (e.g. full stops, hyphens, question marks etc.) as their own values, then they must be prefixed with a backslash. This prefixing of an escape character to overrule their meta-meaning is called "escaping". For example, to search for a full stop, you would use "\.". If you are searching for a backslash then you must also preface the backslash with another backslash "\\".

Examples:

Filter	Match	Description
^RE	RE0001/REC	The beginning of a string must match "RE"
RE$	0001RE	The end of a string must match "RE"
RE	RE001/ABRE001/001RE	Find the letter combination "RE" anywhere
[rR][eE]	re01/AbRe01/01rE	The character string may also contain small letters.
.\d\d\d	A123/0123/0x0123/AA123x	Any character followed by three digits.
\d[AB]	1A4711/2B/012B	Any digit followed by the letter "A" or "B"
^1234$	1234	Exact match from start to finish
^[0-8]*$	12345678/456	Only digits from 0 to 8 in any frequence are allowed.
^[^A-Z]*$	abc0123	All characters except uppercase characters are allowed at any length.
\D*\.\d*	A Test.000xx	Any number of non-numeric characters, a period followed by any number of digits.
REC(350|380)	REC350	REC350 or REC380

3.2.5.4 **Barcode Settings**

To get to the barcode settings, you must either create a new recognition run or you have to copy or edit an existing recognition run.

In this settings dialog, you set all details in the barcode recognition for the recognition run. These include the name of the recognition run, as well as all control parameters for the barcode recognition.

Most of the settings relate to the traditional linear bar codes. For the 2D barcodes (PDF417, Data Matrix, QR Code or Aztec) almost all of the settings are defined automatically.

"Recognition Run"

> Enter a descriptive name for the recognition run.

"Barcode Recognition"

> Depending on the program licence acquired, you can also select between searches for linear barcodes, PDF 417, DataMatrix, QR Code or Aztec Code. The various barcode types cannot be recognised at the same time in one run. You will need to apply a separate recognition run for each type of barcode. Various types of linear barcodes can however be combined in one recognition run. Exceptions to this are Pharma Code, Patch Code and the Codablock F.

Hint: The following settings primarily apply to linear barcodes. For 2D barcode settings see below. Which parameters can be set for 2D barcodes can be found below the following detailed description.

"Orientation"

Specify the directions in which you want to search for barcodes. If you do not know the direction of the barcode (for example, if a barcode label has been attached to the paper upside-down) you can search in the opposite direction

"Select Types"

When selecting the button <Select Types…>", a small dialog appears (see below) in which the desired linear barcode types can be selected.

"Parameters"

Here the default parameters for the barcode recognition can be adapted. (see chapter 3.2.4.3, Parameter).

"Threshold"

Barcode recognition is in principle carried out on black and white images. The threshold value (1-255) defines the grey shade at which a pixel is converted to a white pixel and to a black pixel.

The higher the value, the darker the overall image will be. This is only true however for greyscale and colour images. Alternatively, the threshold value can be determined dynamically. You can call up the dialogue box for the configuration of the threshold value using the "Threshold…" button (see chapter 3.2.4.4, Threshold)

"Advanced"

The "Advanced Search" offers the possibility to vary some parameters during barcode recognition, without starting another recognition run. This increases the probability that a barcode is recognised. The "Advanced Search" can only be used for linear barcodes (see chapter 3.2.4.5, Advanced Search).

"Checksum"

Modern barcodes often contain a checksum. Calculation of the checksum is defined in the barcode symbol definition.

For barcodes without internal checksum, an extra checksum is often calculated and added. We have programmed some frequently used checksum calculations. In this case, select the type: Standard Modulo 10, Modulo 10 Extended or Standard Modulo 43.

If checksum is activated, barcodes without right checksum are classified as erroneous and no result is reported.

Please observe that the length of the barcode means the total number of characters that means the user data plus the check digits.

"Report Checksum in Barcode Result"

Normally the check digit will not appear in the result. When the check box is activated, the check digit is appended to the result.

"Length"

Here you specify the length of the barcode (number of characters in the barcode) you want to detect. Barcodes with different length are ignored. Length 0 means unknown.

"*unknown -> from to*"

> If you want to recognize barcodes of different lengths, you can specify a length range here.

"Count"

> Here is defined how many found barcodes are registered and how finding a number of barcodes with the same contents on one page should be dealt with.
> In the vast majority of cases, the BC_MULTI setting is the correct choice, even if there is only one barcode present on the page. BC_MULTI in the event of multiple barcodes with identical contents only returns one of these barcodes.
> You should choose BC_MULTI_MULTI if multiple barcodes with the same data content are in the documents and all of them should be read. BC_ONE_BREAK can be selected in special cases, when the recognition should happen very quickly. Recognition will be terminated as soon as a barcode has been read in a scan line in the image.
> **WARNING:** As a result of the immediate termination, the read in data will not be verified in other scan lines and recognition errors can occur as a result, particularly if the quality of the barcodes is not ideal. You should only consider this option in combination with a format check and a check digit in the barcode.
> You should only select the other options (BC_ONE, BC_MULTI_ONE, BC_MULTI_BESTGUESS) after consultation with our support team.

"Check for Barcode Existence"

> Sometimes the documents are scanned in as being too light or too dark, so that the expected number of barcode lines cannot be recognised clearly and as a result, the barcode cannot be read.
> The number of "expected lines" arises from the type of barcode and the number of characters encoded. If fewer lines are found, then the "Barcode Exists" notification can be activated in the results using this button. In this way, a "suspicion" of barcodes is flagged.
> The sensitivity of the notification of suspicion can be controlled in the Percent field in the extended parameters (= the percentage threshold of expected number of lines that must be present to trigger the suspicion. That means that the value should be under 100 because, at 100 percent, all of the lines must be present.)

"Maximum Skew Angle"

> Specifies up to which skew angle a barcode is still recognised. Normally 8° is sufficient. However, if you expect barcodes to be more skewed, you should select a higher value. Please note that the higher the angle is, the longer the recognition takes.

"*Intensive Search*"

> For very flat barcodes (aspect ratio from 1:7), it can happen that distorted barcodes slip through the search criteria. This option can also be implemented for barcodes of a poor quality. Using this option, you can refine the search criteria. Warning: This will slow down the barcode recognition!

"*Use configuration file* "

> Using the configuration file "qsbc.ini", you can enhance images before scanning barcodes. You can assign a separate "qsbc.ini" to each recognition run. You can thus configure runs in even more detail. The name of the file selected must, however, not be "qsbc.ini", as it then will be used automatically for every recognition run.

For more information about the configuration of image enhancement, please refer to our document "bcTipps.pdf".

"Pass graphics command directly "

The commands in the "qsbc.ini" can also be transferred directly to the barcode search. This eliminates the need for a separate configuration file. The individual commands must be entered separated by a semicolon (without quotes) in the input field, e.g. "FreeImage:GrayScale,8;FreeImage:Threshold,160;"
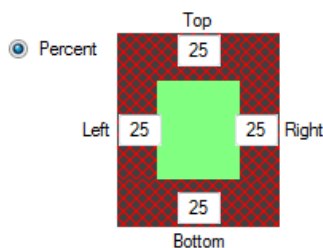
"Recognition Area"

Either the "entire image", an "area" in absolute pixel values or a percentage area can be selected to be the recognition area. If the "Area" option is selected, the area co-ordinates in pixels must be specified. X/Y designates the upper left corner of the area, DX=width, DY=height.

When specifying the percentages, enter the distances in percent from the respective margins, whereby T=Top, B = Bottom, L = Left, R = Right. If you want to look for a bar-code in the upper left quarter in percentage terms, then you should enter Left=0, Top=0, Bottom=50, Right=50.
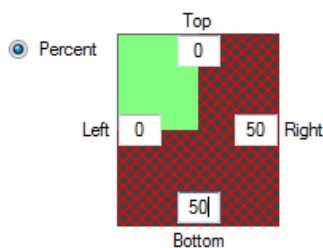
For the lower right quarter, you should use Left=50, Top=50, Bottom=0, Right=0 and for an area of the size of a quarter of the page directly in the middle, you should use Left=25, Top=25, Bottom=25, Right=25.

Settings for 2D barcodes

In the case of two-dimensional barcode types, many of the characteristics of the barcode are prescribed in the specifications, so that many recognition parameters are set internally to the ideal settings when the type is selected, or are not necessary for the reading.

In the case of real 2D barcodes (Data Matrix, QR Code and Aztec), the search is performed at all angles (also interim values). For 2D barcodes, the check digit procedure for the recognition of barcode errors will be replaced by internal error detection and correction algorithms (ECC Error Correction Codes), with which errors cannot only be detected, but can frequently also be automatically corrected.

After selection of the types PDF417, Data Matrix, QR Code or Aztec, the non-configurable setting options will be greyed out and cannot be altered.

The specification values (default values) for 2D barcode types differ from the default values for linear barcodes. The default values that are ideal in the majority of cases will be set upon selection of the type.

The following settings are possible for two dimensional barcodes:

PDF417

Orientation (0°, 90°, 180° and 270°) and Light Margin, Search Distance, Min. Height, Max. Height.

Data Matrix

Light Margin, Search Distance, Tolerance, Min. Height, Max. Height.

For the **QR Code** and the **Aztec** no settings are available.

Note:
Instead of entering all of the above parameters, you could simply select all of the linear types in the type selection and set the length to be unknown between 1 and 64 and activate all rotational alignments. Please observe that entering the type and length precisely will lead to better and more rapid results. If the type and length are known, then you should select the appropriate settings.

3.2.5.4.1 Barcode – Select Linear Barcode Types

Select the linear barcode types that are relevant to your documents. Codablock F, Pharmacodes and Patchcodes can not be combined with the other codes in the same recognition run.

3.2.5.4.2 **Parameter**

The pre-set default settings have proven themselves for linear barcodes in many areas. In special cases, departing from these values may however be necessary to get to the results.

Parameters			
Light Margin	30	Percent	100
Scan Distance	3	Search Distance	15
Tolerance	20	Max. Gap	10
Min. Height	15	Max. Height	1000

OK   Default   Help   Cancel

"Light margin"
>Number of white pixels that are at least expected around a barcode. When white gaps between barcode bars are wider than the light margin, the barcode will not be recognized.

"*Percent*"
>Specifies what percentage of a barcode must be detected for it to be recorded as a barcode.
>Attention, this is a special setting for cases in which the barcode is very uncertain. A "Suspicion of Barcode" will be registered if you have activated the suspicion of barcode in the settings of the recognition run.

"*Scan Distance*"
>Line spacing between the individual scans of the image in the Y-direction.
>It will be automatically changed over to the scan distance as soon as the search has found sufficient lines. Larger values accelerate the search, because fewer lines will be scanned.

"*Search Distance*"
>Distance between the individual scans within a barcode in the Y-direction.
>In the case of "bad" barcodes, the scan distance can be set to 1. If an insufficient number of lines are found in a row for a barcode, then it will be switched back to the scan distance for the search.

"Tolerance"
>Tolerance of the position (x-direction) of the barcode during the search.

"Max. Gap"
>Allowed distance between two successfully detected lines inside a barcode. If this distance is exceeded, the barcode is marked as completed.

"Min. Height"
>The minimum height of a barcode to be considered as valid. For barcodes with a very poor quality, it may be necessary to set this value to a low value.

"*Max. Height*"
    Maximum height of a barcode. Barcodes larger than the maximum height are ignored.

3.2.5.4.3 Threshold
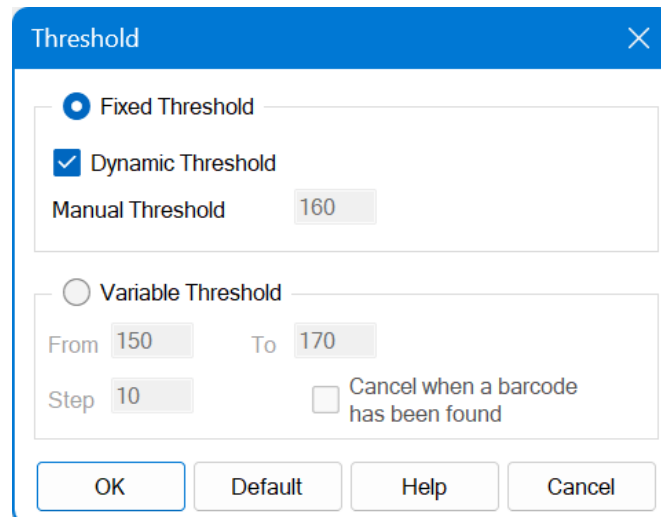
The threshold value (1-255) defines the grey shade at which a pixel is converted to a white pixel and to a black pixel. The threshold value can be determined dynamically or set manually.

In addition, for documents that have different brightness, e.g. because they were generated with different printers or scanners, you can gradually change the threshold value within a range to increase the recognition rate.

"*Dynamic threshold*"
    The dynamic threshold is the default setting. It has the advantage that too light or too dark images are automatically converted to the appropriate b/w image for barcode recognition. On images with heterogeneous appearance concerning the contrast, the dynamic threshold may not achieve good results.

"*Manual threshold*"
    If the dynamic threshold does not work properly, you can control the conversion manually. You can decide which threshold is used for the image conversion. The default value is 160.

Variable threshold

"*Threshold from*"
    Initial threshold value, which binarizes the images to be processed.

"*Threshold to*"
    Highest threshold value to binarize the images to be processed.

"*Step*"
    The threshold value is gradually increased by this value to the maximum value. When a barcode is read, no further attempts are performed.

3.2.5.4.4 **Advanced Search**

The "Advanced Search" allows search parameters to vary within a recognition run. A second scan is started when no barcode is found in the first scan. The "Advanced Search" can only be used for linear barcodes.

"*Light margin 1/Light margin 2/Light margin 3*"

 The default value for the light margin is 30. In the "Advanced Search", you can specify up to three different light margins, which are used for the barcode recognition. A search is generally done with all light margins. This will help to find barcodes with different quiet zones on an image. If you don´t need a quiet zone, set is value to 0.

"*Remove pixels*"

 It often happens that, in particular through a poor-quality scan or by a bad scan original little white pixel gaps occur in the black bars of the barcode. These small pixel gaps often prevent a recognition of the barcode. The white pixel gaps (as well as black pixels in the white bars of a barcode) can be removed up to a width of two pixels automatically. If you don´t want to remove pixels, set this value to 0.

"*Scan distance 2/Scanline 2*"

 Scan distance is the distance between the lines in the y-direction where a barcode is searched for. When the barcode is detected, a switch to scanline takes place.
 The first recognition run is performed with Scan distance/Scanline (default 15/3) from the barcode search parameters dialog. When no barcode was found with these settings, Scan distance 2/Scanline 2 will be used in the second run. If the second scan should not be started, set these values to 0.

3.2.5.5 **OCR-Settings**

In order to access the OCR-settings you have to create a new OCR recognition run or copy and edit an existing one.

Within the setting dialog, you can determine alle details concerning the OCR recognition which include: name of the run, the creation of zones that specify the desired text, the zone's reading settings and general parameters concerning the OCR recognition.

When opening the dialog, only the name of the detection run will be displayed. First, you have to upload a document which you need in order to create reading zones within which the text can be recognized. If you have already created zones for the recognition runs, they will will be automatically loaded when opening a document.
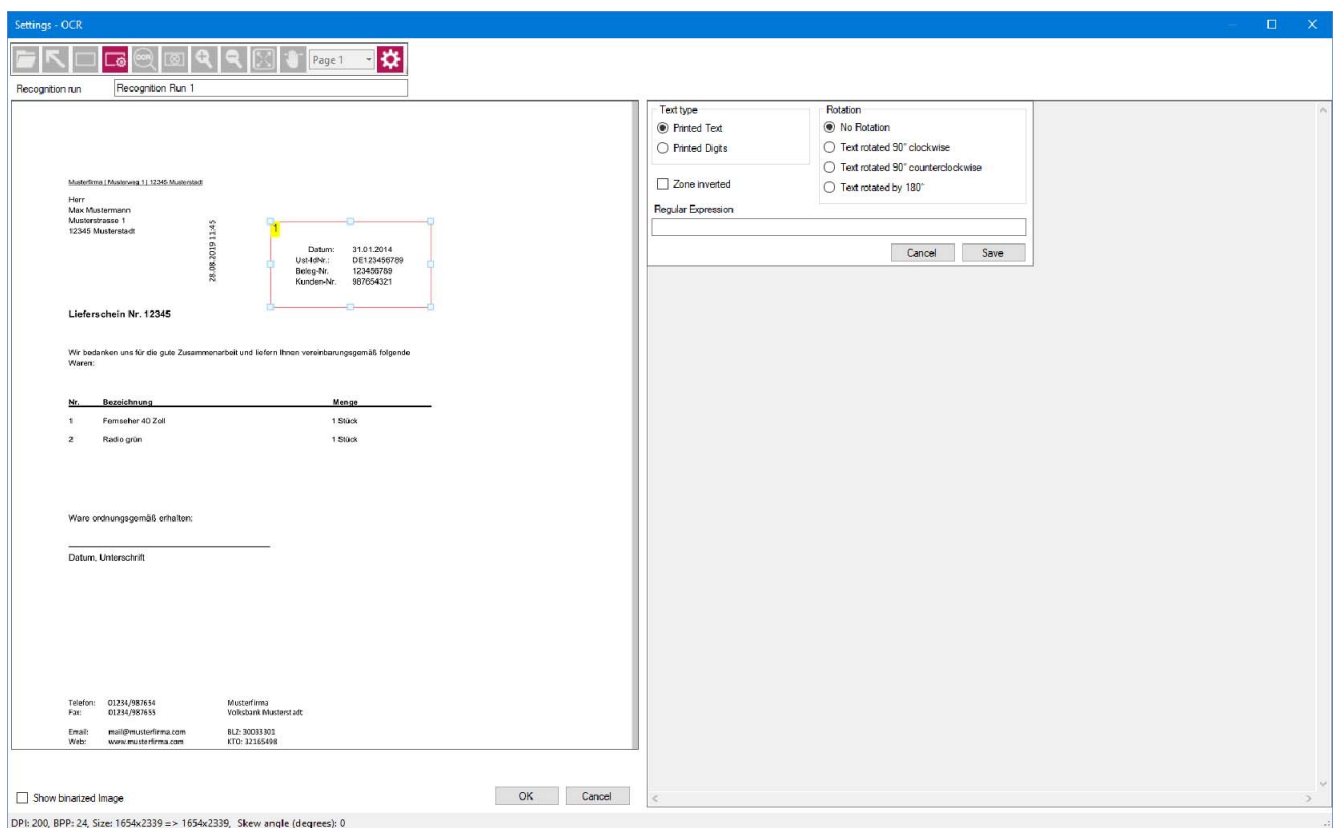
Images that are loaded will be automatically binarised which means that it will be converted into a black and white image, since the OCR recognition only works with this format. Horizontal and vertical lines that may disturb the OCR process will be automatically deleted although the original document will be displayed first.

You can switch between the original image and the binarized image at any time via the "Show Binarized Image" checkbox and check whether image artifacts may hinder the reading of the text.

The processing is done via the horizontal button bar above the loaded document.

"**Open document**". Opens a document in which the zones for reading text can be created.

"**Select zone**". Switches to the selection mode for zones.

"**Create zone**". Switches to the creation mode for zones.

"**Zone settings** ". Opens the zone settings for a selected zone.

"**Read text** ". Reads the text within the defined zones.

"**Delete all zones** ". Deletes all existing zones.

"**Zoom plus**". Enlarges the image display.

"**Zoom minus**". Reduces the image display.

"**Original** d**isplay size** ". Restores the original image size.

"**Pan**". Moves the image within the display.

Page 1 ▾ "**Page selection**". Selection of a page for multi-page documents.

"**Settings**". Selection of special settings of the OCR engine including language selection.

3.2.5.5.1 **The creation and configuration of a zone**

To create a reading zone, click on the "Create zone" button in the toolbar. Then use the mouse to draw a rectangle around the area from which text is to be read during processing.

To check which text is being read in the zone, you can carry out an OCR reading using the "Read text" button. The text read in the zone is displayed in the window to the right of the image.

```
Zone 1:
Datum: 31.01.2014
Ust-IdNr.: DE123456789
Beleg-Nr. 123456789
Kunden-Nr. 987654321
```

You can set various settings for a zone. To call up the settings, activate the desired zone with the mouse and then click on the "Zone settings" button in the toolbar.

"Text type"

    Here you can specify whether it is alphanumeric text or just numbers. The standard case is printed, i.e. alphanumeric text. If you select "Printed digits" here, the text will only be interpreted as a number.

"Zone inverted"

    If you want to read text in inverted documents or areas. The area is then inverted again before reading to make the text readable.

"Rotation"

    f you want to read rotated text on the document, you can specify the direction of rotation. The text is then rotated in the selected direction before reading. When creating a zone, an attempt is already made to determine the direction of rotation, so that the correct direction of rotation is selected when you call these settings.

"Regular Expression"

    If this field is left blank, the text is adopted as it was read in the zone. Regular expressions give you a complex way to filter the text of the zone according to different criteria.

You can apply the changes made using the "Save" button. If you cancel the dialog, any changes will be discarded without confirmation.

Reading text in a zone

Regular expressions enable the very fine and flexible selection of text in the zones. Since the scanning process does not scan the documents in an identical position and the documents cannot be aligned using markers, it is necessary to make the reading zones larger. This can lead to more text being read than required. In order to achieve the desired result anyway, regular expressions can be used flexibly here.

Suppose we created the following zone on a document from which we only need the nine-digit document number to rename the document. A zone only only drawn around the receipt number is prohibited here for the above reasons.

In the zone, however, there are a total of 3 nine-digit digits, so finding a nine-digit number alone does not help us. In this case, we would create a regular expression with a lookbehind function.

Die Lookbehind-Funktion ermöglicht das Finden von komplexerem Text, wobei bestimmte Textbereiche aus der Suchbedingung nicht mit in das Ergebnis übernommen werden. Wenn ich also die neustellige Beleg-Nr. suche, nehme ich den Text "Beleg-Nr." mit in die Suche auf, um ihn dann aus dem Ergebnis auszuschließen.

The lookbehind function allows you to find more complex text, where certain areas of text from the search condition are not included in the result. So if we search for the nine-digit "Beleg-Nr.", we include the text " Beleg-Nr." in the search and then exclude it from the result.

The regular expression would look like this: (?<=Beleg-Nr.\s)\d{9}

We are looking for a nine-digit number: "\d{9}".
This number must be preceded by the string "Beleg-Nr.\s". The "\ s" stands for a tab or a space. The text to be removed after the match is removed using the look-behind function: (?<= x), where the "x" stands for the pattern to be excluded (Beleg-Nr.\s).

The result after reading the zone would look like this:

For this document, the result "123456789" would be used for the renaming of the document.

The regular expressions are based on the rules of the Microsoft .NET Framework. For further information about the regular expression in the Microsoft.NET Framework please see https://docs.microsoft.com/de-de/dotnet/standard/base-types/regular-expression-language-quick-reference

3.2.5.5.2 Managing Zones

In principle, any number of zones can be created on a document and the results are then transferred to processing based on the zone order.

So it would be possible to define two zones whose results would be further processed via the defined parameters of the variable file name.

In principle, the results from OCR recognition are subsequently treated as read barcodes and thus all criteria apply, which also apply to the further processing of barcode results.

If a single zone is to be deleted, activate it with the mouse or draw a frame around the zone with the mouse and delete it using the keyboard by pressing the "Del" key. To move the zone it must also be selected and can then be moved with the mouse or with the arrow keys on the keyboard. To apply this to several zones, you can select the desired zones by "CTRL" + mouse click. Alternatively, you can also use the mouse to draw a frame around the desired zones.

3.2.5.5.3 **Settings OCR-Engine**

Here you can select the language to read the document. Multiple languages can be selected in parallel.

All other default settings of the OCR engine are specifically optimized for the recognition of text in scanned documents. Please change the settings only if you are aware of the consequences. If in doubt, contact support.

3.2.6   Tab "Imprinter"

Software Imprinter

An "imprinter", often also called an "endorser", applies an imprint/stamp to the document during scanning. It serves as a marker, and signifies that the document has been processed; it is also used for later document tracing.

A "software imprinter" or "digital endorser" applies an ID to the digital copy created, i.e. the image file.

QS-DocumentAssembler does the same for processed documents, and can also include pages on which no barcode was recognised.

The position, layout (incl. font) and the information to be contained in that "stamp" are user-definable.

The following variables are available when collating the text, which **QS-DocumentAssembler** will fill at run time with the relevant data, and the software imprinter then applies to the processed documents:

"Target files are provided with an imprinter stamp on each page with barcode"
   Whenever a barcode is found on a page, a imprinter stamp is applied.

"Imprinter stamp on page even if no barcodes was found"
>   An imprinter stamp is applied on pages without barcode.

"Stamp"
>   When creating the text that is applied to processed documents, the following variables can be used:

The following variables are available

%SOURCEFILE%	Original file name
%BARCODE	The first recognized barcode
%BARCODEn	The following recognized barcodes
%BC_REGEX%	Result of a regular expression applied to the first barcode.
%BC_REGEX_REPLACE%	Result of a regular expression replaced in the first barcode
%PAGE_NO%	The current page number of the source document
%PAGE_NOn%	The current page number of the source document, formatted to n digits, e.g. %PAGE_NO3% corresponds to "001"
%SEQNO%	Number of occurrences of a file name in the target directory (sequential number)
%SEQNOn%	The sequential number formatted to n digits, e.g. %SEQNO4% corresponds to "0001"
%YYYY%	four-digit year, e.g. 2014
%YY%	two-digit year, e.g. 14
%MMMM%	Month in long form,e.g. "September"
%MMM%	Month in short form, e.g. "Sep"
%MM%	two-digit month with leading zero, e.g. "07"
%M%	one-digit month
%DDDD%	Day in long form, e.g. "Tuesday"
%DDD%	Day in short form, e.g. "Tu"
%DD%	two-digit day with leading zero, e.g."01"
%D%	one-digit day
%DOY%	Day of the year, e.g. 125
%HH%	two-digit hour
%H%	one-digit hour
%mm%	two-digit minute
%m%	one-digit minute
%SS%	two-digit second
%S%	one-digit second
%AM/PM	AM/PM
%DATE_SHORT%	The current date formatted by the country settings of the OS
%DATE_LONG%	The current date formatted by the country settings of the OS
%TIME	The current time formatted by the country settings of the OS

All barcodes found on a page can be used for creating the imprinter stamp using the variables %BARCODEn%, %BARCODE2%, etc.

The barcodes are internally listed in the order in which they are found in the document. This order does not necessarily match the real order of the barcodes in the document.
Therefore, the barcodes should be sorted according to their position (see chapter 3.2.4, Tab "Barcode").

If there are more barcode parameters than actually existing barcodes, these parameters are ignored.

You can add any additional text to the imprinter stamp.

"Position"

Select a margin for the positioning of the imprinter stamp.

"Alignment"

At the respective edge of the image, the text can be presented left-aligned, right-aligned or centred.

"Format"

The format specifies whether the lower edge text (text footer) appears at the respective image border or in the middle of the image.

"Font"

All installed fonts can be used as typeface to create the imprinter stamp.

"Font Size"

Specify the font size, in which the text is to be displayed.
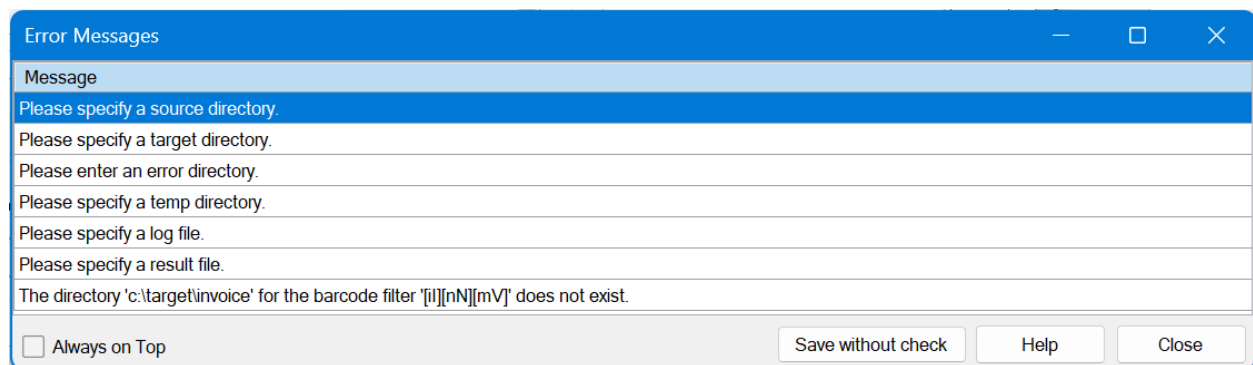
"Font Style"

Specify the font style, in which the text is to be displayed.

4   Plausibility check

When all of these settings are saved, the data will undergo a plausibiltity check. In this check, it will be fundamentally checked whether the data has been input completely and if the contents match each other.

If any errors are found, then a dialogue box with the error messages will be displayed. From this dialogue box, you can jump directly to the tab in which the error has occurred by double clicking on the error message. You can keep this window permanently in the foreground using the "Always on Top" checkbox. Doing this will mean that the irritating search for the window with the error message is not necessary.

Error Messages	—	□	✕
Message			
Please specify a source directory.			
Please specify a target directory.			
Please enter an error directory.			
Please specify a temp directory.			
Please specify a log file.			
Please specify a result file.			
The directory 'c:\target\invoice' for the barcode filter '[iI][nN][mV]' does not exist.			

☐ Always on Top   Save without check   Help   Close

If you are not able to, or do not want to enter the data completely, then you also have the option to save the data from the error dialogue box without checking. Jobs that have not been completely recorded will be presented as "Not Checked" in the overview of jobs. These jobs cannot be activated and as a result, cannot be carried out. Only after the data has been entered completely will these jobs be displayed as being checked and will then as a result be able to be activated.

Errors that have occurred will be labelled with the corresponding error mark 🔴 in the dialogue box. If you move the mouse over an error mark, then the corresponding error message will be displayed in a tooltip. You now have the opportunity to deal with all of the errors that have occurred.

5   Integration of QS-DocumentAssembler in batch processing

QS-DocumentAssembler provides you with the option of integration in your batch processing. You can obtain this with the "docasmcmd.exe" command line version of the QS-DocumentAssembler.

The QS-DocumentAssembler supports several command-line switches. The following rules apply:

-   The switches are Case-insensitive

-   The order of switches doesn't matter, the first occurrence is used in double occurrence of the same switch.

-   Every switch starts with a "/"

/j:\<Name of Job\>

Name of the job to be executed. If the specified job name does not exist, the application terminates.

/s:\<Source Directory\>

By specifying a separate source directory, the setting of the job can be overridden. If the specified source directory does not exist, the application terminates.

/t:\<Target Directory\>

By specifying a separate target directory, the setting of the job can be overridden. If the specified target directory does not exist, the application terminates.

/e:\<Error Directory\>

By specifying a separate error directory, the setting of the job can be overridden. If the specified error directory does not exist, the application terminates.

If the passed parameter contains spaces, it must be enclosed by quotation marks.

The specification of the job name is mandatory, all other parameters are optional. Thus, you can override the directories of the job individually.

If the submitted job was successfully executed by the command line version of the QS-DocumentAssembler, the return value is "0". In case of an error, the return value is "1". Further information to the cause of the error can be found in the error file "\qsMsgLog\qsMsgLog.QS-DocumentAssembler.txt".

6 QS-DocumentAssembler Service

The **QS-DocumentAssembler Service** offers you the option of processing documents as a service without any user interaction, offering you the following advantages:

- No user needs to be logged on.
- Automatic restart by the Windows Service Manager
- Once the server has been restarted, the service starts up automatically.
- Continuous processing of documents without the need for any user interaction
- The service can run in a local system account or a user account.
- Error logging in the Windows event log
- Dispatch of an e-mail in the event of an error

The service can, of course, also process documents for and on network drives, provided the drives in question are available in user mode.

6.1 Service administration

In order to manage the service, you must have the necessary administrator rights for executing the relevant functions.

6.1.1 Service administration via user interface

The service can be managed on the user interface of QS-DocumentAssembler.
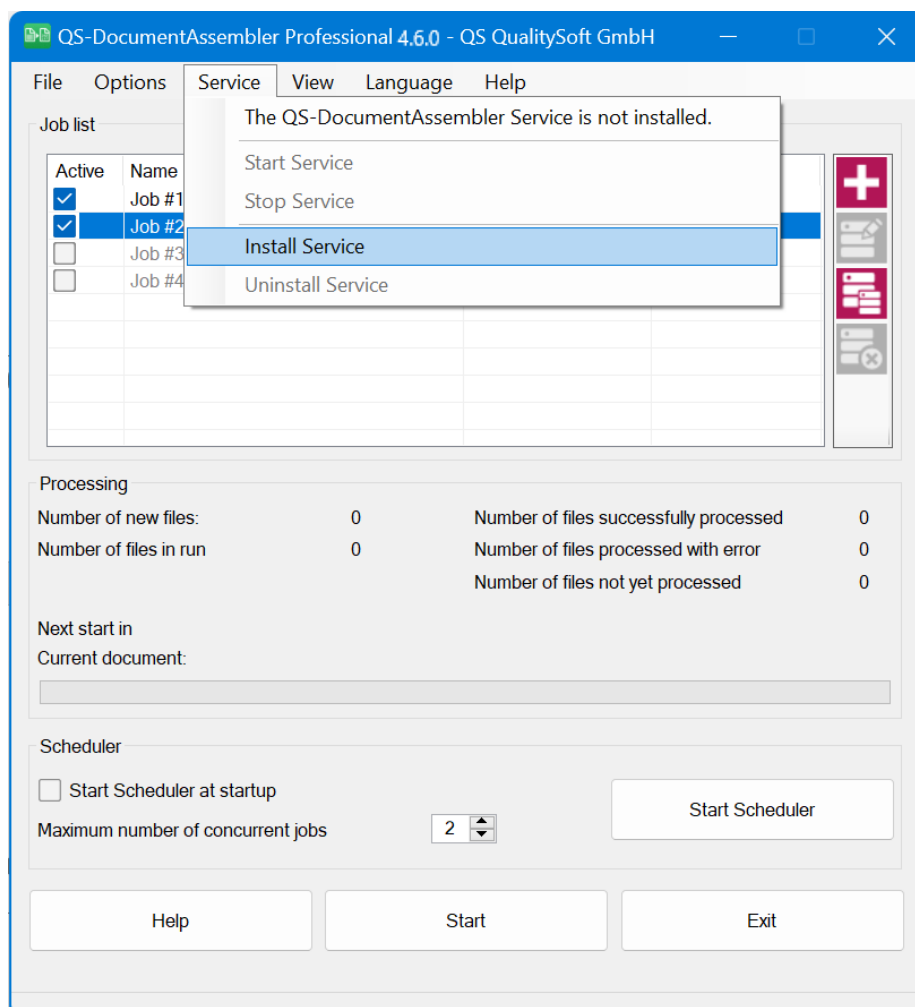
Before you can use the service, you first need to install it via the <Install service> menu.

When installing the service, you need to define the mode in which you want the service to run: in the "local system account" or in a specific user account.

If the QS-DocumentAssembler Service runs in the local system account, you do not need to enter a password. If the service runs in a specific user account, you need to enter ".\" in front of the relevant user name, e.g. ".\Administrator".

If the service runs in a specific user account, the account must have the right authorisation to be executed as a service.

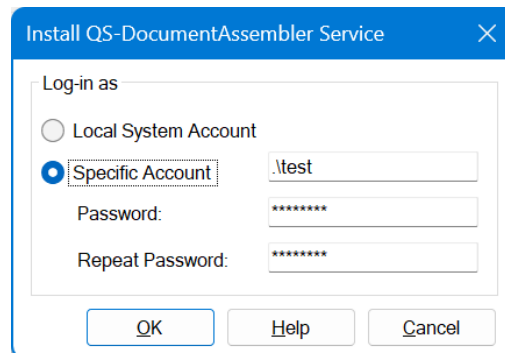If you want the service to access network drives/shares/volumes, the account must be assigned the relevant authorisations for these network drives/shares/volumes.

You can access the service directly via the <Service> menu. You can start, stop and pause the service, for example. You can also uninstall the service via this menu.

Once you have registered the service, it will be added to the list of executable services. Here is an excerpt from the Windows Service Manager.
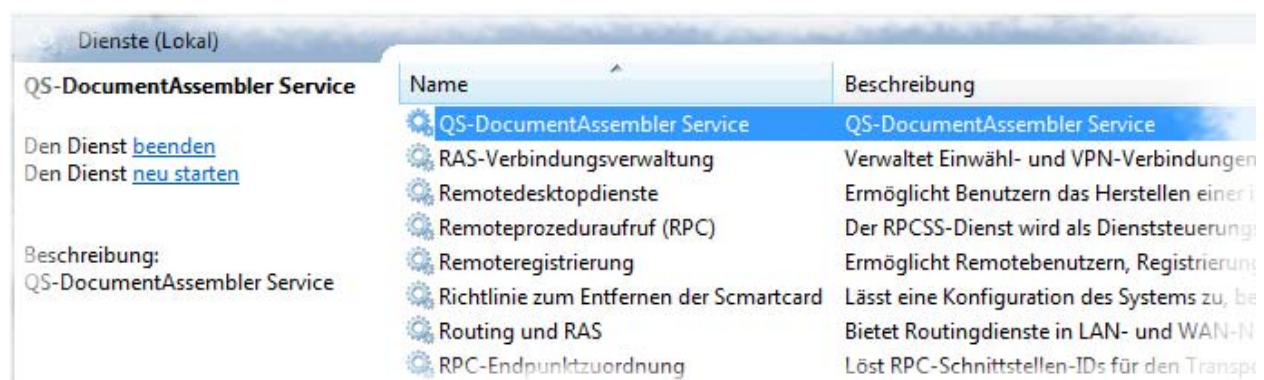
6.1.2 Service administration via Windows service manager

With the exception of the functions for registering/uninstalling the service, you can also access all of the functions in the <Service> menu in the Windows Service Manager. In the Service Manager, you can also configure the procedure for recovering the service in the event of an error.

6.1.3   Service administration via command line

Registration and uninstallation of the service can be run from the command line.

Registration is done by calling
"C:\Windows\Microsoft.NET\Framework\v4.0.30319\installutil /i DocAsmSvc.exe"

the uninstall is done by calling
"C:\Windows\Microsoft.NET\Framework\v4.0.30319\installutil /u DocAsmSvc.exe".

Both calls are to be run with administrator rights from the program directory of the QS DocumentAssembler.

6.2   Service configuration

The QS-DocumentAssembler service draws upon the available configured jobs that have been set up and activated for the scheduler (see chapter 6, QS-DocumentAssembler Service).

As the service does not however carry out any direct interaction with the user, options with an interactive element, for example "Display a Message for Errors" or the inputting of a prefix during processing will only be logged by the service and will not be carried out.

If you want to access network drives during service operation, you must enter the relevant paths in the "general options" in UNC notation, e.g. "\\Server name\Share name\Directory". The desktop version naturally also supports UNC notation. Local drives can be accessed via drive letters.

We recommend testing the settings of a job with the user interface of the QS-DocumentAssembler as a job that is to be carried out manually. After successful manual testing, you can specify start time points and intervals for this job. If the service is already running, this job will be transferred automatically into processing.

The scheduler cannot be activated while the service is started as otherwise, a job could be started a number of times and this could lead to access conflicts for the source files. Conversely, the service cannot be started when the scheduler is running. All jobs that have been defined as manual can however be started while the service is in operation, as there is no possibility of a conflict with the service here.

7   Appendix

7.1   **Log File**

The log file contains information about the processing of all pages of documents or only of pages without barcodes.

The example below shows the structure of the log file. One line is created per page in a document. Individual information fields are separated by semicolon.

09.10.2014 09:21:02: OK; C:\DocAsm\source\01.tif;C:\DocAsm\target\021079.tif;021079;;BARCODES:1/FILECOPY:OK /FILEKILL:OK

Overview of log file fields:

Position	Description	Example
1	Time of processing	09.10.2014 09:21:02:
2	Overall status	OK or ERROR
3	Source file name	C:\DocAsm\source\01.tif
4	Target file name	C:\DocAsm\target\021079.tif
5	First recognized barcode on the page; will remain empty if no barcode exists or no barcode was recognized	021079
6	Status informationen	BARCODES:1/FILECOPY:OK/FILEKILL: OK

7.2   **Supported File Formats**

File formats supported by QS-DocumentAssembler:

- BMP – Standard Windows Bitmap Format
- CUT – Dr. Halo/Dr.Genius Clipboard Format
- DDS – Microsoft DirectDraw Surface Format
- DIB – Standard Windows Bitmap Format
- EMF – Enhanced Windows Metaformat
- EXIF – Exchangeable Image Format
- EXR – OpenEXR Format
- FAX, G3 – Group 3 Raw Fax Format
- GIF, Interlaced GIF – Graphics Interchange Format
- HDR – High Dynamic Range Format
- IFF – Interchange Format
- J2K, J2C – JPEG-2000 Codestream

- JIF, JFIF – JPEG File Interchange Format
- JNG – JPEG Network Graphics
- JP2 – JPEG-2000 Format
- JPEG, JPG, JPE – Joint Photographic Expert Group
- JPEG progressive
- KOA – KOALA Format
- LBM – Interchange File Format-Interleaved Bitmap
- MNG – Multiple-image Network Graphics *
- PBM – Portable Bitmap File
- PBM RAW – Portable Bitmap BINARY
- PCD – Kodak Photo-CD file
- PCT, PICT, PIC – Macintosh PICT Format
- PCX – PC Paintbrush Format
- PDF – Portable Document Format
- PDF/A – Document Format for long term preservation
- PDF Multi-page – Portable Document Format
- PFM – Portable Float Map
- PGM – Portable Graymap File
- PGM RAW – Portable Graymap BINARY
- PSD – Photoshop File
- PNG – Portable Network Graphics Format
- PNM – Portable Any Map
- PPM – Portable Pixmap File
- PPM RAW – Portable Pixmap BINARY
- RAS – Sun Raster Format
- RAW camera image
- RAW memory bits – RAW bitmap
- RLE – Standard Windows Bitmap format
- SGI – Silicon Graphics Image Format
- TGA, TARGA – TARGA Image Format
- TIFF, TIF – Tagged Image Format
- TIFF Multi-page – Multi-page Tagged Image Format
- WBMP, WAP, WBM – Wireless Bitmap
- WEBP – WebP Image Format
- WMF – Standard Windows Metaformat
- XBM – X-Bitmap Format
- XPM – X Pixmap Format

7.3 **Notes to the Adobe PDF file format**

Today, the Adobe PDF file format has established itself as standard for the PC.

A PDF file can contain different components such as text, images, bookmarks and signatures etc.

In the document scanner environment as well, the Adobe PDF format is used alongside the traditional file formats for image files such as BMP, TIF and JPG.

In the most simple form (PDF type 1), these PDF files contain **only the images** from the scanned documents.

Additionally, an optical character recognition **(OCR)** can be applied to the scanned images in the PDF file, so that the **text in the images** will be read and a "searchable PDF2" results (**PDF type 2**). Searches for these texts can then be carried out in the PDF file. It must be noted here that, depending on the quality of the writing and of the image, the character recognition may be afflicted by a number of different read errors. As there is generally no manual correction afterwards, both incorrect characters and gaps in the "searchable" text may ensue.

Furthermore, there are of course PDF files that are created by the "**printing" of documents** (**PDF type 3**). Here, the different components from the documents such as text, typefaces, logos and images are present in the PDF file. The barcodes can be present on the pages either as special typefaces or as small graphics.

Important	**Embedded Fonts**
	If you use typefaces for the creation of barcodes in Adobe PDF documents, then you must "embed" these typefaces in the Adobe PDF document, so that QS-DocumentAssembler can carry out the barcode recognition and processing.
	It is not sufficient for the typeface merely to be installed on the workstation.

QS-DocumentAssembler supports the processing of all listed PDF types.

For the barcode recognition, the pages to be read will be converted into image files, from which the barcode will then be read.

Information: The reading of barcodes from PDF documents is usually possible. As the programs for the creation of PDF files function very differently, you should in any case test the different types of PDF files that you are intending to have read in, before routine use.