
Autobahn DX 5.01

Reference Guide



Version 5.01
March 2020

CONTENTS

1	PRODUCT OVERVIEW AND CONCEPTS	5
1.1	OVERVIEW.....	5
1.1.1	<i>Edition Comparison</i>	<i>5</i>
1.1.2	<i>Autobahn DX Architecture and Concepts</i>	<i>7</i>
1.2	AUTOBAHN WINDOWS SERVICE	7
1.3	XML JOB DEFINITION FILES.....	7
1.4	AUTOBAHN ADMINISTRATION.....	8
1.4.1	<i>Quick Job</i>	<i>8</i>
1.4.2	<i>Job Manager.....</i>	<i>8</i>
1.4.3	<i>Designer</i>	<i>8</i>
1.4.4	<i>Monitor</i>	<i>8</i>
1.5	DOCUMENT DIRECTORIES	8
1.6	DOCUMENT PROCESSING.....	8
1.7	JOB STATUS	8
1.8	AUTOBAHN DX .NET JOB API	8
1.9	FILE ACCESS PERMISSIONS	8
2	INSTALLATION AND LICENSING.....	9
2.1	SYSTEM REQUIREMENTS.....	9
2.2	PRODUCT LICENSING.....	9
2.3	ENTERING LICENSE KEYS	10
2.4	SUPPORT FOR MULTIPLE CPU CORES	10
3	AUTOBAHN DX QUICK START	11
3.1	THE QUICK JOB TAB.....	11
3.2	THE JOB MANAGER TAB	11
3.3	THE DESIGNER TAB.....	12
3.3.1	<i>General Job Properties</i>	<i>12</i>
3.3.2	<i>Scheduling a job.....</i>	<i>13</i>
4	USING THE QUICK JOB TAB	14
4.1.1	<i>OCR Options Section.....</i>	<i>15</i>
4.1.2	<i>OCR PDF Options Section.....</i>	<i>16</i>
4.1.3	<i>PDF to TIFF Options Section.....</i>	<i>17</i>
4.1.4	<i>Split Options Section</i>	<i>17</i>
4.1.5	<i>Other Options Section</i>	<i>17</i>
4.1.6	<i>Barcode Options</i>	<i>18</i>
4.1.7	<i>Task Log Section.....</i>	<i>19</i>
4.2	SECURITY SETTINGS	20
4.3	DOCUMENT METADATA	21
4.4	DOCUMENT OPTIONS	22
4.4.1	<i>Display Modes</i>	<i>22</i>
4.4.2	<i>Settings.....</i>	<i>23</i>
4.5	PDF CONVERSION SETTINGS.....	23
4.5.1	<i>Conversion Control.....</i>	<i>24</i>
4.5.2	<i>Microsoft Office Conversion Settings.....</i>	<i>24</i>
4.5.3	<i>Image and Font Conversion Settings.....</i>	<i>26</i>
4.6	ADVANCED FLAGS	27

5	THE AUTOBAHN DX ADMINISTRATION TOOL	28
5.1	JOB MANAGER	28
5.2	JOB DESIGNER	29
5.2.1	<i>Menu Items</i>	29
5.2.2	<i>Fields</i>	31
5.2.3	<i>Job Scheduling</i>	34
5.2.4	<i>Alerts</i>	35
5.2.5	<i>Workflow Processing versus In-Place Processing</i>	37
5.2.6	<i>Example In-Place Job Setup</i>	37
5.2.7	<i>Step Types</i>	39
5.2.8	<i>Step Type Properties</i>	93
5.3	MONITOR.....	94
5.4	MODULES AND OPTIONS	95
5.5	RUNNING JOBS VIA JOB MANAGER OR DESIGNER	96
6	SCRIPTING CUSTOM STEPS	97
6.1	CUSTOM SCRIPT EXAMPLE	97
6.2	CUSTOM EXAMPLE SCRIPT	98
6.3	FURTHER EXAMPLES	99
6.3.1	<i>Postscript to PDF: custom_ps.wsf</i>	99
6.3.2	<i>MSG to PDF: custom_MSG.wsf</i>	100
7	JOB DEFINITION XML FILES	101
7.1	OVERVIEW OF JOB DEFINITION CREATION AND PROCESSING	101
7.2	JOB IDS.....	101
7.3	SAMPLE JOB DEFINITION FILE (EXCLUDING STEP DETAILS).....	101
7.4	SAMPLE STEP DETAILS.....	103
8	EXTENDED STEP TYPES: STEP TYPE DEFINITION XML FILES	105
8.1	SAMPLE STEPTYPE DEFINITION.....	105
9	AUTOBAHN DX COMMAND LINE INTERFACE	107
9.1	GENERAL JOB OPTIONS	108
9.2	SPLIT PARAMETERS [USED WITH SPLITTIFF AND SPLITPDF OPERATIONS]	109
9.3	PDF CONVERSION OPTIONS [USED WITH TOPDF OPERATIONS].....	110
9.4	[PDF TO TIFF [USED WITH PDFTOTIFF]	115
9.5	OCR [USED WITH OCRIMAGEPDF OR TIFFTOPDF/ MERGETIFFTOPDF].....	116
9.6	PDF OUTPUT FILE SETTINGS – [USED WITH PDFPROPS]	119
9.7	SHAREPOINT SETTINGS – [USED WITH SHAREPOINTDOWNLOAD AND SHAREPOINTUPLOAD]	121
9.8	COMPRESS PDF SETTINGS – [USED WITH COMPRESSPDF]	122
9.9	EXTENDED OCR STEPS	123
9.9.1	<i>Extended OCR Languages</i>	128
9.10	CREATE XML PROPERTY FILE – [USED WITH GETXMLPDF]	133
9.11	OCR ANY FILE TO PDF– [USED WITH OCRANYFILE OR OCRANYFILEEX]	134
9.11.1	<i>Standard</i>	134
9.11.2	<i>Extended</i>	134
9.12	TIFF/PDF BARCODE – [USED WITH BARCODE].....	134
9.13	STAMP PDF – [USED WITH STAMPPDF]	136
9.14	AZURE UPLOAD/DOWNLOAD – [USED WITH AZUREDOWNLOAD AND AZUREUPLOAD]	138
9.15	EXTRACT TEXT DOWNLOAD – [USED WITH EXTRACTTEXT]	138
9.16	MERGE PDF – [USED WITH MERGEPDF].....	139

10	AUTOBAHN DX .NET API.....	140
10.1	API EXAMPLE (LOCAL JOB)	140
10.2	API EXAMPLE (REMOTE JOB)	141
11	OCR PROPERTIES FILE AND THE ADVANCED PRE PROCESSING OPTION	144
11.1	PROPERTIES FILE.....	144
11.2	ADVANCED PRE PROCESSING.....	146
12	EXTENDED OCR MODULE	148
12.1	JOB MANAGER STEPS	148
12.2	IHQC MODULE.....	149
12.2.1	<i>IHQC properties.....</i>	<i>149</i>
12.2.2	<i>Command Line Interface</i>	<i>150</i>
13	MULTICORE MODULE.....	151
13.1	USING THE MULTICORE MODULE	151
13.2	MULTICORE THROTTLING	152
13.3	COMMAND LINE INTERFACE.....	152
14	BARCODE SUPPORT.....	153
14.1	BARCODE FORMATS	153
15	HIGH AVAILABILITY	154
15.1	BACKGROUND INFORMATION	154
15.2	CONFIGURING THE JOB ON HOST ONE:.....	154
15.3	CONFIGURING THE JOB ON HOST TWO:.....	155
15.4	CONFIGURING THE SCHEDULER	156
16	ANY FILE TO PDF.....	157
16.1	OCR ANY FILE TO PDF	157
16.1.1	<i>Configuration.....</i>	<i>157</i>
16.2	CONVERT ANY FILE TO PDF	158
16.2.1	<i>Overview of the Different Conversions Methods.....</i>	<i>158</i>
16.2.2	<i>Comparison of the features provided by the different methods</i>	<i>159</i>
16.2.3	<i>Mappings Between Configuration File and the Conversion Methods.....</i>	<i>160</i>
17	DISTRIBUTED POLLING	162
17.1	JOB SETUP	163
17.1.1	<i>Distributed Polling Job.....</i>	<i>164</i>
17.1.2	<i>Autobahn Job.....</i>	<i>165</i>
18	CLOUD OCR	166
18.1	MICROSOFT COMPUTER VISION.....	170
18.1.1	<i>Pricing</i>	<i>171</i>
18.2	GOOGLE CLOUD VISION	171
18.2.1	<i>Pricing</i>	<i>171</i>
19	KINGFISHER JOB STEP.....	172
20	FILE NAME LENGTH	173
20.1.1	<i>File Name Length - UNC Paths.....</i>	<i>173</i>
20.1.2	<i>File Name Length – Autobahn DX Working Folder Paths.....</i>	<i>173</i>
20.1.3	<i>Long File Names Support.....</i>	<i>173</i>

21	AUTOBAHN DX CONFIGURATION FILES	175
21.1	AUTOBAHN.CONFIG	175
21.1.1	<i>The Input Delay</i>	175
21.1.2	<i>Multi Core Throttle Settings</i>	175
21.1.3	<i>Output Trigger Files</i>	175
21.1.4	<i>The Merge Flag Feature</i>	176
21.1.5	<i>Allow Long File Names</i>	176
21.2	TIFFJUNCTION.EXE.CONFIG.....	177
21.3	PJ.EXE.CONFIG	178
21.4	OTHERS.....	178
22	AUTOBAHN DX DIRECTORIES	179
23	ACKNOWLEDGMENTS	180

1 PRODUCT OVERVIEW AND CONCEPTS

1.1 Overview

Autobahn DX is a Document Processing product designed to fit into an organization's document workflow. It can process and convert a variety of different types of document including TIFF images, Microsoft Office documents, and HTML pages, convert those documents to searchable PDFs and perform a variety of other customizable activities.

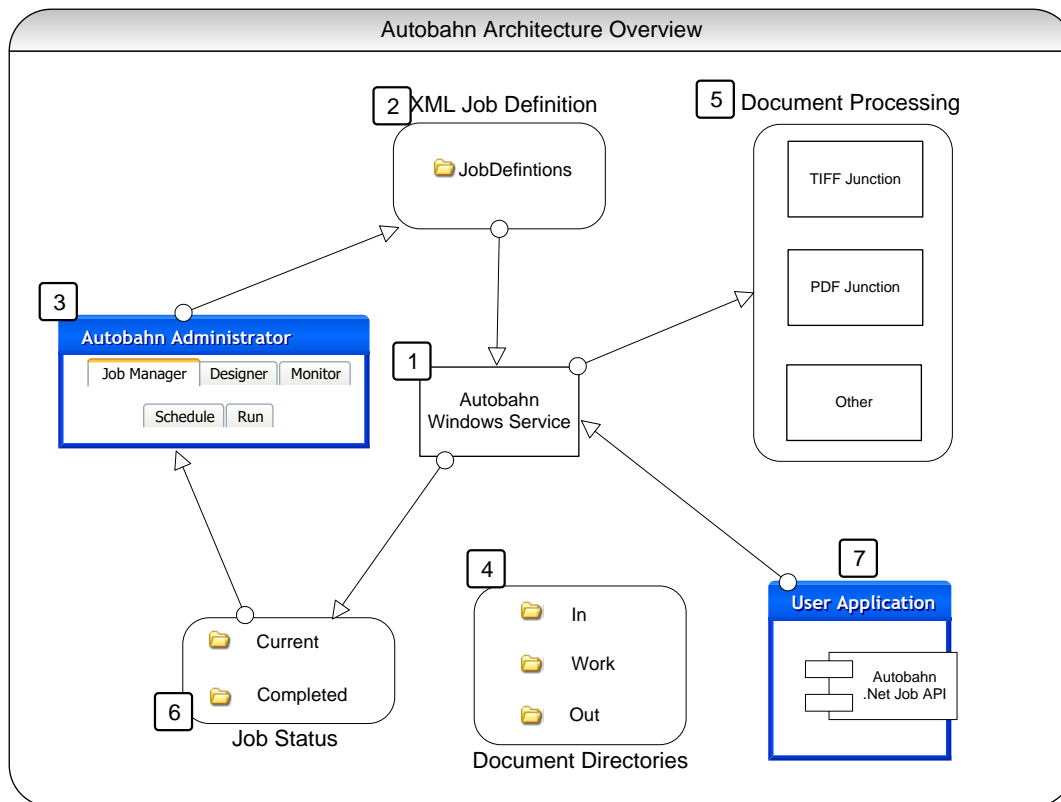
1.1.1 Edition Comparison

The Table below shows you the difference between the different editions of Autobahn DX

Edition Comparison	Workstation	Server with Standard OCR	Server with Extended OCR
Convert TIFF, JPEG, BMP, PNG, GIF to PDF + OCR	✓	✓	✓
OCR Image Only PDF	✓	✓	✓
Convert MS Office & Open Files to PDF	✓	✓	✓
OCR Image Only PDF	✓	✓	✓
Merge TIFF, JPEG, BMP, PNG, GIF to TIFF to TIFF/PDF	✓	✓	✓
Split/Merge PDF Files	✓	✓	✓
Extract Text from PDF Files	✓	✓	✓
Set PDF Properties	✓	✓	✓
Pre-Processing (Deskew, Despeckle, Line Removal) and Auto-Rotation	✓	✓	✓
Support for 23 Languages	✓	✓	✓
TXT, RTF & HTML Output options	✓	✓	✓
Advanced MRC & JBIG2 Compression on output PDFs	✓	✓	✓
Processing Source PDFs with Passwords	✓	✓	✓
Setting PDF Security on Output Files	✓	✓	✓

Split/Rename by Barcode	✓	✓	✓
Multicore (up to a maximum of 64 Cores)		✓	✓
Watched Folders, Scheduled Tasks, and Windows Service Support		✓	✓
Multistep Jobs, allowing the creation of workflows		✓	✓
SharePoint/Office 365 Upload/Download		✓	✓
Read/Send Emails using IMAP4		✓	✓
Use of Custom Scripts		✓	✓
High Availability		✓	✓
Distributed Polling across Multiple Autobahn DX Instances		✓	✓
Integration with Aquaforest Kingfisher		✓	✓
XML-based Job Definitions		✓	✓
.Net API		✓	✓
Run Jobs via Command Line		✓	✓
Support for 129 languages			✓
Better at dealing with poorer quality documents than Standard OCR			✓
Support for multiple languages within a single document from the same character set			✓
Asian Language Support (Optional add-on)			✓
Arabic Language Support (Optional add-on)			✓
Intelligent High-Quality Compression (Optional add-on)			✓
Multiple document output formats: PDF, DOCX, WORDML, RTF, CSV, XLSX, EXCELML, TXT, HTML and XPS			✓

1.1.2 Autobahn DX Architecture and Concepts



1.2 Autobahn Windows Service

This is the heart of the product and controls the execution of both scheduled jobs and ad-hoc jobs whether submitted via the Autobahn Manager or via the Autobahn Job API. The service analyses the XML Job Definition files on start-up and when new files are created in the Job Definition directory by the Autobahn Manager, or via the Job API. The XML Job Definition Files describe the steps to be carried out to complete the job and the Autobahn Service will spawn sub-jobs (such as TIFF Junction or PDF Junction) where required. Job Status records and logs are maintained and can be reviewed in the Job Monitor and Job Manager.

1.3 XML Job Definition Files

Each Job has one or more XML Job Ticket files that define many aspects of the job including:

- Job ID
- Job Name
- Scheduling Information
- Input Directory
- Work Directory
- Output Directory
- Etc.

Check the [Job Definition XML Files](#) section for more details

1.4 Autobahn Administration

This is a windows application that provides the primary administration interface and comprises the following main sections:

1.4.1 Quick Job

Jobs can be defined and run interactively which provides a convenient method for testing the product's capabilities and running simple jobs.

1.4.2 Job Manager

This provides a method of managing the scheduled jobs previously defined in the Process Designer they can be selected and then deleted, edited (in the process designer) or scheduled.

1.4.3 Designer

This provides a graphical interface to allow the creation of a series of steps that make up a job. The process specification can then be saved (as an XML Job File) and run or scheduled via the Job Manager section of the application.

1.4.4 Monitor

This shows the current service status and provides a monitor interface to currently executing jobs. It also allows the service to be stopped or started.

1.5 Document Directories

Each job will have a set of directories containing the source documents, and directories for work, errors, and output.

1.6 Document Processing

Each step in a Job will involve a separate Job Element process being spawned by the Autobahn Service. This may involve its own Sub-Job XML File and always will in the case where the step will execute TIFF Junction or PDF Junction.

1.7 Job Status

To support the Job Monitor, a mechanism is required to indicate the progress of current jobs, start time and the current log file output.

1.8 Autobahn DX .Net Job API

A .Net API is provided to allow a user application to create and execute ad-hoc jobs. See the [AUTOBAHN DX.NET API](#) section for more details.

1.9 File Access Permissions

Administrators of Autobahn DX should be aware that "Quick Job" operations are run in the context of the current logged on user so rely on the permissions granted to that user and may to process files on remote file systems may make use of either UNC's or mapped drives visible to that user. Conversely, Jobs in the Job Manager are run by the Autobahn DX windows service, so when accessing remote file systems UNC paths should be used rather than mapped drives. The job will run as the user specified in the "Log On" property page of the service so it may be necessary to change the user ID is required for security purposes.

2 INSTALLATION AND LICENSING

2.1 System Requirements

Supported Operating Systems	Windows 7 (64bit) Windows 8.1 (64bit) Windows 10 (64bit) Windows Server 2008 (64bit) Windows Server 2012 Windows Server 2016
Disk Space	1.1 GB
.Net Framework	4.5.2
Visual C++ Redistributables	Visual C++ 2015 (x64) Visual C++ 2017 (x64) Visual C++ 2017 (x86)
Additional tools	SharePoint Server Client Components SDK (x64)

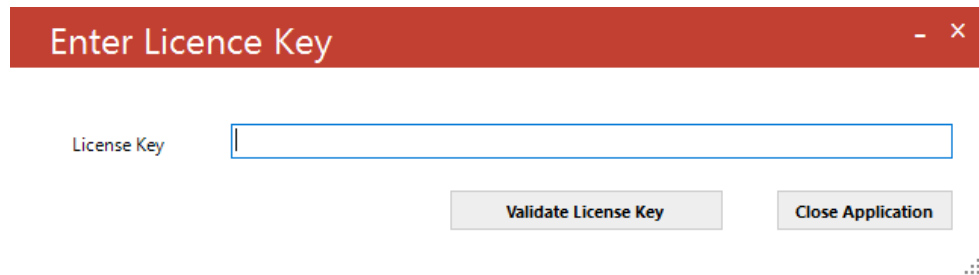
2.2 Product Licensing

Autobahn DX has four license levels which are described below:

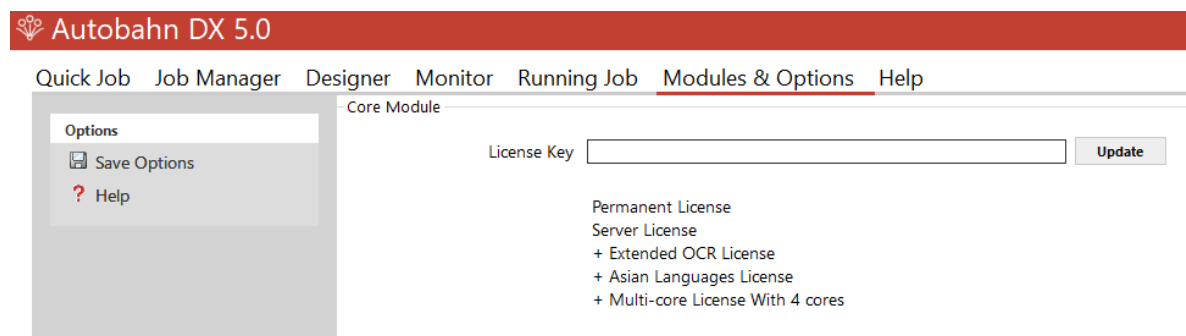
License	Description
Workstation	This license only allows the use of the Quick Job tab.
Server	This license allows the use of all the tabs.
Extended OCR	This module is applied on top of the SERVER license and enables access to the IRIS OCR engine for enhanced recognition.
IHQC	This module is applied to the SERVER + EXTENDED OCR license and enables access to IRIS COMPRESSION.
Asian Language Support	This module is applied to the SERVER + EXTENDED OCR license and enables Asian Language support.
Arabic Language Support	This module is applied to the SERVER + EXTENDED OCR license and enables Arabic Language support.
Multicore	This module is applied on top of the SERVER license and enables the use of a specified number of CPU cores when running jobs.

2.3 Entering License Keys

Autobahn DX cannot start without a valid license key, any time you start Autobahn without a valid license key, it shows you the windows below.



License keys may be entered by using the “Modules & Options” tab.



2.4 Support for Multiple CPU Cores

Your license key will support a specific number of CPU cores. The product will limit the number of concurrent file processing operations to this number and will “throttle” jobs accordingly.

For example, if a 4 core licensed server is currently running a 2 core job and a new job starts that is configured for 4 cores the number of cores allocated to the second job will be reduced accordingly :

Autobahn DX using 2 cores out of 4 allowed.

We will reduce the number of cores in this job from 4 to 2 allowed.

As another example if a 4 core licensed server is currently running a 4 core job and a new job starts that is configured for 2 cores then the second job will not be able to start until cores are freed up :

Autobahn DX using 4 processors out of 4 allowed.

We will attempt to start the job 18 time(s) over the next 180 seconds.

The retry interval and number of tries is determined by these two config file settings in Autobahn.config (by default this file is in C:\Aquaforest\Autobahn DX\config)

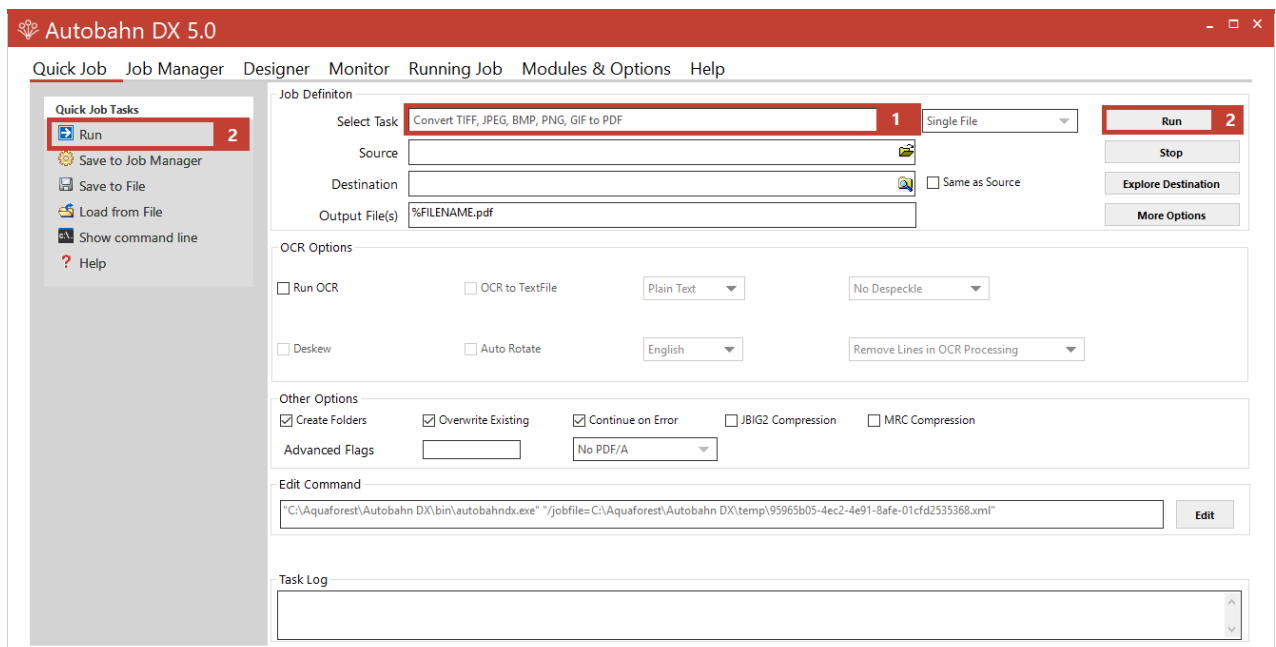
```
<add key="jobqueueinterval" value="180" />
<add key="jobqueueinterval" value="10"/>
```

3 AUTOBAHN DX QUICK START

Autobahn DX is designed to allow the definition of document processing jobs (“Job Definitions”) and then scheduling them to run continuously on watched folders or on a daily scheduled basis. The annotated screenshots in this section are designed to enable rapid familiarization with the product.

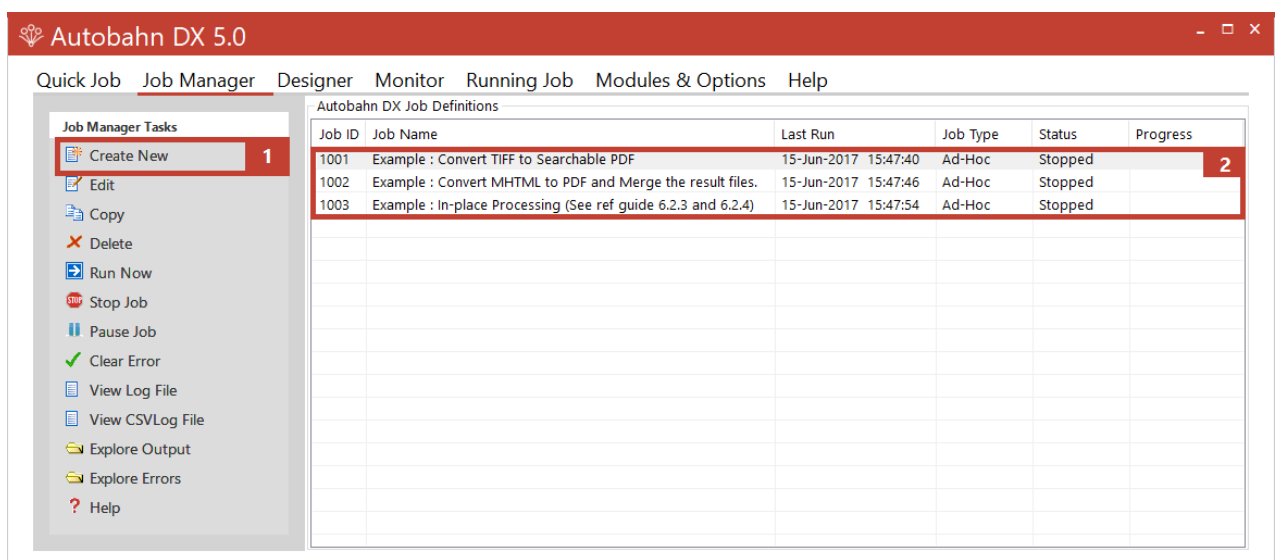
3.1 The Quick Job Tab

The product also allows “Quick Jobs” to be defined and run interactively which provides a convenient method for testing the product’s capabilities and running simple jobs.



1. Use the “Select Task” Combo Box to choose an operation you want to run.
2. Use the “Run” Button to run these tasks

3.2 The Job Manager Tab



The **JOB Manager** tab can be used to create and save jobs that can be executed frequently.

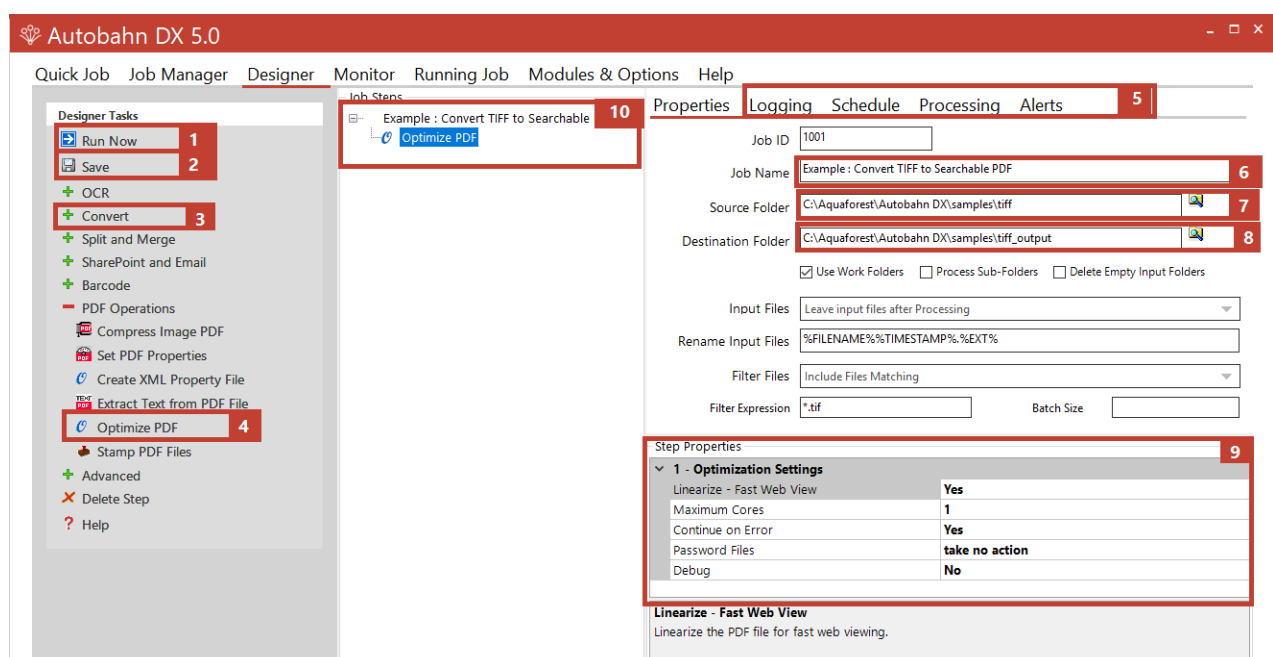
1. Click on **“Create New”** to create your own Job.
2. We provided 3 sample jobs with sample files, this is to show you how to set the jobs up. You can edit these jobs to suit your needs.

Note: Clicking **Create New** or **Edit** will take you to the **Designer** tab

3.3 The Designer Tab

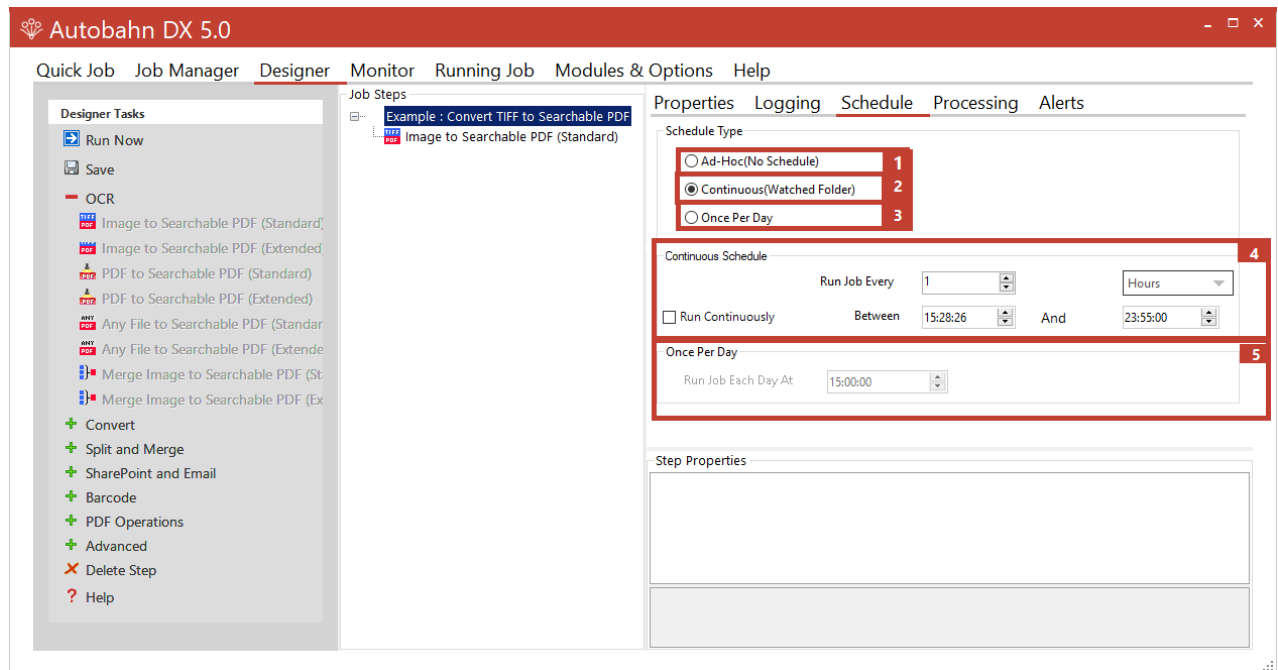
This section will give you a brief overview of the job designer and how jobs are created.

3.3.1 General Job Properties



1. Click the **“Run Now”** Button to run a job that has been created.
2. Click the **“Save”** Button to save any changes made to a job.
3. We arranged similar Operations/Tasks into groups with a **“+”** Button in front of the header name. Click the **“+”** Button to expand the steps under the group.
4. Click on the **“Task”** you want to add to the job and it will show up in the **“Job Steps”** box, one job can have multiple steps, i.e. the output of the first step will be input in the second step.
5. Switch between the tabs to edit the advanced settings in the job.
6. Edit the Job Name using the **“Job Name”** text box.
7. Set the **Source Folder** here.
8. Set the **Destination Folder** here.
9. Set the **Step Properties** (These are the properties of the operation you selected in point 4) here.
10. You can select a step in the **“Job Steps”** list to edit the steps, change its position and delete the step.

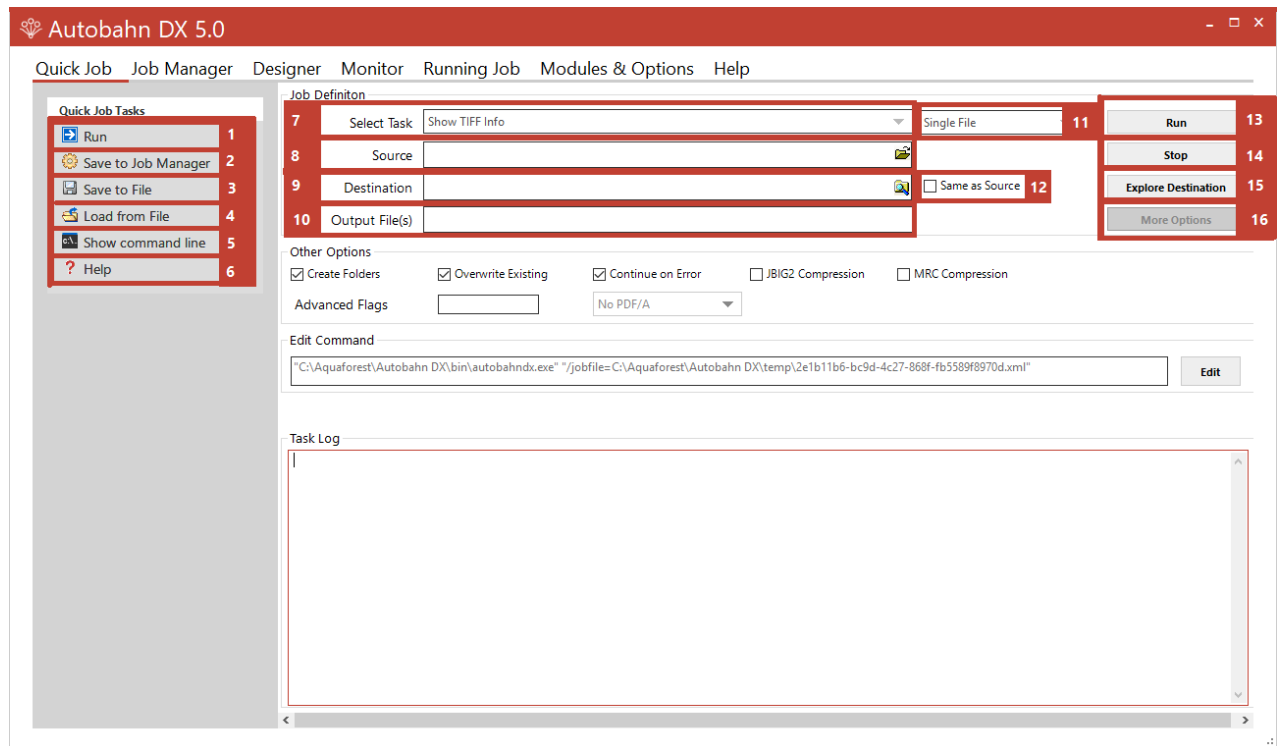
3.3.2 Scheduling a job



1. Choose this option to run the job manually.
2. Choose this option to run the job **Continuously** between a time range.
3. Choose this option to run a job **Once Per Day** at a certain given time.
4. Set the properties for the **Continuous** schedule option here.
5. Set the properties for the **Once Per Day** schedule option here.

4 USING THE QUICK JOB TAB

“Quick Jobs” can be defined and run interactively which provides a convenient method for testing the product’s capabilities and running simple jobs. The job definitions can be saved to the Job Manager for scheduling and further refinement.



1. **Run:** Runs the job using the current Quick Job definition.
2. **Save to Job Manager:** Creates a new job in the Autobahn DX job manager using the current Quick Job definition.
3. **Save to File:** Save the current Quick Job definition to an XML file.
4. **Load from File:** Loads a Quick Job definition from an XML file.
5. **Show Command Line:** Shows the command line equivalent of the current Quick Job definition. [Requires the Server license]
6. **Help:** Clicking this Button will take you to the **Help Tab**
7. **Select Task:** This defines which task is to be run and may be one of:
 - a. Convert a TIFF file to PDF
 - b. OCR an Image-Only PDF File

- c. Convert any file to PDF
 - d. Set PDF Properties
 - e. Merge TIFF Files
 - f. Merge TIFF Files to PDF
 - g. Split a TIFF File
 - h. Show Information for a TIFF File
 - i. Split PDF
 - j. Merge PDF
 - k. Convert a PDF file to TIFF
 - l. Extract Text from PDF
 - m. Convert Any File to Searchable PDF
 - n. TIFF/PDF Barcode
- 8. Source:** The source file or directory.
- 9. Destination:** The location where the generated TIFF or PDF file(s) will be placed. This can be set to be the same as the source location.
- 10. Output File(s):**
- 11. Source Type:** Use this Combo Box to choose if you want to process a file, folder or a tree.
- 12. Same as Source:** Check this box to set the target directory to be the same value as the source directory.
- 13. Run:** Runs the current job. Job output will appear in the “task log section of the screen”.
- 14. Stop:** Halts processing of the currently running job.
- 15. Explore Destination:** Launches Windows Explorer using the destination as the starting point.
- 16. More Options:** This launches a set of property sheets which can be used to set properties of the generated PDF files, such as security and metadata values.

4.1.1 OCR Options Section

Screen Field / Button	Description
Searchable PDF	PDFs generated from TIFF files include searchable hidden text recognized by Autobahn DX's OCR Engine.
OCR to Text File (Professional Edition Only)	Create a separate text file of text recognized by Autobahn DX's OCR Engine.
Other Options (Deskew, Autorotate, Despeckle)	See the standard engine section for further explanation of the OCR options.

4.1.2 OCR PDF Options Section

Screen Field / Button	Description
Image Method	<p>This allows control over the method used to extract images from PDF files for OCR processing when using the “OCR an Image-Only PDF”. The options are:</p> <ul style="list-style-type: none">• Auto – Autobahn will select the most appropriate method. This will be “Convert to TIFF” for image-only PDFs and “Native” otherwise.• Convert to TIFF – The PDF is rasterized using conversion to PDF.• Native - This method places the OCR'd text directly into a copy of the original PDF rather than creating an entirely new PDF. <p>Note: methods other than “Native” change the non-image parts of the PDF to image + text. IMPORTANT NOTE: As the “Native” method does not make any changes to the images in the PDF file, it is not compatible with the JBIG2 or MRC compression, PDF/A compliance option. If these features are required, then it is recommended that the “Convert to TIFF” setting for the PDF OCR image extraction method.</p>
Non-Image PDFs	<p>This allows control over the treatment of non-image PDFs, i.e. PDFs that have some text in them as well as images. The options are:</p> <ul style="list-style-type: none">• OCR: The document will be OCR'd using the image method defined by “Image Method”• Raise Error: The task will terminate with an error. If “On Error Continue” is set this then behaves as Skip. This is the default.• Skip: The document will not be processed.• Pass Through: The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.
DPI	<p>Image PDFs are converted to TIFF for OCR processing. This dropdown allows the DPI for the converted TIFF to be specified. If left on “Auto” the DPI will be determined from the original images in the PDF file.</p>
Retain Bookmarks	<p>Bookmarks from the original PDF are copied to the searchable result PDF.</p>

Retain Metadata	Metadata from the original PDF are copied to the searchable result PDF.
-----------------	---

4.1.3 PDF to TIFF Options Section

Screen Field / Button	Description
TIFF Compression	For black & white documents “Group 4” should be selected. For color, “LZW” should be used.
TIFF Resolution	This specifies the DPI resolution.

4.1.4 Split Options Section

Screen Field / Button	Description				
Split Type	<i>One of:</i> Split into single pages Split by ranges (See below) Split by repeating ranges (See below)				
Page Ranges	Set of page ranges separated by commas that define which pages from the original should be extracted. The following types of page ranges are allowed : <table border="1"> <tr> <td>1</td><td>Specifies a single page</td></tr> <tr> <td>1-3</td><td>Specifies a range of pages</td></tr> </table>	1	Specifies a single page	1-3	Specifies a range of pages
1	Specifies a single page				
1-3	Specifies a range of pages				
Repeating Range	Apply the page range to each set of <i>Page Ranges</i> within the document. For example, if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages. Hence the file is split such that the first output file contains pages 2-4 from the original document; the second contains pages 6-8 and so on.				

4.1.5 Other Options Section

Screen Field / Button	Description
Create Directories (where required)	If checked, when processing a tree, subdirectories will be created in the target directory tree when required.
Continue on Error	When checked, folder and tree processing jobs will continue even if there is an error processing the individual file.

PDF/A-1b PDF/A-2b PDF/A-3b	Processes the output PDF file to ensure it is compliant with ISO 19005 / PDF/A-1b, PDF/A-2b or PDF/A-3b, See section 6 for more details.
JBIG2 Compression	If this option is chosen, bitonal images in the output PDF will be compressed using JBIG2 compression rather than the default of Group 4. This will result in smaller files for PDFs with black and white images.
MRC Compression	This enables Mixed Raster Compression which can dramatically reduce the output size of PDFs comprising color scans. Note that this option cannot be used in conjunction with the “Native” Image Method for OCRing PDFs as the Native method does not adjust the images within the PDF.
Advanced Flags	Additional TIFF Junction advanced command-line flags may be entered here (see section 3 of the TIFF Junction reference guide in the distribution/tj/docs folder)

4.1.6 Barcode Options

Screen Field / Button	Description
Split By Barcode	Choose this option to split the TIFF/PDF file by Barcode
Rename By Barcode	Choose this option to rename the TIFF/PDF file based on Barcode.
Barcode Type	Barcode formats supported.
Split options	Various Options for splitting Files by Barcode <ul style="list-style-type: none"> • Barcode on First Page • Barcode on Last Page • Remove Barcode Page
Try Harder	Spend more time to try to find a barcode; optimize for accuracy, not speed. The default is true.
Output File(s) With Barcode	The output file path template where the split files will be saved. <ul style="list-style-type: none"> - %VALUE%: Replaced by the barcode value found. - %INDEX%: Replaced by the current split index. - %FILENAME: Replaced by the file name

Output File(s) No Barcode	<p>The renaming template to use for page ranges where no barcodes were identified.</p> <p>Allowed templates:</p> <ul style="list-style-type: none"> - %INDEX%: Replaced by the current split index. - %FILENAME: Replaced by the filename of the source file.
Advanced Flags	<p>This enables advanced PDF Junction and TIFF Junction flags to be used, even if there is no direct equivalent in autobahndx.exe.</p>

4.1.7 Task Log Section

Screen Field / Button	Description
Command	<p>This is the operating system command that is used to run the job. It can be edited after using the "Edit" button, although there is normally no need for this.</p>
Task Output	<p>The output of the job appears here.</p>

4.2 Security Settings

The screenshot shows the 'Document Options' dialog box with a red header bar containing a close button (X). Below the header, there are four tabs: 'Security', 'Metadata', 'Document Options', and 'PDF Conversion Settings'. The 'Document Options' tab is selected. The dialog is divided into two main sections: 'Source Document Security' and 'Target Document Security'. The 'Source Document Security' section has a 'Password' field with a red box and the number '1' next to it. The 'Target Document Security' section has three fields: 'Encryption Strength' (a dropdown menu showing 'No Security' with a red box and the number '2'), 'Owner Password' (a text field with a red box and the number '3'), and 'User Password' (a text field with a red box and the number '4'). Below these fields is a section titled 'Target Document Permissions' with a red box and the number '5'. This section contains nine checkboxes arranged in a 3x3 grid: 'Allow All Permissions', 'Printing', 'Copy', 'Fill', 'Assembly', 'Modify Contents', 'Modify Annotations', 'Screen Readers', and 'Degraded printing'. At the bottom of the dialog are 'OK' and 'Cancel' buttons.

1. **Source Password:** A password that will be required to open the source document.
2. **Encryption Strength:** Set this to 128 bits if you want to set security attributes
3. **User Password:** A password that will be required to open the target document.
4. **Owner Password:** A PDF owner password is a password used to set certain document restrictions on the target PDF.
5. **Target Document Permissions:** Document restrictions put in place with an owner password can include the following.
 - a. **Allow ALL Permissions:** All the permissions below.
 - b. **Printing:** Allow high-quality printing
 - c. **Modify Contents:** Allow assembly (see below) and other document medications
 - d. **Copy:** Allow text and graphics copying and extraction
 - e. **Modify Annotations:** Allow modification of annotations
 - f. **Fill:** Allow filling of form fields
 - g. **Screen Readers:** Allow extraction of text and graphics in support of accessibility.
 - h. **Assembly:** Allow rotation, insertion or deletion of pages.
 - i. **Degraded Printing:** Allow low-quality printing

4.3 Document Metadata

This screen allows setting metadata to the generated PDF.

Document Options

Security

Metadata

Document Options

PDF Conversion Settings

Document Metadata

Author

Creator

Title

Subject

Keywords

OK

Cancel

4.4 Document Options

Document Options

Security Metadata Document Options PDF Conversion Settings

Display Modes

Page LayoutSingle Page

Page ModeNeither Bookmarks nor Thumbnails Open

Non-Full Screen ModeNeither Bookmarks nor Thumbnails

Settings

☐ Hide Menu Bar

☐ Hide Tool Bar

☐ Hide Window UI

☐ Fit Window

☐ Center Window

Page Settings and Margins

Width

Left

Top

Height

Right

Bottom

Paper OrientationDefault

OK

Cancel

4.4.1 Display Modes

Option	Description
Page Layout	The setting for the initial document page display. One of : <ul style="list-style-type: none">• Single Page• Continuous• Continuous Facing (odd pages left)• Continuous Facing (odd pages right)
Page Mode	The setting for initial viewer mode. One of : <ul style="list-style-type: none">• Neither Bookmarks nor Thumbnails Open• Bookmarks Open• Thumbnails Open• Bookmarks & Thumbnails Open• Full Screen
Non-Full Screen Mode	Only applicable where Page Mode=Full Screen. The setting for document page display when exiting Full-Screen mode. One of : <ul style="list-style-type: none">• Neither Bookmarks nor Thumbnails Open• Bookmarks Open• Thumbnails Open

4.4.2 Settings

Option	Description
Hide Tool Bar	The viewer's toolbar will be hidden
Hide Menu Bar	The viewer's menu bar will be hidden
Hide Window UI	The viewer's UI elements (scrollbars etc.) will be hidden
Fit Window	The viewer will resize the document's window to fit the size of the first displayed page.
Centre Window	The document window will be positioned in the center of the screen.


4.5 PDF Conversion Settings

Document Options

Security Metadata Document Options PDF Conversion Settings

Conversion Control

Conversion Timeout(Seconds)

Conversion Error Folder 

Microsoft Office Conversion Settings

☐ Convert Hyperlinks

☐ Merge PDF Attachment

☐ Print BG Color(IE)

☐ Retain Attachments

☐ Retain Properties

☐ Print All Sheets(Excel)

☐ Convert Bookmarks

☐ Frame Slides

Depth

Zoom (Excel)

Output Type (PP)

Print Scale %(Visio)

Color Type (PP)

Header(IE)

Handout Order (PP)

Footer(IE)

Print Graphics (Pub)

Image and Font Conversion Settings

Image Compression

☐ Embed Fonts as Type 0

☐ Image Downsizing

Font Embedding

Image JPEG Quality(5-100)

Font Substitution

DownSize Resolution(DPI)

OK

Cancel

4.5.1 Conversion Control

Option	Description
Conversion Timeout	Set timeouts for converting files in seconds.
Conversion Error Folder	The folder where all the error files will be copied to.

4.5.2 Microsoft Office Conversion Settings

Option	Description
Convert Bookmarks	For MS Word, convert bookmarks
Depth	<p>This property will take effect only when the ConvertBookmarks property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, an empty string is functionally equivalent to</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use a non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.</p>
Convert Hyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
Print All Sheets	The flag that indicates whether to print all Excel worksheets or not.
Print BG Color	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
Print Scale	For Visio files, sets the print scale
Header	This property modifies Internet Explorer's header setting.
Footer	This property modifies Internet Explorer's footer setting.
Zoom	Sets printing zoom of the worksheet. The allowed value range is from 10 to 400.
FitToPagesWide	Sets number of pages wide the worksheet will be scaled to. This property is ignored if the Zoom property is set.

FitToPagesTall	Sets number of pages tall the worksheet will be scaled to. This property is ignored if the Zoom property is set.
Merge PDF Attachments	Set this flag to true if you want if you want converted pdf attachments to merged to the output pdf file. Otherwise, the converted files will be merged back to the pdf.
Retain Original Attachment	Set this flag to true if you want the original attachments to be retained in the pdf file after they are merged. This flag only gets picked up if the MergeAttachments flag above is used.
Retain Properties	Set this flag if you want the MS Office properties to be transferred to the target pdf document.
Handout Order	Sets the handout order, this flag only applies to PowerPoint jobs. The possible values are: <ul style="list-style-type: none"> • Vertical First • Horizontal First
Output Type	Sets the output type, it only works with the PowerPoint files. The possible values are: <ul style="list-style-type: none"> • Slides • Build slides • Two slides handouts • Three slides handouts • Four slides handouts • Six slides handouts • Nine slides handouts • Notes • Outline
Print Graphics	Sets the graphics setting for printing. <ul style="list-style-type: none"> • Print Full Resolution • Print Low Resolution • Print Graphics

4.5.3 Image and Font Conversion Settings

Option	Description
Font Embedding	The option PRN_FONT_EMBED_FULLSET (embedding a full set of font) will cause a significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed the font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
Font Substitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from the Control Panel).
Embed Fonts as Type 0	This option is recommended if you have non-standard fonts like barcode font.
Top Margin	Sets top margin.
Bottom	Sets bottom margin.
Left	Sets left a margin.
Right	Sets right margin.
Width	Sets a custom page width.
Height	Sets a custom page height.
Paper Orientation	Sets paper orientation e.g. PRN_PAPER_ORIENT_LANDSCAPE
Image Compression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
Image Downsizing	If this property is set to True, then the resolution of images is reduced to the DPI value specified in the ImageDownsizeResolution property.
Image JPEG Quality	The allowed value range is from 5 to 100 with 100 being the highest quality.
Downsize Resolution	If the ImageDownsizing property is set to True, then the resolution of images is reduced to the DPI value specified in this property.

4.6 Advanced Flags

The Advanced Flags field allows entry of certain options that cannot be entered via the GUI.

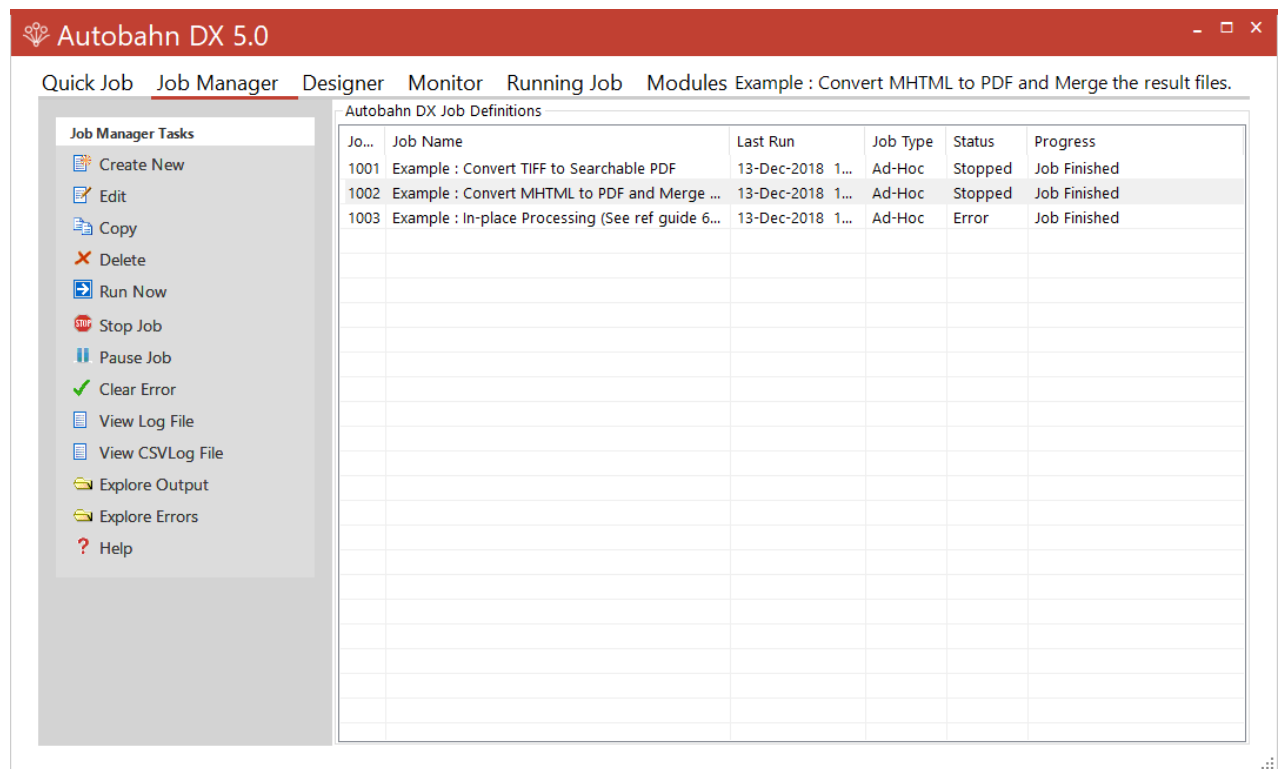
Option	Description
-B 100	<p>Blank Page Removal</p> <p>This option can be used when converting TIFF files to Searchable PDFs. A value should be provided which specifies the pixel threshold to be used to determine whether a page is blank or not. If a page is deemed to be blank it is omitted from the output file. A suggested value is 100 i.e. using the new –B advanced flag: -B 100</p>
-D	<p>Dot Matrix Processing</p> <p>This option will significantly improve recognition for dot matrix documents but should not be used when processing other document types as it will have a negative impact on non-dot matrix documents.</p>
-0 [zero]	<p>Custom Despeckle</p> <p>When used in combination with the despeckle option, this has the effect of de-speckling the page for OCR processing, but the image retained in the final file is not de-speckled.</p>
-q	<p>Custom Quantization</p> <p>This command line option should generally only be used under guidance from technical support. It can control the way that color images are processed and force binarization with a particular threshold. (For example -q 127).</p>

5 THE AUTOBAHN DX ADMINISTRATION TOOL

The Autobahn administration tool is a Windows Forms application designed to enable system administrators to be able to create and configure Job Definitions that control document processing. Several sample Job Definitions are included with the product which can be tailored to meet a user's precise requirements.

5.1 Job Manager

This section of the application displays a list of all current Job Definitions and the set of operations to be performed shown in the menu table below.

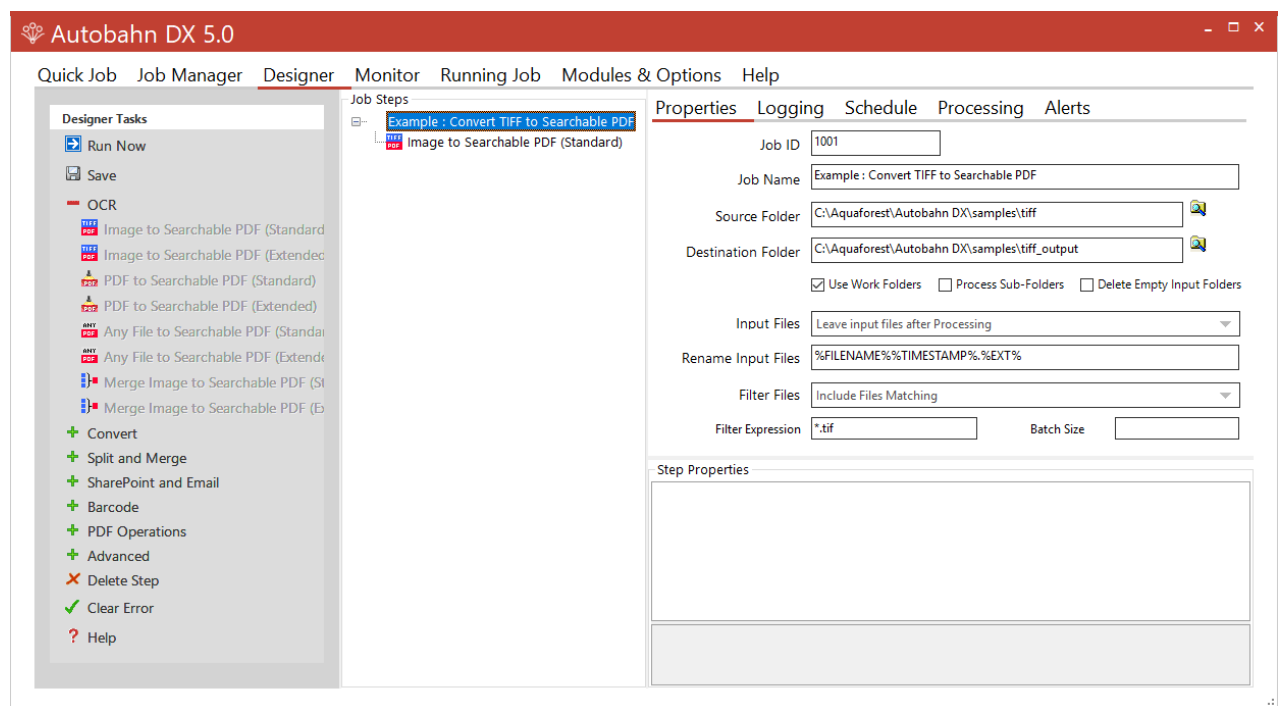


Menu	Item	Action
Create New		Creates a new, skeleton job definition from jobtmplt.xml and then switches the display to the Job Designer.
Edit		Switches the view to Edit on the selected Job.
Copy		Creates a copy of the selected job. The job name will be Job %JOBID% (Copy of %JOBNAME%)
Delete		Deletes the selected job definition after confirmation.
Run Now		Executes the selected job via the interface in the Run tab window.

Stop Job	Stops a running job
Pause Job	Pauses a job that is running
Clear Error	Clears Jobs in error
View Log File	Displays the most recent log file for the job.
Explore Output	Launches windows explorer on the job output folder.
Explore Errors	Launches windows explorer on the job error folder.
Help	Takes you to the help tab in the Autobahn DX GUI

5.2 Job Designer

This allows definition and editing of a job definition, using a tree-list type model coupled with a Visual Studio – style property list. The different step types are listed on the left under the **Designer Task** group box. The step types have been grouped into sub categories, each step type will have its own icon. Drag and drop can be used to allow reordering of steps.



5.2.1 Menu Items

Menu	Item	Action
	Run Now	Executes the job that is being edited, the output is displayed in the Run tab screen.
	Save	Validates the current job and if valid, save the current job definition to %JOBID%.xml in the %JOBDEFDIR% directory.

OCR	<p>This expander contains the steps that perform OCR, autobahn will grey out the invalid steps. The steps types in these groups are:</p> <ul style="list-style-type: none"> • Image to Searchable PDF (Standard) • Image to Searchable PDF (Extended) • PDF to Searchable PDF (Standard) • PDF to Searchable PDF (Extended) • Any File to Searchable PDF (Standard) • Any File to Searchable PDF (Extended) • Merge Image to Searchable PDF (Standard) • Merge Image to Searchable PDF (Extended)
Convert	<ul style="list-style-type: none"> • Convert PDF to TIFF • Convert any File to PDF
Split and Merge	<ul style="list-style-type: none"> • Merge PDF • Split PDF • Merge TIFF • Spilt TIFF
Connectors	<ul style="list-style-type: none"> • Read Mailbox • Send Documents • SharePoint Download • SharePoint Upload • Azure Storage Download • Azure Storage Upload
Barcode	<ul style="list-style-type: none"> • Barcode TIFF/PDF
PDF Operations	<ul style="list-style-type: none"> • Compress Image PDF • Set PDF Properties • Create XML Property File • Extract Text from PDF File • Optimize PDF • Stamp PDF Files
Advanced	<ul style="list-style-type: none"> • Custom Script Step • High Availability • Kingfisher Job • Distributed Polling
Delete Step	Deletes the currently selected step node.
Clear Error	Click this before you can run a job that is in an error state
Help	Returns to the job manager with the current job selected.

5.2.2 Fields

Menu Item	Description
Job ID	A sequential Job ID is allocated for the Job by Autobahn DX. This cannot be changed.
Job Name	A descriptive title for the job.
Source Folder	The folder containing the documents to be processed.
Target Folder	The folder where the processed files will be placed if “Move input files to target folder after processing” is chosen.
Process Sub-Folders	If checked, all sub-folders will be recursively processed.
Always Use Work Folders	By default, Autobahn DX processes job steps by using a separate folder for each step. Hence files from the source folder are copied to a work folder, processed for each step to another work folder and then finally to the target folder. This approach ensures integrity (e.g. correctly processing files that are added to the source folder after a job has started) but can slow down large jobs.
Delete Empty Input Folders	Checking this property will delete empty folders under the source folder after we move or delete your input files.
Input files	<p>This option determines what happens to the input files once processing has been completed. The options are :</p> <p>Leave input files after processing: Files are left in the Source Folder.</p> <p>Move to archive after processing: Files are moved to the Archive Folder.</p> <p>Copy to archive after processing: Files are copied to the Archive Folder.</p> <p>Move input files to target folder after processing: Input files are placed in the same folder as the output files.</p> <p>Delete input files after successful processing: Input files are deleted.</p>
Input Files renaming	<p>This determines how input files will be renamed when moved to the Target or Archive folder. The default is :</p> <p>%FILENAME%%TIMESTAMP%.%EXT%</p>

Filter Files	<table><tr><th>Filter Files Option</th><th>Description</th></tr><tr><td>Include Files Matching</td><td>Only files matching the Filter Expression are included.</td></tr><tr><td>Exclude</td><td>Files matching the Filter Expression are excluded.</td></tr><tr><td>Include with Document Count Limit</td><td>For example “*.pdf; 3000” would limit the job to 3000 PDF files.</td></tr><tr><td>Include Unprocessed PDFs Only</td><td><p>This would limit files selected to PDFs that have not been OCR'd.</p><p>A file is deemed to have been OCR'd if either</p><ul style="list-style-type: none">(a) It has a custom metadata tag “AQUAFORESTOCR”(b) It has one image per page and only has “invisible” text.<p>This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</p></td></tr><tr><td>Include Unprocessed PDFs Only – with Document Count Limit</td><td>As above, but limited to the number of files specified in the filter.</td></tr></table>	Filter Files Option	Description	Include Files Matching	Only files matching the Filter Expression are included.	Exclude	Files matching the Filter Expression are excluded.	Include with Document Count Limit	For example “*.pdf; 3000” would limit the job to 3000 PDF files.	Include Unprocessed PDFs Only	<p>This would limit files selected to PDFs that have not been OCR'd.</p> <p>A file is deemed to have been OCR'd if either</p> <ul style="list-style-type: none">(a) It has a custom metadata tag “AQUAFORESTOCR”(b) It has one image per page and only has “invisible” text. <p>This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</p>	Include Unprocessed PDFs Only – with Document Count Limit	As above, but limited to the number of files specified in the filter.
	Filter Files Option	Description											
	Include Files Matching	Only files matching the Filter Expression are included.											
	Exclude	Files matching the Filter Expression are excluded.											
	Include with Document Count Limit	For example “*.pdf; 3000” would limit the job to 3000 PDF files.											
	Include Unprocessed PDFs Only	<p>This would limit files selected to PDFs that have not been OCR'd.</p> <p>A file is deemed to have been OCR'd if either</p> <ul style="list-style-type: none">(a) It has a custom metadata tag “AQUAFORESTOCR”(b) It has one image per page and only has “invisible” text. <p>This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</p>											
	Include Unprocessed PDFs Only – with Document Count Limit	As above, but limited to the number of files specified in the filter.											
N.B.: Work Folders must be used to enable the use of filters.													
Filter Expression	<p>One or more search options used to determine the files in the source folder that should be processed. Multiple expressions may be used, separated by spaces.</p> <p>Examples :</p> <p>*.pdf *.doc *.ppt *.xls</p>												
Batch Size	Limits the number of documents to be processed to the given size. To use this feature, you must use a “Filter File Option” with “Document Count Limit”.												
Log File	Path of the job log file. This will include %DATESTAMP%, which is the date of the day the job started. A new log file will be created for each day.												

CSV Log File	<p>Path of the job log file. This will include %DATESTAMP%, which is the date of the day the job started. A new CSV file will be created for each day. The columns in the CSV file are :</p> <ul style="list-style-type: none"> • Job Start – Time Job Started • Source Files – Full path to the source file • Target File – Full path to the target file • Job Stopped – Time Job Finished • Success – True or False; Files that could not be processed will have a value of False.
Retention Period	<p>This is an integer value representing the number of days the log file will be kept for before being deleted. Leaving it blank or setting it to a number less than one will keep the log files indefinitely.</p>
Stop Processing on Error	<p>If checked, the job will stop if it returns an error, and will not run again until the error is cleared from the Monitor screen.</p>
Skip Long File Names	<p>Check this box to make Autobahn DX skip long filename, if this box is not checked, Autobahn DX will throw an error if it encounters one of these files.</p>
Skip Folders That Autobahn Can't Access	<p>Check this box to make Autobahn DX Folders it has no permissions to access, if this box is not checked, Autobahn DX will throw an error if it encounters one of these folders.</p>
Archive Folder	<p>The folder where the processed files will be placed if “Move to archive after processing” is chosen.</p>
Work Folder	<p>The folder where files will be temporarily stored during conversion and processing.</p>
Error Folder	<p>Source documents that have errors during processing will be placed in the specified folder.</p>
Temp Folder	<p>Some job steps can require a significant amount of temporary storage, particularly those steps involving OCR. This folder defines the location of the temporary space.</p>
Trigger File	<p>You can find this setting under the Processing tab, if you provide a Trigger File value, Autobahn DX will not process a folder until the Trigger File is present, the file will be deleted after each job run.</p>

5.2.3 Job Scheduling

To use the Job Schedule, you will need to click the Schedule tab under the Designer Tab.

The screenshot shows a configuration window with five tabs: Properties, Logging, Schedule, Processing, and Alerts. The 'Schedule' tab is selected and underlined. Below the tabs, there are three sections for scheduling options:

- Schedule Type:** Contains three radio buttons: 'Ad-Hoc(No Schedule)', 'Continuous(Watched Folder)' (which is selected), and 'Once Per Day'.
- Continuous Schedule:** Contains a 'Run Job Every' field with a value of '1' and a unit dropdown set to 'Hours'. Below this is a checkbox labeled 'Run Continuously' which is checked, followed by a 'Between' time range from '00:00:01' to '23:55:00'.
- Once Per Day:** Contains a 'Run Job Each Day At' field with a time value of '13:00:00'.

The product supports three types of scheduling which are implemented via the Autobahn DX service:

5.2.3.1 Ad-Hoc

This means that the job does not have any fixed schedule, but maybe explicitly run via the management GUI or via one of the API methods.

5.2.3.2 Watched Folder / Continuous Scheduling

This allows the job to be scheduled to run periodically between a start time and end time each day. The periods may be seconds, minutes or hours. For example, a job may be specified to run every 30 seconds between 9:00 and 17:00.

If you check the “Run Continuously” checkbox, the job will run for 24 hours a day. This option is the default for all continuous jobs.

5.2.3.3 Daily Scheduling

This allows the job to be scheduled to run at a specified time each day.

5.2.4 Alerts

This allows to send Emails to your mailbox when the job is successful or fails, to get to the Alerts tab, you will need to click the Alerts tab under the Designer Tab.

Note: You will need to enter your SMTP setting in the [modules and options](#) tab before the email alerts will work properly.

☒ Send Email Alerts on Job Completion

☐ Attach Log File ☐ Attach Job Report

From Email Address

To Email Address

Email Title

Email Message

Job: '%JOBNAME%' Status: '%JOBSTATUS%'.

Log: %LOGFILE%.
Time: %DATESTAMP% %TIMESTAMP%
Source: %JOBSOURCE%
Target: %JOBTARGET%

Only Send Email Alerts if ☐ At least one file was processed ☐ At least one file error occurred

☐ Job Terminated Prematurely

Menu Item	Action
Send Email Alerts on Job Completion	If checked, Autobahn DX will send an email if the job ends naturally or pre maturely. This alert can be further tailored using the properties In the section below.
Only Send Email Alerts if	
At least one file was processed	If you check this option, Autobahn DX will not any email until process at least one file in the job. This is meant to reduce the number of irrelevant messages you get.
Job Terminated Prematurely	Check this if you only want to receive emails when an error occurs during the processing of a job. Note: Individual file errors will not put the job in error, a job error occurs in a more fatal circumstance.
At least one file error occurred	Check this option if you only want to receive emails when individual file errors occur.
Attach Log File	Check this option if you want Autobahn DX to attach the Log file of the job to the email alert.

Attach Job Report	Check this option if you want Autobahn DX to attach a report/summary of the job to the email alert.
From Email Address	The “from” email address that will be used for the message.
To Email Address	The email address that the message will be sent to.
Email Title	The title of the email.
Email Message	The body of the email, this can be HTML content.

When sending emails, there are several variables that can be used to customize the alerts you send out, these variables are enclosed in two percent signs “%%”. Autobahn DX will replace any occurrences of the variables with an appropriate value at run time. The table below shows the possible variables that can be used.

Variable	Meaning
%JOBID%	The Job ID, this works with both the email title and email message.
%JOBNAME %	The Job Name, this works with both the email title and email message.
%JOBSTATUS%	The Job Status, this works with both the email title and email message.
%LOGFILE%	The location of the log file, this works with both the email title and email message.
%JOBSOURCE%	The Source Directory of the job, this works with the email message only.
%JOBTARGET%	The Destination Directory of the job, this works with the email message only.
%DATESTAMP%	The date that the alert was generated, this works with both the email title and email message.
%TIMESTAMP%	The time the alert was generated, this works with both the email title and email message.

5.2.5 Workflow Processing versus In-Place Processing

By “in-place” processing, we generally mean processing PDF documents that have already been added to a document repository or system and need to be turned into searchable PDFs *in-place*. By contrast “workflow” processing is where documents pass through Autobahn DX on their way into a document repository via watched folders.

5.2.6 Example In-Place Job Setup

Autobahn DX can be used for in-place processing, but we have a newer product names [Aquaforest Searchlight](#) that is created solely for in-place conversions, it will handle the Use-Case a lot better out of the box.

The job shown below will convert PDFs under the tree “C:\Aquaforest\Autobahn DX\samples\pdf” to searchable PDFs, processing up to 5 files each time the job is run.

Properties	Logging	Schedule	Processing	Alerts
Job ID	1003			
Job Name	Example : In-place Processing (See ref guide 6.2.3 and 6.2.4)			
Source Folder	C:\Aquaforest\Autobahn DX\samples\tiff			1
Destination Folder	C:\Aquaforest\Autobahn DX\samples\tiff_2			
<input checked="" type="checkbox"/> Use Work Folders <input checked="" type="checkbox"/> Process Sub-Folders <input type="checkbox"/> 2 Create Empty Input Folders				
Input Files	Copy to Archive after Processing			3
Rename Input Files	%FILENAME%%TIMESTAMP%.%EXT%			
Filter Files	Include Unprocessed PDFs Only - with Document Count Limit			4
Filter Expression	*.pdf		5	Batch Size 5 6
Step Properties				
▼ 1 - Conversion Settings				
Output File Name	%FILENAME.pdf			7
Create Directories if Required	Yes			
Continue on Error	Yes			
OCR	Yes			
Box / Graphics Options	None			
Output File Name				
Target file template which can include %FILENAME (original filename without the extension) and...				

1. The **Source Folder** and the **Target Folder** should be the same.
2. Enable the processing of Sub-Folders.
3. For Audit Purposes, the original files should be copied to the archive folder.
4. To Avoid re-processing files, select the **Include Unprocessed PDFs Only – with Document Count Limit** option in the **Filter Files** combo box.
5. Only process pdf files.
6. Limit the job to process 5 files per run (You can increase this to a suitable batch size).
7. The Output file should be configured to **%FILENAME.pdf** so that it will replace the input file.

5.2.7 Step Types

This section will explain each of the step types listed in the table in [section 5.2.1](#).

5.2.7.1 Image To Searchable PDF

This step can be found under the **OCR Expander**. It creates a searchable PDF file from input image types e.g. **.png, .tiff, .jpg, .gif, .bmp**.

Depending upon the Step Type Properties chosen, a separate text, HTML and Office files may be produced from the OCR process.

Autobahn DX is shipped out with OCR engines.

- **Standard Engine**
- **Extended Engine**

See [Standard OCR vs Extended OCR](#) for the differences.

5.2.7.1.1 Standard Engine

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Continue on Error	Continue processing TIFF files after an error occurs.
OCR	Choose "No" to generate an image-only PDF. Choose "Yes" to generate searchable PDF and/or text files.
OCR Language	Select the language the original file is written in. This will determine the dictionary that is used.
Deskew	Straighten the image.
Auto-Rotate	Automatically rotate pages so that text flows left to right.
Despeckle	Remove specks below the specified pixel size from the image.
OCR to Text File	Choose "Yes" to Generate text Output
Output File	<ul style="list-style-type: none">• Plain Text (txt).• Plain Text (txt) No PDF

	<ul style="list-style-type: none"> • MS Word (rtf) • HTML
PDF/A Options	<p>Select the output PDF/A compliant version you would like the output PDF to be.</p> <ul style="list-style-type: none"> • PDF/A1-b • PDF/A2-b • PDF/A3-b
Validate PDF/A	Whether or not to validate the PDF/A document after conversion
JBIG2 Compression	This option will compress bitonal images in generated PDFs using JBIG2 compression rather than the default Group 4 compression scheme. This will result in smaller PDF file sizes, at a cost of increased processing time.
Box/Graphics Options	<p>By default, if an area of the document is identified as a graphic area then no OCR processing is run on that area. However, certain documents may include areas or boxes that are identified as “graphics” or “picture” areas but that actually do contain useful text.</p> <p>To ensure that the OCR engine can be forced to process such areas there are two options :</p> <p><i>“Treat all Graphics Areas as Text”.</i> This option will ensure the entire document is processed as text.</p> <p><i>“Remove Box Lines in OCR Processing”.</i> This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the OCR engine. It does not remove boxes from the final image. Technically, this option removes connected elements with a minimum area (by default 100 pixels).</p>
Line Removal in OCR Processing	This removes lines and boxes during OCR processing to improve recognition – particularly in cases where characters “touch” lines.
MRC	This enables Mixed Raster Compression which can dramatically reduce the output size of PDFs comprising Color scans.
Save Pre-Despeckle	This will use the original image (i.e. before applying pre-processing) in the output PDF. The default value is true.
StampName	This has been deprecated, use the Stamp PDF Files step.

StampValue	This has been deprecated, use the Stamp PDF Files step.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	This specifies the number parallel files you want to be processed at a given time. Note: You need the multi-core license for this.
Debug	Set this to true to execute the job in debug mode.

5.2.7.1.2 Extended Engine

Parameter	Notes
Output File Name	The output filename excluding the extension (which will be added according to the output file type).
Output File Type	One or more of the following, separated by commas if more than one is required. RTF PDF DOCX CSV* SML (Spreadsheet ML XML file)* HTM TXT *These output formats are suitable for table-oriented pages that can be mapped onto a spreadsheet format.
Create Folders If Required	Create an output folder if it does not exists. Default true.
OCR Engine	The OCR engine to use. This must be set to use the IRIS engine.
OCR Language 1-5	You can set up to five different languages for OCR recognition in one page as long as they are in the same character set.
Automatic language detection	Property that enables or disables the Auto Language Detection feature. The aim of this feature is to detect the most probable language of a single-language page. If at least one language has been detected, recognition will be performed in the first language candidate that has been detected, and not in the language(s) set through Language or Languages. If it fails to detect a language, recognition will be performed using the language(s) set through Language or Languages.

Auto rotate	Detect page orientation and correct if required
Deskew	Rotates the image to correct its skew angle.
Advanced Deskew	Set this to true to defined advanced deskew properties.
Force Deskew	<p>Under certain circumstances, rotating the image to correct its skew angle may decrease the OCR accuracy. The extended engine is able to analyze the image and detect from an OCR accuracy point of view whether it's better to rotate the image or not. Because the skew angle may be visible in the output document (e.g. if KeepDeskew is set to 'true'), you can choose to force the deskew to rotate the image, even if it affects the accuracy.</p> <p>If turned off, the image is analyzed before rotation and the engine may choose not to rotate the image depending on the analysis result.</p> <p>If turned on, the image is rotated to correct skew angle.</p>
Adjustment Mode	Set the behavior regarding dimension adjustment for deskew operation.
Despeckle	Removes all the groups of connected pixels with a few pixels below the parameter. Suggested range: 1-20.
Advanced Despeckle	Set the advanced despeckle settings, advanced despeckle provides advanced image noise reduction features by the image despeckle filter.
Remove White Pixels	By default, Advanced Despeckle removes black pixels. If this setting is set to 'true', white pixels will be removed instead of black pixels.
Dilate	Despeckle removes all the groups of connected pixels with a few pixels below the SpeckleSize parameter. Those connected pixels are not removed if the distance to a larger connected component is below this parameter. As a result, only the isolated pixels get deleted. The maximum value for this property is 20 pixels. The default value is '0'.
Layout	<p>The layout for the dock or rtf document</p> <ul style="list-style-type: none"> • Standard • Flow
PDFVersion	<p>This determines the PDF version of the generated PDF :</p> <p>1.4</p>

	1.5 1.6 1.7 PDF/A1B (<i>PDF/A-1b</i>)
Remove Blank Page	Set this to true to remove blank pages from Tiff or PDF documents. Value needs to be set for sensitivity (see below).
Sensitivity	The sensitivity, from 1 to 100. With high sensitivity, fewer blank pages are detected.
Work Depth	This parameter (0 – 255) defines how deeply the OCR engine will analyze a page with 255 being the deepest. For poorer quality documents, higher values can give better recognition results.
JPEG Quality	This parameter (0 – 255) determines the compression/quality of color JPEG images in generated PDFs. 0 gives the smallest file size whilst 255 gives the best quality. The default value is 128.
JPEG2000 Compression	Enable/Disable JPEG2000 Compression.
JPEG2000 Compression Mode	The JPEG2000 Compression Mode to use.
JPEG2000 Compression Value	The Value to set for the selected Compression Mode.
IHQC Compression	Apply Intelligent High-Quality Compression
IHQC Compression Level	Level1 is the basic compression level while level 3 is the most advanced Intelligent High-Quality Compression Mode.
IHQC Quality Factor	The quality Factor for IHQC
No OCR	Whether are not to perform OCR on the document (Yes to not perform OCR, No to perform OCR).
Binarization	Whether or not to perform binarization on the document.
Brightness	The brightness (higher values will darker the result).
Contrast	The contrast (lower values will darker the result).
Smoothing Level	Smoothing may be useful to binarize text with a colored background to avoid noisy pixels (0 disables smoothing, higher values smooth more).

Undithering	<p>Whether or not to use automatic undithering while processing a page. NOTE: Automatic undithering will be applied only if smoothing is also activated (Smoothing Level)</p> <p>Dithering is a scanning technique which consists in representing a color or greyscale image using only a limited color palette. This allows reducing file size while maintaining the general aspect of the image. This technique is known to create images more difficult to handle for OCR technology; therefore specific image preprocessing is needed to detect and revert it.</p>
Threshold	Sets the threshold for fixed threshold binarization (0 for automatic threshold computation).
Remove Lines	Whether or not to remove lines from an image (The image must be black and white).
Horizontal Clean X	The parameter for cleaning noisy pixels attached to the horizontal lines.
Horizontal Clean Y	The parameter for cleaning noisy pixels attached to the horizontal lines.
Vertical Clean X	The parameter for cleaning noisy pixels attached to the vertical lines.
Vertical Clean Y	The parameter for cleaning noisy pixels attached to the vertical lines.
Horizontal Dilate	The dilate parameter that helps the detection of horizontal lines.
Vertical Dilate	The dilate parameter that helps the detection of vertical lines.
Horizontal Max Gap	The maximum horizontal line gap to close. It is useful to remove broken lines.
Vertical Max Gap	The maximum vertical line gap to close. It is useful to remove broken lines.
Horizontal Max Thickness	The maximum thickness of the horizontal lines to remove. It is useful to keep vertical lines larger than this parameter. Can be also useful to keep vertical letter strokes.
Vertical Max Thickness	The maximum thickness of the vertical lines to remove. It is useful to keep horizontal lines larger than this parameter. Can be also useful to keep horizontal letter strokes.

Horizontal Min Length	The minimum length of the horizontal lines to remove.
Vertical Min Length	The minimum length of the vertical lines to remove.
Remove Dark Borders	Removes the dark surrounding from bitonal, grayscale or color images. The dark surrounding of the image is whitened (Note: The dark border should be touching the edge of the page for this to work).
Punch Hole Removal	Attempts to remove punch holes from pages. Note: The punch hole algorithm can be used on images with the following minimum dimensions width: 300px, height: 100px (computed for 300 DPI). The minimum height and width can vary with the image resolution.
Interpolation	Interpolates the source image to the given resolution. This value (the target resolution) must be greater than the source image's resolution.
Interpolation Mode	Sets the interpolation mode.
Keep Original Image	Set this to true if you want to use the pre-processed image for OCR but keep the original image in the output document. The default value is 'true'. Note: This property only applies when processing image files or when processing PDF files with the Extract Image Method set to ConvertToTiff
Keep Deskewed Image	Set this to true if you want to use the deskewed image in the output document. Note: This property only applies when Keep Original Image is set to No
Keep Despeckled Image	Set this to true if you want to use the despeckled image in the output document. This requires the source image to be black and white. Note: This property only applies when Keep Original Image is set to No
Keep Dark Border Removal	Set this to true if you want to use the image after dark borders have been removed, in the output document. Note: This property only applies when Keep Original Image is set to No
Keep Punch Hole Removal	Set this to true if you want to use the image after punch holes have been removed, in the output document. Note: This property only applies when Keep Original Image is set to No

5.2.7.2 PDF to Searchable PDF

Creates a searchable PDF file from the set of images from an image-only PDF file.

Depending upon the Step Type Properties chosen, a separate text, HTML and Office files may be produced from the OCR process.

Autobahn DX is shipped out with OCR engines.

- **Standard Engine**
- **Extended Engine**

See [Standard OCR vs Extended OCR](#) for the differences.

5.2.7.2.1 Standard Engine

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Continue on Error	Continue processing TIFF files after an error occurs.
OCR	Choose "No" to generate an image-only PDF. Choose "Yes" to generate searchable PDF and/or text files.
OCR Language	Select the language the original file is written in. This will determine the dictionary that is used.
Deskew	Straighten the image.
Auto-Rotate	Automatically rotate pages so that text flows left to right.
Despeckle	Remove specks below the specified pixel size from the image.
OCR to Text File	Choose "Yes" to Generate text Output
Output File	<ul style="list-style-type: none">• Plain Text (txt).• Plain Text (txt) No PDF• MS Word (rtf)• HTML
Non-Image PDFs	This allows control over the treatment of non-image PDFs, i.e. PDFs that have some text in them as well as images. The options are:

	<ul style="list-style-type: none"> • OCR: The document will be OCR'd using the image method defined by "Image Method" • Raise Error: The task will terminate with an error. If "On Error Continue" is set this then behaves as Skip. This is the default. • Skip: The document will not be processed. • Pass Through: The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.
Remove Hidden Text	This applies only when a PDF is being used as the source for OCR. When set to true this will not include any searchable text that already exists text from the source document. Such functionality might be useful if the source document was created by OCR of an image only PDF or other image file and the quality of the text from the previous OCR is poor. NOTE: There is no way to distinguish text added as a result of OCR from text added by other means and as a result, this option should be used with care.
Extract PDF Images Via	Choose the method for PDF image extraction. <ul style="list-style-type: none"> • Auto (default) • Native • Convert to TIFF
DPI	When OCR'ing a PDF, the PDF is rasterized to produce a TIFF file which is then OCR'd. By default, the TIFF image resolution is determined from the images embedded in the source PDF but this flag can be used to override default processing and specify the DPI of the TIFF that will be generated.
TIFF Compression	Sets the Compression for the TIFF file used if the "Convert To TIFF" Option above is used. <ul style="list-style-type: none"> • Auto (Selects Group 4 if the page is Black AND White else it uses LZW Compression) • Group 4 (Black and White) • LZW (Colored)
Retain Metadata	Copy metadata from the source PDF to the Searchable result PDF.
Retain Bookmarks.	Copy bookmarks from the source PDF to the Searchable result PDF
Retain Viewer Preferences	Retains any PDF Viewer Preferences, Page Mode and Page Layout from the source file in the output when using Extract Image Method='Convert To Tiff' .

PDF/A Options	<p>Select the output PDF/A compliant version you would like the output PDF to be.</p> <ul style="list-style-type: none"> • PDF/A1-b • PDF/A2-b • PDF/A3-b
Validate PDF/A	Whether or not to validate the PDF/A document after conversion
Box/Graphics Processing	<p>By default, if an area of the document is identified as a graphic area then no OCR processing is run on that area. However, certain documents may include areas or boxes that are identified as “graphics” or “picture” areas but that actually do contain useful text.</p> <p>To ensure that the OCR engine can be forced to process such areas there are two options:</p> <p><i>“Treat all Graphics Areas as Text”</i>. This option will ensure the entire document is processed as text.</p> <p><i>“Remove Box Lines in OCR Processing”</i>. This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the OCR engine. It does not remove boxes from the final image. Technically, this option removes connected elements with a minimum area (by default 100 pixels).</p>
Line Removal in OCR Processing	This removes lines and boxes during OCR processing to improve recognition – particularly in cases where characters “touch” lines.
JBIG2 Compression	This option will compress bitonal images in generated PDFs using JBIG2 compression rather than the default Group 4 compression scheme. This will result in smaller PDF file sizes, at a cost of increased processing time.
MRC Compression	Applies Mixed Raster Compression which can drastically reduce the size of PDF documents.
Save Pre-Despeckle	This will use the original image (i.e. before applying pre-processing) in the output PDF. The default value is true.
StampName	This has been deprecated, use the Stamp PDF Files step.
StampValue	This has been deprecated, use the Stamp PDF Files step.
Advanced Flags	Command line flags to be passed through to the underlying executable.

Maximum Cores	This specifies the number parallel files you want to be processed at a given time. Note: You need the multi-core license for this.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder. <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.2.2 Extended Engine

Parameter	Notes
Output File Name	The output filename excluding the extension (which will be added according to the output file type).
Output File Type	One or more of the following, separated by commas if more than one is required. RTF PDF DOCX CSV* SML (Spreadsheet ML XML file)* HTM TXT *These output formats are suitable for table-oriented pages that can be mapped onto a spreadsheet format.
OCR Engine	The OCR engine to use. This must be set to use the IRIS engine.
OCR Language 1-5	You can set up to five different languages for OCR recognition in one page as long as they are in the same character set.
Automatic language detection	Property that enables or disables the Auto Language Detection feature. The aim of this feature is to detect the most probable language of a single-language page. If at least one language has been detected, recognition will be performed in the first language candidate that has been detected, and not in the language(s) set through Language or Languages. If it fails to detect a language, recognition will be performed using the language(s) set through Language or Languages.

Auto rotate	Detect page orientation and correct if required
Deskew	Rotates the image to correct its skew angle.
Advanced Deskew	Set this to true to defined advanced deskew properties.
Force Deskew	<p>Under certain circumstances, rotating the image to correct its skew angle may decrease the OCR accuracy. The extended engine is able to analyze the image and detect from an OCR accuracy point of view whether it's better to rotate the image or not. Because the skew angle may be visible in the output document (e.g. if KeepDeskew is set to 'true'), you can choose to force the deskew to rotate the image, even if it affects the accuracy.</p> <p>If turned off, the image is analyzed before rotation and the engine may choose not to rotate the image depending on the analysis result.</p> <p>If turned on, the image is rotated to correct skew angle.</p>
Adjustment Mode	Set the behavior regarding dimension adjustment for deskew operation.
Despeckle	Removes all the groups of connected pixels with a number of pixels below the parameter. Suggested range: 1-20.
Advanced Despeckle	Set the advanced despeckle settings, advanced despeckle provides advanced image noise reduction features by the image despeckle filter.
Remove White Pixels	By default, Advanced Despeckle removes black pixels. If this setting is set to 'true', white pixels will be removed instead of black pixels.
Dilate	<p>Despeckle removes all the groups of connected pixels with a few pixels below the SpeckleSize parameter. Those connected pixels are not removed if the distance to a larger connected component is below this parameter. As a result, only the isolated pixels get deleted. The maximum value for this property is 20 pixels.</p> <p>The default value is '0'.</p>
Retain Bookmark	<p>This option allows you to retain the bookmarks in the new PDF if the old PDF was Converted to TIFF before it was OCR'd.</p> <p>Note: This Will Only work if : "Extract Images Method = Convert to TIFF"</p>
Retain Metadata	This option allows you to retain the metadata in the new PDF if the old PDF was Converted to TIFF before it was OCR'd.

	Note: This Will Only work if : “Extract Images Method = Convert to TIFF”
Layout	The layout for the dock or rtf document <ul style="list-style-type: none"> • Standard • Flow
PDFVersion	This determines the PDF version of the generated PDF : 1.4 1.5 1.6 1.7 PDF/A1B (<i>PDF/A-1b</i>) Note: This Will Only work if : “Extract Images Method = Convert to TIFF”
Extract Images Method	Whether to convert the images in a PDF document to TIFF or not. <ul style="list-style-type: none"> • Convert to TIFF – The pages in the PDF document are rasterized and saved as TIFF images • Native - This method places the OCR’ed text directly into a copy of the original PDF rather than creating an entirely new PDF.
Remove Blank Page	Set this to true to remove blank pages from Tiff or PDF documents. Value needs to be set for sensitivity (see below).
Sensitivity	The sensitivity, from 1 to 100. With high sensitivity, fewer blank pages are detected.
Work Depth	This parameter (0 – 255) defines how deeply the OCR engine will analyze a page with 255 being the deepest. For poorer quality documents, higher values can give better recognition results.
JPEG Quality	This parameter (0 – 255) determines the compression/quality of Color JPEG images in generated PDFs. 0 gives the smallest file size whilst 255 gives the best quality. The default value is 128.
JPEG2000 Compression	Enable/Disable JPEG2000 Compression.
JPEG2000 Compression Mode	The JPEG2000 Compression Mode to use.
JPEG2000 Compression Value	The Value to set for the selected Compression Mode.

IHQC Compression	Apply Intelligent High-Quality Compression
IHQC Compression Level	Level1 is the basic compression level while level 3 is the most advanced Intelligent High-Quality Compression Mode.
IHQC Quality Factor	The quality Factor for IHQC
Binarization	Whether or not to perform binarization on the document.
Brightness	The brightness (higher values will darker the result).
Contrast	The contrast (lower values will darker the result).
Smoothing Level	Smoothing may be useful to binarize text with a colored background to avoid noisy pixels (0 disables smoothing, higher values smooth more).
Undithering	Whether or not to use automatic undithering while processing a page. NOTE: Automatic undithering will be applied only if smoothing is also activated (Smoothing Level) Dithering is a scanning technique which consists in representing a color or greyscale image using only a limited color palette. This allows reducing file size while maintaining the general aspect of the image. This technique is known to create images more difficult to handle for OCR technology; therefore specific image preprocessing is needed to detect and revert it.
Threshold	Sets the threshold for fixed threshold binarization (0 for automatic threshold computation).
Remove Lines	Whether or not to remove lines from an image (The image must be black and white).
Horizontal Clean X	The parameter for cleaning noisy pixels attached to the horizontal lines.
Horizontal Clean Y	The parameter for cleaning noisy pixels attached to the horizontal lines.
Vertical Clean X	The parameter for cleaning noisy pixels attached to the vertical lines.
Vertical Clean Y	The parameter for cleaning noisy pixels attached to the vertical lines.
Horizontal Dilate	The dilate parameter that helps the detection of horizontal lines.
Vertical Dilate	The dilate parameter that helps the detection of vertical lines.

Horizontal Max Gap	The maximum horizontal line gap to close. It is useful to remove broken lines.
Vertical Max Gap	The maximum vertical line gap to close. It is useful to remove broken lines.
Horizontal Max Thickness	The maximum thickness of the horizontal lines to remove. It is useful to keep vertical lines larger than this parameter. Can be also useful to keep vertical letter strokes.
Vertical Max Thickness	The maximum thickness of the vertical lines to remove. It is useful to keep horizontal lines larger than this parameter. Can be also useful to keep horizontal letter strokes.
Horizontal Min Length	The minimum length of the horizontal lines to remove.
Vertical Min Length	The minimum length of the vertical lines to remove.
Remove Dark Borders	Removes the dark surrounding from bitonal, grayscale or color images. The dark surrounding of the image is whitened (Note: The dark border should be touching the edge of the page for this to work).
Punch Hole Removal	Attempts to remove punch holes from pages. Note: The punch hole algorithm can be used on images with the following minimum dimensions width: 300px, height: 100px (computed for 300 DPI). The minimum height and width can vary with the image resolution.
Interpolation	Interpolates the source image to the given resolution. This value (the target resolution) must be greater than the source image's resolution.
Interpolation Mode	Sets the interpolation mode.
Keep Original Image	Set this to true if you want to use the pre-processed image for OCR but keep the original image in the output document. The default value is 'true'. Note: This property only applies when processing image files or when processing PDF files with the Extract Image Method set to ConvertToTiff
Keep Deskewed Image	Set this to true if you want to use the deskewed image in the output document. Note: This property only applies when Keep Original Image is set to No
Keep Despeckled Image	Set this to true if you want to use the despeckled image in the output document. This requires the source image to be black and white.

	Note: This property only applies when Keep Original Image is set to No
Keep Dark Border Removal	Set this to true if you want to use the image after dark borders have been removed, in the output document. Note: This property only applies when Keep Original Image is set to No
Keep Punch Hole Removal	Set this to true if you want to use the image after punch holes have been removed, in the output document. Note: This property only applies when Keep Original Image is set to No

5.2.7.2.3 Cloud OCR

Check the [Cloud OCR](#) section.

5.2.7.3 Merge TIFFs to PDF

This step first merges the input images in a folder into a multi-page PDF file, then performs an OCR on the file. Depending upon the Step Type Properties chosen, a separate text, HTML and Office files may be produced from the OCR process.

Autobahn DX is shipped out with OCR engines.

- **Standard Engine**
- **Extended Engine**

See [Standard OCR vs Extended OCR](#) for the differences.

5.2.7.3.1 Standard Engine

Parameter	Notes
Output File Name	Target file template which can include %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
OCR Options	Choose "No OCR" to generate an image-only PDF. Choose "OCR" to generate searchable PDF and/or text files.
Continue on Error	Continue processing TIFF files after an error occurs.
OCR Language	Select the language the original file is written in. This will determine the dictionary that is used.

Deskew	Straighten the image.
Auto-Rotate	Automatically rotate pages so that text flows left to right.
Despeckle	Remove specks below the specified pixel size from the image.
Save Pre-Despeckle	This will use the original image (i.e. before applying pre-processing) in the output PDF. The default value is true.
Output PDF	Choose “Yes” to Generate a PDF file.
Output TXT	Choose “Yes” to generate a .txt file (only applicable if OCR is specified).
Output RTF	Choose “Yes” to generate a .rtf file (only applicable if OCR is specified).
Output HTML	Choose “Yes” to generate a .htm file (only applicable if OCR is specified).
Advanced Flags	Command line flags to be passed through to the underlying executable.
PDF/A1-b PDF/A2-b PDF/A3-b	Select the output PDF/A compliant version you would like the output PDF to be.
Validate PDF/A	Whether or not to validate the PDF/A document after conversion

5.2.7.3.2 Extended Engine

Check [Extended Engine](#) above.

5.2.7.3.3 Cloud OCR

Check the [Cloud OCR](#) section.

5.2.7.4 Convert Any File to PDF

This converts any printable document to PDF, such as Microsoft Word, Excel, PowerPoint, HTML, etc. subject to the native application being available on the server. See [Section 18.2](#) for more details.

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)

Continue on Error	Continue processing files after an error occurs.
Conversion Timeout (ms)	Limits the amount of time in milliseconds that can be spent on conversion. A value of zero means there is no time limit.
Convert Bookmarks	For MS Word, convert bookmarks
Bookmark Depth	<p>This property will take effect only when the Convert Bookmarks property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, an empty string is functionally equivalent to</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use a non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.</p>
Convert Hyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
Print All Sheets (Excel)	The flag that indicates whether to print all Excel worksheets or not.
Print Background Color (IE)	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
Print Scale % (Visio)	For Visio files, sets the print scale
Header (IE)	This property modifies Internet Explorer's header setting.
Footer (IE)	This property modifies Internet Explorer's footer setting.
Image Compression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
Image Downsizing	If this property is set to Yes, then the resolution of images is reduced to the DPI value specified in the Downsize Resolution DPI property.

Downsize Resolution DPI	If the Image Downsizing property is set to True, then the resolution of images is reduced to the DPI value specified in this property.
Image JPEG Quality	The allowed value range is from 5 to 100 with 100 being the highest quality.
Font Embedding	The option PRN_FONT_EMBED_FULLSET (embedding a full set of font) will cause a significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed the font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
Font Substitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from the Control Panel).
Embed Fonts as Type 0	This option is recommended if you have non-standard fonts like barcode font.
Top Margin	Sets top margin. (Inches)
Bottom Margin	Sets bottom margin. (Inches)
Left Margin	Sets left a margin. (Inches)
Right Margin	Sets right margin. (Inches)
Page Width	Sets a custom page width. (Inches)
Page Height	Sets a custom page height. (Inches)
Paper Orientation	<p>Sets paper orientation to</p> <ul style="list-style-type: none"> • Default (Maintain Source Orientation) • Landscape • Portrait
PDF Compliance	<p>Allows the User To choose a PDF/A or PDF/X Compliant files</p> <ul style="list-style-type: none"> • None (No PDF/A Output) • PDF/A-1b (PDF/A-1b compliant) • PDF/X-1a (PDF/X-1a compliant) • PDF/X-3 (PDF/X-3 compliant)
Convert MSG Attachments	If you set this to true, Autobahn DX will convert the both MSG files and their Attachments to a single PDF file.

Attach MSG Attachments to PDF	<p>If set to true, Autobahn DX will Attach Msg Attachments that are converted as PDF Attachments.</p> <p>If set to false, Autobahn DX will merge Msg Attachments that are converted to the PDF file generated by the body.</p>
Preserve Word Attachments	<p>Determines whether embedded and linked files will be preserved during conversion. Default value: False (disabled).</p> <p>Note: This will work with WordExtentionEX only</p>
Convert PDF Attachments (PDF)	Convert PDF Attachments to create a combined PDF file.
Merge PDF Attachments (PDF)	Set this flag to true if you want to convert pdf attachments and merged them into the output pdf file. Otherwise, the converted files will be merged back to the pdf.
Retain PDF Attachment (PDF)	Switch this on to Retain the Original PDF attachments if you set Merge PDF Attachments to true.
Retain Properties (Office)	Set this flag if you want the MS Office properties to be transferred to the target pdf document.
Color Type (PowerPoint)	Use this property to set PowerPoint to print with either color, grayscale, or black and white.
Handout Order (PowerPoint)	<p>Sets the handout order, this flag only applies to PowerPoint jobs.</p> <p>The possible values are:</p> <ul style="list-style-type: none"> • Vertical First • Horizontal First
Output Type (PowerPoint)	<p>Sets the output type, it only works with the PowerPoint files. The possible values are:</p> <ul style="list-style-type: none"> • Slides • Build slides • Two slides handouts • Three slides handouts • Four slides handouts • Six slides handouts • Nine slides handouts • Notes • Outline
Print Graphics (Publisher)	<p>Sets the graphics setting for printing.</p> <ul style="list-style-type: none"> • Print Full Resolution • Print Low Resolution

	<ul style="list-style-type: none"> • Print Graphics
Frame Slides (PowerPoint)	Indicate whether to draw a frame around the border of the slides.
Zoom (Excel)	Sets printing zoom of the worksheet. The allowed value range is from 10 to 400.
Fit to Pages Wide (Excel)	Sets number of pages wide the worksheet will be scaled to. This property is ignored if the Zoom property is set.
Fit to Pages Tall (Excel)	Sets number of pages tall the worksheet will be scaled to. This property is ignored if the Zoom property is set.
Include Document Markups	Determines whether document markups are retained. When this property is False (the default), document markups are omitted. When this property is True, markups are included.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	<p>This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder.</p> <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.5 Set PDF Properties

This is used to set PDF Metadata properties (such as Author, Title, etc.), Security settings and Document Display properties.

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension), %DIRNAME (directory name of the original file), %UNIQUEn (e.g. %UNIQUE4 for 4 digits), %BOOKMARK and %PAGE n (e.g. %PAGE4 for 4 digits)

Encryption Strength	Must be set to 128 bits if security attributes are to be set.
User Password	A password that will be required to open the document.
Owner Password	A password that will be required to change the document permissions.
Allow Printing	Allow high-quality printing
Allow Modify Contents	Allow assembly and other document modifications
Allow Copy	Allow text and graphics copying and extraction
Allow Modify Annotations	Allow modification of annotations
Allow Filling	Allow filling of form fields
Allow Screen Readers	Allow extraction of text and graphics in support of accessibility.
Allow Assembly	Allow rotation, insertion or deletion of pages.
Allow Degraded Printing	Allow low-quality printing
Author	Sets the Author property
Title	Sets the Title property
Subject	Sets the Subject property
Keywords	Sets the Keywords property
Creator	Sets the Creator property
Page Layout	The setting for the initial document page display.
Page Mode	The setting for initial viewer mode
Non-Full Screen Mode	Only applicable where Page Mode=Full Screen. The setting for document page display when exiting Full-Screen mode.
Hide Menu Bar	The viewer's menu bar will be hidden
Hide Window UI	The viewer's UI elements (scrollbars etc.) will be hidden
Hide Tool Bar	The viewer's toolbar will be hidden
Fit Window	The viewer will resize the document's window to fit the size of the first displayed page.

Center Window	The document window will be positioned in the center of the screen.
---------------	---

5.2.7.6 Custom Script

This can be used to support a custom scripted step in the process. See [section 6](#) for more details.

Parameter	Notes
Custom Script File	Name of the custom script file to be run located in the Autobahn custom folder.

5.2.7.7 Stamp PDF Files

This step can be used to add stamps to PDF pages, we have given the user the ability to customize these stamps extensively in a very simple manner. See the step properties below.

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension), %DIRNAME (directory name of the original file).
Stamp Operation	<p>Autobahn DX has different ways to apply stamps to a page, this gives the user some level of flexibility.</p> <ul style="list-style-type: none"> • StampTextAsString: When this operation has selected the text passed as the StampObject will be stamped on the PDF document as text. • StampPDFText: When this operation is selected the text passed as the StampObject will be stamped on the PDF document as an image. • StampPageNumber: When this operation is selected, every page in the PDF file will be stamped with a page number, starting from the start number. E.g. if StartNumber = 6 the first-page number will start from 6. • StampPageNumberBates: When this operation is selected, every page in the PDF file will be stamped with a bates number, starting from the start number. E.g. if StartNumber = 6 the first-page number will start from 000006. • StampVariable: This option allows a user to specify a variable like a date, filename or time. The variable specified by the StampObject will be stamped on the document. Check the table below for different Stamp variables provided.

	<ul style="list-style-type: none"> • StampPDFImage: When this operation is selected the text passed as the StampObject is the address of the image to be stamped on the PDF document.
Stamp Placement	<p>The property specifies the location in a page a stamp can be placed. Below is a list of options available.</p> <ul style="list-style-type: none"> • Bottom Center • Bottom Left • Bottom Right • Center • Center Left • Center Right • Top Center • Top Left • Top Right
Stamp Direction	<p>This represents the direction of the stamp on the output PDF</p> <ul style="list-style-type: none"> • Normal • Diagonal Up • Diagonal Down
Stamp Text	Enter any static text to be stamped on a PDF page, this works with the StampPDFText stamp operation.
Stamp Variable	Enter a stamp variable to be stamped on a PDF page, this works with the StampVariable stamp operation. See table below for more details.
Image Path	The path to the image if you are using the StampPDFImage operation.
Page Range	Set of page ranges separated by commas that define which pages from the original should be stamped.
Start Number	The number that the page numbering will start with, works with StampPageNumber and StampPageNumberBates .
Start Page	Specifies the page that the stamping should start
End Page	Specifies the page that the stamping should stop
Bates Prefix	Specifies the prefix of the Bates stamp
Bate Suffix	Specifies the suffix of the Bates stamp
Bate Length	Specifies the length of the Bates stamp
Stamp Color	The color of non-image stamps. Enter a valid color name or black will be used.

Stamp Opacity	The opacity of non-image stamps. Enter a valid color name or black will be used.
Font Name	The font name of non-image stamps. Choose the font you want from a drop-down list of different fonts.
Font Size	The font size of non-image stamps, default value = 20.
Stamp Text as Image	Set this to Yes if you want Autobahn DX to convert text-based stamps to images before applying it to the PDF page.
Image Background Color	When you set Stamp Text as Image to yes, use this property to set the background color of the image(rectangle) that the text is converted to.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder. <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.7.1 Stamp Variables

The table below shows the different Stamp variables supported by Autobahn DX, the idea is that Autobahn will replace an occurrence of the variable with the appropriate value in a text string before applying the stamp. E.g. to Stamp **Today is Monday** on a PDF page, use the following Stamp variable "Today is %A".

Variable	Stamp
%a	Short Day (Mon)
%A	Long Day (Monday)
%b	Short Month (Jan)
%B	Long Month (January)
%c	Date and time (30 October 2013 17:21)
%C	Date and Time with seconds (30 October 2013 17:21:50)
%d	Month and Year (October 2013)

%D	Day and Month (30 October)
%e	Short Year (13)
%E	Long Year (2013)
%f	Short Time of Day (17:21)
%F	Time of Day with Seconds (17:21:20)
%G	Full Date and time (Wed, 30 October 2013 17:21:50 GMT)
%Y	File Name

5.2.7.8 Merge PDF

Merges a folder of PDF files into a single file.

Parameter	Notes
Output File Name	Target file template which can include %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Retain Bookmarks	Generated files will include bookmarks from the original file.
Retain Metadata	Generated files will include metadata (such as Author and Title) from the original file.
File Names as Bookmarks	Generate bookmarks in the output PDF using filenames of source PDF files.
Continue on Error	Continue processing if an error occurs.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.9 Split PDF

Splits each input PDF file into a set of files, either a single page per file or by page ranges.

Parameter	Notes
Output File Name	The target file template which can include %UNIQUEn (a unique number starting at 1, zero padded to n digits) %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Retain Bookmarks	Generated files will include bookmarks from the original file.
Retain Metadata	Generated files will include metadata (such as Author and Title) from the original file.
Split Type	
Ranges (e.g. 1,3-10)	Set of page ranges separated by commas that define which pages from the original should be extracted.
Repeat Every (Pages)	Apply the page range to each set of Page Ranges within the document. For example, if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages.
Continue on Error	Continue processing if an error occurs.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder. <ul style="list-style-type: none">• Take no action.• Move to Error Folder• Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.10 Merge TIFFs

Merges a folder of TIFF files into a single file.

Parameter	Notes
Output File Name	Target file template which can include %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Continue on Error	Continue processing if an error occurs.
Debug	Set this to true to execute the job in debug mode.

5.2.7.11 Split TIFF

Splits each input TIFF file into a set of files, either a single page per file or by page ranges.

Parameter	Notes
Output File Name	The target file template which can include %UNIQUE (a unique number starting at 1, zero padded to n digits) %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Split Type	
Ranges (e.g. 1,3-10)	Set of page ranges separated by commas that define which pages from the original should be extracted.
Repeat Every (Pages)	Apply the page range to each set of Page Ranges within the document. For example, if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Continue on Error	Continue processing if an error occurs.
Debug	Set this to true to execute the job in debug mode.

5.2.7.12 Read Inbox

This can read mailboxes and extract attachments using IMAP4, in accordance with the parameters specified below. Use of this step type requires a Server License.

Check with your System Administrator and ensure the following:

- IMAP4 is enabled for the mail server and your account.
- You have the IMAP address of the mail server.

Note: The files will be downloaded in the following format, [name@timestamp@email@.ext](#) where:

- Name = Filename
- timestamp= date of the email.
- email= The from address

The following Microsoft article provides information on how to verify basic IMAP connectivity by using Telnet: <http://support.microsoft.com/kb/189326>

Parameter	Notes
imap_server	The IMAP server address e.g. imap.company.co.uk
email_account	The email account to be read e.g. pdf@company.com
email_password	Password for the account. This is held encrypted.
in_mailbox	Mailbox to read e.g. Inbox
processed_mailbox	Mailbox to move processed email to e.g. Deleted Items. If left blank, the emails will be left in the inbox which can be useful for testing.
Target	The folder in which to place the attachments extracted from the email. <i>Files are named name@timestamp@email@.ext where name.ext is the original filename, email is the user's email address and timestamp is an integer timestamp to ensure uniqueness.</i>
Include	Regular expression. If specified, only files matching the expression will be processed. E.g. *.tif. This allows alternate jobs to be created for different file types.
Exclude	Regular expression. If specified, files matching the expression will not be processed. E.g. *.pdf
Debug	Set this to true to execute the job in debug mode.

5.2.7.13 Send Documents

Use of this step type requires a Server License.

Note: The input file of this step must be in the format of [name@timestamp@email@.ext](#) where:

- Name = Filename
- timestamp= date of the email.
- email= the address where we will send the output files.

Parameter	Notes
Domain	The sending domain E.g. aquaforest.com
SMTP Server	SMTP Server address E.g. smtp.aquaforest.com
Send Logon	Sending user login e.g. admin.
Send User Password	Password for the account.
Sending User	Sending user e.g. admin@aquaforest.com
Email Title	The title of the Email
Email Body	The body of the Email
Debug	Set this to true to execute the job in debug mode.

5.2.7.14 Convert PDF to TIFF

Rasterizes a PDF file, converting into a multi-page TIFF file.

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension)
Compression	Group 4 (For bitonal images) or LZW (for color).
Resolution	The DPI of the resulting TIFF File.
Continue on Error	Continue processing if an error occurs.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder.

	<ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.15 Extract Text from PDF

Extracts the raw text from a searchable PDF. [NB this does not perform an OCR process, it just extracts the existing text from the PDF file.]

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension)
Continue on Error	Continue processing if an error occurs.
Page From	The start of the range of pages from which to extract text. If not specified, a start page of 1 is assumed.
Page To	The end of the range of pages from which to extract text. If not specified, the last page is assumed.
Page Separator	This allows the definition of an optional page separator string in the output text file.
Page Separator Placement	Specifies whether the Page Separator will appear at the beginning or the end of the page.
Extract Text Engine.	The Extract Text Engine to use: <ul style="list-style-type: none"> • 0 = PDFBox with Formatting, • 1 = BCL • 2 = PDFBox
Copy Input PDF to Target Folder	Set to true if you want Autobahn DX to copy the input PDF file to the target folder.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder. <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.16 SharePoint Download

This step downloads documents from the specified SharePoint document library ready for processing

Parameter	Notes
SharePoint Site URL	Site, the URL of the SharePoint site that you want to access, e.g. http://localhost/testsite
SharePoint Online (Office 365)	Whether or not the upload location is in SharePoint Online (Office 365).
Use ADFS	Switch this on if you use Active Directory for your SharePoint User Management.
Username	The username used to connect to the SharePoint site. Leave empty to use Windows Credentials (for local SharePoint only).
Password	The password used to connect to the SharePoint site. Leave empty to use Windows Credentials (for local SharePoint only).
ADFS Host	Provide the name of the Active Directory server.
ADFS Relying Party Identifier	Provide the Relying Party Trust identifier for your SharePoint.
SharePoint Library	Library, the name of the library that you want to access, e.g. "Test Library"
SharePoint Sub Folder	Download documents from the specified subfolder in the SharePoint library only.
Extension Filter	An optional extension mask that limits those files to manipulate, e.g. "pdf,tiff"
Recurse SharePoint Library	If set to "Yes" sub-folders of the SharePoint Library are handled.
Include Pattern	Autobahn will only include the files that match this pattern.
Exclude Pattern	Any file that matches this pattern will be excluded.
Debug	If set to "Yes" to see more processing information on the console.
Continue on Error	Continue processing if an error occurs.

5.2.7.17 SharePoint Upload

This step uploads documents to the specified SharePoint document library.

Parameter	Notes
SharePoint Site URL	Site, the URL of the SharePoint site that you want to access, e.g. http://localhost/testsite
SharePoint Online (Office 365)	Whether or not the upload location is in SharePoint Online (Office 365).
Use ADFS	Switch this on if you use Active Directory for your SharePoint User Management.
Username	The username used to connect to the SharePoint site.
Password	The password used to connect to the SharePoint site.
ADFS Host	Provide the name of the Active Directory server.
ADFS Relying Party Identifier	Provide the Relying Party Trust identifier for your SharePoint.
SharePoint Library	Library, the name of the library that you want to access, e.g. "Test Library"
SharePoint Sub Folder	The subfolder inside the SharePoint library to upload the files into. The subfolder should be present in the library or else the following message will be displayed: "The remote server returned an error: (409) Conflict."
Extension Filter	An optional extension mask that limits those files to manipulate, e.g. ".pdf,tiff"
Recurse Source Folder	Recurse the source folder and its subfolders for files to upload and create the folders in SharePoint if they do not already exist. Note: If "Use Work Folders" is checked, then "Process Sub-Folders" must also be checked for this to work.
Create Directories if Required	Force creation of any output directories if they do not already exist.
Include Pattern	Only file that matches this pattern will be included.
Exclude Pattern	Any file that matches this pattern will be excluded.
Debug	If set to "Yes" the user will see more processing information on the console.
Continue on Error	Continue processing if an error occurs.

5.2.7.18 Azure Storage Download

This step will download files to your local machine from an Azure storage Container.

Parameter	Notes
Storage Account Name	The name of the Azure blob container you want to download files from.
Azure Account Key	Key 1 under the accesskeys section of the storage account in the portal.
Container Name	The name of the Azure blob container you want to download files from.
Extension Filter	File extension filters separated by commas (e.g. .tif,.pdf)
Recurse Azure Storage	Download documents from folders and subfolders in the SharePoint Library
Debug	If set to “Yes” the user will see more processing information on the console.

5.2.7.19 Azure Storage Upload

This step will upload files from your local machine to an Azure storage Container.

Parameter	Notes
Storage Account Name	The name of the Azure blob container you want to upload files to.
Azure Account Key	Key 1 under the accesskeys section of the storage account in the portal.
Container Name	The name of the Azure blob container you want to upload files to.
Extension Filter	File extension filters separated by commas (e.g. .tif,.pdf)
Recurse Azure Storage	Upload documents from folders and subfolders of the local folder
Debug	If set to “Yes” the user will see more processing information on the console.

5.2.7.20 Compress Image PDF

This step uses MRC compression to Compress PDF files.

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension)
MRC JPEG Quality	JPEG quality setting (percentage value 1 - 100) for use in saving the background and foreground images. The default value is 75
MRC Background Level	Sampling size for the background portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. The default value is 3
MRC Foreground Level	Sampling size for the foreground portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. The default value is 3
Advanced Flags	Several advanced flags are available for use under guidance from Aquaforest technical support. Contact support@aquaforest.com if needed.
Continue on Error	Continue processing files after an error occurs.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Debug	Set this to true to execute the job in debug mode.

5.2.7.21 Create XML Property File

This step takes a PDF input file and generates an XML output file.

Parameter	Notes
Copy the Source PDF to Target Folder	Set to true if you want Autobahn DX to copy the input PDF file to the target folder.
Continue on Error	Continue processing files after an error occurs.
Debug	Set this to true to execute the job in debug mode.

5.2.7.22 Optimize PDF

This allows the creation of Web Optimized (Linearize) PDFs.

Parameter	Notes
Linearize – Fast Web View	Set to true to true to Linearize a PDF file
Continue on Error	Continue processing files after an error occurs.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Password Files	This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder. <ul style="list-style-type: none">• Take no action.• Move to Error Folder• Copy to Error Folder
Debug	Set this to true to execute the job in debug mode.

5.2.7.23 OCR Any File to PDF

This step attempts to convert all files to searchable PDFs, Autobahn DX is shipped out with OCR engines.

- **Standard Engine**
- **Extended Engine**

See [Standard OCR vs Extended OCR](#) for the differences.

5.2.7.23.1 Standard Engine

Parameter	Notes
General Settings	
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Continue on Error	Continue processing TIFF files after an error occurs.
OverWrite Existing	Over Writes the target document if it exists.
Advanced Flags	Command line flags to be passed through to the underlying executable.

Password Files	<p>This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder.</p> <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Maximum Cores	<p>This specifies the number parallel files you want to be processed at a given time.</p> <p>Note: You need the multi-core license for this.</p>
Debug	Set this to true to execute the job in debug mode.
Standard OCR Settings	
OCR Language	Select the language the original file is written in. This will determine the dictionary that is used.
Deskew	Straighten the image.
Auto-Rotate	Automatically rotate pages so that text flows left to right.
Despeckle	Remove specks below the specified pixel size from the image.
OCR to Text File	Choose “Yes” to Generate text Output
Output File	<ul style="list-style-type: none"> • Plain Text (txt). • Plain Text (txt) No PDF • MS Word (rtf) • HTML
Non-Image PDFs	<p>This allows control over the treatment of non-image PDFs, i.e. PDFs that have some text in them as well as images. The options are:</p> <ul style="list-style-type: none"> • OCR: The document will be OCRred using the image method defined by “Image Method” • Raise Error: The task will terminate with an error. If “On Error Continue” is set this then behaves as Skip. This is the default. • Skip: The document will not be processed. • Pass Through: The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.
Remove Hidden Text	<p>This applies only when a PDF is being used as the source for OCR. When set to true this will not include any searchable text that already exists text from the source document. Such</p>

	<p>functionality might be useful if the source document was created by OCR of an image only PDF or other image file and the quality of the text from the previous OCR is poor. NOTE: There is no way to distinguish text added as a result of OCR from text added by other means and as a result, this option should be used with care.</p>
Extract PDF Images Via	<p>Choose the method for PDF image extraction.</p> <ul style="list-style-type: none"> • Auto (default) • Native • Convert to TIFF
DPI	<p>When OCRing a PDF, the PDF is rasterized to produce a TIFF file which is then OCRed. By default, the TIFF image resolution is determined from the images embedded in the source PDF but this flag can be used to override default processing and specify the DPI of the TIFF that will be generated.</p>
TIFF Compression	<p>Sets the Compression for the TIFF file used if the “Convert To TIFF” Option above is used.</p> <ul style="list-style-type: none"> • Auto (Selects Group 4 if the page is Black AND White else it uses LZW Compression) • Group 4 (Black and White) • LZW (Colored)
Retain Metadata	<p>Copy metadata from the source PDF to the Searchable result PDF.</p>
Retain Bookmarks.	<p>Copy bookmarks from the source PDF to the Searchable result PDF</p>
Retain Viewer Preferences	<p>Retains any PDF Viewer Preferences, Page Mode and Page Layout from the source file in the output when using Extract Image Method='Convert To Tiff'.</p>
PDF/A Options	<p>Select the output PDF/A compliant version you would like the output PDF to be.</p> <ul style="list-style-type: none"> • PDF/A1-b • PDF/A2-b • PDF/A3-b
Validate PDF/A	<p>Whether or not to validate the PDF/A document after conversion</p>
Box/Graphics Processing	<p>By default, if an area of the document is identified as a graphic area then no OCR processing is run on that area. However, certain documents may include areas or boxes that are identified as “graphics” or “picture” areas but that actually do contain useful text.</p>

	<p>To ensure that the OCR engine can be forced to process such areas there are two options:</p> <p><i>“Treat all Graphics Areas as Text”</i>. This option will ensure the entire document is processed as text.</p> <p><i>“Remove Box Lines in OCR Processing”</i>. This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the image used by the OCR engine. It does not remove boxes from the final image. Technically, this option removes connected elements with a minimum area (by default 100 pixels).</p>
Line Removal in OCR Processing	This removes lines and boxes during OCR processing to improve recognition – particularly in cases where characters “touch” lines.
JBIG2 Compression	This option will compress bitonal images in generated PDFs using JBIG2 compression rather than the default Group 4 compression scheme. This will result in smaller PDF file sizes, at a cost of increased processing time.
MRC Compression	Applies Mixed Raster Compression which can drastically reduce the size of PDF documents.
Save Pre-Despeckle	This will use the original image (i.e. before applying pre-processing) in the output PDF. The default value is true.
StampName	This has been deprecated, use the Stamp PDF Files step.
StampValue	This has been deprecated, use the Stamp PDF Files step.
Any File To PDF Conversion Settings	
Conversion Timeout (ms)	Limits the amount of time in milliseconds that can be spent on conversion. A value of zero means waits indefinitely.
Convert Bookmarks	For MS Word, convert bookmarks
Bookmark Depth	<p>This property will take effect only when the Convert Bookmarks property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, an empty string is functionally equivalent to</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use a non-English version of Microsoft Word,</p>

	then you may need to replace the word "Heading" with its localized version.
Convert Hyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
Print All Sheets (Excel)	The flag that indicates whether to print all Excel worksheets or not.
Print Background Color (IE)	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
Print Scale % (Visio)	For Visio files, sets the print scale
Header (IE)	This property modifies Internet Explorer's header setting.
Footer (IE)	This property modifies Internet Explorer's footer setting.
Image Compression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
Image Downsizing	If this property is set to Yes, then the resolution of images is reduced to the DPI value specified in the Downsize Resolution DPI property.
Downsize Resolution DPI	If the Image Downsizing property is set to True, then the resolution of images is reduced to the DPI value specified in this property.
Image JPEG Quality	The allowed value range is from 5 to 100 with 100 being the highest quality.
Font Embedding	The option PRN_FONT_EMBED_FULLSET (embedding a full set of font) will cause a significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed the font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
Font Substitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from the Control Panel).
Embed Fonts as Type 0	This option is recommended if you have non-standard fonts like barcode font.
Top Margin	Sets top margin. (Inches)
Bottom Margin	Sets bottom margin. (Inches)

Left Margin		Sets left margin. (Inches)
Right Margin		Sets right margin. (Inches)
Page Width		Sets a custom page width. (Inches)
Page Height		Sets a custom page height. (Inches)
Paper Orientation		<p>Sets paper orientation to</p> <ul style="list-style-type: none"> • Default (Maintain Source Orientation) • Landscape • Portrait
PDF Compliance		<p>Allows the User To choose a PDF/A or PDF/X Complaint files</p> <ul style="list-style-type: none"> • None (No PDF/A Output) • PDF/A-1b (PDF/A-1b compliant) • PDF/X-1a (PDF/X-1a compliant) • PDF/X-3 (PDF/X-3 compliant)
Convert Attachments	MSG	If you set this to true, Autobahn DX will convert the both MSG files and their Attachments to a single PDF file.
Attach Attachments to PDF	MSG	<p>If set to true, Autobahn DX will Attach Msg Attachments that are converted as PDF Attachments.</p> <p>If set to false, Autobahn DX will merge Msg Attachments that are converted to the PDF file generated by the body.</p>
Preserve Attachments	Word	<p>Determines whether embedded and linked files will be preserved during conversion. Default value: False (disabled).</p> <p>Note: This will work with WordExtentionEX only</p>
Convert Attachments (PDF)	PDF	Convert PDF Attachments to create a combined PDF file.
Merge Attachments (PDF)	PDF	Set this flag to true if you want to convert pdf attachments and merged them into the output pdf file. Otherwise, the converted files will be merged back to the pdf.
Retain PDF Attachment (PDF)		Switch this on to Retain the Original PDF attachments if you set Merge PDF Attachments to true.
Retain Properties (Office)		Set this flag if you want the MS Office properties to be transferred to the target pdf document.
Color Type (PowerPoint)		Use this property to set PowerPoint to print with either color, grayscale, or black and white.
Handout Order (PowerPoint)		Sets the handout order, this flag only applies to PowerPoint jobs.

	<p>The possible values are:</p> <ul style="list-style-type: none"> • Vertical First • Horizontal First
Output Type (PowerPoint)	<p>Sets the output type, it only works with the PowerPoint files.</p> <p>The possible values are:</p> <ul style="list-style-type: none"> • Slides • Build slides • Two slides handouts • Three slides handouts • Four slides handouts • Six slides handouts • Nine slides handouts • Notes • Outline
Print Graphics (Publisher)	<p>Sets the graphics setting for printing.</p> <ul style="list-style-type: none"> • Print Full Resolution • Print Low Resolution • Print Graphics
Frame Slides (PowerPoint)	<p>Indicate whether to draw a frame around the border of the slides.</p>
Zoom (Excel)	<p>Sets printing zoom of the worksheet.</p> <p>The allowed value range is from 10 to 400.</p>
Fit to Pages Wide (Excel)	<p>Sets number of pages wide the worksheet will be scaled to.</p> <p>This property is ignored if the Zoom property is set.</p>
Fit to Pages Tall (Excel)	<p>Sets number of pages tall the worksheet will be scaled to.</p> <p>This property is ignored if the Zoom property is set.</p>
Include Document Markups	<p>Determines whether document markups are retained.</p> <p>When this property is False (the default), document markups are omitted.</p> <p>When this property is True, markups are included.</p>

5.2.7.23.2 Extended Engine

Parameter	Notes
General Settings	
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Continue on Error	Continue processing TIFF files after an error occurs.
Over Write Existing	Over Writes the target document if it exists.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Password Files	<p>This option specifies what Autobahn does when it encounters a password protected PDF file. The file will be copied to the password sub directory in the Error Folder.</p> <ul style="list-style-type: none"> • Take no action. • Move to Error Folder • Copy to Error Folder
Maximum Cores	<p>This specifies the number parallel files you want to be processed at a given time.</p> <p>Note: You need the multi-core license for this.</p>
Debug	Set this to true to execute the job in debug mode.
Extended OCR Settings	
Output File Type	<p>One or more of the following, separated by commas if more than one is required.</p> <p>RTF PDF DOCX CSV* SML (Spreadsheet ML XML file)* HTM TXT</p> <p>*These output formats are suitable for table-oriented pages that can be mapped onto a spreadsheet format.</p>
OCR Engine	The OCR engine to use. This must be set to use the IRIS engine.
OCR Language 1-5	You can set up to five different languages for OCR recognition in one page as long as they are in the same character set.

Automatic language detection	Property that enables or disables the Auto Language Detection feature. The aim of this feature is to detect the most probable language of a single-language page. If at least one language has been detected, recognition will be performed in the first language candidate that has been detected, and not in the language(s) set through Language or Languages. If it fails to detect a language, recognition will be performed using the language(s) set through Language or Languages.
Auto rotate	Detect page orientation and correct if required
Deskew	Rotates the image to correct its skew angle.
Advanced Deskew	Set this to true to defined advanced deskew properties.
Force Deskew	Under certain circumstances, rotating the image to correct its skew angle may decrease the OCR accuracy. The extended engine is able to analyze the image and detect from an OCR accuracy point of view whether it's better to rotate the image or not. Because the skew angle may be visible in the output document (e.g. if KeepDeskew is set to 'true'), you can choose to force the deskew to rotate the image, even if it affects the accuracy. If turned off, the image is analyzed before rotation and the engine may choose not to rotate the image depending on the analysis result. If turned on, the image is rotated to correct skew angle.
Adjustment Mode	Set the behavior regarding dimension adjustment for deskew operation.
Despeckle	Removes all the groups of connected pixels with a few pixels below the parameter. Suggested range: 1-20.
Advanced Despeckle	Set the advanced despeckle settings, advanced despeckle provides advanced image noise reduction features by the image despeckle filter.
Remove White Pixels	By default, Advanced Despeckle removes black pixels. If this setting is set to 'true', white pixels will be removed instead of black pixels.
Dilate	Despeckle removes all the groups of connected pixels with a few pixels below the SpeckleSize parameter. Those connected pixels are not removed if the distance to a larger connected component is below this parameter. As a result,

	only the isolated pixels get deleted. The maximum value for this property is 20 pixels. The default value is '0'.
Layout	The layout for the dock or rtf document <ul style="list-style-type: none"> • Standard • Flow
PDFVersion	This determines the PDF version of the generated PDF: 1.4 1.5 1.6 1.7 PDF/A1B (<i>PDF/A-1b</i>)
Remove Blank Page	Set this to true to remove blank pages from Tiff or PDF documents. Value needs to be set for sensitivity (see below).
Sensitivity	The sensitivity, from 1 to 100. With high sensitivity, fewer blank pages are detected.
Work Depth	This parameter (0 – 255) defines how deeply the OCR engine will analyze a page with 255 being the deepest. For poorer quality documents, higher values can give better recognition results.
JPEG Quality	This parameter (0 – 255) determines the compression/quality of color JPEG images in generated PDFs. 0 gives the smallest file size whilst 255 gives the best quality. The default value is 128.
JPEG2000 Compression	Enable/Disable JPEG2000 Compression.
JPEG2000 Compression Mode	The JPEG2000 Compression Mode to use.
JPEG2000 Compression Value	The Value to set for the selected Compression Mode.
IHQC Compression	Apply Intelligent High-Quality Compression
IHQC Compression Level	Level1 is the basic compression level while level 3 is the most advanced Intelligent High-Quality Compression Mode.
IHQC Quality Factor	The quality Factor for IHQC
No OCR	Whether are not to perform OCR on the document (Yes to not perform OCR, No to perform OCR).
Binarization	Whether or not to perform binarization on the document.

Brightness	The brightness (higher values will darker the result).
Contrast	The contrast (lower values will darker the result).
Smoothing Level	Smoothing may be useful to binarize text with a colored background to avoid noisy pixels (0 disables smoothing, higher values smooth more).
Undithering	Whether or not to use automatic undithering while processing a page. NOTE: Automatic undithering will be applied only if smoothing is also activated (Smoothing Level) Dithering is a scanning technique which consists in representing a color or greyscale image using only a limited color palette. This allows reducing file size while maintaining the general aspect of the image. This technique is known to create images more difficult to handle for OCR technology; therefore specific image preprocessing is needed to detect and revert it.
Threshold	Sets the threshold for fixed threshold binarization (0 for automatic threshold computation).
Remove Lines	Whether or not to remove lines from an image (The image must be black and white).
Horizontal Clean X	The parameter for cleaning noisy pixels attached to the horizontal lines.
Horizontal Clean Y	The parameter for cleaning noisy pixels attached to the horizontal lines.
Vertical Clean X	The parameter for cleaning noisy pixels attached to the vertical lines.
Vertical Clean Y	The parameter for cleaning noisy pixels attached to the vertical lines.
Horizontal Dilate	The dilate parameter that helps the detection of horizontal lines.
Vertical Dilate	The dilate parameter that helps the detection of vertical lines.
Horizontal Max Gap	The maximum horizontal line gap to close. It is useful to remove broken lines.
Vertical Max Gap	The maximum vertical line gap to close. It is useful to remove broken lines.

Horizontal Max Thickness	The maximum thickness of the horizontal lines to remove. It is useful to keep vertical lines larger than this parameter. Can be also useful to keep vertical letter strokes.
Vertical Max Thickness	The maximum thickness of the vertical lines to remove. It is useful to keep horizontal lines larger than this parameter. Can be also useful to keep horizontal letter strokes.
Horizontal Min Length	The minimum length of the horizontal lines to remove.
Vertical Min Length	The minimum length of the vertical lines to remove.
Remove Dark Borders	Removes the dark surrounding from bitonal, grayscale or color images. The dark surrounding of the image is whitened (Note: The dark border should be touching the edge of the page for this to work).
Punch Hole Removal	Attempts to remove punch holes from pages. Note: The punch hole algorithm can be used on images with the following minimum dimensions width: 300px, height: 100px (computed for 300 DPI). The minimum height and width can vary with the image resolution.
Interpolation	Interpolates the source image to the given resolution. This value (the target resolution) must be greater than the source image's resolution.
Interpolation Mode	Sets the interpolation mode.
Keep Original Image	Set this to true if you want to use the pre-processed image for OCR but keep the original image in the output document. The default value is 'true'. Note: This property only applies when processing image files or when processing PDF files with the Extract Image Method set to ConvertToTiff
Keep Deskewed Image	Set this to true if you want to use the deskewed image in the output document. Note: This property only applies when Keep Original Image is set to No
Keep Despeckled Image	Set this to true if you want to use the despeckled image in the output document. This requires the source image to be black and white. Note: This property only applies when Keep Original Image is set to No
Keep Dark Border Removal	Set this to true if you want to use the image after dark borders have been removed, in the output document.

	Note: This property only applies when Keep Original Image is set to No
Keep Punch Hole Removal	Set this to true if you want to use the image after punch holes have been removed, in the output document. Note: This property only applies when Keep Original Image is set to No
Any File To PDF Conversion Settings	
Conversion Timeout (ms)	Limits the amount of time in milliseconds that can be spent on conversion. A value of zero means waits indefinitely.
Convert Bookmarks	For MS Word, convert bookmarks
Bookmark Depth	<p>This property will take effect only when the Convert Bookmarks property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, an empty string is functionally equivalent to</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use a non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.</p>
Convert Hyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
Print All Sheets (Excel)	The flag that indicates whether to print all Excel worksheets or not.
Print Background Color (IE)	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
Print Scale % (Visio)	For Visio files, sets the print scale
Header (IE)	This property modifies Internet Explorer's header setting.
Footer (IE)	This property modifies Internet Explorer's footer setting.
Image Compression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).

Image Downsizing	If this property is set to Yes, then the resolution of images is reduced to the DPI value specified in the Downsize Resolution DPI property.
Downsize Resolution DPI	If the Image Downsizing property is set to True, then the resolution of images is reduced to the DPI value specified in this property.
Image JPEG Quality	The allowed value range is from 5 to 100 with 100 being the highest quality.
Font Embedding	The option PRN_FONT_EMBED_FULLSET (embedding a full set of font) will cause a significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed the font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
Font Substitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from the Control Panel).
Embed Fonts as Type 0	This option is recommended if you have non-standard fonts like barcode font.
Top Margin	Sets top margin. (Inches)
Bottom Margin	Sets bottom margin. (Inches)
Left Margin	Sets left margin. (Inches)
Right Margin	Sets right margin. (Inches)
Page Width	Sets a custom page width. (Inches)
Page Height	Sets a custom page height. (Inches)
Paper Orientation	Sets paper orientation to <ul style="list-style-type: none"> • Default (Maintain Source Orientation) • Landscape • Portrait
PDF Compliance	Allows the User To choose a PDF/A or PDF/X Complaint files <ul style="list-style-type: none"> • None (No PDF/A Output) • PDF/A-1b (PDF/A-1b compliant) • PDF/X-1a (PDF/X-1a compliant) • PDF/X-3 (PDF/X-3 compliant)

Convert MSG Attachments	If you set this to true, Autobahn DX will convert the both MSG files and their Attachments to a single PDF file.
Attach MSG Attachments to PDF	If set to true, Autobahn DX will Attach Msg Attachments that are converted as PDF Attachments. If set to false, Autobahn DX will merge Msg Attachments that are converted to the PDF file generated by the body.
Preserve Word Attachments	Determines whether embedded and linked files will be preserved during conversion. Default value: False (disabled). Note: This will work with WordExtentionEX only
Convert PDF Attachments (PDF)	Convert PDF Attachments to create a combined PDF file.
Merge PDF Attachments (PDF)	Set this flag to true if you want to convert pdf attachments and merged them into the output pdf file. Otherwise, the converted files will be merged back to the pdf.
Retain PDF Attachment (PDF)	Switch this on to Retain the Original PDF attachments if you set Merge PDF Attachments to true.
Retain Properties (Office)	Set this flag if you want the MS Office properties to be transferred to the target pdf document.
Color Type (PowerPoint)	Use this property to set PowerPoint to print with either color, grayscale, or black and white.
Handout Order (PowerPoint)	Sets the handout order, this flag only applies to PowerPoint jobs. The possible values are: <ul style="list-style-type: none"> • Vertical First • Horizontal First
Output Type (PowerPoint)	Sets the output type, it only works with the PowerPoint files. The possible values are: <ul style="list-style-type: none"> • Slides • Build slides • Two slides handouts • Three slides handouts • Four slides handouts • Six slides handouts • Nine slides handouts • Notes • Outline
Print Graphics (Publisher)	Sets the graphics setting for printing. <ul style="list-style-type: none"> • Print Full Resolution • Print Low Resolution

	<ul style="list-style-type: none"> • Print Graphics
Frame Slides (PowerPoint)	Indicate whether to draw a frame around the border of the slides.
Zoom (Excel)	Sets printing zoom of the worksheet. The allowed value range is from 10 to 400.
Fit to Pages Wide (Excel)	Sets number of pages wide the worksheet will be scaled to. This property is ignored if the Zoom property is set.
Fit to Pages Tall (Excel)	Sets number of pages tall the worksheet will be scaled to. This property is ignored if the Zoom property is set.
Include Document Markups	Determines whether document markups are retained. When this property is False (the default), document markups are omitted. When this property is True, markups are included.

5.2.7.24 Barcode TIFF/PDF

This step can detect barcodes in TIFF/PDF files and either Split/Rename the file based on the barcodes detected.

Screen Field / Button	Description
Output File Name	<p>The output file path template where the split files will be saved.</p> <ul style="list-style-type: none"> • %VALUE%: Replaced by the barcode value found. • %INDEX%: Replaced by the current split index. • %FILENAME: Replaced by the file name
Output File Name (No Barcode)	<p>The renaming template to use for page ranges where no barcodes were identified.</p> <p>Allowed templates:</p> <ul style="list-style-type: none"> • %INDEX%; Replaced by the current split index. • %FILENAME: Replaced by the filename of the source file.
Barcode Operation	<p>Select between Split by Barcode or Rename by Barcode</p> <ul style="list-style-type: none"> • Split by Barcode: Choose this option to split the TIFF/ PDF file by Barcode. • Rename by Barcode: Choose this option to rename the TIFF/PDF file based on Barcode.
Split Mode	Various Options for splitting Files by Barcode

	<ul style="list-style-type: none"> • Barcode on First Page • Barcode on Last Page • Remove Barcode Page
Barcode Format	Barcode formats supported.
Try Harder	Spend more time to try to find a barcode; optimize for accuracy, not speed. The default is true.
Overwrite Existing	Overwrites any file that exists with the same name in the output folder. Note: If you have the same barcode in different pages or files, they will be overwritten if this is set to true.
Metadata Name	<p>Sets the extract barcode value as metadata in the target PDF file. Chose a Metadata you want to set in the output file</p> <ul style="list-style-type: none"> • None • Author • Creator • Keywords • Producer • Subject • Title • Trapped • Custom
Custom Metadata Name	If you want to add custom metadata instead, you can do this by setting Metadata Name to Custom , then enter the name here.
Metadata Value	<p>Enter a template for the Metadata Value, you can use the file naming variables here too.</p> <ul style="list-style-type: none"> • %VALUE%: Replaced by the barcode value found. • %INDEX%: Replaced by the current split index. • %FILENAME: Replaced by the file name
Perform Pre-processing	Do not enable this option unless instructed by Aquaforest support.
Binarize	Set this to true to get better results from colored files.
Deskew	Straighten the image.
Remove Lines	Whether or not to remove lines from an image.

Despeckle	Remove specks below the specified pixel size from the image.
Box Size	This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the barcode reader. Technically, this option removes connected elements with a minimum area (in pixels and defined by this property). This option is currently only applied for bitonal images.
Zones	Only examine the region specified for barcode(s). Note to specify the zone you need to set the following in the step properties: <ul style="list-style-type: none"> • Left • Top • Width • Height
PDF DPI	The DPI of TIFF images generated from the source PDF file. These images are then used for barcode recognition.
TIFF Compression	The compression to set to the TIFF images generated or converted from the source PDF file. These images are then used for barcode recognition
Advanced Flags	Additionally advanced command-line flags may be entered here (see section 3)
Continue on Error	Continue processing TIFF/PDF files after an error occurs.
Maximum Cores	The number of parallel files Autobahn DX will attempt to process at the same time.
Debug	Set this to true to execute the job in debug mode.

5.2.7.25 High Availability

The high availability step in Autobahn DX is designed to utilize two instances of the product running on separate hosts.

Screen Field / Button	Description
Current Job ID	The Job ID on the current host.
Default Status	Select the Default status of the current host (Master Slave)
Shared Status File	Enter the shared.txt file location – this needs to be on a shared network location accessible to both hosts.
Hostname	Name of the paired host.
ADX Install Path	Install path of Autobahn DX on the paired host.
Job ID	The Job ID on the paired host

5.2.7.26 Distributed Polling

This step can be used to implement load balancing in Autobahn DX, it achieves this by copying a fraction of the files from a central input location to the local system where Autobahn DX is running. Multiple Autobahn DX servers can point to one input folder, as a result, the files will be shared across several servers and the processing will be more optimized.

See the [Distributed Polling Section](#) for more details

Screen Field / Button	Description
Autobahn Job ID	The Job ID of the Job that will be processing your input files. Note: The Source Folder of this job will be the Destination Folder of the Distributed Polling Job
Limit	The maximum number of files to be copied to the per run.
Extensions	Enter the file extensions you want us to copy separated by a comma. E.g. “.pdf,.tif,tiff”
Process Sub Folder	Select true if you want to copy subfolders.
Debug	Select true if you want to see more debug output.

5.2.7.27 Kingfisher Job

This step is meant to work with Kingfisher, you can set a kingfisher job up in Kingfisher and add it as a step in Autobahn DX, see [Kingfisher Job Step Section](#) for more details.

Screen Field / Button	Description
Kingfisher Job ID	The Kingfisher Job ID

5.2.8 Step Type Properties

Each of the Step Types referred to in the previous section will have a set of properties such as that shown below for “Convert any File to PDF”. Each property has a description associated with it which is displayed when the property is highlighted.

Step Properties

1 - Conversion Settings

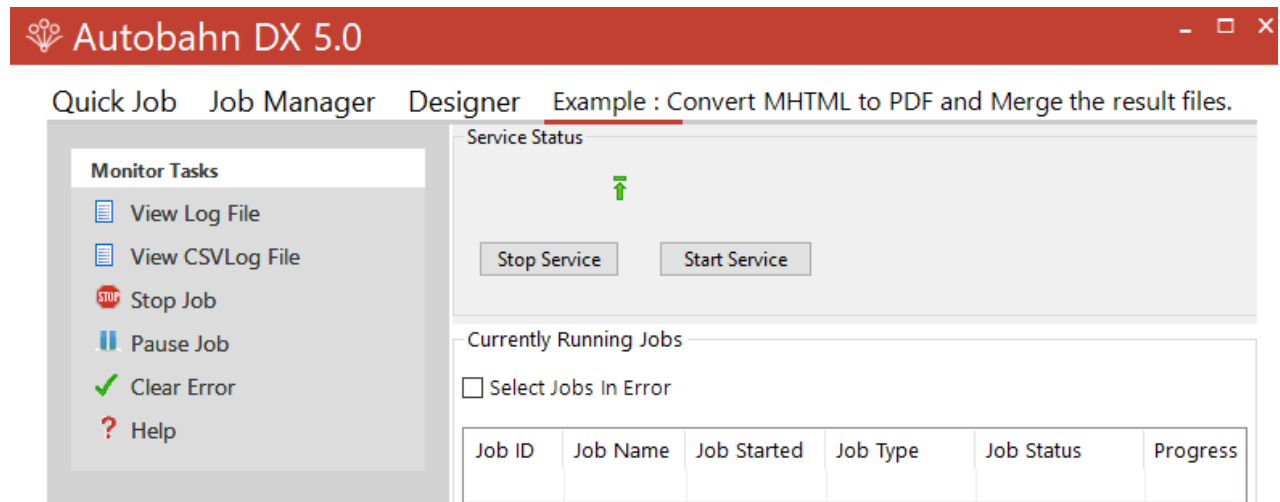
Output File Name	%FILENAME.pdf
Continue on Error	Yes
Conversion Timeout (ms)	
Convert Bookmarks	No
Bookmark Depth	
Convert Hyperlinks	No
Print All Sheets (Excel)	No
Print Background Color (IE)	No
Print Scale % (Visio)	
Header (IE)	
Footer (IE)	
Image Compression	COMPRESS JPEG

Output File Name

Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of original file)

5.3 Monitor

This provides an interface to stop and start the Autobahn service, and to see which Jobs are currently running.



Menu	Item	Action
	View Log File	View the log file of the currently selected job.
	View CSV Log File	View the log file of the currently selected job.
	Stop Job	Halt execution of the selected job.
	Pause Job	Pauses the execution of a job, you continue this job later. Note: If you make any changes to the Job when it is in a Paused state the job will start from the beginning.
	Clear Error	Clears an error flag associated with a job, allowing it to be rerun.
	Stop Service	Stop the Autobahn service.
	Start Service	Start the Autobahn service.

5.4 Modules and Options

The Modules and options tab serves two purposes. The first use is for defining the license key which was explained in [section 3.3](#). The second use is to define the information of the Email server. This is used to support the “Send Email on Error” functionality defined in the previous section. The information entered is help in Autobahn.config in the config directory. The password is held in an encrypted form.

Autobahn DX 5.0

Quick Job Job Manager Designer Monitor Running Job **Modules & Options** Help Example : Convert MHTML to PDF and Merge

Options

Save Options

Help

Core Module

License Key

Permanent License

Server License

+ Extended OCR License

+ Asian Languages License

+ IHQC Licence

+ Multi-core License With 16 cores

Barcode License

SMTP Settings

Server ☐ Need Authentication

SMTP Port

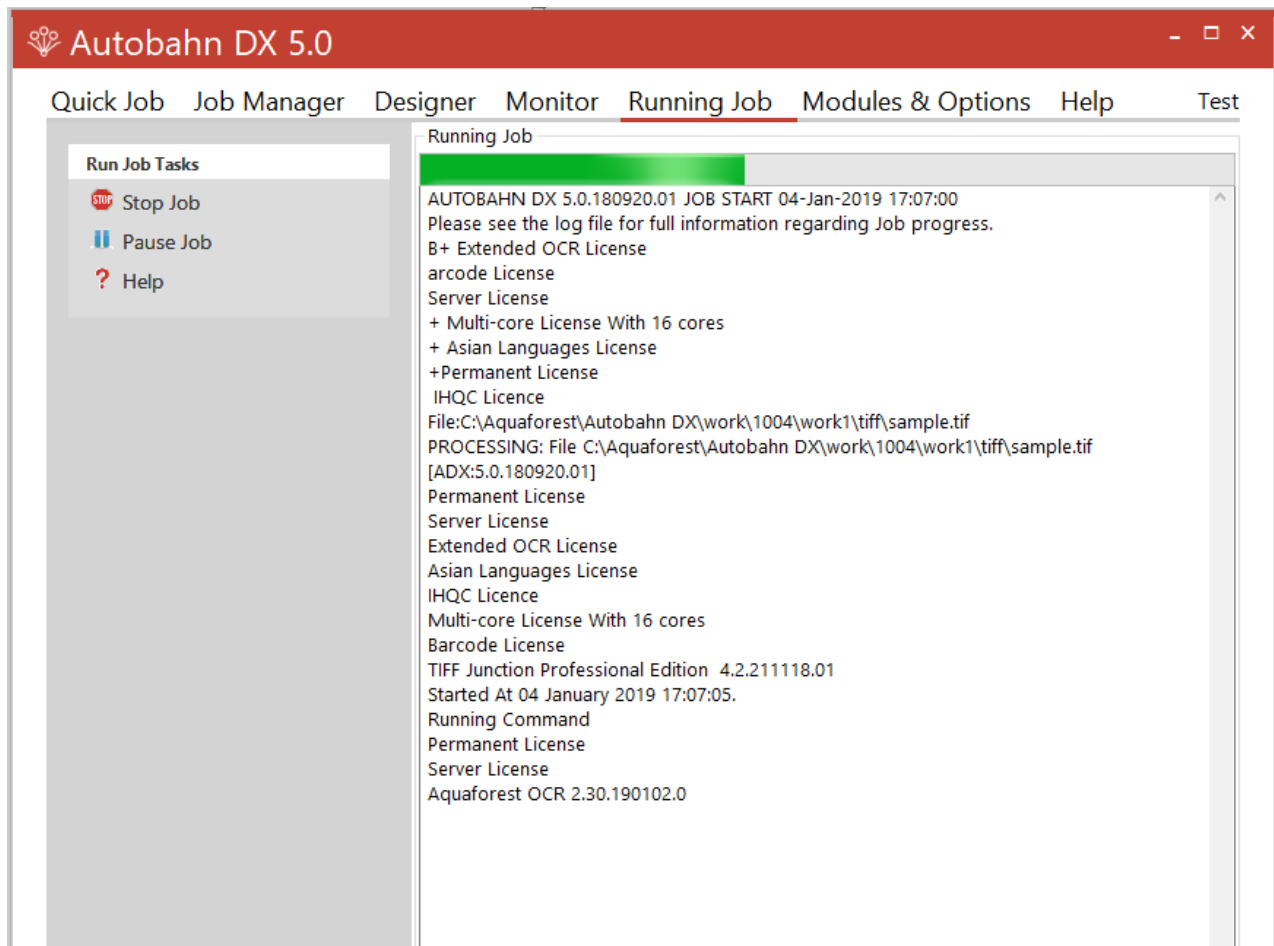
Username

Password

Field	Description
Server	Address of the Server hosting the SMTP server.
SMTP Port	SMTP Server port. Default is 25.
Username	Username for authentication by the server.
Password	Password for the username.
Need Authentication	If checked, authentication will be used and the username and password can be filled in.

5.5 Running Jobs via Job Manager or Designer

Choosing “Run Now” from the menu executes the current job definition and all output is placed in the text area window, as well as the Job log files.



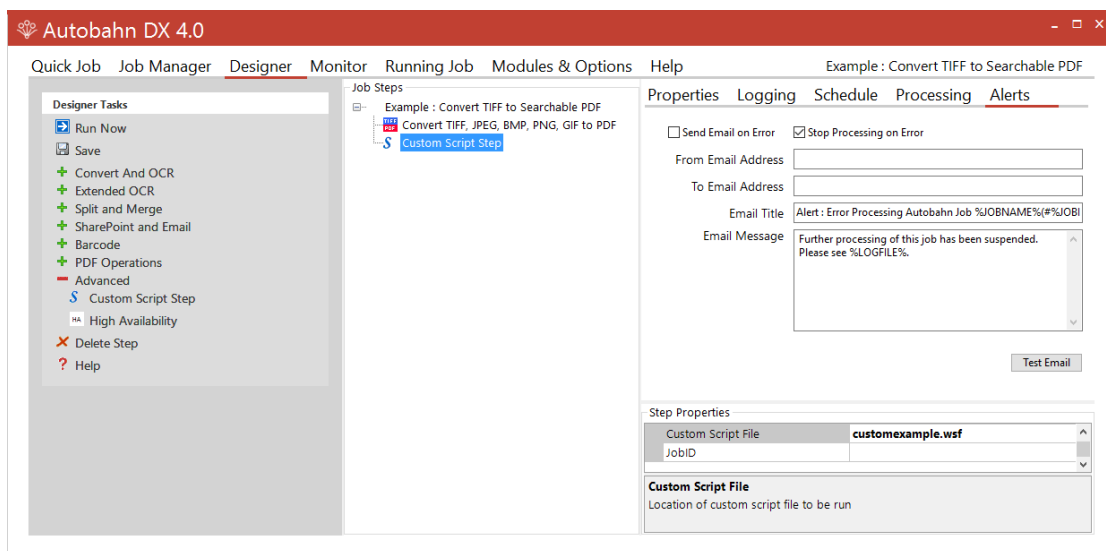
Menu	Item	Action
	Stop	Stops the currently running job.
	Pause Job	Pauses the execution of a job, you continue this job later. Note: If you make any changes to the Job when it is in a Paused state the job will start from the beginning.
	Help	Take you to the Help tab.

6 SCRIPTING CUSTOM STEPS

Autobahn DX allows custom job steps to be defined that can be included in a Job Definition in the same way as any other step type. The custom step itself can be any form of executable file or script. Some template wsh scripts are included which demonstrates calling a command-line executable from within the script.

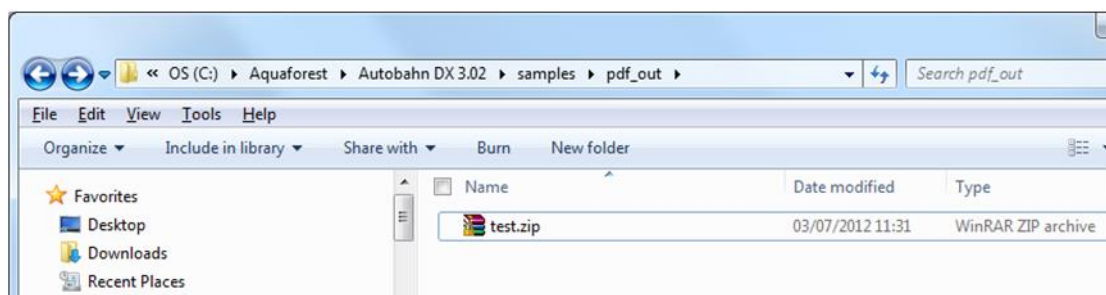
6.1 Custom Script Example

Included with the product is an example custom script (in the Autobahn DX custom folder) that will split a PDF file into single pages and then create a single ZIP file that contains the pages. The example makes use of the Freebyte Zip application which you can download from here <http://www.freebyte.com/download/fbzipack.exe>, run and unpack into C:\Aquaforest\Autobahn DX\custom.



Custom script files must be placed in C:\Aquaforest\Autobahn DX\custom folder. In this case, the custom script file is customexample.wsf.

The results of the Job Definition will be a zipped file as shown:



6.2 Custom Example Script

The contents of Customexample.wsf are shown below. This file can be used as a template for other custom job steps.

Note that by default Autobahn DX processes job steps by using a separate folder for each step. Files to be processed will be in the folder defined by the infolder variable and the resultant files placed in the folder defined by the outFolder variable. This is required even if the custom step does not actually change the files.

Jobs can be configured to run without using Work Folders, in which case the infolder and will be the same.

```
<job>
<runtime>
</runtime>
<script language="JScript">

/*****
**
**      (C) Aquaforest Limited 2006
**
**      FILE
**      customexample.wsf
**
**      DESCRIPTION
**      Example Custom Script for Autobahn DX
**
*****/

var WshShell = WScript.CreateObject("WScript.Shell");

/*****
      Arguments
*****/

var inFolder="";
var outFolder="";

if(WScript.Arguments.Named.Exists("in"))
{
    inFolder=WScript.Arguments.Named("in");
}
if(WScript.Arguments.Named.Exists("out"))
{
    outFolder=WScript.Arguments.Named("out");
}

doOut("In Custom Script Example! /in="+inFolder+" /out="+outFolder);

doCmd('C:\\Program Files\\Aquaforest\\Aquaforest Autobahn\\custom\\fbzip.exe -a "'+outFolder+"\\customexample.zip" "'+inFolder+' ');

/* Write Output */

function doOut(s)
{
    WScript.StdOut.WriteLine(s);
    if(loggingOn)
    {
        logOut.WriteLine(s)
    }
}

/* Execute Command */

function doCmd(cmd)
{
```

```

doOut(cmd);
var oExec = WshShell.Exec(cmd);
while(oExec.Status == 0)
{
    WScript.Sleep(100);
    if(verbose)
    {
        while(!oExec.Stdout.AtEndOfStream)
            doOut(oExec.Stdout.ReadLine())
    }
    else
    {
        while(!oExec.Stdout.AtEndOfStream)
            oExec.Stdout.ReadLine()
    }
    while(!oExec.Stderr.AtEndOfStream)
        doOut(oExec.Stderr.ReadLine())
}

if(verbose)
{
    while(!oExec.Stdout.AtEndOfStream)
        doOut(oExec.Stdout.ReadLine())
}
else
{
    while(!oExec.Stdout.AtEndOfStream)
        oExec.Stdout.ReadLine()
}
while(!oExec.Stderr.AtEndOfStream)
    doOut(oExec.Stderr.ReadLine())

if(oExec.ExitCode==0)
{
}
else
{
    doOut("Task failed with error code : "+oExec.ExitCode);
    WScript.Quit(1);
}
}
</script>
</job>

```

6.3 Further Examples

6.3.1 Postscript to PDF: custom_ps.wsf

This script included in the custom folder can be used to convert Postscript files to PDF. It makes use of the ps2pdf.bat script included in Ghostscript 8.63. See <http://pages.cs.wisc.edu/~ghost/>

The script should be placed in the Autobahn DX custom folder (by default C:\Aquaforest\Autobahn DX\custom) and can be used by creating a job with a custom script step and entering *custom_ps.wsf* as the value of Custom Script File.

If required, changes can be made to the allowableExtensions variable and the genCommand() function.

As the ps2pdf script calls other Ghostscript components, the Ghostscript bin and lib folders should be added to the system PATH.

6.3.2 MSG to PDF: custom_MSG.wsf

As Microsoft Outlook is limited in the way that it can be automated in the way required by the easyPDF Driver used by Autobahn DX it can be more pragmatic to employ a custom script approach,

This script can be used to convert msg files to PDF. This example makes use of TotalMailConverter (see <http://www.coolutils.com/TotalMailConverter>) but other products could equally be used.

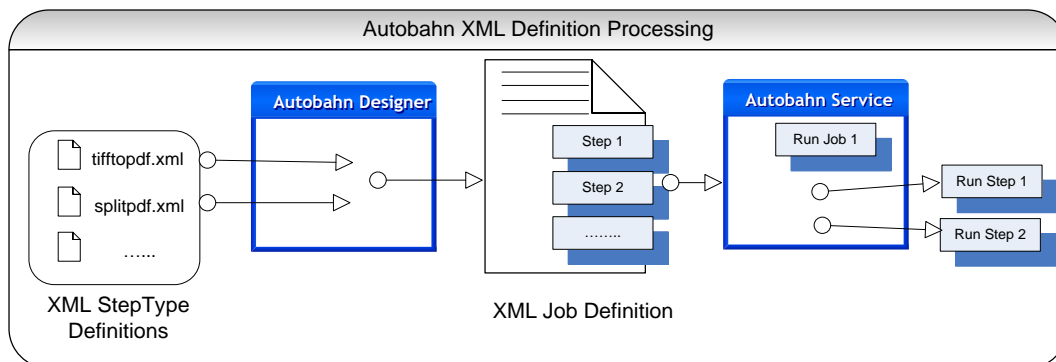
The script should be placed in the Autobahn DX custom folder (by default C:\Aquaforest\Autobahn DX\custom) and can be used by creating a job with a custom script step and entering *custom_msg.wsf* as the value of Custom Script File.

If required, changes can be made to the allowableExtensions variable and the genCommand() function.

The Autobahn DX service should be configured to allow interaction with the desktop and to run as a standard user account rather than the Local System account.

7 JOB DEFINITION XML FILES

7.1 Overview of Job Definition Creation and Processing



Autobahn Job Definitions are represented as XML files which can be created “by hand”, or more likely by using the Autobahn Administrator (see [section 6](#)). A job definition file contains certain standard pieces of information (source folder for example) and also has a “StepType” Definition for each step in the process. It is these step definitions that are executed via the Autobahn service.

7.2 Job IDs

Autobahn uses a sequential integer job id that starts at 1001. The “next job id” value is held in the config/next_job_id/xml file and is updated each time a new job is created or copied using the administration tool.

The initial contents of the file are:

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<next_job_id>1001</next_job_id>
```

7.3 Sample Job Definition File (Excluding Step Details)

Below is a simple example of a job definition which is designed to continuously (every 30 seconds) monitor a directory (c:\faxes) for new .tif files and convert them to searchable PDF, and place the resulting files in c:\processed faxes.

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<autobahnjob>
  <jobid>1002</jobid>
  <jobname>Watch Incoming Faxes</jobname>
  <schedulertype>continuous</schedulertype>
  <scheduleevery>30</scheduleevery>
  <scehduleeveryunits>seconds</scehduleevery>
  <schedulefrom>00:00</schedulefrom>
  <scheduleto>23:59</scheduleto>
  <scheduleat></scheduleat>
</joblogfile>
<jobsteps>1</jobsteps>
<jobsourcetype>folder</jobsourcetype>
<jobsource>c:\faxes</jobsource>
<jobtarget>c:\processedfaxes</jobtarget>
<joberrors>c:\errorfaxes</joberrors>
<jobdeleteonsuccess>yes</jobdeleteonsuccess>
<jobwork>C:\autobahnwork\1002</jobwork>
```

```

<jobstep>
  <stepsequence>1</stepsequence>
  <steptype>splitpdf</steptype>
  <stepdetails>
    (see section 8.4 below)
  </stepdetails>
</jobstep>
</autobahnjob>

```

XML	Element	Description
jobid		The job ID number (see 3.2 above).
jobname		Job Description. Default is "Job %JOBID%"
scheduletype		<ul style="list-style-type: none"> • Ad-hoc • Continuous • Onceperday
scheduleevery		For continuous schedule type, interval between runs
scheduleeveryunits		For the continuous schedule type, unit of the interval between runs
schedulefrom		For continuous, start time for runs
scheduleat		For continuous, end time for runs
scheduleto		For onceperday, the time at which to run
job logfile		Location of the job log file. By default the log file is logs/%JOBID%/TIMESTAMP%.txt
jobsteps		The number of job steps
jobsourcetype		File, folder or tree
jobsource		The source file or folder
jobtarget		The target folder
Joberrors		Folder for job errors. If not specified files that cannot be processed will be placed in jobwork/errors (this is also the default).
jobdeleteonsuccess		If "yes", when a job has successfully completed, all work files (hence input files) are deleted.
jobwork		The root of the temporary work directories used by the job. The work directories themselves are named work1, work2 etc.
jobstep		Contains the definition of a job step
stepsequence		Step sequence number starting at 1.
steptype		The steptype identifies which job element defines the step (see section 8.4 below)
stepdetails		Holds additional step definitions which will be used as the step definition file.

7.4 Sample Step Details

The details of the step are provided here and provide sufficient information for the Autobahn service to execute the step, in conjunction with the information in the StepType definitions. Files are moved from the jobsource directory into the working directory specified by <sourcefiles> and the result files are placed in the <target> directory. Upon completion of all the steps, the service will move the files from the final work directory to the <jobtarget> directory.

```
<stepdetails>
  <operation>pdf</operation>
  <sourcetype>folder</sourcetype>
  <sourcefiles>C:\autobahnwork\1002\work1</sourcefiles>
  <target>C:\autobahnwork\1002\work2</target>
  <errors>c:\errorfaxes</errors>
  <joboptions>-d -k 1 -h 0 -o %FILENAME.pdf -p </joboptions>
  <advancedflags></advancedflags>
  <metadata></metadata>
  <security></security>
  <dcoptions></dcoptions>
  <custom_script></custom_script>
  <custom_script_pre_job>N</custom_script_pre_job>
  <custom_script_post_job>N</custom_script_post_job>
  <custom_script_pre_folder>N</custom_script_pre_folder>
  <custom_script_post_folder>N</custom_script_post_folder>
  <custom_script_pre_file>N</custom_script_pre_file>
  <custom_script_post_file>N</custom_script_post_file>
  <custom_script_on_error>N</custom_script_on_error>
  <logfile></logfile>
</stepdetails>
```

XML	Element	Description
Operation		The operation (e.g. split). This is defined in the step definition file for the steptype.
Sourcetype		Folder, file or tree.
Sourcefiles		Source file or folder.
Target		Target folder.
Errors		Files that cannot be processed are placed in this directory. Inherited from the Job definition.
Joboptions		These are steptype-specific parameters that are derived from the options selected in the Job Designer.
advanced Flags		Additional advanced steptype-specific parameters that can be entered manually only.
Metadata		PDF file metadata settings derived from the options selected in the Job Designer.
Security		PDF file security settings derived from the options selected in the Job Designer.
Dcoptions		PDF file open options derived from the options selected in the Job Designer.
custom_script		If not left blank, specifies the name of a custom windows script file in %PDFJUNCTIONDIR%\custom or %TIFFJUNCTIONDIR%\custom that will be called in accordance with the custom_script_* settings below.

custom_script_pre_job	If =Y, the custom script will be called at the start of the job
custom_script_post_job	If =Y, the custom script will be called at the end of the job
custom_script_pre_folder	If =Y, the custom script will be called at the start of processing each folder
custom_script_post_folder	If =Y, the custom script will be called at the end of processing each folder
custom_script_pre_file	If =Y, the custom script will be called at the start of processing each file
custom_script_post_file	If =Y, the custom script will be called at the end of processing each file
custom_script_on_error	If =Y, the custom script will be called when a processing error occurs.
Logfile	If specified, the output will be logged to a file with this name in %PDFJUNCTIONDIR%\logs or %TIFFJUNCTIONDIR%\logs

Notes

- Note that step-level custom scripts referred to above as supported by TIFF Junction and PDF Junction are different from “custom scripts” that may be used as a step.
- The step-level log files as referred to above as supported by TIFF Junction and PDF Junction are separate from the overall log file used by Autobahn, and can only be specified manually.

8 EXTENDED STEP TYPES: STEP TYPE DEFINITION XML FILES

Autobahn is designed to be extensible, and as such is driven by StepType Definitions which are used to define each of the possible steps that can be included in a job definition.

A step type definition will include information such as:

- Type of Document Accepted (e.g. TIFF, Word, PDF)
- Supported Source Types (e.g. Folder, Email...)
- Type of Document Output (e.g. PDF)
- Process to Call
- Command Line Parameters

8.1 Sample StepType Definition

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<steptype>tifftopdf</steptype>
<command>cscript %AUTOBAHNROOT%/tj/bin/tiffjunction.wsf /jobid=%AUTOBAHNTEMP%/TEMPJOBFILE%</command>
<operation>pdf</operation>
<joboptions>-p</joboptions>
<wrapper>tiffjunction_job</wrapper>
<multithread>yes</multithread>
<validinputtype>tif</validinputtype>
<outputtype>pdf</outputtype>
<iconfile>tifftopdf.bmp</iconfile>
<attributes>
  <attribute>
    <attributeid>SPDF</attributeid>
    <attributename>Searchable PDF</attributename>
    <attributetype>option</attributetype>
    <attributedefault_value>Yes</attributedefault_value>
    <attributeoptions>
      <option>
        <value>Yes</value>
        <joboptions>-k 3</joboptions>
      </option>
      <option>
        <value>No</value>
        <attributejoboptions></attributejoboptions>
      </option>
    </attributeoptions>
  </attribute>
  <attribute>
    <attributeid>ofn</attributeid>
    <attributename>Output File Name</attributename>
    <attributetype>value</attributetype>
    <defaultvalue>%FILENAME.pdf</defaultvalue>
    <attributejoboptions>-o%ATTRIBUTEVALUE%</attributejoboptions>
  </attribute>
</attributes>
```

XML	Element	Description
	steptype	The name that uniquely identifies the step type
	command	The command line to be executed.
	operation	The “internal” step type name used by the command
	joboptions	Setting for the <joboptions> element in <stepdetails>. This value may be added to by attribute settings.
	wrapper	Wrapper element to be used when creating temp step job file.
	multithread	Yes or no – determines whether multiple instances of the steptype may run concurrently.
	validinputtype	Type(s) of files accepted by the step. (NB this is NOT the same as a file extension). * means all types. Multiple values may be separated by commas.
	outputtype	Type of file output by the step. (NB this is NOT the same as a file extension)
	iconfile	Bitmap file with the steptype icon. The file is assumed to be in the %STEPTYPEDEFDIR% directory.
	attribute	The definition of an attribute which the user may set a value for
	attributeid	The unique internal name for the attribute
	attributename	The display name for the attribute
	attributetype	Option or value
	attributeoptions	Only applies where attributetype=value. Note that the default option is the first listed
	value	Display value
	valuejoboptions	String to be appended to <joboptions>
	defaultvalue	Default value (for value types)

9 AUTOBAHN DX COMMAND LINE INTERFACE

A command line interface is available to run any job that could be run within the Quick Job section of the product. The set of parameters is comprehensive and command line syntax is generally best built by using the Quick Job “Command Line” menu option.

The stand-alone command-line interface can only be used with the Server licenses.

A sample Visual Studio project showing calling of the command line interface via C# is provided in the AUTOBAHN \code examples\CommandLineCall folder where AUTOBAHN is the folder where Autobahn DX has been installed – by default this is C:\Aquaforest\Autobahn DX.

```
autobahndx.exe /operation=[operation name] /source=[tiff file or folder] /output=[output file]  
/target=[target folder] [/option=value]...
```

Examples

1. Generate a searchable PDF c:\destination\2.pdf from a TIFF file

```
autobahndx.exe /source=C:\source\2.tif /sourcetype=file /target=C:\destination\2 /outputtype=pdf  
/operation=tifftopdf /ocr=true
```

2. Generate a searchable PDF file from a folder of TIFF and JPEG files, with Deskew and page orientation detection and correction.

```
autobahndx.exe /source=C:\Source /sourcetype=folder /target=C:\destination /output=outfilef  
/outputtype=pdf /ocr=true /autorotate /deskew /operation=mergetifftopdf
```

3. Generate searchable PDF files from image PDF files found in a folder and subfolders.

```
autobahndx.exe /source=c:\C:\Users\khalil\Desktop\2 /sourcetype=tree  
/target=c:\Users\khalil\Desktop\3 output=%FILENAME /ocr=true /outputtype=pdf  
/operation=ocrimagepdf
```

4. Run an Autobahn job that is in the Job Manager, e.g to run job 1001.

```
autobahndx.exe /jobId=1001
```

The flags below work with the **autobahndx.exe** file in the bin folder.

9.1 General Job Options

Parameter	Notes
/jobid	Use this flag to run jobs that you have created in the job manager.
/source	Source file or folder
/sourcetype	One of: file, folder, tree
/target	The Target folder
/output	The output filename mask e.g %FILENAME.xxxx where xxxx is the file extension.
/logfile	i.e. /logfile="C:\logfiles\log.txt"
/debug	Output additional logging
/onerrorcontinue	When processing folders or folder trees, the process will by default halt when an error is found. If this is set to true, processing will continue.
/errors	Define the folder where documents are placed if their conversion raised an error.
/createfolders	If set to true, destination folders will be created if required.
/overwrite	If set to true, existing destination files will be overwritten.
/advancedflags	This enables advanced PDF Junction and TIFF Junction flags to be used, even if there is no direct equivalent in autobahndx.exe.
/MaxCores	The number of parallel files Autobahn DX will attempt to process at the same time.
/nopassword	Tells autobahn what to do with pdf files that are password protected and no password is provided. <ul style="list-style-type: none">• take no action• move to error folder• copy to error folder

/operation	<p>This represents the operation you want to carry out, below is the list of operation available in autobahndx, they correspond to the step types in section 6.2.5.</p> <ul style="list-style-type: none"> • tifftopdf • ocrimagepdf • topdf • pdftotiff • pdfprops • mergetiff • mergetifftopdf • splittiff • tiffinfo • mergepdf • splitpdf • extracttext • optimizepdf • sharepointdownload • sharepointupload • compresspdf • pdftopdfa • getxmlpdf • ocranyfile • ocranyfileex • barcode • highavailability • kingfisher • polling • azureupload • azuredownload • stamppdf
------------	---

9.2 Split Parameters [used with splittiff and splitpdf operations]

Parameter	Notes				
/splitby	Pages, ranges, repeatingrange, bookmarks (PDF Only)				
/splitrage	<p>Set of page ranges separated by commas that define which pages from the original should be extracted. The following types of page ranges are allowed :</p> <table border="1"> <tr> <td>1</td><td>Specifies a single page</td></tr> <tr> <td>1-3</td><td>Specifies a range of pages</td></tr> </table>	1	Specifies a single page	1-3	Specifies a range of pages
1	Specifies a single page				
1-3	Specifies a range of pages				
/splitrepeatingrange	Apply the page range to each set of <i>Page Ranges</i> pages within the document. For example, if 2-4 is specified for page ranges,				

	<p>and 4 is specified as the repeating range, then the range is re-applied every 4 pages.</p> <p>Hence the file is split such that the first output file contains pages 2-4 from the original document, the second contains pages 6-8 and so on.</p>
/RetainBookmarks	<p>Generated files will include bookmarks from the original file.</p> <p>Note: This works with the spit pdf operation only</p>
/RetainMetadata	<p>Generated files will include metadata (such as Author and Title) from the original file.</p> <p>Note: This works with the spit pdf operation only</p>

9.3 PDF Conversion Options [used with topdf operations]

Parameter	Notes
/fileconversiontimeout	Limits the amount of time in seconds that can be spent on conversion. A value of zero means waits indefinitely.
/Zoom	Sets printing zoom of the worksheet. The allowed value range is from 10 to 400.
/FitToPagesWide	Sets number of pages wide the worksheet will be scaled to. This property is ignored if the Zoom property is set.
/FitToPagesTall	Sets number of pages tall the worksheet will be scaled to. This property is ignored if the Zoom property is set.
/MergeAttachments	Set this flag to true if you want converted pdf attachments to merged to the output pdf file. Otherwise, the converted files will be merged back to the pdf.
/RetainPDFAttachments	Set this flag to true if you want the original attachments to be retained in the pdf file after they are merged. This flag only gets picked up if the MergeAttachments flag above is used.
/IncludeDocProps	Set this flag if you want the MS Office properties to be transferred to the target pdf document.
/HandoutOrder	Sets the handout order, this flag only applies to PowerPoint jobs. The possible values are: PRN_PP_HANDOUT_VERTICAL_FIRST, PRN_PP_HANDOUT_HORIZONTAL_FIRST

/PPOutPutType	<p>Sets the output type, it only works with the PowerPoint files. The possible values are:</p> <p>Slides = PRN_PP_OUTTYPE_SLIDES Build slides = PRN_PP_OUTTYPE_BUILDSLIDES Two slides handouts = PRN_PP_OUTTYPE_TWOSLIDEHANDOUTS Three slides handouts = PRN_PP_OUTTYPE_THREESLIDEHANDOUTS Four slides handouts = PRN_PP_OUTTYPE_FOURSLIDEHANDOUTS Six slides handouts = PRN_PP_OUTTYPE_SIXSLIDEHANDOUTS Nine slides handouts = PRN_PP_OUTTYPE_NINESLIDEHANDOUTS Notes= PRN_PP_OUTTYPE_NOTESPAGES Outline = PRN_PP_OUTTYPE_OUTLINE</p>
/PrintGraphics	<p>Sets the graphics setting for printing.</p> <p>Print Full Resolution = PRN_PB_PRINT_GRAPHICS_HIGHRES Print Low Resolution = PRN_PB_PRINT_GRAPHICS_LOWRES Print Graphics = PRN_PB_PRINT_GRAPHICS_NONE</p>
/convertbookmarks	For MS Word, convert bookmarks
/bookmarkdepth	<p>This property will take effect only when the <u>ConvertBookmarks</u> property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, an empty string is functionally equivalent to.</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use a non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.</p>
/converthyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
/printallsheets	The flag that indicates whether to print all Excel worksheets or not.
/printbgcolor	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
/Printscale	For Visio files, sets the print scale
/ieheader	This property modifies Internet Explorer's header setting.
/iefooter	This property modifies Internet Explorer's footer setting.
/imagecompression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
/imagedownsizing	If this property is set to True, then the resolution of images is reduced to the DPI value specified in the <u>ImageDownsizeResolution</u> property.

/imagejpegquality	The allowed value range is from 5 to 100 with 100 being the highest quality.																		
/imagedownsize resolution	If the <u>ImageDownsizing</u> property is set to True, then the resolution of images is reduced to the DPI value specified in this property.																		
/fontembedding	The option PRN_FONT_EMBED_FULLSET (embedding full set of font) will cause a significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed the font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.																		
/fontsubstitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from the Control Panel).																		
/fontembedastype0	This option is recommended if you have non-standard fonts like barcode font.																		
/margintop	The margin in Inches.																		
/marginbottom	The margin in Inches.																		
/marginleft	The margin in Inches.																		
/marginright	The margin in Inches.																		
/pagewidth	In Inches.																		
/pageheight	In Inches.																		
/paperorientation	PORTRAIT or LANDSCAPE																		
/Papersize	<p>Word does not use the paper size setting from the printer driver. If you need to change the paper size, use this property instead. If you need to use a custom paper size, set this property to PRN_MSO_PAPER_CUSTOM, and set the <u>PageWidth</u> and <u>PageHeight</u> properties.</p> <p>Below is a list of Paper sizes</p> <table border="1"> <tr> <td>PRN_MSO_PAPER_10X14 = 0</td><td>10 x 14 in</td></tr> <tr> <td>PRN_MSO_PAPER_11X17 = 1</td><td>11 x 17 in</td></tr> <tr> <td>PRN_MSO_PAPER_A3 = 6</td><td>A3 297 x 420 mm</td></tr> <tr> <td>PRN_MSO_PAPER_A4 = 7</td><td>A4 210 x 297 mm</td></tr> <tr> <td>PRN_MSO_PAPER_A4SMALL = 8</td><td>A4 Small 210 x 297 mm</td></tr> <tr> <td>PRN_MSO_PAPER_A5 = 9</td><td>A5 148 x 210 mm</td></tr> <tr> <td>PRN_MSO_PAPER_B4 = 10</td><td>B4 (JIS) 250 x 354 mm</td></tr> <tr> <td>PRN_MSO_PAPER_B5 = 11</td><td>B5 (JIS) 182 x 257 mm</td></tr> <tr> <td>PRN_MSO_PAPER_CSHEET = 12</td><td>C size sheet 17 x 22 in</td></tr> </table>	PRN_MSO_PAPER_10X14 = 0	10 x 14 in	PRN_MSO_PAPER_11X17 = 1	11 x 17 in	PRN_MSO_PAPER_A3 = 6	A3 297 x 420 mm	PRN_MSO_PAPER_A4 = 7	A4 210 x 297 mm	PRN_MSO_PAPER_A4SMALL = 8	A4 Small 210 x 297 mm	PRN_MSO_PAPER_A5 = 9	A5 148 x 210 mm	PRN_MSO_PAPER_B4 = 10	B4 (JIS) 250 x 354 mm	PRN_MSO_PAPER_B5 = 11	B5 (JIS) 182 x 257 mm	PRN_MSO_PAPER_CSHEET = 12	C size sheet 17 x 22 in
PRN_MSO_PAPER_10X14 = 0	10 x 14 in																		
PRN_MSO_PAPER_11X17 = 1	11 x 17 in																		
PRN_MSO_PAPER_A3 = 6	A3 297 x 420 mm																		
PRN_MSO_PAPER_A4 = 7	A4 210 x 297 mm																		
PRN_MSO_PAPER_A4SMALL = 8	A4 Small 210 x 297 mm																		
PRN_MSO_PAPER_A5 = 9	A5 148 x 210 mm																		
PRN_MSO_PAPER_B4 = 10	B4 (JIS) 250 x 354 mm																		
PRN_MSO_PAPER_B5 = 11	B5 (JIS) 182 x 257 mm																		
PRN_MSO_PAPER_CSHEET = 12	C size sheet 17 x 22 in																		

PRN_MSO_PAPER_CUSTOM = 41	Custom Paper Size
PRN_MSO_PAPER_DSHEET = 13	D size sheet 22 x 34 in
PRN_MSO_PAPER_ENVELOPE10 = 25	Envelope #10 4 1/8 x 9 1/2 in
PRN_MSO_PAPER_ENVELOPE11 = 26	Envelope #11 4 1/2 x 10 3/8 in
PRN_MSO_PAPER_ENVELOPE12 = 27	Envelope #12 4 3/4 x 11 in
PRN_MSO_PAPER_ENVELOPE14 = 28	Envelope #14 5 x 11 1/2 in
PRN_MSO_PAPER_ENVELOPE9 = 24	Envelope #9 3 7/8 x 8 7/8 in
PRN_MSO_PAPER_ENVELOPEB4 = 29	Envelope B4 250 x 353 mm
PRN_MSO_PAPER_ENVELOPEB5 = 30	Envelope B5 176 x 250 mm
PRN_MSO_PAPER_ENVELOPEB6 = 31	Envelope B6 176 x 125 mm
PRN_MSO_PAPER_ENVELOPEC3 = 32	Envelope C3 324 x 458 mm
PRN_MSO_PAPER_ENVELOPEC4 = 33	Envelope C4 229 x 324 mm
PRN_MSO_PAPER_ENVELOPEC5 = 34	Envelope C5 162 x 229 mm
PRN_MSO_PAPER_ENVELOPEC6 = 35	Envelope C6 114 x 162 mm
PRN_MSO_PAPER_ENVELOPEC65 = 36	Envelope C65 114 x 229 mm
PRN_MSO_PAPER_ENVELOPEDL = 37	Envelope DL 110 x 220 mm
PRN_MSO_PAPER_ENVELOPEITAL Y = 38	Envelope Italy 110 x 230 mm
PRN_MSO_PAPER_ENVELOPEMONARCH = 39	Envelope Monarch 3.875 x 7.5 in
PRN_MSO_PAPER_ENVELOPEPERSONAL = 40	6 3/4 Envelope 3 5/8 x 6 1/2 in
PRN_MSO_PAPER_ESHEET = 14	E size sheet 34 x 44 in
PRN_MSO_PAPER_EXECUTIVE = 5	Executive 7 1/4 x 10 1/2 in
PRN_MSO_PAPER_FANFOLDLEGAL GERMAN = 15	German Legal Fanfold 8 1/2 x 13 in
PRN_MSO_PAPER_FANFOLDSTDGERMAN = 16	German Std Fanfold 8 1/2 x 12 in
PRN_MSO_PAPER_FANFOLDUS = 17	US Std Fanfold 14 7/8 x 11 in
PRN_MSO_PAPER_FOLIO = 18	Folio 8 1/2 x 13 in
PRN_MSO_PAPER_LEDGER = 19	Ledger 17 x 11 in
PRN_MSO_PAPER_LEGAL = 4	Legal 8 1/2 x 14 in
PRN_MSO_PAPER_LETTER = 2	Letter 8 1/2 x 11 in

	PRN_MSO_PAPER_LETTERSMALL = 3	Letter Small 8 1/2 x 11 in		
	PRN_MSO_PAPER_NOTE = 20	Note 8 1/2 x 11 in		
	PRN_MSO_PAPER_QUARTO = 21	Quarto 215 x 275 mm		
	PRN_MSO_PAPER_STATEMENT = 22	Statement 5 1/2 x 8 1/2 in		
	PRN_MSO_PAPER_TABLOID = 23	Tabloid 11 x 17 in		
/PrintColorType	With this property, you can set the PowerPoint to print either with <ul style="list-style-type: none">• Color : PRN_PP_PRINT_COLOR• Greyscale: PRN_PP_PRINT_GRAYSCALE• Black and White: PRN_PP_PRINT_BLACK_AND_WHITE			
/queuewaittimeout	Since PowerPoint cannot perform multiple conversions at the same time, each conversion request will be put into a queue, and they are converted one at a time. Awaiting print job will timeout if it waits for more than the value specified in QueueWaitTimeout property. The timeout value is in milliseconds. Set to zero if you want to wait indefinitely.			
/disablescriptdebugger	Enables/Disables Internet Explorer's script debugger. It is highly recommended that Internet Explorer's script debugger is turned off during printing so that the PDF conversion process is not interrupted.			
/DisplayErrorDialogOnEveryError	Enables/Disables Internet Explorer's error dialog displayed on every error. It is highly recommended that this feature is turned off from Internet Explorer so that the PDF conversion process is not interrupted.			
/pdfa	The output file will be PDF/A-1b compliant.			
/pdfx1a	The output file will be PDF/X-1a compliant.			
/pdfx3	The output file will be PDF/X-3 compliant.			
/ConvertAttachments	For msg files, email attachments are processed so that the resulting PDF is composed of the email plus each page of each attachment.			
/IncludeDocumentMarkups	The “Include Document Markups” option allows Word comments and markup to be included in the generated PDF.			
/RevisionMode	The default selection is PRN_DEFAULT_REVISION, meaning the current revision mode that is set inside Word will be used. Specify PRN_INLINE_REVISION to use inline/strikethrough revision mode and use PRN_BALLOON_REVISION to choose balloon revision mode. Used in conjunction with Include Document Markup Word revision mode constants. <table><tr><td>0</td><td>Use the revision mode currently set inside Word</td></tr></table>		0	Use the revision mode currently set inside Word
0	Use the revision mode currently set inside Word			

	<table> <tr> <td>1</td><td>Inline/strikethrough revision mode (strikethrough used for deletions and underlining used for insertions)</td></tr> <tr> <td>2</td><td>Balloon revision mode (displays revisions in balloons in the margins)</td></tr> </table>	1	Inline/strikethrough revision mode (strikethrough used for deletions and underlining used for insertions)	2	Balloon revision mode (displays revisions in balloons in the margins)
1	Inline/strikethrough revision mode (strikethrough used for deletions and underlining used for insertions)				
2	Balloon revision mode (displays revisions in balloons in the margins)				
/SmartShrinking	Enables/disables intelligent content shrinking. This is only valid when using the HTMLExtension method.				
/AutoAdjustPaperSize	A value of True means that easyPDF ensures that no content gets cropped. If needed, it automatically adjusts the paper width, depending on the layout. The paper size will never get smaller than what you specify. If you enable AutoAdjustPaperSize, SmartShrinking is ignored. You may either shrink the content to fit the page, or extend the page to fit the content, but not both.				
/PrintView	Gets/Sets the flag that indicates whether to render the document the same way as the Print Preview when converting HTML to PDF.				
/WaitTimeAfterPageIsLoaded	When an HTML page contains JavaScript, it is often necessary to wait for the JavaScript execution to complete before the document is converted. The purpose of this property is to introduce such a wait. Default is 200ms				
/HighQualityModeForGraphics	The default value is False, which means Microsoft Office prints certain images dithered, as if the output device was an actual ink and paper printer. Set this property to True to request high quality RGB graphics, with no dithering, which is more appropriate for PDF files intended for computer screens.				

9.4 [PDF to TIFF [used with pdftotiff]

Parameter	Notes
/password	Source PDF password
/TIFFResolution	Sets the DPI of images in the output file.
/TIFFCompression	<ul style="list-style-type: none"> Group 4 (Black and White) LZW (Colored)

9.5 OCR [used with ocrimagepdf or tifftopdf/ mergetifftopdf]

Parameter	Notes
/pdfa	Use this flag to generate a PDF/A compliant file, below are the possible values it can take. <ul style="list-style-type: none">• 1 (Generates a PDF/A-1b file e.g. /pdfa=1)• 2 (Generates a PDF/A-2b file)• 3 (Generates a PDF/A-3b file)
/ocr	Set to true if a searchable PDF is required. Note: This will work with the tifftopdf and mergetifftopdf steps only.
/password	Source PDF password Note: This will work with the ocrimagepdf step only.
/NonImagePDF	This allows control over the treatment of non-image PDFs, i.e. PDFs that have some text in them as well as images. The options are: <ul style="list-style-type: none">• OCR: The document will be OCR'd using the image method defined by "Image Method"• Raise Error: The task will terminate with an error. If "On Error Continue" is set this then behaves as Skip. This is the default.• Skip: The document will not be processed.• Pass Through: The file will not be processed, but a copy of the document will be made and named as if the processing had occurred. Note: This will work with the ocrimagepdf step only.
/RemoveHiddenText	When set to true this will not include any searchable text that already exists text from the source document. Such functionality might be useful if the source document was created by OCR of an image only PDF or other image file and the quality of the text from the previous OCR is poor. NOTE: There is no way to distinguish text added as a result of OCR from text added by other means and as a result, this option should be used with care. Note: This will work with the ocrimagepdf step only.
/deskew	When set to true this will straightens the image

/despeckle	<p>Removes dots below the specified pixel size, the possible input values range from 0-9.</p> <p>Note: To carry over the despeckled image to the output file, set the SavePreDespeckle flag explained below to false.</p>																																																		
/SavePreDespeckle	<p>If set to true and despeckle is also set to true, the pre-despeckled image is added in the output PDF document. Other pre-processing settings such as line removal and morph are not maintained in the output PDF document if enabled.</p>																																																		
/autorotate	<p>Rotates the page image if required</p>																																																		
/language	<p>OCR Language dictionary to be used. The default is English.</p> <table border="1"> <thead> <tr> <th>LANGUAGE</th><th>/language Flag Value</th></tr> </thead> <tbody> <tr><td>English</td><td>0</td></tr> <tr><td>German</td><td>1</td></tr> <tr><td>French</td><td>2</td></tr> <tr><td>Russian</td><td>3</td></tr> <tr><td>Swedish</td><td>4</td></tr> <tr><td>Spanish</td><td>5</td></tr> <tr><td>Italian</td><td>6</td></tr> <tr><td>Russian English</td><td>7</td></tr> <tr><td>Ukrainian</td><td>8</td></tr> <tr><td>Serbian</td><td>9</td></tr> <tr><td>Croatian</td><td>10</td></tr> <tr><td>Polish</td><td>11</td></tr> <tr><td>Danish</td><td>12</td></tr> <tr><td>Portuguese</td><td>13</td></tr> <tr><td>Dutch</td><td>14</td></tr> <tr><td>Czech</td><td>15</td></tr> <tr><td>Roman</td><td>16</td></tr> <tr><td>Hungar</td><td>17</td></tr> <tr><td>Bulgar</td><td>18</td></tr> <tr><td>Slovenian</td><td>19</td></tr> <tr><td>Latvian</td><td>20</td></tr> <tr><td>Lithuanian</td><td>21</td></tr> <tr><td>Estonian</td><td>22</td></tr> <tr><td>Turkish</td><td>23</td></tr> </tbody> </table>	LANGUAGE	/language Flag Value	English	0	German	1	French	2	Russian	3	Swedish	4	Spanish	5	Italian	6	Russian English	7	Ukrainian	8	Serbian	9	Croatian	10	Polish	11	Danish	12	Portuguese	13	Dutch	14	Czech	15	Roman	16	Hungar	17	Bulgar	18	Slovenian	19	Latvian	20	Lithuanian	21	Estonian	22	Turkish	23
LANGUAGE	/language Flag Value																																																		
English	0																																																		
German	1																																																		
French	2																																																		
Russian	3																																																		
Swedish	4																																																		
Spanish	5																																																		
Italian	6																																																		
Russian English	7																																																		
Ukrainian	8																																																		
Serbian	9																																																		
Croatian	10																																																		
Polish	11																																																		
Danish	12																																																		
Portuguese	13																																																		
Dutch	14																																																		
Czech	15																																																		
Roman	16																																																		
Hungar	17																																																		
Bulgar	18																																																		
Slovenian	19																																																		
Latvian	20																																																		
Lithuanian	21																																																		
Estonian	22																																																		
Turkish	23																																																		
/graphicsproc	<p>“all”: Treat all Graphics Areas as Text.</p> <p>“box”: Remove Box Lines in OCR Processing.</p>																																																		
/extractimages	<p>One of the following:</p> <ul style="list-style-type: none"> Auto – TIFF Junction will select the most appropriate method. 																																																		

	<ul style="list-style-type: none"> • Convert to TIFF – The PDF is rasterized using conversion to PDF. • Native 										
/PDFDPI	<p>An integer value representing the resolution of the rasterized Tiff file if “Convert to TIFF” is used above.</p> <p>Note: This setting will only work if extractimages is set to Convert to TIFF</p>										
/TIFFCompression	<p>Sets the Compression for the TIFF file used if the “Convert To TIFF” Option above is used.</p> <ul style="list-style-type: none"> • Auto (Selects Group 4 if the page is Black AND White else it uses LZW Compression) • Group 4 (Black and White) • LZW (Colored) <p>Note: This setting will only work if extractimages is set to Convert to TIFF</p>										
/RetainBookmarks	<p>Set to true to retain bookmarks from the original PDF are copied to the searchable result PDF.</p> <p>Note: This setting will only work if extractimages is set to Convert to TIFF and operation is set to ocrimagepdf</p>										
/RetainMetadata	<p>Set to true to retain metadata from the original PDF are copied to the searchable result PDF.</p> <p>Note: This setting will only work if extractimages is set to Convert to TIFF and operation is set to ocrimagepdf</p>										
/mrc	<p>This enables Mixed Raster Compression which can dramatically reduce the output size of PDFs comprising color scans.</p> <p>Note: This setting will only work if extractimages is set to Convert to TIFF</p>										
/JBIG2	<p>This is a compression used for Bitonal images.</p> <p>Note: This setting will only work if extractimages is set to Convert to TIFF</p>										
/OCRTextFile	<p>This option tells Autobahn DX that text found in the document should also be written to a text output file. The output file can be of 4 different types, use the “/TextFile” flag below to specify.</p> <ul style="list-style-type: none"> • true: produce text output • false: do not produce text output 										
/TextFile	<p>This flag must be used together with the “/OCRTextFile” flag above, it specifies the type of text file to be produced</p> <table border="1"> <thead> <tr> <th>Value</th><th>Output files</th></tr> </thead> <tbody> <tr> <td>“plain text”</td><td>txt, pdf</td></tr> <tr> <td>“plain text (no pdf)”</td><td>txt only</td></tr> <tr> <td>“rtf”</td><td>rtf, pdf</td></tr> <tr> <td>“html”</td><td>html, pdf</td></tr> </tbody> </table>	Value	Output files	“plain text”	txt, pdf	“plain text (no pdf)”	txt only	“rtf”	rtf, pdf	“html”	html, pdf
Value	Output files										
“plain text”	txt, pdf										
“plain text (no pdf)”	txt only										
“rtf”	rtf, pdf										
“html”	html, pdf										

/optimiseocr	See the Advanced Pre-Processing Section for a full explanation of this option.
--------------	--

9.6 PDF Output File Settings – [used with pdfprops]

Parameter	Notes																				
/userpassword	Sets the user password for the output document																				
/ownerpassword	Sets the owner document for the output document.																				
/permissions	<p>A comma separated list of document permissions</p> <table> <tr> <th>Permission</th><th>Description</th></tr> <tr> <td>allowprinting</td><td>Allow high-quality printing</td></tr> <tr> <td>allowmodifycontents</td><td>Allow assembly (see below) and other document modifications</td></tr> <tr> <td>allowcopy</td><td>Allow text and graphics copying and extraction</td></tr> <tr> <td>allowmodifyannotations</td><td>Allow modification of annotations</td></tr> <tr> <td>allowfillin</td><td>Allow filling of form fields</td></tr> <tr> <td>allowscreenreaders</td><td>Allow extraction of text and graphics in support of accessibility.</td></tr> <tr> <td>allowassembly</td><td>Allow rotation, insertion or deletion of pages.</td></tr> <tr> <td>allowdegradedprinting</td><td>Allow low-quality printing</td></tr> <tr> <td>allowall</td><td>Allow all the permissions</td></tr> </table>	Permission	Description	allowprinting	Allow high-quality printing	allowmodifycontents	Allow assembly (see below) and other document modifications	allowcopy	Allow text and graphics copying and extraction	allowmodifyannotations	Allow modification of annotations	allowfillin	Allow filling of form fields	allowscreenreaders	Allow extraction of text and graphics in support of accessibility.	allowassembly	Allow rotation, insertion or deletion of pages.	allowdegradedprinting	Allow low-quality printing	allowall	Allow all the permissions
Permission	Description																				
allowprinting	Allow high-quality printing																				
allowmodifycontents	Allow assembly (see below) and other document modifications																				
allowcopy	Allow text and graphics copying and extraction																				
allowmodifyannotations	Allow modification of annotations																				
allowfillin	Allow filling of form fields																				
allowscreenreaders	Allow extraction of text and graphics in support of accessibility.																				
allowassembly	Allow rotation, insertion or deletion of pages.																				
allowdegradedprinting	Allow low-quality printing																				
allowall	Allow all the permissions																				
/strength	Set to 128 to make use of security settings																				
/author	Sets the PDF author																				
/title	Sets the PDF title																				
/subject	Sets the PDF subject																				
/keywords	Sets the PDF keywords																				

/creator	Sets the PDF creator
/pagelayoutSinglepage	Single Page (SinglePage)
/pagelayoutContinuous	Continuous(OneColumn)
/pagelayoutContinuousfacingleft	ContinuousFacing [odd pages left] (TwoColumnLeft)
/pagelayoutContinuousfacingright	ContinuousFacing [odd pages right] (TwoColumnRight)
/pagemodeusenone	No Thumbnails or Bookmarks Visible (UseNone)
/pagemodeuseoutlines	Bookmarks Visible (Use Outlines)
/pagemodeusethumbs	Thumbnails Visible (UseThumbs)
/pagemodefultscreen	Full Screen Mode (FullScreen)
/hidetoolbar	The viewer's toolbar will be hidden
/hidemenubar	The viewer's menu bar will be hidden
/hidewindowui	The viewer's UI elements (scrollbars etc.) will be hidden
/fitwindow	The viewer will resize the document's window to fit the size of the first displayed page.
/centerwindow	The document window will be positioned in the center of the screen.
/nonfullscreenpagemodeusenone	On exiting full-screen mode, neither outline nor thumbnail images are shown.
/nonfullscreenpagemodeuseoutline	On exiting full-screen mode outlines are visible.
/nonfullscreenpagemodeusethumbs	On exiting full-screen mode thumbnails are visible.

9.7 Sharepoint Settings – [used with sharepointdownload and sharepointupload]

Parameter	Notes
/s	Site URL, the URL of the SharePoint site that you want to access, e.g. http://localhost/testsite
/l	Library, the name of the library that you want to access, e.g. "Test Library"
/k	The subfolder in the library to download from or upload to.
/f	The local folder to download documents to or upload a document from.
/e	An optional extension mask that limits those files to manipulate, e.g. "pdf,tiff"
/r	If set to "True" sub-folders of the Sharepoint Library are handled.
/a	The action: <ul style="list-style-type: none">• Download: /a=0• Upload: /a=1
/u	The username used to connect to the SharePoint site.
/p	The password (encrypted) used to connect to the SharePoint site.
/o	Set this if the site is in office 365
/c	Continue processing if an error occurs.
/UseADFS	Set this to true if you use Active Directory for your SharePoint User Management.
/ADFSHost	Provide the name of the Active Directory server
/ADFSRelyingPartyIdentifier	Provide the Relying Party Trust identifier for your SharePoint.
/Createfolders	When uploading and you set recursive to true, switch this on to recreate folder that do not exist in SharePoint
/Maxcores	Number of Parallel processors used to upload or download files to/from SharePoint

/IncludePattern	Only file that matches this pattern will be included.
/ExcludePattern	Any file that matches this pattern will be excluded.

9.8 Compress PDF Settings – [used with compresspdf]

Parameter	Notes
/mrcq	JPEG quality setting (percentage value 1 - 100) for use in saving the background and foreground images. The default value is 75
/mrcb	Sampling size for the background portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. The default value is 3
/mrcf	Sampling size for the foreground portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. The default value is 3
/advancedflags	<p>A number of advanced flags are available for use under guidance from Aquaforest technical support. Contact support@aquaforest.com if needed.</p> <p> /pdfa or /pdfa=1 – convert the compressed document to PDF/A1b format /pdfa=2 – convert the compressed document to PDF/A2b format /pdfa=3 – convert the compressed document to PDF/A3b format /nojb2 – use G4 fax compression rather than JBIG2 for bitonal images /debug – generate additional informational messages /allowtext – allow PDFs containing text to be processed </p>

9.9 Extended OCR Steps

The Autobahn DX command line interface has been enhanced to support the new Extended OCR module. **Using /ocrengine=1 as a parameter is a requirement.**

autobahndx.exe /operation=[operation name] /source=[tiff file or folder] /output=[output file] /target=[target folder] [/option=value]...

Examples

1. Generate a searchable PDF c:\out\outfile.pdf and a Word file c:\out\outfile.docx from a multi-page TIFF file

autobahndx.exe /source=c:\in\in.tif /sourcetype=file /target=c:\out /output=outfile /outputtype=pdf,docx /operation=tifftopdf /ocrengine=1

2. Generate a searchable PDF file from a folder of TIFF and JPEG files, with Deskew and page orientation detection and correction.

autobahndx.exe /source=c:\in\folder /sourcetype=folder /target=c:\out /output=outfile /outputtype=pdf /autorotate /deskew /operation=mergetifftopdf /ocrengine=1

3. Generate searchable PDF files from image PDF files found in a folder and subfolders, while keeping the original file names.

autobahndx.exe /source=c:\in\folder /sourcetype=tree /target=c:\out /output=%FILENAME /outputtype=pdf operation=ocrimagepdf /ocrengine=1

The Extended OCR steps use the parameters listed in the table below.

Parameter	Notes
/operation	The operation that needs to be carried out: <ul style="list-style-type: none">• tifftopdf• mergetifftopdf• ocrimagepdf• ocranyfileex
/ocrengine	The OCR engine to use. This <u>must</u> be set to 1 to use the IRIS engine. /ocrengine=1
/source	Source file or folder

/sourcetype	File (default) or Folder
/target	The Target folder
/output	The output filename excluding the extension (which will be added according to the output file type).
/outputtype	<p>One or more of the following, separated by commas if more than one is required.</p> <p>RTF PDF DOCX CSV* SML (SpreadsheetML XML file)* HTM TXT</p> <p>*These output formats are suitable for table-oriented pages that can be mapped onto a spreadsheet format.</p>
/ExtractImages	<p>Whether to convert the images in a PDF document to TIFF or not.</p> <ul style="list-style-type: none"> • Convert to TIFF – The pages in the PDF document are rasterized and saved as TIFF images • Native - This method places the OCR'ed text directly into a copy of the original PDF rather than creating an entirely new PDF.
/Autorotate	Detect page orientation and correct if required
/RemoveBlankPage	Set this to true to remove blank pages from Tiff or PDF documents. Value needs to be set for sensitivity (see below).
/Sensitivity	The sensitivity, from 1 to 100. With high sensitivity, fewer blank pages are detected.
/Deskew	Rotates the image to correct its skew angle.
/AdvancedDeskew	Set this to true if you want to set the advanced deskew properties below.
/AdjustmentMode	Set the behavior regarding dimension adjustment for deskew operation.
/ForceDeskew	If turned off, the image is analyzed before rotation and the engine may choose not to rotate the image depending on the analysis result. If turned on, the image is rotated to correct skew angle.

/Despeckle	Removes all the groups of connected pixels with a number of pixels below the parameter. Suggested range: 1-20.
/Workdepth	This parameter (0 – 255) defines how deeply the OCR engine will analyze a page with 255 being the deepest. For poorer quality documents, higher values can give better recognition results.
/JPEGQuality	This parameter (0 – 255) determines the compression/quality of color JPEG images in generated PDFs. 0 gives the smallest file size whilst 255 gives the best quality. The default value is 128.
/PDFVersion	This determines the PDF version of the generated PDF : 1.4 1.5 1.6 1.7 PDF/A1B (<i>PDF/A-1b</i>)
/ValidatePDF/A	Set this flag to true if you want to validate the output PDF/A files.
/LanguageDetection	Set this flag to true to enable Auto Language Detection feature. The aim of this feature is to detect the most probable language of a single-language page. If at least one language has been detected, recognition will be performed in the first language candidate that has been detected, and not in the language(s) set through Language or Languages. If it fails to detect a language, recognition will be performed using the language(s) set through Language or Languages. Note: This is set to true by default.
/language	Determines the language to be used for OCR. This may be a comma separated list for multiple languages e.g. /language=1, 2 for German and French. Note that these codes are not the same as those used by the default Aquaforest engine. See the full table of languages and their codes in the next section.
/createfolders	Create an output folder if it does not exists. Default true.
/dpi	Sets the DPI of images in the output file. Set to Auto by default, alternatively can be set to 300, 200 or 150 to force a specific resolution.
/nonimagepdf	This allows control over the treatment of non-image-only PDFs, i.e. PDFs that have some text in them as well as images. The options are :

	<ul style="list-style-type: none"> • OCR. The document will OCR'd using the image method defined by "Image Method" • Raise Error. The task will terminate with an error. If "On Error Continue" is set this then behaves as Skip. This is the default. • Skip. The document will not be processed. • Pass Through. The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.
/noocr	Whether are not to perform OCR on the document (Yes to not perform OCR, No to perform OCR).
/AdvancedDespeckle	Set the advanced despeckle settings, advanced despeckle provides advanced image noise reduction features by the image despeckle filter.
/RemoveWhitePixels	By default, despeckle removes black pixels. If set to true, the despeckle will remove white pixels rather than black pixels.
/Dilate	Despeckle removes all the groups of connected pixels with a few pixels below the SpeckleSize parameter. Those connected pixels are not removed if the distance to a larger connected component is below this parameter. As a result, only the isolated pixels get deleted. The maximum value for this property is 20 pixels. The default value is '0'.
/Binarization	Whether or not to perform binarization on the document.
/Brightness	The brightness (higher values will darker the result).
/Contrast	The contrast (lower values will darker the result).
/SmoothingLevel	Smoothing may be useful to binarize text with a colored background in order to avoid noisy pixels (0 disables smoothing, higher values smooth more).
/Undithering	Whether or not to use automatic undithering while processing a page. NOTE: Automatic undithering will be applied only if smoothing is also activated (SmoothingLevel).
/Threshold	Sets the threshold for fixed threshold binarization (0 for automatic threshold computation).
/RemoveLines	Whether or not to remove lines from an image (The image must be black and white).

/HorizontalCleanX	The parameter for cleaning noisy pixels attached to the horizontal lines.
/HorizontalCleanY	The parameter for cleaning noisy pixels attached to the horizontal lines.
/VerticalCleanX	The parameter for cleaning noisy pixels attached to the vertical lines.
/VerticalCleanY	The parameter for cleaning noisy pixels attached to the vertical lines.
/HorizontalDilate	The dilate parameter that helps the detection of horizontal lines.
/VerticalDilate	The dilate parameter that helps the detection of vertical lines.
/HorizontalMaxGap	The maximum horizontal line gap to close. It is useful to remove broken lines.
/VerticalMaxGap	The maximum vertical line gap to close. It is useful to remove broken lines.
/HorizontalMaxThickness	The maximum thickness of the horizontal lines to remove. It is useful to keep vertical lines larger than this parameter. Can be also useful to keep vertical letter strokes.
/VerticalMaxThickness	The maximum thickness of the vertical lines to remove. It is useful to keep horizontal lines larger than this parameter. Can be also useful to keep horizontal letter strokes.
/HorizontalMinLength	The minimum length of the horizontal lines to remove.
/VerticalMinLength	The minimum length of the vertical lines to remove.
/RemoveDarkBorders	Removes the dark surrounding from bitonal, grayscale or color images. The dark surrounding of the image is whitened (Note: The dark border should be touching the edge of the page for this to work).
/RemovePunchHoles	Attempts to remove punch holes from pages. Note: The punch hole algorithm can be used on images with the following minimum dimensions width: 300px, height: 100px (computed for 300 DPI). The minimum height and width can vary with the image resolution.
/Interpolation	Interpolates the source image to the given resolution. This value (the target resolution) must be greater than the source image's resolution.

/InterpolationMode	Sets the interpolation mode.
/KeepOriginalImage	Set this to true if you want to use the pre-processed image for OCR but keep the original image in the output document. The default value is 'true'. Note: This setting will only work if ExtractImages is set to Convert to TIFF
/KeepDeskew	Set this to true if you want to use the deskewed image in the output document. Note: This property only applies when Keep Original Image is set to No
/KeepDespeckle	Set this to true if you want to use the despeckled image in the output document. This requires the source image to be black and white. Note: This property only applies when KeepOriginalImage is set to No
/KeepDarkBorderRemoval	Set this to true if you want to use the image after dark borders have been removed, in the output document. Note: This property only applies when KeepOriginalImage is set to No
/KeepPunchHoleRemoval	Set this to true if you want to use the image after punch holes have been removed, in the output document. Note: This property only applies when KeepOriginalImage is set to No
/resourcesfolder	By default, the OCR resources folder is a subfolder of the distribution/extendedocr folder. This option allows the resources to be located elsewhere if required.

9.9.1 Extended OCR Languages

Member name	Value	Description
English	0	English (American)
German	1	
French	2	
Spanish	3	
Italian	4	
British	5	
Swedish	6	

Danish	7	
Norwegian	8	
Dutch	9	
Portuguese	10	
Brazilian	11	
Galician	12	
Icelandic	13	
Greek	14	
Czech	15	
Hungarian	16	
Polish	17	
Romanian	18	
Slovak	19	
Croatian	20	
Serbian	21	
Slovenian	22	
Luxemb	23	
Finnish	24	
Turkish	25	
Russian	26	
Byelorussian	27	
Ukrainian	28	
Macedonian	29	
Bulgarian	30	
Estonian	31	
Lithuanian	32	
Afrikaans	33	

Albanian	34	
Catalan	35	
Irish_Gaelic	36	
Scottish_Gaelic	37	
Basque	38	
Breton	39	
Corsican	40	
Frisian	41	
Nynorsk	42	
Indonesian	43	
Malay	44	
Swahili	45	
Tagalog	46	
Japanese	47	You will need to have the Asian OCR license to use this language.
Korean	48	You will need to have the Asian OCR license to use this language.
Schinese	49	You will need to have the Asian OCR license to use this language.
Tchinese	50	You will need to have the Asian OCR license to use this language.
Quecha	51	
Aymara	52	
Faroese	53	
Friulian	54	
Greenlandic	55	
Haitian_Creole	56	
Rhaeto_Roman	57	

Sardinian	58	
Kurdish	59	
Cebuano	60	
Bemba	61	
Chamorro	62	
Fijan	63	
Ganda	64	
Hani	65	
Ido	66	
Interlingua	67	
Kicongo	68	
Kinyarwanda	69	
Malagasy	70	
Maori	71	
Mayan	72	
Minangkabau	73	
Nahuatl	74	
Nyanja	75	
Rundi	76	
Samoan	77	
Shona	78	
Somali	79	
Sotho	80	
Sundanese	81	
Tahitian	82	
Tonga	83	
Tswana	84	

Wolof	85	
Xhosa	86	
Zapotec	87	
Javanese	88	
Pidgin_Nigeria	89	
Occitan	90	
Manx	91	
Tok_Pisin	92	
Bislama	93	
Hiligaynon	94	
Kapampangan	95	
Balinese	96	
Bikol	97	
Ilocano	98	
Madurese	99	
Waray	100	
None	101	No language, Latin alphabet
Serbian_Latin	102	
Latin	103	
Latvian	104	
Hebrew	105	You will need to have the Hebrew OCR license to use this language.
Numeric	114	This language limits recognition to numeric characters.
Esperanto	115	
Maltese	116	
Zulu	117	
Afaan	118	

Asturian	119	
AzeriLatin	120	
Luba	121	
Papamiento	122	
Tatar	123	
Turkmen	124	
Welsh	125	
Arabic	126	You will need to have the Arabic OCR license to use this language.
Farsi	127	You will need to have the Arabic OCR license to use this language.
Mexican	128	
BosnianLatin	129	Bosnian (Latin). CharsetCategory.E
BosnianCyrillic	130	Bosnian (Cyrillic). CharsetCategory.D
Moldovan	131	Moldovan. CharsetCategory.E
SwissGerman	132	German (Switzerland). CharsetCategory.C
Tetum	133	Tetum. CharsetCategory.C
Kazakh	134	Kazakh (Cyrillic). CharsetCategory.D
MongolianCyrillic	135	Mongolian (Cyrillic). CharsetCategory.D
UzbekLatin	136	Uzbek (Latin). CharsetCategory.C

9.10 Create XML Property File – [used with getxmlpdf]

Parameter	Notes
/CopyPDF	Value is true or false, setting this to true will copy the pdf source to the target folder.

9.11 OCR Any File to PDF– [used with ocranyfile or ocranyfileex]

9.11.1 Standard

Parameter	Value
/operation	ocranyfile

This uses the flags in section [Standard OCR](#) and [Any File to PDF](#)

9.11.2 Extended

Parameter	Value
/operation	ocranyfileex

This uses the flags in section [Extended OCR](#) and [Any File to PDF](#)

9.12 TIFF/PDF Barcode – [used with barcode]

Command Line	Description
/SplitByBarcode (bool)	Choose this option to split the PDF/TIFF file by Barcode
/RenameByBarcode (bool)	Choose this option to rename the PDF file based on Barcode.
/BarcodeFormats	Barcode formats supported. See table below for more explanation
/ExcludedBarcodeFormats	Barcode formats not supported.
/SplitMode (bool)	Various Options for splitting Files by Barcode
/TryHarder (bool)	Spend more time to try to find a barcode; optimize for accuracy, not speed. The default is true.
/SplitOutputPath (string)	The output file path template where the split files will be saved. - %VALUE%: Replaced by the barcode value found. - %INDEX%: Replaced by the current split index. - %FILENAME: Replaced by the file name

/SplitNoBarcodeTemplate (string)	<p>The renaming template to use for page ranges where no barcodes were identified.</p> <p>Allowed templates:</p> <ul style="list-style-type: none"> - %INDEX%; Replaced by the current split index. - %FILENAME: Replaced by the filename of the source file.
/Binarize	Switch this on if your input files are colored, it first converts it to black and white before conversion.
/RemoveLines (float)	<p>Remove lines from images.</p> <p>Value = 100.5</p>
/Zones (string)	<p>Only examine the region specified for barcode(s).</p> <p>Values =left(int), top(int), width(int), height(int)</p> <p>e.g. /Zones= 5,10,200,200</p>
/BoxSize	This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the barcode reader. Technically, this option removes connected elements with a minimum area (in pixels and defined by this property). This option is currently only applied for bitonal images.
/PdfToImageCompression	<p>PDF files get converted to TIFF files before the barcode is recognized, this flag represents the compression of the TIFF file.</p> <p>"CCITT4" or "LZW"</p>
/PdfToImageDpi	<p>PDF files get converted to TIFF files before the barcode is recognized, this flag represents the DPI of the TIFF file.</p> <p>150,200,300</p>
/Deskew (bool)	Deskew (straighten) the image. The default value is false
/ Despeckle (int)	Removes specks below the specified pixel size from the image. The default value is 0 and the maximum value is 9.

/TempFolder (string)	This flag can be used to specify a temp location where the processing will be carried out.
/EndPage (int)	The page you want to start recognizing the barcode.
/StartPage (int)	The page you want to start recognizing the barcode.
/BlankPageThreshold (int)	Any page with black pixels lower than the threshold set will be ruled as a blank page.
/ReadMultipleBarcodes (bool)	This is used to read multiple barcodes in a single page
/Morph	<p>Morphological options that will be applied to the binarized image before OCR. If set to empty none is applied. Common options include those listed below but for more options please contact support@aquaforest.com:</p> <ul style="list-style-type: none"> - d2.2 – 2x2 dilation applied to all black pixel areas, useful for faint prints. - e2.2 – 2x2 erosion applied to all black pixel areas, useful for heavy prints. - c2.2 – a closing process that performs a 2x2 dilation followed by a 2x2 erosion with the result that holes and gaps in the characters are filled. <p>This option should generally only be used under guidance from technical support.</p>
/ProductsOnly	Only decode the UPC and EAN families of barcodes.

9.13 Stamp PDF – [used with stamppdf]

Parameter	Notes
/stampoperation	<p>Select the way you want to stamp the PDF pages. Below are the different options.</p> <ul style="list-style-type: none"> • text (Stamp PDF Text) • image (Stamp PDF Text) • pagenumber (Stamp PDF Text) • bates (Stamp PDF Text) • variables (Stamp PDF Text)

/stampposition	<p>Select where you want the stamp to appear, below are the options.</p> <ul style="list-style-type: none"> • bottomcenter • bottomleft • bottomright • center • centerleft • centerright • topcenter • topleft • topright 								
/stampdirection	<p>Select the direction you want the stamp to face</p> <ul style="list-style-type: none"> • none • diagonalup • diagonaldown 								
/stampobject	<p>This represent the text, image path or variable you want to stamp on the pdf. The table below will show you what stampobject stands for in relation to the selected stampoperation.</p> <table> <tr> <th>stampoperation</th><th>Stampobject meaning</th></tr> <tr> <td>text</td><td>The string you want to stamp on the PDF</td></tr> <tr> <td>image</td><td>The path to the image to be used as a stamp on the PDF</td></tr> <tr> <td>variables</td><td>The stamp variable you want to use to stamp the PDF file</td></tr> </table>	stampoperation	Stampobject meaning	text	The string you want to stamp on the PDF	image	The path to the image to be used as a stamp on the PDF	variables	The stamp variable you want to use to stamp the PDF file
stampoperation	Stampobject meaning								
text	The string you want to stamp on the PDF								
image	The path to the image to be used as a stamp on the PDF								
variables	The stamp variable you want to use to stamp the PDF file								
/stamppagerange	Set of page ranges separated by commas that define which pages from the original should be stamped.								
/stampstartnumber	The number that the page numbering will start with.								
/StartPage	The page where the stamping should start from.								
/EndPage	The page where the stamping should end.								
/batesprefix	Prefix for Bates Numbering.								
/batessuffix	Suffix for Bates Numbering.								
/bateslength	Length of Bates Numbering								
/stampcolor	Enter a valid color name. If the color name is not valid, we will use black								

/stampopacity	Sets the transparency of image stamps, the value is from 0-100. 0 stands for transparent, while 100 stands for solid.
/fontname	Font Name for all text-based stamps, use a type1 font name or provide a path to a true type font(.ttf) file.
/fontsize	Font Size for all text-based stamps
/stampasimage	If set to true, autobahn will convert text values to an image before stamping.
/stampbackgroundcolor	Enter a valid color name. If the color name is not valid, we will use Transparent

9.14 Azure Upload/Download – [used with azuredownload and azureupload]

Parameter	Notes
/storagename	The name of the Azure blob container you want to download files from.
/azureaccesskey	Connection string of key 1 under the accesskeys section of the storage account in the portal.
/ContainerName	The name of the Azure blob container you want to download files from.
/AzureExt	File extension filters separated by commas (e.g. .tif,.pdf)
/AzureRecurse	Download documents from folders and subfolders in the SharePoint Library

9.15 Extract Text Download – [used with extracttext]

Parameter	Notes
/PageSep	This allows the definition of an optional page separator string in the output text file.
/PlaceTop	Specifies whether the PageSep will appear at the beginning or the end of the page.

/ExtractTextEngine	<p>The Extract Text Engine to use, below is a list of possible values.</p> <ul style="list-style-type: none"> • 0 (PDFBox with Formatting) • 1 (BCL) • 2 (PDFBox)
/FromPage	The start of the range of pages from which to extract text. If not specified, a start page of 1 is assumed.
/ToPage	The end of the range of pages from which to extract text. If not specified, the last page is assumed.
/copypdf	Set to true if you want Autobahn DX to copy the input PDF file to the target folder.

9.16 Merge PDF – [used with mergepdf]

Parameter	Notes
/filenamebookmark	It will add a bookmark for each file, the bookmark title will be same as the file name.
/RetainBookmarks	Generated files will include bookmarks from the original file.
/RetainMetadata	Generated files will include metadata (such as Author and Title) from the original file.

10 AUTOBAHN DX .NET API

An API is provided which allows a .Net application to allow execution of existing job definitions and creation “on the fly” of new jobs from XML Job Ticket files. The API DLL is located in the AUTOBAHN\bin folder and a sample Visual Studio 2008 project is provided in the AUTOBAHN \code examples\AutobahnAPIExample folder where AUTOBAHN is the folder where Autobahn DX has been installed – by default this is C:\Aquaforest\Autobahn DX.

10.1 API Example (Local Job)

```
using System;
using System.Threading;
using Aquaforest.Autobahn.API;

namespace ConsoleApplication2
{
    class Class1
    {
        [STAThread]
        static void Main(string[] args)
        {
            IJob job = null;
            string JobFinished;
            string Jobstatus;
            try
            {
                /* Run Job and Wait to Complete */

                job = new Job(1001);
                job.Start();
                job.WaitForExit();
                JobFinished=job.JobFinished();
                Jobstatus=job.Jobstatus();

                if(Jobstatus=="Error")
                {
                    Console.WriteLine("Job Failed");
                }
                else if(Jobstatus=="Stopped")
                {
                    if(JobFinished=="True")
                        Console.WriteLine("Job Completed");
                    else
                        Console.WriteLine("Job Was Terminated");
                }
                job.Dispose();

                /* Run Job and allow 10 seconds to Complete */

                job = new Job(1001);
                job.Start();
                job.WaitForExit(10000);
                JobFinished=job.JobFinished();
                Jobstatus=job.Jobstatus();

                if(Jobstatus=="Error")
                {
                    Console.WriteLine("Job Failed - Check Job Log for Details");
                }
                else if(Jobstatus=="Stopped")
```

```

        {
            if(JobFinished=="True")
                Console.WriteLine("Job Completed");
            else
                Console.WriteLine("Job Was Terminated due to Timeout");
        }
        job.Dispose();
    }
    catch (Exception oExp)
    {
        Console.WriteLine(oExp.Message);
        Console.ReadLine();
    }
}
}

```

10.2 API Example (Remote Job)

If you have remote access to the machine where autobahn is installed, you can use the API to invoke jobs remotely.

```

using System;
using System.Threading;
using Aquaforest.Autobahn.API;
namespace AutobahnRemoteAPIExample
{
    internal class Program
    {
        private static int Main(string[] args)
        {
            IJob loJob = null;
            Console.WriteLine("Enter the Job ID Below.");
            int liJobID = ParseInt(Console.ReadLine()); // Read Job ID To Run

            Console.WriteLine("Enter the Job Timeout in Ms (The job will exit if this value elapses in)");
            int liTimeOut = ParseInt(Console.ReadLine()); // MS to wait (0 to wait until job completion)
            /*
             * Parse the location of the config file.
             * If you want to use the "Autobahn.config" in this project, follow the steps below.
             * 1.Replace all the "remotehost" values with the host name of the remote machine. including file paths C: to //remotehost
             * 2.Change the "remote.port" value to the port number of the ADX instance running on the remote machine.
             * you can find this in the config file of ADX on the remote machine in "ADXInstallation\config\autobahn.config"
             */
            Console.WriteLine("The location of the config file.");
            string configFile = Console.ReadLine();
            int liRetVal = 0;

            try
            {
                loJob = new Job(liJobID, configFile);
            }
        }
    }
}

```

```

        loJob.Start();

        if (liTimeOut > 0)
        {
            loJob.WaitForExit(liTimeOut);
        }
        else
        {
            loJob.WaitForExit();
            Console.WriteLine("Job execution finished:" + loJob.JobFinished());
            Console.WriteLine("Exit code is:" + loJob.Jobstatus());
        }

        if (loJob.Jobstatus().ToUpper().Equals("ERROR"))
        {
            liRetVal = -1;
        }
        else if (loJob.Jobstatus().ToUpper().Equals("STOPPED"))
        {
            if (loJob.JobFinished().ToUpper().Equals("TRUE"))
            {
                liRetVal = 0;
            }
            else
            {
                liRetVal = -1;
            }
        }
    }
    catch (Exception oExp)
    {
        liRetVal = -1;
        Console.WriteLine(oExp.Message);
    }
    finally
    {
        if (loJob != null)
        {
            loJob.Dispose();
        }
    }
    Console.WriteLine(liRetVal.ToString());
    Console.ReadLine();
    return liRetVal;
}

private static int ParseInt(string psValue)
{
    try
    {
        return int.Parse(psValue);
    }
    catch
    {
        //---[ Ignore ]---
    }
    return 0;
}
}
}

```

Constructors	<p>public Job(int jobid) <i>Create a Job object using an existing jobid.</i></p>
	<p>public Job(string jobdeffile) <i>Create a Job object using a temporary job definition file.</i></p>
	<p>public Job(int jobid, string config) <i>Create a Job object using an existing jobid and config file of with the configuration of the remote machine</i></p>
	<p>public Job(string jobdeffile, string config) <i>Create a Job object using a temporary job definition file and config file of with the configuration of the remote machine</i></p>
Methods	<p>public Start() <i>Starts the Job.</i></p>
	<p>public void WaitForExit(int ms) <i>Waits for the job to exit, for up to ms milliseconds. The job is stopped if it has not completed.</i></p>
	<p>public void WaitForExit() <i>Waits indefinitely for the job to compete.</i></p>
	<p>public virtual void Dispose() <i>Disposes of the resources associated with the Job.</i></p>
	<p>Public string Jobstatus() <i>Returns the status of the job which may be either Stopped or Error. This should be checked in conjunction with the value of JobFinished() to determine whether the job completed before being stopped.</i></p>
	<p>Public string JobFinished() <i>Indicates whether the job has completed (return value True) or has been stopped (return value False) as a result of timeout being exceeded.</i></p>
	<p>Public void ClearError(string JobStatusPath) <i>When a job is error, ClearError(string jobstatusPath) will clear the error job and set the job status to stopped.</i></p>
	<p>Public string GetLastRunDate() <i>Returns the last Date and Time the job executed.</i></p>
	<p>Public List<DocumentResults> GetDocumentsStatus() <i>Returns a list of files processed and their status.</i></p>

11 OCR PROPERTIES FILE AND THE ADVANCED PRE PROCESSING OPTION

11.1 Properties File

The following are descriptions of those properties in the file Properties.xml that are most likely to be changed to improve engine performance. If you require further information regarding any properties in the file then please contact Aquaforest via support@aquaforest.com for assistance.

Binarize – This setting determines how the image will be converted into a bitonal one for OCR. The following are valid options:

- 1 – This utilizes a technique whereby those parts of the image that have certain characteristics indicative of characters are extracted from the underlying image. This approach can give the best results on pages such as magazine images, newsprint, etc and will handle light text on darker backgrounds. This approach can cause an increase in processing time with certain images.
- 0 – This utilizes the binarization capabilities built into the OCR engine and whilst it can give good results in limited situations it is not generally recommended.
- >0 – A value greater than 0 (the recommended default is 200) will use a simple threshold technique comparing the intensity of the pixel to the threshold value to determine whether it should be set to black or white. This simple approach is the fastest option.

BoxSize – Setting a value above 0 will cause the removal of enclosing boxes from the image used for the OCR processing. The default recommended is 100, i.e. where the box edges are 100 pixels or greater.

BackgroundFactor - Sampling size for the background portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. The default value is 3

DotMatrix - Set this to True to improve recognition of dot-matrix fonts. The default value is False. If set to true for non-dot-matrix fonts then the recognition can be poor

ForegroundFactor - Sampling size for the foreground portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. The default value is 3

Jbig2EncFlags – These are the flags that will be passed to the application used to generate JBIG2 versions of images used in PDF generation (assuming this compression is enabled). Options are as follows:

- b <basename>: output file root name when using symbol coding
- d --duplicate-line-removal: use TPGD in generic region coder
- p --pdf: produce PDF ready data
- s --symbol-mode: use text region, not generic coder
- t <threshold>: set classification threshold for symbol coder (def: 0.85)
- T <bw threshold>: set 1 bpp threshold (def: 188)
- r --refine: use refinement (requires -s: lossless)
- O <outfile>: dump thresholded image as PNG
- 2: upsample 2x before thresholding
- 4: upsample 4x before thresholding
- S: remove images from mixed input and save separately

-j --jpeg-output: write images from mixed input as JPEG

-v: be verbose

Language – The acceptable values are as follows:

- 0 - English
- 1 - German
- 2 - French
- 3 - Russian
- 4 - Swedish
- 5 - Spanish
- 6 - Italian
- 7 - Russian English
- 8 - Ukrainian
- 9 - Serbian
- 10 - Croatian
- 11 - Polish
- 12 - Danish
- 13 - Portuguese
- 14 - Dutch
- 19 - Czech
- 20 - Roman
- 21 - Hungar
- 22 - Bulgar
- 23 - Slovenian
- 24 - Latvian
- 25 - Lithuanian
- 26 - Estonian
- 27 - Turkish

MaxDeskew – Maximum angle by which a page will be deskewed

Morph – Morphological options that will be applied to the binarized image before OCR. If left blank none is applied. Common options include those listed below but for more options please contact support@aquaforest.com:

d2.2 – 2x2 dilation applied to all black pixel areas, useful for faint prints.

e2.2 – 2x2 erosion applied to all black pixel areas, useful for heavy prints.

c2.2 – a closing process that performs a 2x2 dilation followed by a 2x2 erosion with the result that holes and gaps in the characters are filled.

NoPictures - By default, if an area of the document is identified as a graphic area then no OCR processing is run on that area. However, certain documents may include areas or boxes that are identified as “graphics” or “picture” areas but that actually do contain useful text. Setting NoPictures to True will cause it to ignore areas identified as pictures whilst setting it to False will force OCR of areas identified as pictures.

OneColumn - The default value for this is true which improves the handling of single column text. Better handling of multi-column text such as a magazine or newsprint can be achieved.

PdfToImageIncludeText – When set to False this will prevent the conversion of real text (i.e. electronically generated as opposed to text that is part of a scanned image) from being rendered in the page images extracted from the PDF. This is because the text is already searchable and so generally does not require OCR. The value can be set to True however if the OCR is required on this real text.

Quality - JPEG quality setting (percentage value 1 - 100) for use in saving the background and foreground images. The default value is 75

RemoveLines – The value used in Line removal. If blank no line removal will occur. The normal value to use to enable line removal is 100.5 but if you are experiencing difficulties with this value or have any questions then please contact support@aquaforest.com.

11.2 Advanced Pre Processing

This option can be used to run each page OCR with 2 or more different settings and then chose the best set of results.

When the /optimiseocr is specified in the advanced flag field. The OCR and image processing engines will use the settings in the ImagePreProcessingDefaults section of the file Properties.xml modified by any properties set on the OCR and PreProcessing objects.

This will enable the use of these default settings first (without modification by the properties set on the OCR and PreProcessing objects) followed by the same defaults modified by the values in the ImagePreProcessing sections from ID="1" to ID="n" where n is the last consecutive set defined in Properties.xml.

Using heuristics and dictionary lookup the quality of the OCR output is then compared in order to determine the optimum set to output. In this way, it is possible to define different sets of OCR and pre-processing conditions that are suited to different types of source documents. This approach can also improve the handling of documents that contain different types of pages, e.g. scanned at different qualities, containing different languages, containing standard and dot matrix prints, etc.

Sample Section of Properties.xml

```
<ImagePreProcessing ID="1">
  <Binarize>-1</Binarize>
  <Morph>c2.2</Morph>
</ImagePreProcessing>
<ImagePreProcessingDefaults>
  <RemoveLines></RemoveLines>
  <Binarize>200</Binarize>
  <BlackPixelLimit>0.65</BlackPixelLimit>
  <BoxSize>0</BoxSize>
  <GrayscaleQuality>0</GrayscaleQuality>
  <Jbig2EncFlags>-s</Jbig2EncFlags>
  <Language>0</Language>
  <MaxDeskew>10</MaxDeskew>
  <MinDeskewConfidence>3.0</MinDeskewConfidence>
  <Morph></Morph>
  <Mrc>
    <ForegroundFactor>3</ForegroundFactor>
```

```

    <BackgroundFactor>3</BackgroundFactor>
    <Quality>75</Quality>
</Mrc>
<Ocr>
    <DotMatrix>False</DotMatrix>
    <OneColumn>False</OneColumn>
    <NoPictures>False</NoPictures>
    <Tables>False</Tables>
</Ocr>
<PdfImageExtraction>
    <PdfToImage>0</PdfToImage>
    <PdfToImageMinRes>200</PdfToImageMinRes>
    <PdfToImageMaxRes>300</PdfToImageMaxRes>
    <PdfToImageIncludeText>False</PdfToImageIncludeText>
</PdfImageExtraction>
<RemoveLines></RemoveLines>
<SavePredespeckle>False</SavePredespeckle>
<TextLayerExtraction>
    <MaxBoxes>0</MaxBoxes>
    <FilterWidth>130</FilterWidth>
    <FilterHeight>130</FilterHeight>
    <FilterRatio>1</FilterRatio>
    <FilterPercentage>0.6</FilterPercentage>
    <FilterWidthInverted>40</FilterWidthInverted>
    <FilterHeightInverted>40</FilterHeightInverted>
    <FilterRatioInverted>0.95</FilterRatioInverted>
    <FilterPercentageInverted>0.5</FilterPercentageInverted>
</TextLayerExtraction>
</ImagePreProcessingDefaults>

```

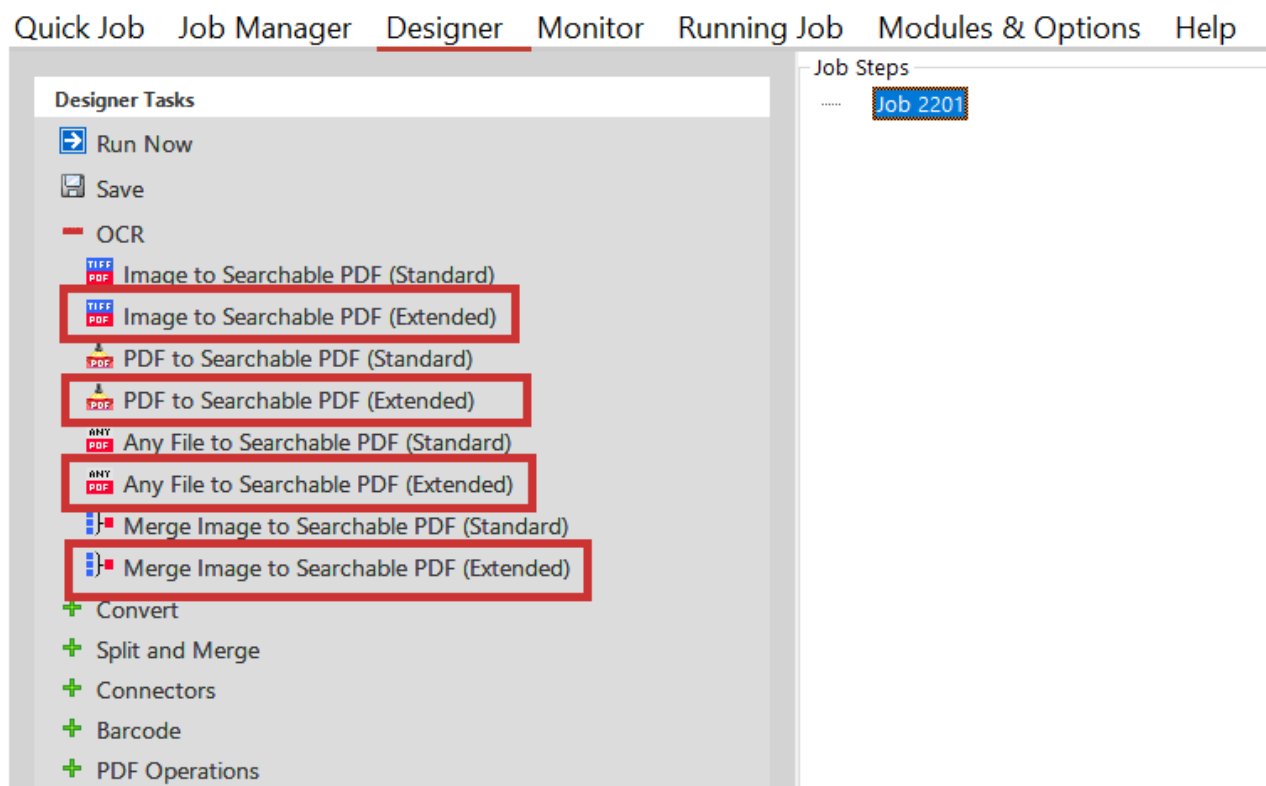
12 EXTENDED OCR MODULE

The optional Extended OCR module extends Autobahn DX with an additional OCR engine and has the following benefits over and above the standard Aquaforest OCR engine:

- IRIS OCR Engine providing enhanced recognition
- New PDF Rasterizer component
- DOCX Output option
- Improved RTF Output
- CSV and SpreadsheetML output options
- Additional Western Language Support (see [section 9.9.1](#) for details)
- Optional Asian Language Support
- Support for multiple languages within a single document from the same alphabet - e.g. French + German + Italian

12.1 Job Manager Steps

The Extended OCR Job steps can be found under the **OCR** group of the Job designer tab, the job step section gives you more details of the steps highlighted below.



These job steps mirror the default equivalents (Convert TIFF to PDF etc.) but will have some different options (e.g. different language and output formats) as shown below.

Step Properties

1 - Conversion Settings

Output File Name	%FILENAME_searchable
Create Directories if Required	Yes
Continue on Error	Yes
OCR Engine	IRIS
OCR Language 1	English
OCR Language 2	
OCR Language 3	
OCR Language 4	
OCR Language 5	
Deskew	No
Auto-Rotate	No
Despeckle	0
Output File	PDF,RTF,DOCX
PDF Version	1.7
Advanced Flags	
Workdepth	
JPEG Quality	128
Resources Folder	

Output File
Choose the type of output file. One or more of the following, separated by commas if more than one is required: RTF,PDF, DOCX, CSV, SML (SpreadsheetML XML file)...

New IRIS OCR Engine

You can select up to 5 different

Different output file

12.2 IHQC Module

The IHQC Module is an optional module in Extended OCR. It enables the use of IRIS' new Intelligent High-Quality Compression technology for powerful PDF compression without compromising visual quality, text resolution and legibility of documents.

12.2.1 IHQC properties

Step Properties

Advanced Flags	
Workdepth	
JPEG Quality	128
iHQC Compression	Yes
iHQC Compression Level	1
iHQC Quality Factor	Medium
No OCR	No
Advanced Despeckle	
Remove White Pixels	No
Binarization	No
Brightness	
Contrast	
Smoothing Level	

iHQC Compression
Apply intelligent High Quality Compression.

IHQC

12.2.2 Command Line Interface

autobahndx.exe /operation=[operation name] /source=[tiff file or folder] /output=[output file] /target=[target folder] /mrc=[true or false] /ihqcleve=[1 to 4] /ihqcqf=[1 to 9] [/option=value]...

Example

Generate a compressed searchable PDF file from a folder of TIFF and JPEG files using intelligent high-quality compression.

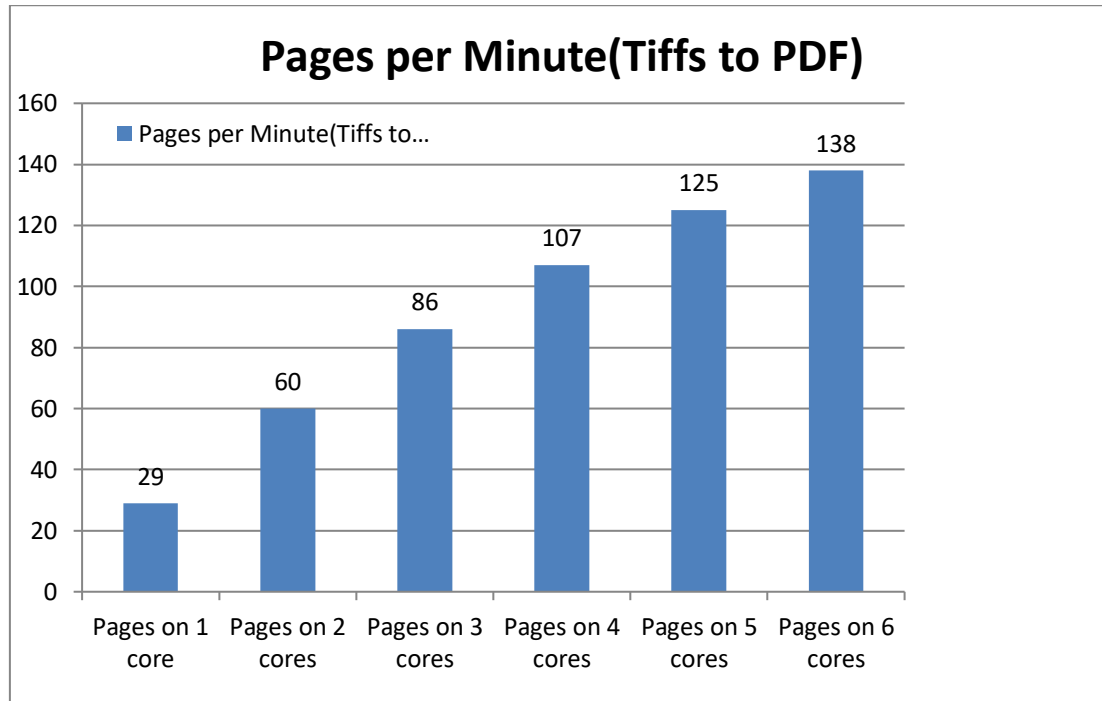
autobahndx.exe /source=c:\in\folder /sourcetype=folder /target=c:\out /output=outfilef /outputtype=pdf operation=mergetifftopdf /mrc=true /ihqcleve=4 /ihqcqf=5 /ocrengine=1

The following parameters are needed to use IHQC:

Parameter	Notes																				
/mrc	Apply intelligent High-Quality Compression. True or False																				
/IHQCLevel	The compression level to be used. Level 1 is the basic compression level. Level 3 is the most advanced intelligent High-Quality Compression mode. <table><tr><th>Compression Level</th><th>Code</th></tr><tr><td>3</td><td>4</td></tr><tr><td>2b</td><td>3</td></tr><tr><td>2a</td><td>2</td></tr><tr><td>1</td><td>1</td></tr></table>	Compression Level	Code	3	4	2b	3	2a	2	1	1										
Compression Level	Code																				
3	4																				
2b	3																				
2a	2																				
1	1																				
/IHQCQF	The IHQC quality factor: <table><tr><th>Quality</th><th>Code</th></tr><tr><td>Maximal quality</td><td>9</td></tr><tr><td>Very high quality</td><td>8</td></tr><tr><td>High quality</td><td>7</td></tr><tr><td>Favor quality over size</td><td>6</td></tr><tr><td>Medium</td><td>5</td></tr><tr><td>Favor size over quality</td><td>4</td></tr><tr><td>Small size</td><td>3</td></tr><tr><td>Very Small Size</td><td>2</td></tr><tr><td>Minimal size</td><td>1</td></tr></table>	Quality	Code	Maximal quality	9	Very high quality	8	High quality	7	Favor quality over size	6	Medium	5	Favor size over quality	4	Small size	3	Very Small Size	2	Minimal size	1
Quality	Code																				
Maximal quality	9																				
Very high quality	8																				
High quality	7																				
Favor quality over size	6																				
Medium	5																				
Favor size over quality	4																				
Small size	3																				
Very Small Size	2																				
Minimal size	1																				

13 MULTICORE MODULE

This module is used to take full advantage of the number of processors available on a computer. The current release allows users to process up to 64 files in parallel thereby reducing the time needed to execute a job. Below is a chart showing the number of pages converted from TIFFs to PDF per minute.



13.1 Using the MultiCore Module

The MultiCore module is enabled on all job step types, but you might not notice the improvement on fast job steps like Merge TIFFs, Split TIFFs, Merge PDFs and Split PDFs. To take advantage of the MultiCore module, the number of processors needed to run a job can be set from the MaxCores property from the step properties section as shown below. Just bear in mind that setting this value to the number cores on your system will cause a 100% usage of CPU resources by the process. The Max Available option allows .NET to decide the best number of processors to be used.

Step Properties	
1 - Conversion Settings	
Output File Name	%FILENAME.tif
Create Directories if Required	Yes
Continue on Error	Yes
TIFF Compression	Group 4 (Black and White)
Resolution (DPI)	200
Maximum Cores	8
Password Files	take no action
Debug	No

Note: Setting the Maximum Cores to Max Available will set the number of cores to the Minimum between the number of cores on your system and number of cores on your license.

13.2 Multicore Throttling

Your license key will support a specific number of CPU cores. The product will limit the number of concurrent file processing operations to this number and will “throttle” jobs accordingly.

For example, if a 4 core licensed server is currently running a 2 core job and a new job starts that is configured for 4 cores the number of cores allocated to the second job will be reduced accordingly:

Autobahn DX using 2 cores out of 4 allowed.

We will reduce the number of cores in this job from 4 to 2 allowed.

As another example if a 4 core licensed server is currently running a 4 core job and a new job starts that is configured for 2 cores then the second job will not be able to start until cores are freed up :

Autobahn DX using 4 processors out of 4 allowed.

We will attempt to start the job 18 time(s) over the next 180 seconds.

The retry interval and number of tries is determined by these two config file settings in Autobahn.config (by default this file is in C:\Aquaforest\Autobahn DX\config)

```
<add key="jobqueuetimeout" value="180" />
<add key="jobqueueinterval" value="10"/>
```

13.3 Command Line Interface

The Autobahn DX command line interface has been enhanced to support the new MultiCore module. Using **/maxcores=x** as a parameter is a requirement, where x is the number of cores.

```
autobahndx.exe /operation=[operation name] /source=[tiff file or folder] /output=[output file]
/target=[target folder] [/option=value]...
```

Example

Generate searchable PDF files from image PDF files found in a folder and subfolders, while keeping the original file names (using 4 processors).

```
autobahndx.exe /source=c:\in\folder /sourcetype=tree /target=c:\out /output=%FILENAME
/outputtype=pdf operation=ocrimagepdf /ocrengine=1 /maxcores=4
```

The MultiCore module uses the parameter listed below.

Parameter	Notes
/maxcores	An integer representing the number of cores needed.

14 BARCODE SUPPORT

Autobahn DX now can Split and rename PDF/TIFF Files based on barcodes present in them.

14.1 Barcode Formats

The table below gives us details about the Barcode Formats supported by Autobahn DX

Barcode Format	Code	Group
AZTEC	1	2D
CODABAR	2	1D
CODE_39	4	1D
CODE_93	8	1D
CODE_128	16	1D
DATA_MATRIX	32	2D
EAN_8	64	1D
EAN_13	128	1D
ITF (CODE_25)	256	1D
MAXICODE	512	2D
PDF_417	1024	Others
QR_CODE	2048	2D
RSS_14	4096	Others
RSS_EXPANDED	8192	Others
UPC_A	16384	1D
UPC_E	32768	1D
UPC_EAN_EXTENSION	65536	Others
MSI	131072	Others
PLESSEY	262144	Others

15 HIGH AVAILABILITY

15.1 Background information

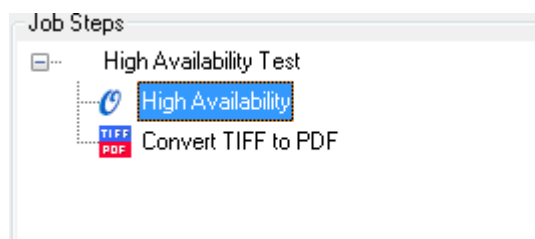
The high availability step in Autobahn DX is designed to utilize two instances of the product running on separate hosts, each instance will be configured with a job which will be scheduled to run at given intervals using the Autobahn scheduler.

The concept is based on the master host initially running the conversion whilst the slave host continuously monitors the master host, the slave host will take over the conversions if the master host:

- Is NOT accessible (based on a ping).
- Autobahn Service is NOT running.
- Job ID on the Master host is in NOT 'Running'.

15.2 Configuring the job on host one:

1. Create a new job and add the 'High Availability' step as the first step.
2. Add any additional required step for your conversion.



3. Next define the Source and Target folders for the Job, these need to be on a shared network location accessible to both hosts.

The screenshot shows a 'Job Properties' window with the following fields:

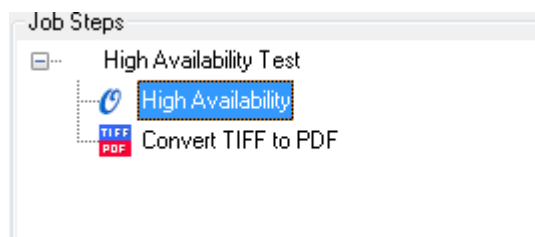
Job ID	1012
Job Name	High Availability Test
Source Folder	\\HMNHD-TI2RX1\qa\adx_high_availability_in
Target Folder	\\HMNHD-TI2RX1\qa\adx_high_availability_out

4. In the 'High Availability' step properties section define the following:
 - a. Enter the Job ID on the current host.
 - b. Select the Default status of the current host (Master | Slave)
 - c. Enter the shared.txt file location – this needs to be on a shared network location accessible to both hosts.
 - d. Name of the paired host.
 - e. Install path of Autobahn DX on the paired host.

Step Properties	
1 - Shared Settings	
Current Job ID	1012
Default Status	Master
Shared Status File	\\HMNHD-TI2RX1\Backups\Shrevin\qa\adx_high_availability_shared\shared.txt
2 - Pair Settings	
Hostname	aquaforestqa
ADX Install Path	C:\Aquaforest\Autobahn DX
Job ID	1007

15.3 Configuring the job on host two:

1. Create a new job and add the 'High Availability' step as the first step.
2. Add any additional required step for your conversion.



3. Next define the Source and Target folders for the Job, these need to be the same as those configured in the job on the first host.

Job Properties	
Job ID	1007
Job Name	High Availability Test
Source Folder	\\HMNHD-TI2RX1\Backups\qa\adx_high_availability_in
Target Folder	\\HMNHD-TI2RX1\Backups\qa\adx_high_availability_out

4. In the 'High Availability' step properties section define the following:
 - a. Enter the Job ID on the current host.
 - b. Select the Default status of the current host (Master | Slave)
 - c. Enter the shared.txt file location set in the job on the first host.
 - d. Name of the paired host.
 - e. Install path of Autobahn DX on the paired host.

Step Properties	
1 - Shared Settings	
Current Job ID	1007
Default Status	Slave
Shared Status File	\\HMNHD-TI2RX1\Backups\Shrevin\qa\adx_high_availability_shared\shared.txt
2 - Pair Settings	
Hostname	fileserver
ADX Install Path	D:\Aquaforest\Autobahn DX
Job ID	1012

15.4 Configuring the scheduler

The schedule is job specific and will need to be configured for each job on each host.

By default this is set to Ad-Hoc, change this to run continuously and set the desired frequency. In the following example, the scheduler will attempt to start the job every minute between 12:00 and 23:59.

If the job is already running on the master host the scheduler will attempt to start the job after another minute.

When the scheduler attempts to start the job on the slave host, the slave will check to see if the master host is up AND if the Autobahn Service is running AND if the job is running. If these conditions are met the job will not run and the scheduler will try after another minute.

16 ANY FILE TO PDF

This is divided into two steps OCR any file to PDF and Convert Any File To PDF:

16.1 OCR Any File to PDF

The OCR Any file step is a combination of the “Convert Any File to PDF” and OCR image PDF. The idea behind this step is to allow users to point this job to a folder with different file types and then Autobahn will decide on the operations to carry out to make these files searchable PDF files.

The main advantage of this step is that users can go from any file type to a searchable PDFs in an atomic process, hence allowing them to uncheck the “Use Work Folder” Which makes the processed files available in the target folder as soon as it is processed.

16.1.1 Configuration

A Configuration file is available for this step in the location below,
“Autobahn Dx DIR\bin\ AnyFileToSearchablePDF.exe.config”.

This allows the user to have some control over the step, Additional Configuration variables that can be set via the configuration file are as follows:

```
<configuration>
  <appSettings>
    <add key="ConvertToPDF"
value=".docx.doc.rtf.xls.xlsx.ppt.pptx.vsd.txt.xml.html.pub.puz.htm.mht.msg.odt.swx.w
pd.ods.sxc.odp.sxi.odg.sxd"/>
    <add key="OCRFile" value=".tiff.tif.pdf.jpg.jpeg.png.bmp.gif"/>
    <add key="ConvertToPDFAndOCR" value=".ps"/>
  </appSettings>
  <startup>
    <supportedRuntime version="v4.0" sku=".NETFramework,Version=v4.5" />
  </startup>
</configuration>
```

Command Line / Config File Name	Description
ConvertToPDF	A list of file extensions to be converted to PDF Only. Note: these should be text-based documents.
OCRFile	A list of file extensions to OCR'd Only. Note: These should be Image PDFs and Image Files only.
ConvertToPDFAndOCR	A list of file extensions to be Converted to PDF and OCR'd.

16.2 Convert Any File to PDF

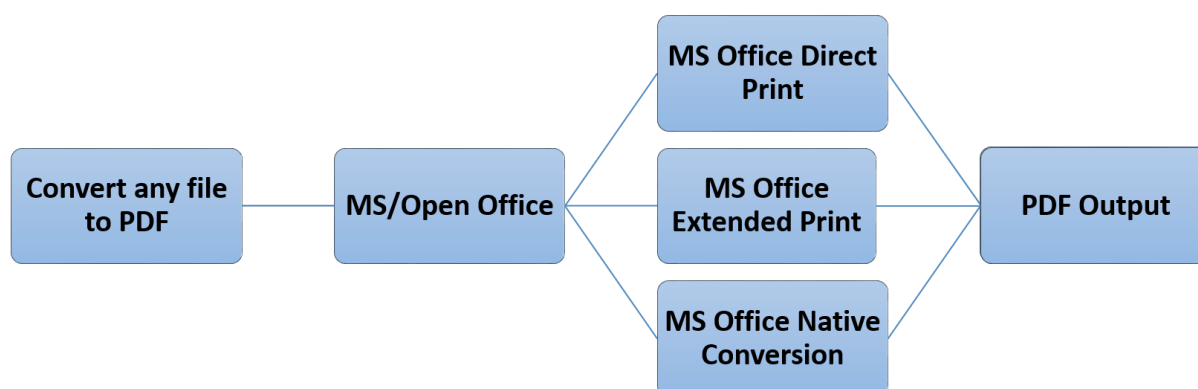
This converts any file type to PDF, the searchability status of the output PDF depends on the source document, E.G MS Office Input file.

This step uses a virtual printer to print files to a PDF file instead of a physical printer. Due to the printer and Microsoft office automation, there are a lot of technical and security Limitations. Thus, we have introduced different methods of converting files to PDF so that our users can have a robust solution to their requirements.

We have provided a configuration file to allow users to switch between these various methods. This blog will provide a detailed explanation of the different configuration options provided by Aquaforest.

16.2.1 Overview of the Different Conversions Methods

Aquaforest uses three different methods to convert office files to PDF files, these methods are explained in the section below.



16.2.1.1 MS Office Native Conversion

This step uses the “Save as PDF” option in MS Office to save the PDF file directly. This requires Microsoft Office 2007 or newer. Microsoft Office 2007 requires the free "Save as PDF or XPS" add-in for Office 2007 to be installed. We recommend you use the 2010 and new versions.

If you are not interested in changing properties like Image Compression, Image Downsizing, Font embedding and you are ok with the default PDF/A output from your version of office, we recommend using this approach.

16.2.1.2 MS Office Direct Print

This method uses the BCL ‘easyPDF SDK x’ printer to print out PDF files directly from Microsoft Office. This is the recommended option for producing PDF/A files and PDF files with font embedding.

16.2.1.3 MS Office Extended Print

In this method, Office automation is used to create an intermediate XPS file, which is then printed with automatic hyperlink extraction. This approach has some technical differences to the first method that allows the users to use the software in a different way. This requires Microsoft Office 2007 or newer. Microsoft Office 2007 requires the free "Save as PDF or XPS" add-in for Office 2007 to be installed. If you are performing conversions in server environments e.g., via IIS, Windows Services, Session 0 and ASP, this method does not require an interactive user to be logged on for server conversions to succeed, thus we recommend this setting for server environments.

16.2.2 Comparison of the features provided by the different methods

The table below shows you the features available in the different methods.

Property	Direct Print	Extended Print	Native Conversion
Bookmark depth	✓	✓	✓
Convert Hyperlinks	✓	✓	✓
Convert Bookmarks (Word)	✓	✓	✓
IncludeDocumentMarkups (Track Changes Word)	✓	✓	✓
Paper Orientation	✓	✓	✓
Paper Size	✓	✓	✓
Margin	✓	✓	✓
PDFA1b	✓	✓	Depends on the version of Microsoft Office
PDFX1a	✓	✓	
PDFX3	✓	✓	
Image Compression	✓	✓	✗
Image Downsizing	✓	✓	✗
Font Embedding	✓	✓	✗
Print All Sheets(Excel)	✓	✓	✓
Include Document Properties	✗	✓	✓
Fit to Page (Excel)	✓	✓	✓
MSG Files	✓	✗	✗
Frame Slides (PowerPoint)	✓	✓	✓
Print Color Type (PowerPoint)	✓	✓	✓
Output Type (PowerPoint)	✓	✓	✓
Handout Order (PowerPoint)	✓	✓	✓
Print Graphics (Pub)	✓	✗	✗

16.2.2.1 In-depth Look at the Configuration file

The configuration file can be found in:

"<Autobahn DX Installation directory>\distribution\pj\bin\topdf.exe.config"

Below are the contents of the configuration file, to use a method for a file type, just write the extension in the value property of the conversion Method as shown below.

Note: Do not write an extension in more than one method

```
<!-- Extension Mapping -->
<add key="AutoExtension" value="" />
<add key="AutoExtensionEx" value=".jpeg.jpg" />
<add key="AutoExtensionOpenNative" value="" />
<add key="AutoExtensionNative" value="" />
<add key="WordExtension" value="" />
<add key="WordExtensionEx" value=".docx.doc.rtf" />
<add key="ExcelExtension" value="" />
<add key="ExcelExtensionEx" value=".xls.xlsx" />
<add key="GenericExtension" value=".txt.dwg" />
<add key="PowerPointExtension" value="" />
<add key="PowerPointExtensionEx" value=".ppt.pptx" />
<add key="VisioExtension" value="" />
<add key="VisioExtensionEx" value=".vsd" />
<add key="IEExtension" value=".xml" />
<add key="PublisherExtension" value="" />
<add key="PublisherExtensionEx" value=".pub.puz" />
<add key="IEExtendedExtension" value=".mht" />
<add key="HTMLExtension" value=".htm.html" />
<add key="OutlookExtension" value=".msg" />
<add key="OpenOfficeExtension"
value=".odt.swx.wpd.ods.sxc.odp.sxi.odg.sxd" />
<add key="ExcludedExtensions" value=".zip.exe.pps.ps.chm" />
<add key="PDFExtension" value=".pdf" />
```

16.2.3 Mappings Between Configuration File and the Conversion Methods

The table below maps the configuration file with the Conversion Methods discussed earlier.

Note: The Configuration options ending with “Ex” work with the “NativeOfficePDF” config option to select a suitable conversion method.

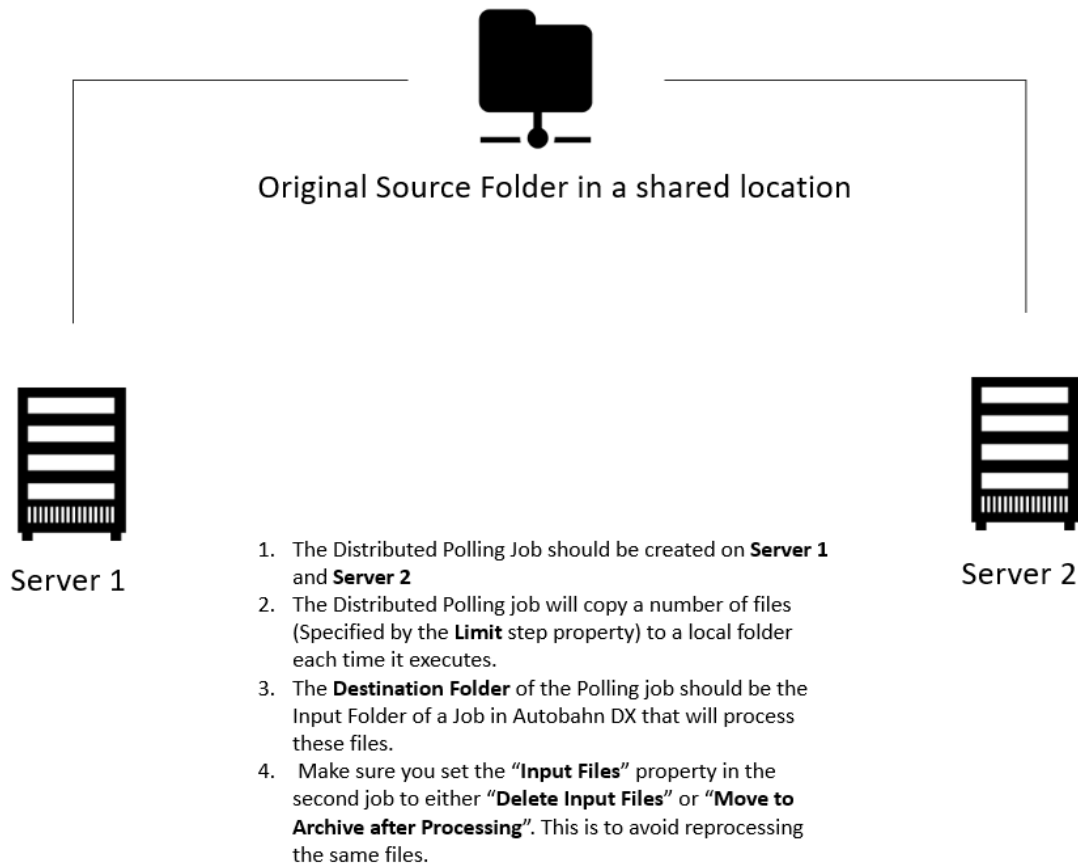
Configuration	Conversion Method	File Types	Comment
AutoExtension	Direct Print	All file types	This option works with any file extension that has an application that is compatible with the BCL printer
AutoExtensionEx	Extended Print	All file types	
AutoExtensionNative	Native Conversion	MS Office files	
OpenOfficeExtension	Direct Print	Open Office files	This option works with all files that can be opened in open office.
AutoExtensionOpenNative	Native Conversion	Open Office files	
WordExtension	Direct Print	.doc, .docx, .rtf... (All files that can be opened in MS Word)	
WordExtensionEx NativeOfficePDF=true	Native Conversion		
WordExtensionEx NativeOfficePDF=false	Extended Print		
ExcelExtension	Direct Print	.xls, .csv, .xlsx... (All files that open in excel)	
ExcelExtensionEx NativeOfficePDF=true	Native Conversion		
ExcelExtensionEx NativeOfficePDF=false	Extended Print		
PowerPointExtension	Direct Print	.ppt, .pptx (All power point files)	
PowerPointExtensionEx NativeOfficePDF=true	Native Conversion		

PowerPointExtensionEx NativeOfficePDF=false	Extended Print		
VisioExtension	Direct Print	.vsd	
VisioExtensionEx NativeOfficePDF=true	Native Conversion		
VisioExtensionEx NativeOfficePDF=false	Extended Print		
IEExtension	Direct Print	.html, .xml, .mht All IE files.	
IEExtendedExtension	Extended Print		
HTMLExtension	Direct Print	.htm, .html	
PublisherExtension	Direct Print	.pub, .puz	
PublisherExtensionEx NativeOfficePDF=true	Native Conversion		
PublisherExtensionEx NativeOfficePDF=false	Extended Print		
OutlookExtension	Direct Print	.msg files	
PDFExtension		PDF files	We usually convert PDF attachments if you select the option
GenericExtension	Direct Print	.txt, .dwg	The GenericExtension print job is normally only used as a last resort, when a file type is totally unrecognized by the ToPDF component. This defaults to use Windows Notepad for output.
ExcludedExtensions		others	Skips all the extensions present here.

17 DISTRIBUTED POLLING

The distributed polling step is used to provide a method of distributing processing load between a number of servers.

The Distributed Polling job copies a number of files based on the **Limit** step property from a shared location to a local folder. The local folder serves as an input file for another job to process.



This step can be used to implement load balancing in Autobahn DX. Multiple Autobahn DX servers can point to one input folder, as a result, the files will be shared across several servers and the processing will be more optimized.

The next section will work us through setting up a Distributed Polling job.

17.1 Job Setup

Before we start setting up the jobs, make sure you perform the steps below.

- Open “C:\Aquaforest\Autobahn DX\config\Autobahn.config” and change the value of “inputdelay” to 30.

```
<?xml version="1.0" encoding="utf-8"?>
<configuration>
  <appSettings>
    <add key="autobahnroot" value="C:\Aquaforest\Autobahn DX" />
    <add key="pdfjunctiondir" value="C:\Aquaforest\Autobahn DX\pj" />
    <add key="tiffjunctiondir" value="C:\Aquaforest\Autobahn DX\tj" />
    <add key="autobahntemp" value="C:\Aquaforest\Autobahn DX\temp" />
    <add key="autobahnwork" value="C:\Aquaforest\Autobahn DX\work" />
    <add key="jobdefdir" value="C:\Aquaforest\Autobahn DX\jobdef" />
    <add key="steptypedefdir" value="C:\Aquaforest\Autobahn DX\steptype" />
    <add key="jobstatusdir" value="C:\Aquaforest\Autobahn DX\jobstatus" />
    <add key="imgdir" value="C:\Aquaforest\Autobahn DX\img" />
    <add key="inputdelay" value="30" />
    <add key="logsdir" value="C:\Aquaforest\Autobahn DX\logs" />
```

- Make sure the Autobahn service is configured to run as a user with enough privileges to access all the shared locations.
- Restart the Autobahn service

Note: The effectiveness of the distribution of the documents from the central location to the local servers as well as the OCR will depend on the schedule intervals of the Job below, the “**Limit**” step property of the **Distributed Polling Job** and the “**inputdelay**”. You May need use trial and error to get the optimum settings. Do not set the two jobs to run at the same time, because the first job needs to run at least once for the second job to have any input files to process.

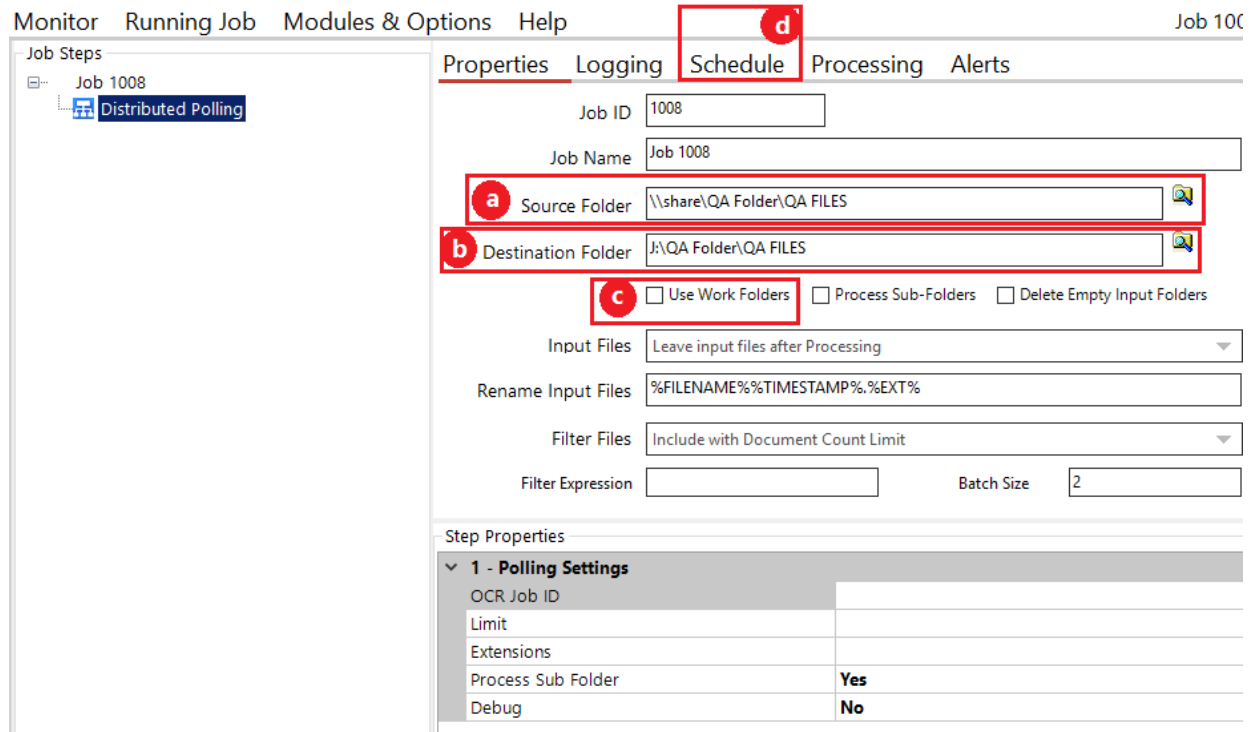
To set up a distributed polling job on an Autobahn Server you will have to create two jobs:

1. **Distributed Polling Job:** To copy the files from a shared folder to a local folder.
2. **Autobahn Job:** This job will process the files that were copied from the Job above, i.e. the **Destination Folder** of the first job will serve as the **Source Folder** of this job.

17.1.1 Distributed Polling Job

The **Distributed Polling Job** will copy files from the central shared location to the local computer's input location to be process by an **Autobahn Job**.

- After copying the files to the local location (b)
- They will be deleted from the central location (a).



- The central shared location containing all the input files.
- The folder to where the files will be copied to.
- Do Not use work folders.
- Set the scheduler to run every x number of seconds/minutes. This must be different from the 1(e).

Screen Field	Description
Autobahn Job ID	The Job ID of the Job that will be processing your input files. Note: The Source Folder of this job will be the Destination Folder of the Distributed Polling Job
Limit	The maximum number of files to be copied to the per run.
Extensions	Enter the file extensions you want us to copy separated by a comma. E.g. ".pdf,.tif,tiff"
Process Sub Folder	Select true if you want to copy subfolders. Note: If this is set to true, then "Process Sub-folders" needs to be ticked in the OCR job.
Debug	Select true if you want to see more debug output.

17.1.2 Autobahn Job

In this section we will show you how to set up an **Autobahn Job** to process the files that were copied from the **Distributed Polling Job**.

- Set the source folder to a local location on the server.
- The destination folder can either be local to the server or a shared location.
- Set "Input Files" to either "Delete Input Files" or "Move to Archive after Processing".
- Make a note of the job ID. This will be used in the next step.
- Set the scheduler to run every x number of seconds/minutes.

Monitor Running Job Modules & Options Help Job 1008

Job Steps

- Job 1008
 - Image to Searchable PDF (Standard)

Properties Logging **Schedule** Processing Alerts

d Job ID 1008

Job Name Job 1008

a Source Folder J:\QA Folder\QA FILES

b Destination Folder J:\QA Folder\Ouput Files

☒ Use Work Folders ☐ Process Sub-Folders ☐ Delete Empty Input Folders

c Input Files Move to Archive after Processing

Rename Input Files %FILENAME%%TIMESTAMP%.%EXT%

Filter Files Include with Document Count Limit

Filter Expression Batch Size 2

Step Properties

1 - Conversion Settings	
Output File Name	%FILENAME.pdf
Create Directories if Required	Yes
Continue on Error	Yes
OCR	Yes

18 CLOUD OCR

The optional Cloud OCR module extends Autobahn DX with additional OCR engines from Microsoft and Google, the main advantages of these OCR engines is their Hand writing recognition capabilities. These OCR engines are available as a SAAS model provided by both vendors. Before you can start using these steps in Autobahn DX, you will need to have a subscription first.

We have added two step types to the **Advanced** section of the **Job Designer** tab of Autobahn DX, the steps are named:

- Image to Searchable PDF (Cloud OCR)
- PDF to Searchable PDF (Cloud OCR)

The table below will explain the step properties of the Cloud OCR job steps.

Step Property	Description
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Continue on Error	Continue processing files after an error occurs.
OCR Service	We have two options for OCR service. <ul style="list-style-type: none">• Google• Microsoft
End Point (Microsoft Only)	The URL to the cognitive services endpoint where the OCR will be performed
Subscription Key	The subscription key to the above endpoint if you are using Microsoft or the path to the json subscription key file if you are using google. See section 18.1 and 18.2 for more.
Text Recognition Mode	Types of text to recognize <ul style="list-style-type: none">• Handwritten• Printed Default is Printed
OCR Language	Select the language to use for OCR processing. This will determine the dictionary that is used. Auto-Detect will automatically detect the language for each page. Printed text (see Text Recognition Mode) OCR supports 25 languages. Handwritten text OCR only supports English. 0 – Auto-Detect (default) 1 – Chinese (simplified) 2 – Chinese (traditional)

	3 – Czech 4 – Danish 5 – Dutch 6 – English 7 – Finnish 8 – French 9 – German 10 – Greek 11 – Hungarian 12 – Italian 13 – Japanese 14 – Korean 15 – Norwegian 16 – Polish 17 – Portuguese 18 – Russian 19 – Spanish 20 – Swedish 21 – Turkish 22 – Arabic 23 – Romanian 24 – Serbian Cyrillic 25 – Serbian Latin 26 – Slovak
Autorotate (Microsoft Only)	Auto-rotate the image – this will ensure all text oriented normally. The default value is false (disabled). Note: When using a PDF source, auto-rotation will be disabled on any pages already containing text.
Deskew	Deskew (straighten) the image. The default value is No (disabled).
Despeckle	The method removes all disconnected elements within the image that have height or width in pixels less than the specified figure. The maximum value is 9 and the default value is 0.
Line Removal in OCR Processing	Removes lines during OCR for improved results
Save Pre-Despeckle	This will use the original image (i.e. before applying pre-processing) in the output PDF. The default value is true.
Output File	PDF and/or TXT (separated by commas)
PDF/A Options	Select the output PDF/A compliant version you would like the output PDF to be. <ul style="list-style-type: none"> • PDF/A1-b • PDF/A2-b

	<ul style="list-style-type: none"> PDF/A3-b 						
Validate PDF/A	Whether or not to validate the PDF/A document after conversion						
JBIG2 Compression	This option will compress bitonal images in generated PDFs using JBIG2 compression rather than the default Group 4 compression scheme. This will result in smaller PDF file sizes, at a cost of increased processing time.						
MRC	This enables Mixed Raster Compression which can dramatically reduce the output size of PDFs comprising Color scans.						
Remove Blank Pages	Set this to true to remove blank pages from Tiff or PDF documents. Value needs to be set for the Blank Page Threshold (see below).						
Blank Page Threshold	Use this to set the minimum number of "On Pixels" that must be present in the image for a page not to be considered blank. A value of -1 will turn off blank page detection.						
Handwritten Results Retries	The number of times to wait for the handwritten OCR results						
Handwritten Results Wait	The amount of time (in milliseconds) to wait between each retry						
Advanced Flags	Command line flags to be passed through to the underlying executable.						
Maximum Cores	This specifies the number parallel files you want to be processed at a given time. Note: You need the multi-core license for this.						
Debug	Set this to true to execute the job in debug mode.						
PDF to Searchable PDF Only Properties							
Non-Image PDFs	<p>This allows control over the treatment of non-image only PDFs, i.e. PDFs that have some text in them as well as images.</p> <table> <tr> <th>Member name</th><th>Description</th></tr> <tr> <td>OCR</td><td>The document will be OCR'd using the image extraction method defined by "ExtractImageMethod".</td></tr> <tr> <td>Raise Error</td><td>The task will terminate with an error.</td></tr> </table>	Member name	Description	OCR	The document will be OCR'd using the image extraction method defined by "ExtractImageMethod".	Raise Error	The task will terminate with an error.
Member name	Description						
OCR	The document will be OCR'd using the image extraction method defined by "ExtractImageMethod".						
Raise Error	The task will terminate with an error.						

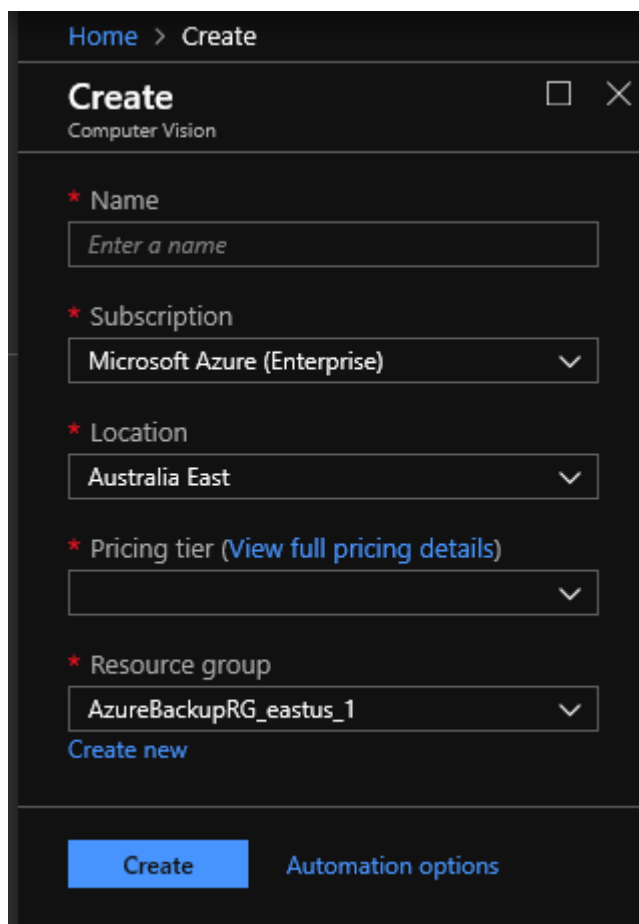
	Skip	The document will not be processed.
	Pass Through	The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.
Remove Hedden Text	<p>This applies only when a PDF is being used as the source for OCR. When set to true this will not include any searchable text that already exists from the source document. Such functionality might be useful if the source document was created by OCR of an image only PDF or other image file and the quality of the text from the previous OCR is poor.</p> <p>Note: There is no way to distinguish text added as a result of OCR from text added by other means and as a result this option should be used with care.</p>	
Extract PDF Images Via	<p>This allows control over the method used to extract images from PDF files for OCR processing. The default value is 'Native'.</p> <ul style="list-style-type: none"> • Native • Convert To TIFF 	
DPI	<p>The DPI to set to the images rasterized from each page of the source PDF file. These images are then OCR'd to create the searchable PDF.</p> <p>The default value for this property is taken from each page in the source PDF file.</p>	
TIFF Compression	<p>The compression to set to the images extracted or rasterized from each page of the source PDF file. These images are then OCR'd to create the searchable PDF.</p> <p>The default value for this property is taken from each page in the source PDF file. Valid values are CCITT4 or LZW</p>	
Retain Bookmark	<p>Retains any bookmarks from the source file in the output.</p> <p>Note: Extract Image Method must be set to Convert to TIFF for this to work.</p>	
Retain Metadata	<p>Retains any metadata from the source file in the output when using.</p> <p>Note: Extract Image Method must be set to Convert to TIFF for this to work.</p>	
Retain Viewer Preferences	<p>Retains any PDF Viewer Preferences, Page Mode and Page Layout from source file in the output when using.</p> <p>Note: Extract Image Method must be set to Convert to TIFF for this to work.</p>	

18.1 Microsoft Computer Vision

[Azure's Computer Vision service](#) provides developers with access to advanced algorithms that process images and return information. The images processing algorithms can analyze content in several different ways, depending on the visual features you're interested in. Computer Vision provides several services that recognize printed or handwritten text that appears in images.

To use this service, you will need a:

- Microsoft Azure account, you can sign up for this using the following [link](#).
- Microsoft Computer Vision API endpoint, you can add this to the azure account you created using the following [link](#).
 - Enter a suitable name for the endpoint.
 - Choose your preferred azure subscription.
 - Choose any location (Using a location that is closer to your files should give better performance).
 - Select a suitable pricing tier depending on your work load.
 - Select or create a new resource group.



The screenshot shows the 'Create' form for the 'Computer Vision' service in the Azure portal. The form is titled 'Create' with a subtitle 'Computer Vision'. It contains several required fields marked with a red asterisk:

- Name:** A text input field with the placeholder text 'Enter a name'.
- Subscription:** A dropdown menu showing 'Microsoft Azure (Enterprise)'.
- Location:** A dropdown menu showing 'Australia East'.
- Pricing tier:** A dropdown menu with a link '(View full pricing details)' next to it.
- Resource group:** A dropdown menu showing 'AzureBackupRG_eastus_1'.

Below the resource group field is a link 'Create new'. At the bottom of the form are two buttons: 'Create' and 'Automation options'.

18.1.1 Pricing

The table below gives you an estimate of the costs involved in using the **Microsoft Computer Vision API** to perform OCR operations. Note that you will consume one transaction per page.

To have a more accurate estimate you can use the following [link](#).

Price	Transactions per month
Free	0 - 5000
\$1 per 1,000 transactions	0 – 5M
\$0.80 per 1,000 transactions	1M – 5M
\$0.65 per 1,000 transactions	5M+

18.2 Google Cloud Vision

[Cloud Vision API](#) allows developers to easily integrate vision detection features within applications, including image labeling, face, and landmark detection, optical character recognition (OCR), and tagging of explicit content. We only use the OCR and Handwriting recognition features in Autobahn DX.

To use the [Cloud Vision API](#) in Autobahn DX, you will need a:

- Google account, you can sign up for one using the following [link](#).
- Subscription key for the [Google Cloud Platform](#) . You can start your free trial using the following [link](#), register for the trial and download your subscription key as a JSON file. Use the location of this JSON file as the value for the **Subscription Key** step property in Autobahn DX.

18.2.1 Pricing

The table below gives you an estimate of the costs involved in using the **Google Cloud Vision API** to perform OCR operations. Note that you will consume 1 unit for each page.

To have a more accurate estimate you can use the following [link](#).

Price	Units per month
Free	0-1000
\$1.50 per 1000 units	1001 – 5M
\$0.60 per 1000 units	5M – 20M

19 KINGFISHER JOB STEP

Aquaforest Kingfisher is one of our products that can perform the operations below on PDF files.

- Separate by Text Patterns
- Separate by Barcodes
- Renaming based on Text Patterns
- Renaming based on barcodes
- Zonal Patterns
- Extract Text to Text Files
- Extract Text to CSV/Excel Files

You can now use the “Kingfisher Job” step to run jobs that you have set up in Kingfisher within Autobahn DX, for more information about kingfisher, visit the link below.

<https://www.aquaforest.com/en/kingfisher.asp>

You can check here for more details on how to set up a kingfisher job in autobahn dx.

<https://www.aquaforest.com/wp/index.php/adding-a-kingfisher-job-as-an-autobahn-dx-step/>

20 FILE NAME LENGTH

Windows has a file name length of 260 characters and a folder name length limit of 248 characters. There are potentially two areas where this could impact Autobahn DX processing.

20.1.1 File Name Length - UNC Paths

If a local file path is at or close to one of the windows limits, it can potentially cause issues. For example this local file

C:\documents\major accounts payable 2010\.....with 260 characters.pdf

When accessed remotely via a UNC, such as the one below will exceed the 260 character limit when Autobahn DX is scanning folders for files to process. This can even affect files that are to be filtered out (e.g. .msg files in the case where *.pdf has been specified as a filter).

\\servername\c\$\documents\major accounts payable 2010\.....with 260 characters.pdf

The best solution for this problem is to create a share such that the UNC path will not exceed the limit.

For example create a share to *C:\documents\major accounts payable 2010* with a short name such as “dp2” will then allow access via a UNC such as *\\servername\dp2\.....with 260 characters.pdf*

In other words, in Designer the source folder is [\\servername\dp2](#) rather than [\\servername\c\\$\documents\major accounts payable 2010](#)

20.1.2 File Name Length – Autobahn DX Working Folder Paths

For integrity reasons, Autobahn DX will (by default) copy source files and writes processed files to work folders before copying to the target output folders. The location of the working folders is defined in the job manager – for example, *C:\Aquaforest\Autobahn DX\work\1008*

As an example, given a file *C:\Accounts\Business Documents 2010\Invoices\.... with 260 characters.pdf* and the source folder defined as *C:\Accounts\Business Documents 2010\Invoices* then the working file will be *C:\Aquaforest\Autobahn DX\work\1008\ Invoices\.... with 260 characters.pdf* which will exceed the limit.

The best solution is to shorten the working folder path via the Designer screen – for example to *C:\dw\1008* instead of *C:\Aquaforest\Autobahn DX\work\1008*

20.1.3 Long File Names Support

Starting in Windows 10 and Windows Server 2016, version 1607, MAX_PATH limitations have been removed from common Win32 file and directory functions. However, you must opt-in to the new behavior. This section will show you how to configure your computer and Autobahn DX to use long file names.

20.1.3.1 Autobahn DX Config

- Open the Autobahn DX config file found in the location below:
 - “**Autobahn Install Folder**\ config\Autobahn.config”
- Set the “**allowlongfilename**” value to “**True**” as shown below:
 - `<add key="allowlongfilename" value="True" />`

20.1.3.2 Windows Configuration

- Start the registry editor (regedit.exe)
- Navigate to HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\FileSystem
- Double click LongPathsEnabled
- Set to 1 and click OK
- Reboot

21 AUTOBAHN DX CONFIGURATION FILES

This section will list the different configuration files we have in Autobahn DX and their uses.

21.1 Autobahn.config

The Autobahn DX configuration file “Autobahn.config” can be found in the “config” directory of Autobahn DX, this file allows the user to make some backend changes to the way Autobahn DX works. Below are the contents of the file.

Note: We strongly advice that users should not change any settings that we do not document here, there are a lot of settings present in the config file, but we left them out of the reference guide. These settings are important for Autobahn DX to function properly and should never bet changed unless guided by the Aquaforest support team.

```
<configuration>
  <appSettings file="">
    <clear />
    <add key="inputdelay" value="10" />
    <add key="jobqueuetimeout" value="180" />
    <add key="jobqueueinterval" value="10" />
    <add key="mergeflag" value="" />
    <add key="trigger1001" value="" />
    <add key="triggerlevel" value="" />
    <add key="allowlongfilename" value="" />
  </appSettings>
</configuration>
```

21.1.1 The Input Delay

This in an integer value that represents the number of seconds a file must exist in the input folder before Autobahn DX attempts to process it. The default value is 0, to change this just update the value of the **inputdelay** key.

```
<add key="inputdelay" value="10" />
```

Note: The value is in seconds.

21.1.2 Multi Core Throttle Settings

See [section 2.4](#) for more information.

```
<add key="jobqueuetimeout" value="180" />
<add key="jobqueueinterval" value="10"/>
```

Note: The **jobqueuetimeout** value is in seconds.

21.1.3 Output Trigger Files

This feature makes Autobahn DX to copy a file (trigger file) to the output folder after the job completes and successfully copies the processed files to the Output folder. This file can be used to trigger a post processing workflow.

The keys below can be used to control the way this feature acts.

Setting Name	Details
trigger%JOBID%	When you add a value to this key, Autobahn DX will create a file with the name of the value in the output folder of the job with ID = %JOBID% To send a trigger file for a job with job id 1002 , use trigger1002 as the key
triggerlevel	This specifies if the trigger file should be copied to the top-level folder only or also the sub folders. The possible values are <ul style="list-style-type: none"> • sub: Copies the file to the top-level folder and the sub folders too. • top: Copies the file to the top-level folder only.

The setting below will copy **test.csv** to the top-level Output Folder of a Job with Job ID 1001

```
<add key="trigger1001" value="test.csv" />
<add key="triggerlevel" value="top" />
```

21.1.4 The Merge Flag Feature

This feature is used to give the user a certain level of control over which folders to be included in the merge operations ("Merge Tiffs", "Merge PDF", "Merge Tiffs to PDF" and "Extended Merge Tiffs to PDF"). It provides the user with a mechanism of flagging folders that are to be included for processing.

To use this feature, the following steps need to be followed:

- Choose a name you want to set as the flag name. For instance, "_ready".
- Open Autobahn.config file located at "<InstallDirectory>\Autobahn DX\config\"
- Change the value of <add key="mergeflag" value="" />
 - to <add key="mergeflag" value="ready.adx.merge" />.
- Note: The file extension must be "*.adx.merge" for this to work.
- Next, create the 0 size byte file named "_ready.adx.merge" and copy and paste in each folder that requires processing.
- Restart the service to load the new configuration.
- Note: restarting the service will cause all jobs that are running through the job manager to terminate.

21.1.5 Allow Long File Names

Starting in Windows 10 and Windows Server 2016, version 1607, MAX_PATH limitations have been removed from common Win32 file and directory functions. However, you must opt-in to the new behavior. [Section 20](#) gives more details on how to enable this setting.

After enabling support for long filenames in the registry, set the value of the **allowlongfilename** key to true.

21.2 tiffjunction.exe.config

The TIFF Junction configuration file “tiffjunction.exe.config” can be found in the “distribution\tj\bin” directory of Autobahn DX, this file allows the user to make some backend changes to the way TIFF Junction works. Below are the contents of the file.

```
<?xml version="1.0"?>
<configuration>
  <appSettings>
    <add key="DetailedValidation" value="false" />
    <add key="CompatabilityMode" value="false" />
    <add key="DateFormat" value="MM dd yyyy" />
    <add key="AllowErrorTiffFiles" value="false" />
    <add key="UseLibTiff4JPEG" value="false" />
    <add key="OrderByInt" value="false" />
    <add key="ConvertToTIFFThreshold" value="50" />
    <!--Extensions in lower case -->
    <add key="AcceptedImages" value=".jpeg,.png,.bmp,.wbmp,.tif..." />
  </appSettings>
</configuration>
```

We are concerned only with the element within the “appSettings” tags, the table below gives a more detailed explanation of what each

Setting Name	Details
DetailedValidation	Set this to true if you want TIFF Junction to carry out extra checks and validations like PDF/A validation after conversion. Setting it to true will make TIFF Junction slower.
CompatabilityMode	Set this to true if the arguments you are using to call tiffjunction were generated before 2015.
DateFormat	You can change the DateFormat of the output file names here, this works in conjunction with the %DATESTAMP%.
AcceptedImages	Tiff junction will treat any extension found here as an acceptable image file.
OrderByInt	When merging files with numbers as their file name, setting this to true will make sure that they are ordered numerically.
AllowErrorTiffFiles UseLibTiff4JPEG ConvertToTIFFThreshold	Only use this when advised from Aquaforest support.

21.3 pj.exe.config

The PDF Junction configuration file “pj.exe.config” can be found in the “distribution\pj\bin” directory of Autobahn DX, this file allows the user to make some backend changes to the way PDF Junction works. Below are the contents of the file.

```
<?xml version="1.0"?>
<configuration>
  <startup useLegacyV2RuntimeActivationPolicy="true">
  </startup>
  <appSettings>
    <add key="OrderByInt" value="true"/>
    <add key="ValidatePDFa" value="true"/>
    <add key="RemoveDuplicateFontFiles" value="true"/>
    <add key="DateFormat" value="MM dd yyyy" />
  </appSettings>
</configuration>
```

We are concerned only with the element within the “appSettings” tags, the table below gives a more detailed explanation of what each

Setting Name	Details
ValidatePDFa	Set this to true if you want PDF Junction to validate PDF/A files after conversion.
DateFormat	You can change the DateFormat of the output file names here, this works in conjunction with the %DATESTAMP%.
RemoveDuplicateFontFiles	When merging PDFs with embedded fonts, this option will make sure there are no duplicate font files in the target file, setting this to false can cause very large output PDF files, while setting it to true might slow down the process a bit.
OrderByInt	When merging files with numbers as their file name, setting this to true will make sure that they are ordered numerically.

21.4 Others

Filename	Location	Comment
topdf.exe.config	distribution\pj\bin	Check section 19.1.1 for more details
AnyFileToSearchablePDF.exe.config	bin	Check section 19.2.2.1 for more details

22 AUTOBAHN DX DIRECTORIES

After installation, the following subdirectories are created in the PDF Junction install directory:

Directory	Contents
bin	Executables, DLLs and WSH files.
code examples	Sample API and command line projects
config	gettingstarted Autobahn.config
custom	Custom Script Files
logs	The default location for log files is logs/%JOBID%/DATESTAMP%.txt
distribution	Contains applications that are used to execute some Autobahn DX jobs.
jobdef	Contains the Job Definition files.
jobstatus	Contains the Job Status files.
steptype	Contains the StepType definition files.
samples	Sample documents
work	Default work directory. Subdirectories are created as required underneath this directory and named 9999/work1, 9999/work2 etc. where 9999 is the jobid.
temp	The default directory for temporary files.
license	Contains key.txt which holds the license key
docs	Includes the Autobahn reference guide and release notes

23 ACKNOWLEDGMENTS

This product makes use of several Open Source components which are included in the binary form. The appropriate acknowledgments and copyright notices are given below.

Name	Homepage
BitMiracle.LibTiff.NET	Homepage GitHub
Cuneiform	n/a (Copyright (c) 1993-2008, Cognitive Technologies)
FreedImage.NET	Homepage
IKVM.NET	Homepage Sourceforge
Jurassic	GitHub
Leptonica	Homepage
Libjpeg	Homepage
Libpng	Homepage
Libtiff	Homepage
log4Net	Homepage
MetroFramework	Homepage GitHub
PDFBox	Homepage
S22.imap	Homepage GitHub
veraPDF	Homepage GitHub
Zlib	Homepage
ZXing.NET	Homepage