

Δquaforest

Autobahn DX 6.0 **Reference Guide**



Autobahn DX **Reference Guide**



Version 6.0
February 2025

Content

1	PRODUCT OVERVIEW AND CONCEPTS	10
1.1	Overview	10
1.1.1	Edition Comparison	10
1.2	Autobahn DX Architecture and Concepts	13
1.2.1	Autobahn Administrator	13
1.2.1.1	Quick Job	13
1.2.1.2	Job Manager	13
1.2.1.3	Designer	14
1.2.1.4	Monitor	14
1.2.1.5	Running Job	14
1.2.1.6	Modules & Options	14
1.2.1.7	Help	14
1.2.2	User Application	14
1.2.3	Autobahn Windows Service	14
1.2.4	Job Definition Files	14
1.2.4.1	Debug mode	15
1.2.5	Document Folders	15
1.2.6	Document Processing	15
1.2.7	Job Status	15
1.2.8	Autobahn DX .NET Job API	15
1.2.9	BCL EasyPDF service	15
1.2.10	File Access Permissions	15
1.2.10.1	Quick Jobs	15
1.2.10.2	Job Manager (Ad-Hoc or Scheduled)	15
1.2.11	What is the definition of a Core?	16
2	INSTALLATION AND LICENSING	17
2.1	System Requirements	17
2.2	Product Licensing	18
2.3	Entering License Keys	18
2.4	Autobahn DX Windows Service	19
2.4.1	Autobahn and BCL easyPDF Windows Services	19
2.4.1.1	Create and initialize an Autobahn DX user account:	19
2.4.1.2	Log on as the Autobahn Service account	19
2.5	Upgrading from previous versions	19
2.5.1	Uninstalling the previous version	19
2.5.2	Previous jobs and license keys	19
2.5.2.1	Job definitions	19
2.5.2.2	License Key	19
2.5.2.3	Configure Windows Service setting	20
2.6	Support for Multiple CPU Cores	21
3	AUTOBAHN DX QUICK START	22
3.1	The Quick Job Tab	22

3.2	The Job Manager Tab	23
3.3	The Designer Tab	24
3.3.1	General Job Properties	24
3.3.2	Scheduling a job	25
4	USING THE QUICK JOB TAB	26
4.1.1	OCR Options Section	28
4.1.2	OCR PDF Options Section	29
4.1.3	PDF to TIFF Options Section	30
4.1.4	Split Options Section	30
4.1.5	Other Options Section	31
4.1.6	Barcode Options	32
4.1.7	Task Log Section	35
4.2	Security Settings	36
4.3	Document Metadata	37
4.4	Document Options	37
4.4.1	Display Modes	38
4.4.2	Settings	38
4.4.3	Page Settings and Margins	38
4.5	PDF Conversion Settings	39
4.5.1	Conversion Control	39
4.5.2	Microsoft Office Conversion Settings	40
4.5.3	Image and Font Conversion Settings	42
4.6	Advanced Flags	43
5	THE AUTOBAHN DX ADMINISTRATION TOOL	44
5.1	Job Manager	44
5.2	Job Designer	45
5.2.1	Menu Items	46
5.2.2	Fields	48
5.2.3	Job Scheduling	52
5.2.3.1	Ad-Hoc	52
5.2.3.2	Watched Folder / Continuous Scheduling	52
5.2.3.3	Daily Scheduling	52
5.2.4	Alerts	53
5.2.4.1	Alert variables	54
5.2.5	Workflow Processing versus In-Place Processing	54
5.2.5.1	Example In-Place Job Setup	55
5.2.6	Step Types	56
5.2.6.1	Image To Searchable PDF	58
5.2.6.1.1	Standard Engine	58
5.2.6.1.2	Extended Engine	60
5.2.6.2	PDF to Searchable PDF	64
5.2.6.2.1	Standard Engine	64
5.2.6.2.2	Extended Engine	67
5.2.6.3	Merge TIFFs to PDF	72
5.2.6.3.1	Standard Engine	72
5.2.6.4	Convert Any File to PDF	73
5.2.6.5	Set PDF Properties	76
5.2.6.6	Custom Script	77

5.2.6.7	Stamp PDF Files	78
5.2.6.7.1	Stamp Variables	80
5.2.6.8	Merge PDF	81
5.2.6.9	Split PDF	82
5.2.6.10	Merge TIFFs	83
5.2.6.11	Split TIFF	83
5.2.6.12	Read Inbox	84
5.2.6.13	Send Documents	86
5.2.6.14	Convert PDF to TIFF	87
5.2.6.15	Extract Text from PDF	87
5.2.6.16	SharePoint Download	88
5.2.6.17	SharePoint Upload	89
5.2.6.18	Azure Storage Download	90
5.2.6.19	Azure Storage Upload	90
5.2.6.20	Create XML Property File	91
5.2.6.21	Optimize PDF	91
5.2.6.22	OCR Any File to PDF	92
5.2.6.22.1	Standard Engine	92
5.2.6.22.2	Extended Engine	97
5.2.6.23	Barcode TIFF/PDF	107
5.2.6.24	High Availability	110
5.2.6.25	Distributed Polling	110
5.2.6.26	Kingfisher Job	111
5.2.6.27	PDF To PDF/A Job	111
5.2.6.28	PDF Recognition to JSON Job	111
5.2.6.29	Modern Compress PDF	113
5.2.6.30	Validate PDF/A	114
5.2.6.31	Linearize PDF	115
5.2.6.32	Convert Any File To PDF (GdPicture)	115
5.2.6.33	Combine Any File To PDF	118
5.2.6.34	Combine PDFs	121
5.2.6.35	PDF To JPEG/PDF To PNG/ PDF To Tiff	122
5.2.6.36	PDF To Text	124
5.2.6.37	PDF To Searchable PDF (GdPicture)	125
5.2.6.38	PDF Portfolio	125
5.2.6.39	Smart Redaction	126
5.2.6.40	Detect Signatures	127
5.2.6.41	Key Value Pair Extraction	128
5.2.6.41.1	CSV Output Warning	128
5.2.6.42	Pattern Redaction / Pattern Highlight	130
5.2.6.43	Split PDF (GdPicture)	131
5.2.6.44	Split by Barcode	132
5.2.6.45	Pattern Enumeration	134
5.2.6.46	Get Document Information	135
5.2.6.47	Convert PDF To Office	136
5.2.6.48	Convert Any File To Office	137
5.2.7	Step Type Properties	141
5.3	Monitor	142
5.4	Modules and Options	143
5.4.1	License key	143
5.4.2	Email settings	143
5.4.2.1	Email settings for SMTP:	143
5.4.2.1.1	Sending emails from Autobahn DX via Microsoft 365 (Office 365) without authentication	144
5.4.2.2	Email settings for OAuth2:	145
5.5	Running Jobs via Job Manager or Designer	146

6	SCRIPTING CUSTOM STEPS	147
6.1	Custom Script Example	147
6.2	Building a Custom Script	148
6.2.1	Use case – Script only	148
6.2.2	The Script	148
6.2.2.1	WSF wrapper	148
6.2.2.2	Jscript code	149
6.2.3	The Autobahn DX Job	151
7	JOB DEFINITION XML FILES	152
7.1	Overview of Job Definition Creation and Processing	152
7.2	Job IDs	152
7.3	Sample Job Definition File	153
7.3.1	Job Step	156
7.3.1.1	Step Details	157
7.3.1.2	Attribute Values	157
8	AUTOBAHN DX COMMAND LINE INTERFACE	158
8.1	Examples	158
8.1.1	Generate a searchable PDF C:\ADX Demo\Output\ US2007246939A1.pdf from a TIFF file	158
8.1.2	Generate a searchable PDF file from a folder of TIFF and JPEG files, with Deskew and page orientation detection and correction.	158
8.1.3	Generate searchable PDF files from image PDF files found in a folder and subfolders.	158
8.1.4	Run an Autobahn job that is in the Job Manager, e.g to run job 1001.	158
8.1.5	Run an Autobahn job from an existing jobfile, e.g to run job 1002.	158
8.1.6	Smart Redaction of Credit Card numbers only	158
8.2	General Job Options	159
8.3	Split Parameters [used with splittiff and splitpdf operations]	161
8.4	PDF Conversion Options [used with topdf operations]	162
8.5	[PDF to TIFF [used with pdftotiff]	167
8.6	OCR [used with ocrimagepdf or tifftopdf/ mergetifftopdf]	168
8.7	PDF Output File Settings – [used with pdfprops]	171
8.8	Sharepoint Settings – [used with sharepointdownload and sharepointupload]	173
8.9	Extended OCR Steps	174
8.9.1	Extended OCR Languages	180
8.10	Create XML Property File – [used with getxmlpdf]	187
8.11	OCR Any File to PDF– [used with ocranyfile or ocranyfileex]	187
8.11.1	Standard	187
8.11.2	Extended	187
8.11.3	GDPicture	187
8.12	TIFF/PDF Barcode – [used with barcode]	188

8.13	Stamp PDF – [used with stamppdf]	190
8.14	Azure Upload/Download – [used with azuredownload and azureupload]	191
8.15	Extract Text Download – [used with extracttext]	192
8.16	Merge PDF – [used with mergepdf]	192
8.17	GDPicture Steps	193
8.17.1	Validate PDF/A – [used with validatepdfa]	194
8.17.2	Linearize PDF – [used with linearizepdf]	195
8.17.3	Convert Any File to PDF – [used with convertany]	196
8.17.4	Combine Any File– [used with combineany]	198
8.17.5	Combine PDFs– [used with combinepdfs]	201
8.17.6	Convert PDF to JPEG – [used with pdftojpeg]	202
8.17.7	Convert PDF to PNG – [used with pdftopng]	204
8.17.8	Convert PDF to TIFF – [used with pdftotiffgd]	205
8.17.9	Convert PDF to Text – [used with pdftotext]	206
8.17.10	OCR PDF to searchable PDF – [used with ocrpdfgd]	207
8.17.11	Create PDF Portfolio – [used with pdfportfolio]	208
8.17.12	Convert PDF to PDF/A – [used with gdpicturepdfa]	209
8.17.13	Compress PDF – [used with gdpicturecompression]	211
8.17.14	Smart Redaction – [used with smartredaction]	213
8.17.15	Detect Signatures – [used with detectsignatures]	214
8.17.16	Extract Key Value Pairs – [used with keyvaluepairs]	215
8.17.17	Pattern Redaction – [used with patternredaction]	216
8.17.18	Pattern Highlight – [used with patternhighlight]	216
8.17.19	Split PDF (GdPicture) – [used with splitpdfgdpicture]	217
8.17.20	Split By Barcode – [used with splitbybarcode]	218
8.17.21	Pattern Enumeration	219
8.17.22	Get Document Information	220
8.17.23	Convert PDF To Office	220
8.17.24	Convert Any File To Office	221
9	AUTOBAHN DX .NET API	225
9.1	API Example (Local Job)	225
9.2	API details	226
10	OCR PROPERTIES FILE AND THE ADVANCED PRE-PROCESSING OPTION	227
10.1	Properties File	227
10.2	Advanced Pre-Processing	229
11	UPGRADE JOB DEFINITIONS	233
11.1	Automatic Upgrades	233
11.2	Upgrading Manually	233
12	EXTENDED OCR MODULE	234
12.1	Job Manager Steps	234
12.2	IHQC Module	235

12.2.1	IHQC properties	235
12.2.2	Command Line Interface	236
12.3	Image Requirements	237
13	BARCODE SUPPORT	238
13.1	Barcode Formats	238
14	HIGH AVAILABILITY	239
14.1	Background information	239
14.2	Configuring the job on host one:	239
14.3	Configuring the job on host two:	240
14.4	Configuring the scheduler	241
15	TOPDF (BCL EASYPDF)	242
15.1.1	Overview of the Different Conversions Methods	242
15.1.1.1	MS Office Native Conversion	242
15.1.1.2	MS Office Direct Print	242
15.1.1.3	MS Office Extended Print	242
15.1.2	Comparison of the features provided by the different methods	243
15.1.2.1	In-depth Look at the Configuration file	244
15.1.3	Mappings Between Configuration File and the Conversion Methods	245
16	DISTRIBUTED POLLING	247
16.1	Job Setup	248
16.1.1	Distributed Polling Job	249
16.1.2	Autobahn Job	250
17	CLOUD OCR	251
17.1	Microsoft Computer Vision	255
17.1.1	Pricing	257
17.2	Google Cloud Vision	257
17.2.1	Pricing	257
18	KINGFISHER JOB STEP	258
19	GDPICTURE STEPS	259
19.1	PDFA Validation	259
19.2	Linearize PDF	259
19.3	Convert Any File To PDF (GdPicture)	260
19.4	Combine Any File To PDF	261
19.5	Combine PDFs	261

19.6	PDF To JPEG	261
19.7	PDF To PNG	261
19.8	PDF To TIFF (GdPicture)	261
19.9	PDF To Text	261
19.10	PDF To Searchable PDF (GdPicture)	261
19.11	PDF Portfolio	263
19.12	Convert to PDF/A	263
19.13	Compression	263
19.14	Detect Signatures	263
19.15	Smart Redaction	263
19.16	Key Value Pair Extraction	263
19.17	Pattern Redaction	263
19.18	Pattern Highlighting	263
19.19	Split PDF (GdPicture)	263
19.20	Split by Barcode	263
19.21	Pattern Enumeration	263
19.22	Get Document Information	263
19.23	Convert PDF To Office	264
19.24	Convert Any File To Office	264
20	PDF RECOGNITION TO JSON JOB STEP	265
21	MULTICORE SUPPORT	266
21.1	Using multiple cores in a job	266
21.2	Maximum number of cores allowed	266
21.3	Multicore Throttling	267
21.4	Command Line Interface	267
22	FILE NAME LENGTH	268
22.1.1	File Name Length - UNC Paths	268
22.1.2	File Name Length – Autobahn DX Working Folder Paths	268
22.1.3	Long File Names Support	268
22.1.3.1	Autobahn DX Config	269
22.1.3.2	Windows Configuration	269

23	AUTOBAHN DX CONFIGURATION FILES	270
23.1	Autobahn.config	270
23.1.1	The Input Delay	270
23.1.2	Multi Core Throttle Settings	270
23.1.3	Output Trigger Files	272
23.1.4	The Merge Flag Feature	272
23.1.5	Allow Long File Names	273
23.2	tiffjunction.exe.config	274
23.3	pj.exe.config	275
23.4	Other configuration files	276
24	AUTOBAHN DX DIRECTORIES	277
25	ACKNOWLEDGMENTS	278

1 Product Overview and Concepts

Our company has rebranded and aligned our vision for the future. With these changes, we have changed our naming and branding of products to have a cohesive suite of offerings. That means that Autobahn DX is getting a new coat of paint, transforming into the **Nutrient Document Automation Server (DAS)**.

This current release has minimal changes, only graphical differences, but future versions will be more impactful. We will give guidance for the upgrade to the future version to make the transition to DAS as smooth as possible.

Both the **Release Notes** and **Reference Guide** will not longer be updated as a file. Instead, you can find all documentation online here:

<https://www.nutrient.io/guides/document-automation-server/document-conversion/>

1.1 Overview

Autobahn DX is a Document Processing product designed to fit into an organization's document workflow.

Autobahn DX provides high performance automated OCR and Conversion of a variety of different input document formats including TIFF images, PDF Files, Microsoft Office documents and HTML pages.

The Server editions expand on this capability by providing the ability to create multi-step jobs – combining different steps together – that can be scheduled automatically to run.

This can provide a “hot folder” capability in Windows File System, SharePoint (Online/Office365/On-Premises) or Azure Blob storage.

The output files can then be saved to Windows File System, SharePoint (Online/Office365/On-Premises) or Azure Blob storage.

1.1.1 Edition Comparison

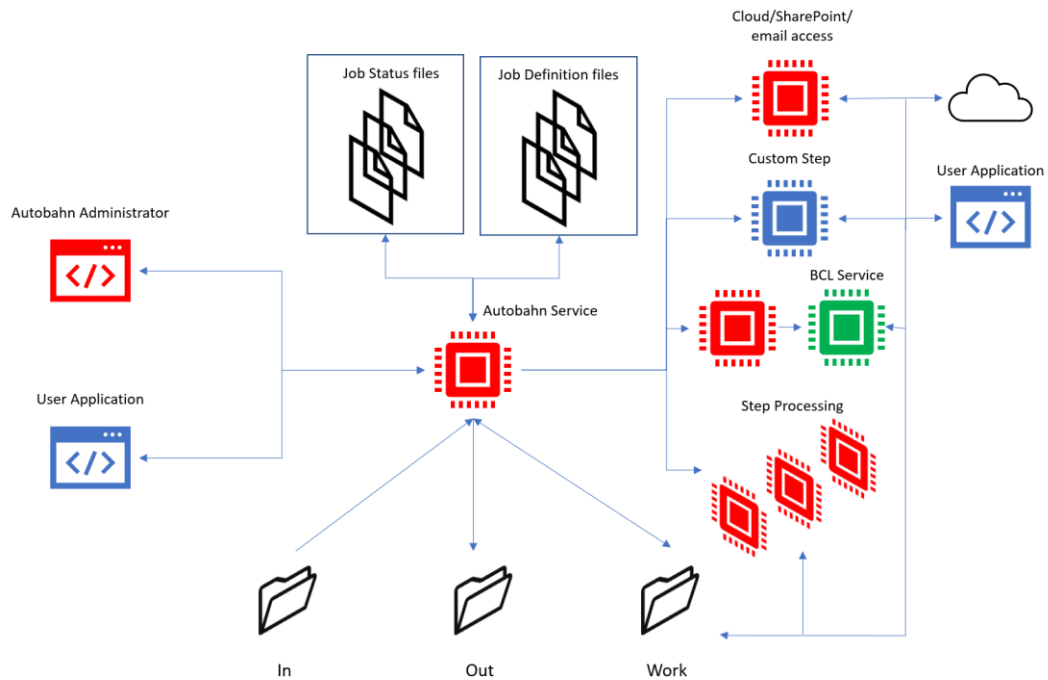
The Table below shows you the difference between the different editions of Autobahn DX

Edition Comparison	Workstation	Server with Standard OCR	Server with Extended OCR
Convert TIFF, JPEG, BMP, PNG, GIF to PDF + OCR	✓	✓	✓
Convert MS Office & Open Files to PDF	✓	✓	✓
OCR Image Only PDF	✓	✓	✓
Merge TIFF, JPEG, BMP, PNG, GIF to TIFF/PDF	✓	✓	✓
Split/Merge PDF Files	✓	✓	✓
Extract Text from PDF Files	✓	✓	✓
Set PDF Properties	✓	✓	✓
Pre-Processing (Deskew, Despeckle, Line Removal) and Auto-Rotation	✓	✓	✓
OCR Language support	23	23	129

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	✓	✓	✓
Handwriting recognition via Google and Microsoft Cloud APIs		✓	✓
Load Balancing/ Distributed Polling across Multiple Autobahn DX Instances		✓	✓
Pause and Restart Jobs		✓	✓
Azure Storage Support		✓	✓
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 and Basic or OAUTH2 authentication		✓	✓
Use of Custom Scripts		✓	✓
High Availability		✓	✓
Integration with Aquaforest Kingfisher		✓	✓
XML-based Job Definitions		✓	✓
.NET API		✓	✓
Run Jobs via Command Line		✓	✓
GDPicture OCR engine		✓	✓
IRIS Extended OCR engine			✓
Improved recognition of poorer quality documents			✓
Support for multiple languages within a single document from the same character set			✓
Asian Language Support			✓
Arabic Language Support			✓
Intelligent High-Quality Compression)			✓

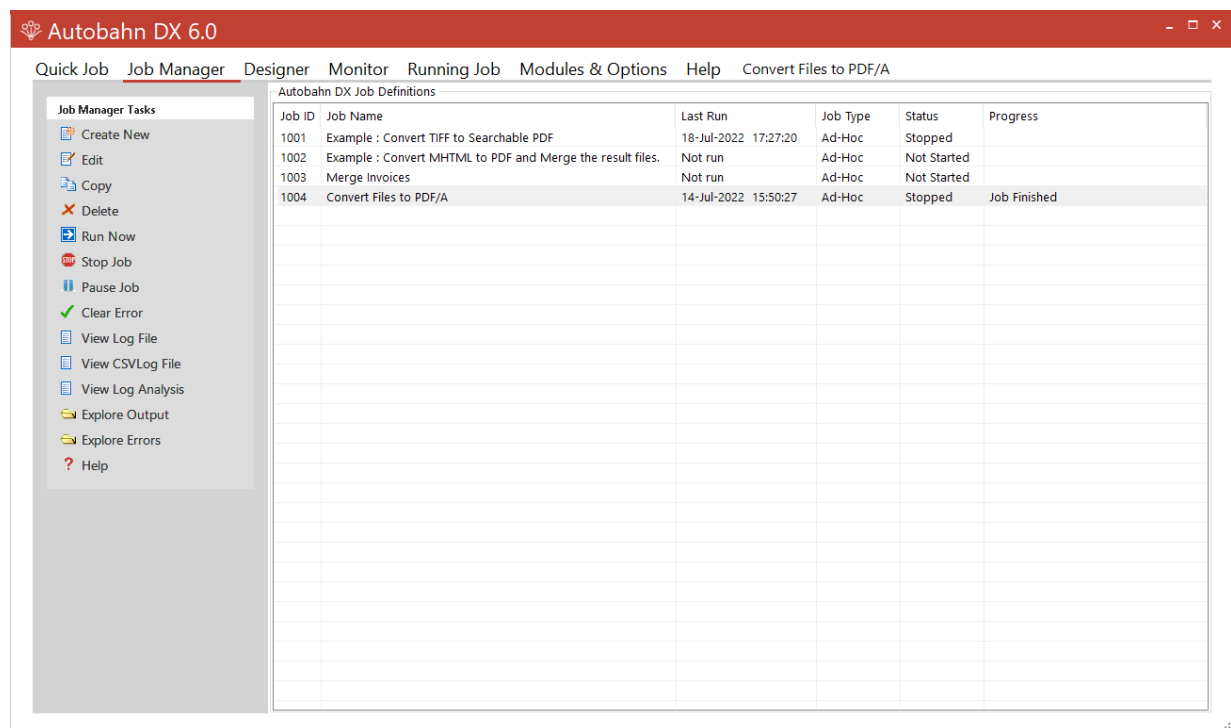
Multiple document output formats: PDF, DOCX, WORDML, RTF, CSV, XLSX, EXCELML, TXT, HTML and XPS			✓
---	--	--	---

1.2 Autobahn DX Architecture and Concepts



1.2.1 Autobahn Administrator

This is a Windows application that provides the primary administration interface.



This application has the following sections.

1.2.1.1 QuickJob

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

1.2.1.2 Job Manager

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

1.2.1.3 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.

1.2.1.4 Monitor

This shows the current service status and displays the Job Status of currently executing jobs. It also allows the service to be stopped or started.

1.2.1.5 Running Job

This displays the logging for the currently selected job. [Multi-core](#) jobs will display logging after the completion of each step.

1.2.1.6 Modules & Options

Entry of the license key, display of licensed options and the email settings

1.2.1.7 Help

Display various help options and advice.

1.2.2 User Application

User applications can communicate with Autobahn either through the .NET API or the Command line.

The .NET API allows a user application to create and execute ad-hoc jobs. See the [AUTOBAHN DX .NET API](#) section for more details.

The Command Line interface can run any job that could be run within the Quick Job section of the product. Multi-step jobs can be created by making consecutive calls to the Command Line Interface. See the [Autobahn Command Line Interface for more details](#).

1.2.3 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 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.

1.2.4 Job Definition Files

Each Job Definition file contains the settings for the job, including where it looks for input files, where output files are saved, intermediate work folders, log files, scheduling information plus settings for each Job Step.

A Job Definition contains one or more Job Steps, each one contains the settings for that step. The Step Types include:

- OCR
- Conversion
- Splitting and Merging
- Barcodes

- PDF operations
- Advanced

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

1.2.4.1 Debug mode

Most step types have an option for setting Debug Mode to Yes.

This will log additional information about the job step, including settings passed to sub-processes and intermediate results. This can be useful for the user to debug problem steps or their interaction with particular files and locations.

When Debug is set to Yes, it will also not delete any temporary files and folders created by that step. This needs to be monitored as this can cause large quantities of files to be left in the temporary location, especially if processing large numbers of files while in Debug mode.

1.2.5 Document Folders

Each job will have a set of directories containing the source documents (In), output documents (Out) and Work directories (including temporary, error, log, and work step folders).

1.2.6 Document Processing

Each step in a Job involves a separate Job Element process being spawned by the Autobahn Service. Details of each step are stored as elements of the Job Definition file. See [Job Definition XML files and Step Types](#) for more details.

1.2.7 Job Status

All jobs have an associated status file. This contains the name, state, progress, log file name, CSV log file name, last run time and cores.

1.2.8 Autobahn DX .NET Job API

1.2.9 BCL EasyPDF service

The BCL EasyPDF service is used for the conversion of some file types, it uses an installation of the relevant product for the input file type for the conversion. and that the product has been set up for the service's login user.

See Convert Any File and File Access Permissions

1.2.10 File Access Permissions

1.2.10.1 Quick Jobs

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 so may process files on remote file systems may make use of either UNC's or mapped drives visible to that user.

1.2.10.2 Job Manager (Ad-Hoc or Scheduled)

Conversely, Jobs in the Job Manager are run by the Autobahn DX windows service (and in some cases the BCL easyPDF 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 respective service, so it is recommended that the Service user is changed. See [Autobahn and BCL easyPDF Windows Services](#) for more details.

1.2.11 What is the definition of a Core?

The number of cores licensed determines how many CPU cores the software can use concurrently. As a rough guide, the software can process 1,000 pages per hour per CPU core although this will vary according to various factors.

Autobahn DX Server can be licensed as single core, four cores or multiples of four cores. The maximum number of cores that can be used by Autobahn DX on an installation is limited to the smaller of:

- the number of cores licensed
- the number of logical processors (not physical cores) in the CPU.

2 Installation and Licensing

2.1 System Requirements

Supported Operating Systems	Windows 10 (64 bit) Windows 11 Windows Server 2012 R2 (64 bit) Windows Server 2016 Windows Server 2019 Windows Server 2022
Supported Document Stores	SharePoint 2013 SharePoint 2016 SharePoint 2019 SharePoint Online (Office 365) OneDrive for Business Azure File Storage Azure Blob Storage Windows File Systems
Disk Space	2.0 GB
.NET Framework	4.7.2
Visual C++ Redistributables	Visual C++ 2017 (x64) Visual C++ 2017 (x86) https://go.microsoft.com/fwlink/?LinkId=746571
Additional tools	SharePoint Server Client Components SDK (x64)

2.2 Product Licensing

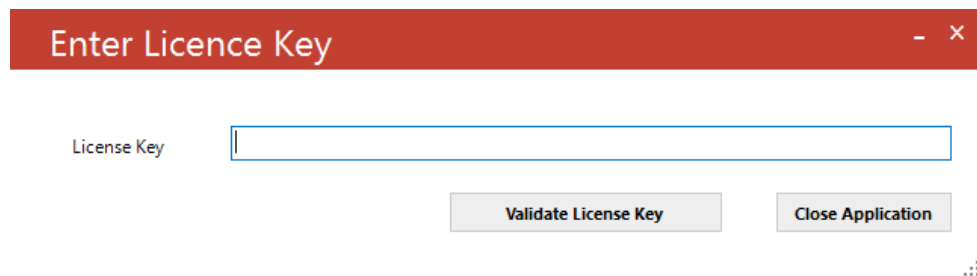
Autobahn DX has three license levels which are described below:

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.

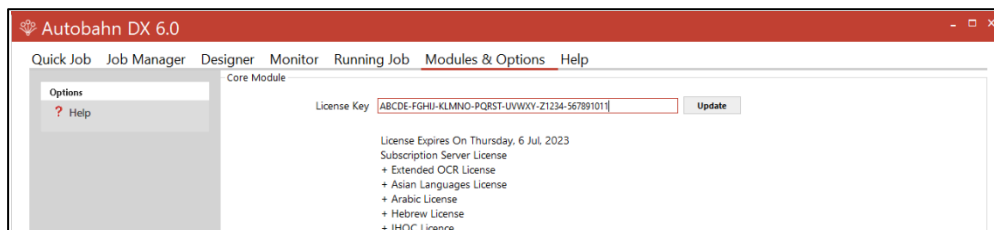
Server and Extended OCR are available with single and four cores, plus additional blocks of four cores can be added to the four core licenses. See [What is the definition of a Core?](#)

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 Autobahn DX Windows Service

2.4.1 Autobahn and BCL easyPDF Windows Services

These are the heart of the Server editions of the product and control the execution of both scheduled jobs and ad-hoc jobs submitted via the Autobahn Manager, command line or the Autobahn Job API.

Once you have installed Autobahn, you need to take a few easy steps described below to configure the Windows Service setting for both Autobahn and the BCL easy PDF services.

2.4.1.1 Create and initialize an Autobahn DX user account:

If this is the first time that Autobahn has been installed on the machine, then a suitable Autobahn service account with the required permissions is required.

- Log-on as an Administrator on the machine
- Create a new user account that will run the Autobahn and BCL easyPDF services. You can have two separate users or use the same one. Subject to your computer policy it is suggested that the account's password is set to "Never Expire".
- Assign the required permissions to the user.

2.4.1.2 Log on as the Autobahn Service account

YOU MUST log on as the Autobahn service account at least once after each installation of Autobahn DX.

This is required to initialize the account, including printer driver setting, or the easyPDF SDK service will not run properly. You will also need to open and initialize all applications used for the conversions: i.e. Word, Excel etc.

If Autobahn DX is installed again (for example as part of an upgrade), it is recommended that this step is repeated.

2.5 Upgrading from previous versions

2.5.1 Uninstalling the previous version

Two versions of Autobahn cannot be installed at the same time. To install a new version of Autobahn, the previous version must first be uninstalled. Before starting the uninstall, first stop the Autobahn DX service and close the UI. The '**...Autobahn DX/jobdef**' folder contains the definitions for all jobs. We advise saving this folder in a separate location before upgrading.

2.5.2 Previous jobs and license keys

Certain files will be left in the Autobahn DX install location after the uninstall. Both the jobdef folder containing the descriptions for all the previous jobs and the license key file will be preserved.

2.5.2.1 Job definitions

Job definitions are now automatically upgraded to the latest version of Autobahn DX on opening the UI. This does not require any action from the user. The jobs before upgrading will be saved in the '**...Autobahn DX/jobdef_copy**' folder.

If you are upgrading to the latest version from a version older than v5.0, we would recommend recreating your jobs instead of relying on the job upgrade.

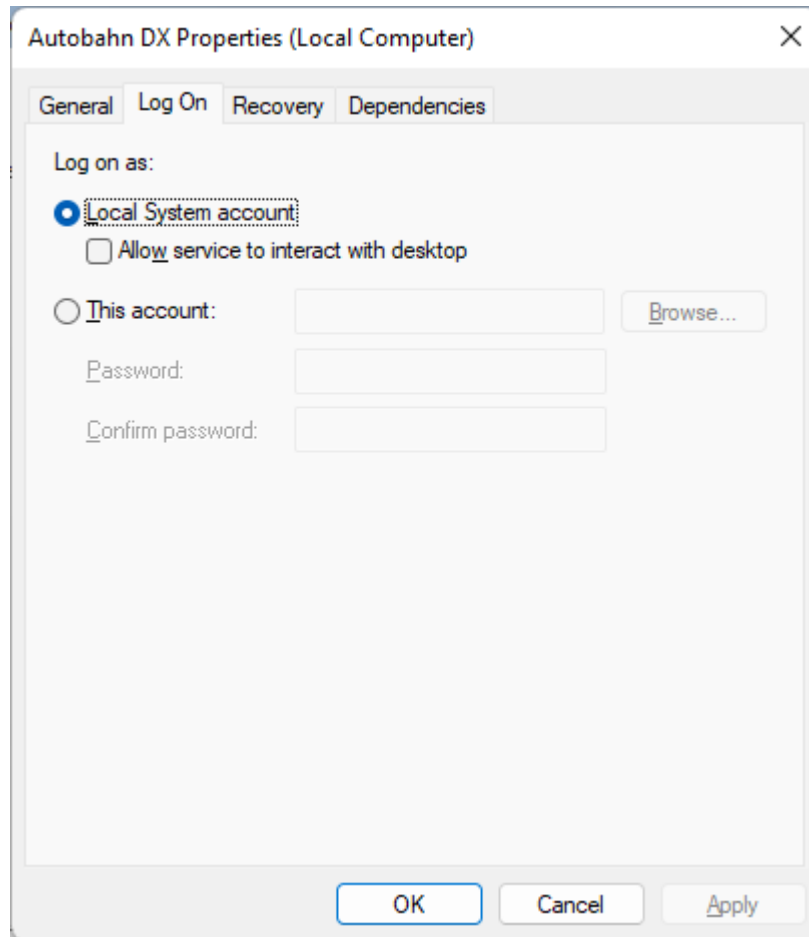
2.5.2.2 License Key

The license key will remain from your previous version. For upgrades between minor releases, the license key should still be valid for the newly installed version.

If you are on a new major version (5.0, 5.5, 6.0), then the license key will no longer be valid. In this case, you should contact support for a new license key.

2.5.2.3 Configure Windows Service setting

- Log-on to the computer as an Administrator on the machine.
- From the Search box on the Task Bar, type Services and select the Services app.
- Select and right-click on the Autobahn DX service and select Properties.
- Click the Log On tab.



- Select "This account" and either type the username and password for the service user or select it via the Browse button and enter the password.
- Click OK to close the property dialog box and return to the main Services window.
- Right click on the Autobahn DX service and select Restart. The service will restart with the new settings.
- Close the Services window.
- Repeat the above for the BCL easyPDF Service

2.6 Support for Multiple CPU Cores

Your Autobahn DX Server Edition license key will support a specific number of CPU cores.

If only one core is licensed, the core count is not displayed on the Modules & Options tab.

License Expires On Thursday, 6 Jul, 2023

Subscription Server License

- + Extended OCR License
- + Asian Languages License
- + Arabic License
- + Hebrew License
- + IHQC Licence

The licensed cores are shown for multi-core licenses.

License Expires On Friday, 2 Nov, 2029

Server License

- + Extended OCR License
- + Asian Languages License
- + Arabic License
- + Hebrew License
- + IHQC Licence
- + Multi-core License With 8 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="10" />
<add key="jobqueueinterval" value="10" />
```

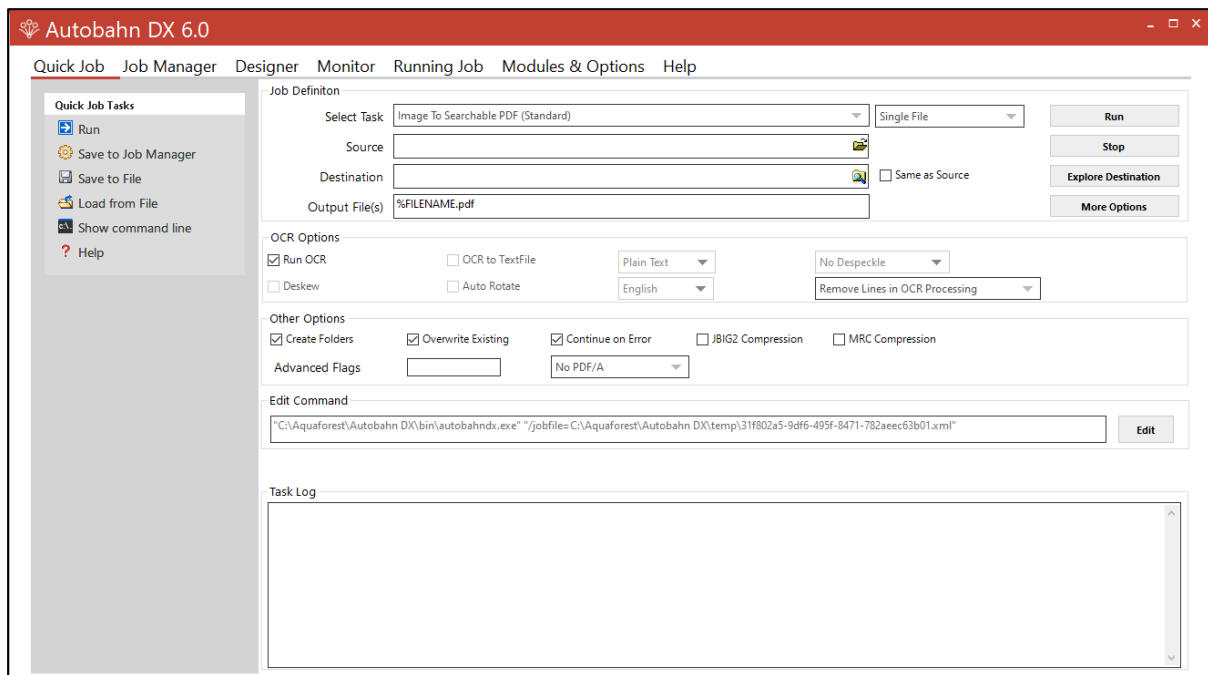
When creating and scheduling jobs, it is recommended that core use is considered. Jobs that need to be run at the same time should not exceed the core count. It might be more efficient to schedule the jobs so that they run in series (at different times).

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.

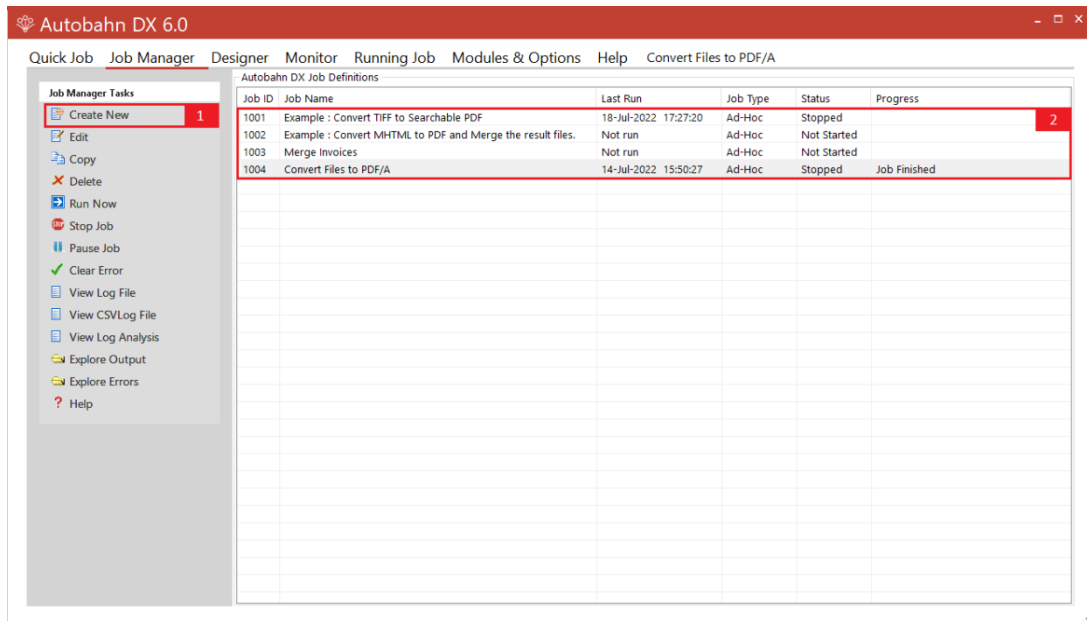
3.1 The Quick Job Tab

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.

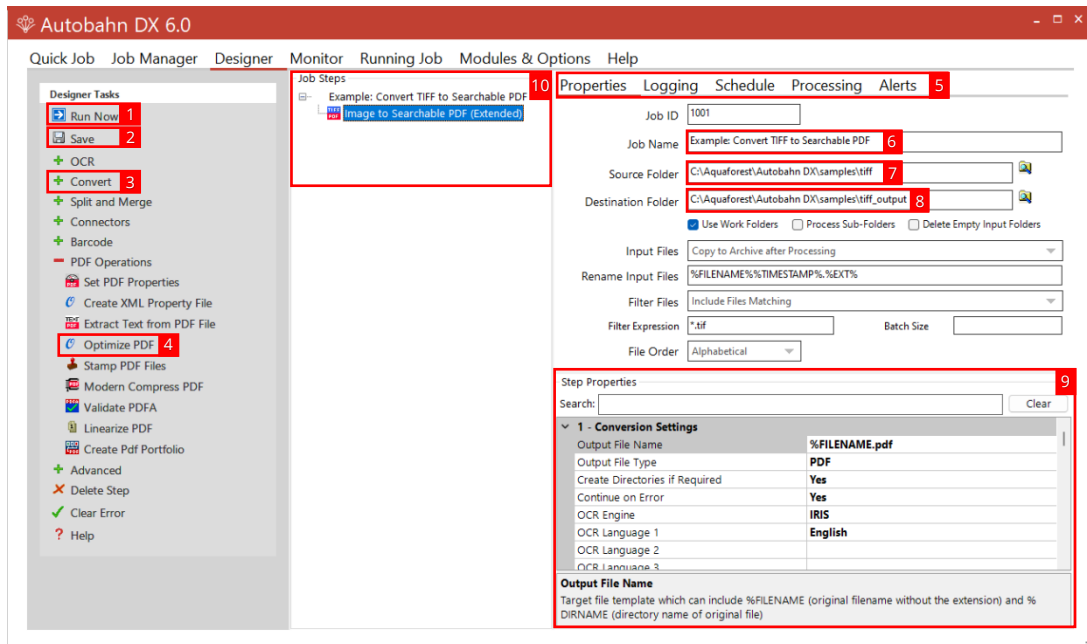
1. Click on **"Create New"** to create your own Job.
2. We provided two sample jobs with sample files. This is to show you how to set the jobs up. You can edit these jobs to suit your needs.

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

3.3 The Designer Tab

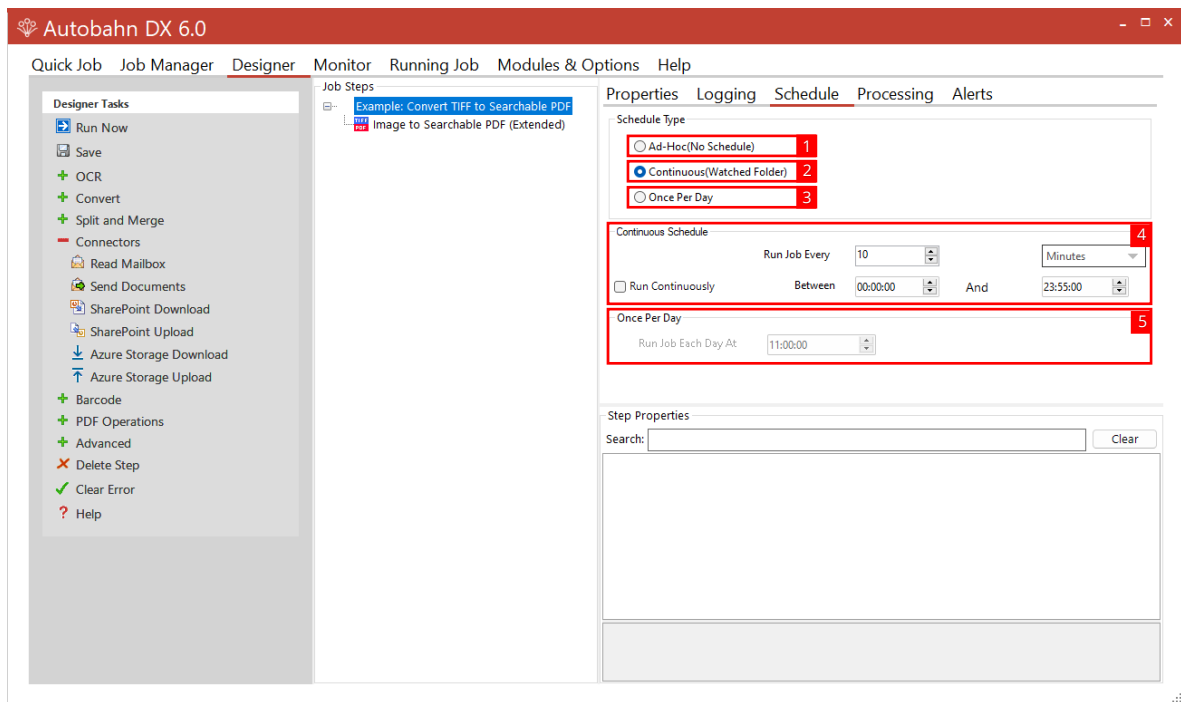
This section will give you a brief overview of the job designer and how jobs are created.

3.3.1 General Job Properties



1. Click the **"Run Now"** Button to save and run the displayed job.
2. Click the **"Save"** Button to save changes made to the displayed 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. A search bar allows you to find specific properties quickly.
10. You can select a step in the **"Job Steps"** list to edit the steps, change its position and delete the step.

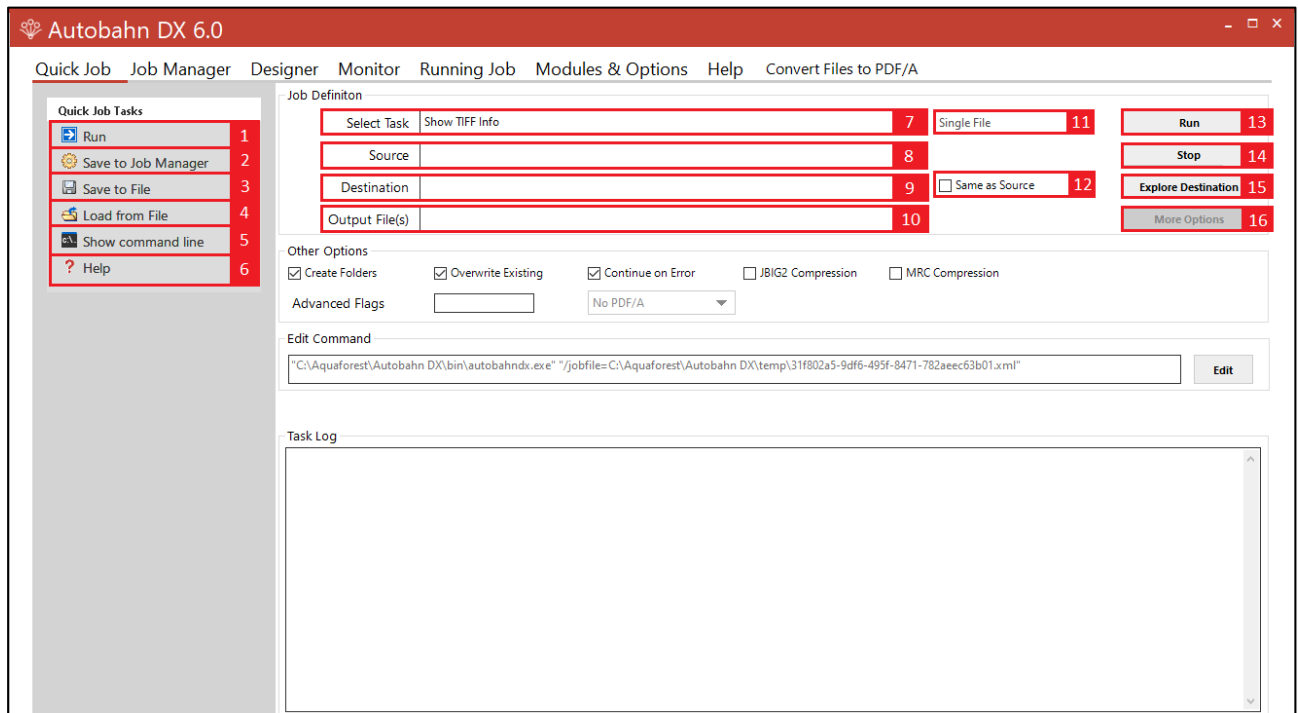
3.3.2 Scheduling a job



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):** Target file template, which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of the original file).
- 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 a conversion to TIFF.• Native - This method places the OCRed 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.</p> <p>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 is used in 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 OCRed 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	<p>Metadata from the original PDF are copied to the searchable result PDF.</p>

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> <ul style="list-style-type: none">• Split into single pages• Split by ranges (See below)• Split by repeating ranges (See below)				
Page Ranges	<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><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	<p>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.</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>				

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 TIFFJunction advanced command-line flags may be entered here (see section 3 of the TIFFJunction reference guide in the distribution/tj/docs folder)

4.1.6 Barcode Options

Screen Field / Button	Description
Output File(s) With Barcode	<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(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.
Barcode Operation	<p>Split By Barcode</p> <ul style="list-style-type: none"> • Choose this option to split the TIFF/PDF file by Barcode <p>Rename By Barcode</p> <ul style="list-style-type: none"> • Choose this option to rename the TIFF/PDF file based on Barcode
Barcode Formats	<p>Various reading options for barcode types. Other barcode types exist and can be set manually. Ask support@aquaforest.com for more details.</p> <p>Codabar – Default, All, All 1D Code128 – Default, All, All 1D Code25 – Default, All, All 1D Code39 – Default, All, All 1D EAN13 – Default, All, All 1D EAN8 – Default, All, All 1D QRCode – All 2D UPCA – Default, All, All 1D Selected – Only searches for the barcode types selected below</p>
Split Mode	<p>Various Options for splitting Files by Barcode</p> <ul style="list-style-type: none"> • Barcode on First Page • Barcode on Last Page • Remove Barcode Page
Metadata Name	The name of the metadata field you want to add to the output file
Metadata Value	The value of the metadata field you want to add to the output file.
Deskew Mode	<p>Controls how the toolkit handles documents that are skewed or rotated.</p> <p>0 – Compatibility mode, no skew, aligned 1 – No skew, barcode must be aligned with page 2 – Skew, aligned 3 – Skew, aligned, two different skew angles can be read 4 – Skew, can be unaligned, three skew angles allowed 5 – Skew, unaligned, four skew angles allowed</p>

Screen Field / Button	Description
	6 – Skew, unaligned, five or more skew angles allowed
Skew Tolerance	<p>Only has effect if Deskew Mode is set to 0.</p> <p>Controls the maximum angle from horizontal or vertical at which a barcode will be recognized.</p> <p>0 – up to 5 degrees 1 – 13 degrees 2 – 21 degrees 3 – 29 degrees 4 – 37 degrees 5 – 45 degrees</p>
Noise Reduction	Runs the image through a noise reduction filter before scanning for barcodes, removing marks that are unlikely to be part of a barcode. A larger value removes larger marks but may destroy barcode information. A typical value for noise reduction is 10.
Despeckle	If noise reduction is nonzero, setting this property will remove white speckles from inside barcodes, and remove black marks in the spaces between bars.
Error Correction	Attempts to correct Code 39 barcodes if able.
Max Length	Max length of the barcode output.
Min Length	Min length of the barcode output.
Min Separation	Minimum distance between two barcodes with the same value and alignment to be considered separate values.
Multiple Read	Read the entire page for barcode values. If false, the reader will stop after the first barcode is found.
Page No	Specifies the page to read the barcode from. A value of 0 will scan every page for barcodes.
Pattern	A regular expression to compare barcodes to. Only barcodes that match this pattern will be returned.
Pref Occurrence	As the barcode reader scans, it assigns scores to barcode reader candidates. After the scan, only candidates above this score are reported.
Quiet Zone Size	On a scan line, the barcode reader ignores parts of the line not preceded by a number of white pixels specified by QuietZoneSize. A value of 0 calculates a zone 1/10th of the image resolution.
Report Unread Barcodes	<p>A mask that controls the reporting of unsuccessfully decoded barcodes. Has the following limitations:</p> <ul style="list-style-type: none"> • Maximum of 1 unread barcode per page • No unread barcodes reported on pages with a barcode successfully read • Linear Barcodes must contain at least 40 black bars • 2D barcodes only reported if they reach and fail the error correction stage of decoding • Any large series of parallel lines will be reported as an unread linear barcode

Screen Field / Button	Description
	<ul style="list-style-type: none"> All unread barcodes are reported as type UNREAD with value UNREAD The appropriate 2D type must be enabled to detect unread barcodes of that type If linear barcode types are disabled, barcodes of that type may still be reported as unread <p>Mask values 1 = Linear Barcodes 2 = Datamatrix 4 = QR-Code 8 = PDF-417</p>
Rotate by 45 if no barcode	Reader will rotate by 45° and rescan if no barcode found.
Scan Direction	<p>A mask that controls all the directions the reader will scan for barcodes.</p> <p>Mask values 1 = Left to Right 2 = Top to Bottom 4 = Right to Left 8 = Bottom to Top</p>
Show Check Digit	When true, the barcode check digit is included in the returned value, if the barcode type contains a check digit.
Timeout	The maximum time in milliseconds that the barcode reader will allow for scanning a page in the document.
Use Over-Sampling	When true, the reader samples 3 lines at a time, taking the average pixel value. Useful against image speckles.
Left, Top, Width, Height	Sets the rectangle zone that will be searched for a barcode. If a value is not set, the whole page will be scanned.
DPI	DPI of Tiff image generated from PDF files. These images are then scanned for barcodes.
BPP	BPP of Tiff image generated from PDF files. The images are then scanned for barcodes.
TIFF Compression	Type of compression used when converting PDF input to Tiff, before scanning for barcodes
Advanced Flags	This enables advanced flags to be used, even if there is no direct equivalent in autobahndx.exe.

4.1.7 Task Log Section

Screen Field / Button	Description
Command	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.
Task Output	The output of the job appears here.

4.2 Security Settings

PDF files can have built in security settings, this Autobahn DX tab allows access to secured source PDFs with the same password, and set the passwords on output PDFs as well as set the target document permissions.

Document Options

Security Metadata Document Options PDF Conversion Settings

Source Document Security

Password 1

Target Document Security

Encryption Strength 2

Owner Password 3

User Password 4

Target Document Permissions 5

☐ Allow All Permissions ☐ Printing ☐ Copy

☐ Fill ☐ Assembly ☐ Modify Contents

☐ Modify Annotations ☐ Screen Readers ☐ Degraded printing

OK Cancel

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. **Owner Password:** A PDF owner password is a password used to set certain document restrictions on the target PDF.
4. **User Password:** A password that will be required to open the target document.
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 modifications
 - 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 tab allows setting the metadata for 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 Layout

Single Page

Page Mode

Neither Bookmarks nor Thumbnails Open

Non-Full Screen Mode

Neither 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 Orientation

Default

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

4.4.3 Page Settings and Margins

Note: For "... Any File to PDF" steps only

Option	Description
Width	The width of the output PDF
Height	The height of the output PDF
Left	The left margin of the output PDF
Right	The right margin of the output PDF
Top	The top margin of the output PDF
Bottom	The bottom margin of the output PDF
Paper Orientation	The orientation of the output PDF


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 timeout 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 Hyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
Merge PDF Attachments	Set this flag to true if you want converted pdf attachments to be merged to the output pdf file. Otherwise, the converted files will be merged back to the pdf.
Print BG Color	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
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.
Print All Sheets	The flag that indicates whether to print all Excel worksheets or not.
Convert Bookmarks	For MS Word, convert bookmarks
Frame Slides	Indicate whether to draw a frame around the border of the slides. This setting is for PowerPoint only.
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>
Output Type	<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
Color Type	Use this property to set PowerPoint to print with either color, grayscale, or black and white.

Option	Description
Handout Order	<p>Sets the handout order, this flag only applies to PowerPoint jobs. The possible values are:</p> <ul style="list-style-type: none"> • Vertical First • Horizontal First
Print Graphics	<p>Sets the graphics setting for printing.</p> <ul style="list-style-type: none"> • Print Full Resolution • Print Low Resolution • Print Graphics
Zoom	<p>Sets the printing zoom of the worksheet.</p> <p>The allowed value range is from 10 to 400.</p>
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.

4.5.3 Image and Font Conversion Settings

Option	Description
Image Compression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
Font Embedding	The option PRN_FONT_EMBED_FULLSET (embedding a full set of fonts) 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 fonts) 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.
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.

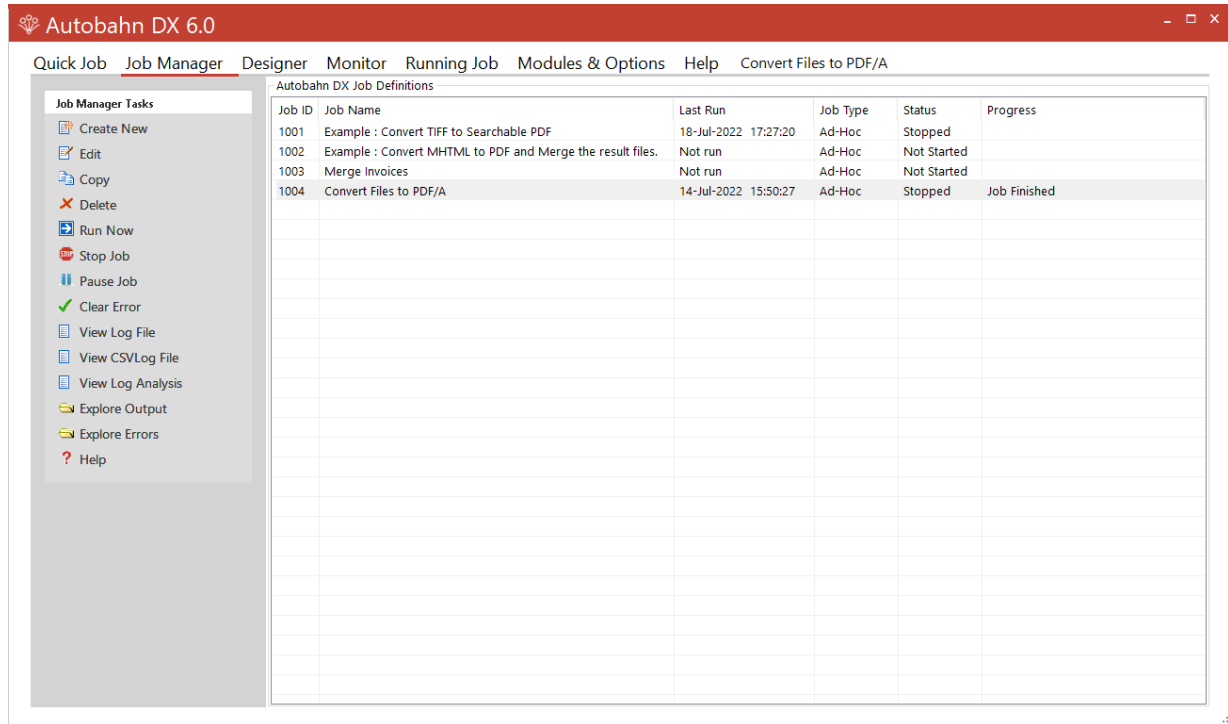
Option	Description
-B 100	<p>Blank Page Removal</p> <p>This option can be used when converting TIFF files to Searchable PDFs. A value should be provided which specifies the pixel threshold to be used to determine whether a page is blank or not. If a page is deemed to be blank, it is omitted from the output file. A suggested value is 100 i.e., using the -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 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

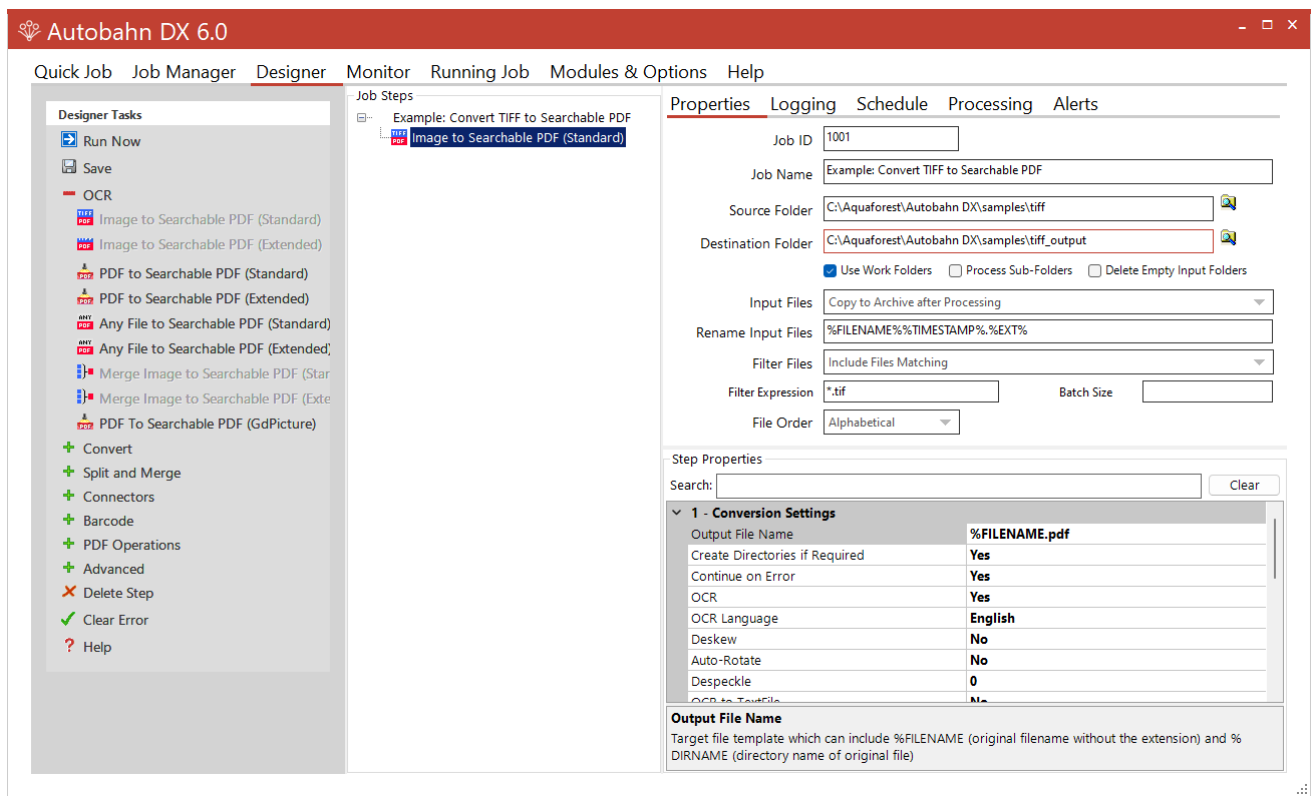
This section of the application displays a list of all current Job Definitions and the set of operations to be performed shown in the menu table below.



Menu Item	Action
Create New	Creates a new, skeleton job definition from jobtmpl.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 gray out the invalid steps. The step 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) • PDF to Searchable PDF (GdPicture)
Convert	<ul style="list-style-type: none"> • Convert PDF to TIFF • Convert Any File to PDF • Convert PDF to PDFA • Convert Any File to PDF (GdPicture) • Combine Any File to PDF • PDF to JPEG • PDF to PNG • PDF to TIFF • PDF to Text • Convert PDF to Office • Convert Any File to Office
Split and Merge	<ul style="list-style-type: none"> • Merge PDF • Split PDF • Merge TIFF • Split TIFF • Combine PDFs • Split PDF (GdPicture)
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"> • Set PDF Properties • Create XML Property File • Extract Text from PDF File • Optimize PDF • Stamp PDF Files • Modern Compress PDF • Validate PDFA

	<ul style="list-style-type: none"> • Linearize PDF • Create PDF Portfolio • Get Document Information
Advanced	<ul style="list-style-type: none"> • Custom Script Step • High Availability • Kingfisher Job • Distributed Polling • PDF Recognition to JSON • Image to Searchable PDF (Microsoft Cloud) • PDF to Searchable PDF (Microsoft Cloud) • Image to Searchable PDF (Google Cloud) • PDF to Searchable PDF (Google Cloud) • Detect Signatures • Key Value Pair Extraction • Smart Redaction • Pattern Redaction • Pattern Highlight • Pattern Enumeration
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	Takes you to the 'Help' tab, which has links to many useful blogs, documents, and other resources. It also has contacts if you need help from our support or sales team.

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.
Destination Folder	The folder where the processed files will be placed if "Move input files to target folder after processing" is chosen.
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.
Process Sub-Folders	If checked, all sub-folders will be recursively processed.
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>
Rename Input Files	<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> <p>You can also use %EMAILNAME% for files named in the email format. This will rename the file to it's original name.</p>

Menu Item	Description												
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> <p>N.B.: Work Folders must be used to enable the use of filters.</p>	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.												
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</p> <p>*.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".												

Menu Item	Description
File Order	<p>The order that the files will be processed. There are UTC and local time variants of the date options, totaling nine options:</p> <ul style="list-style-type: none"> • Alphabetically, • Created Date (Ascending) • Created Date (Descending) • Modified Date (Ascending) • Modified Date (Descending). <p>Note: this setting does not work for "Merge Image to PDF..." steps, the merge and OCR must be done in two separate job steps.</p>
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. • Page counts (not all steps generate page counts and dependent on configuration setting)
Retention Period	<p>This is an integer value representing the number of days the log file will be kept for before being deleted.</p> <p>Leaving it blank or setting it to a number less than one will keep the log files indefinitely.</p>
Max Size	Set the maximum log file size. If a log file is created above this size, it will be split into smaller segments.
Stop Processing on Error	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.
Skip Long File Names	Check this box to make Autobahn DX skip files with long filenames. If this box is not checked, Autobahn DX will throw an error if it encounters one of these files.
Skip Folders That Autobahn Can't Access	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.
Archive Folder	The folder where the processed files will be placed if "Move to archive after processing" is chosen.
Work Folder	The folder where files will be temporarily stored during conversion and processing.
Error Folder	Source documents that have errors during processing will be placed in the specified folder.

Menu Item	Description
Temp Folder	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.
Trigger File	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.

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 software interface with five tabs: Properties, Logging, Schedule, Processing, and Alerts. The 'Schedule' tab is selected and highlighted with a red underline. Below the tabs, there are three main sections for scheduling configuration:

- 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 menu set to 'Hours'. Below this is a checkbox labeled 'Run Continuously' which is checked. To the right of the checkbox are two time fields: 'Between 00:00:01' and 'And 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 you 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 prematurely. 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 send any email until it processes 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.

5.2.4.1 Alert variables

When sending emails, there are several variables that can be used to customize the alerts you send out, these variables are enclosed by 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

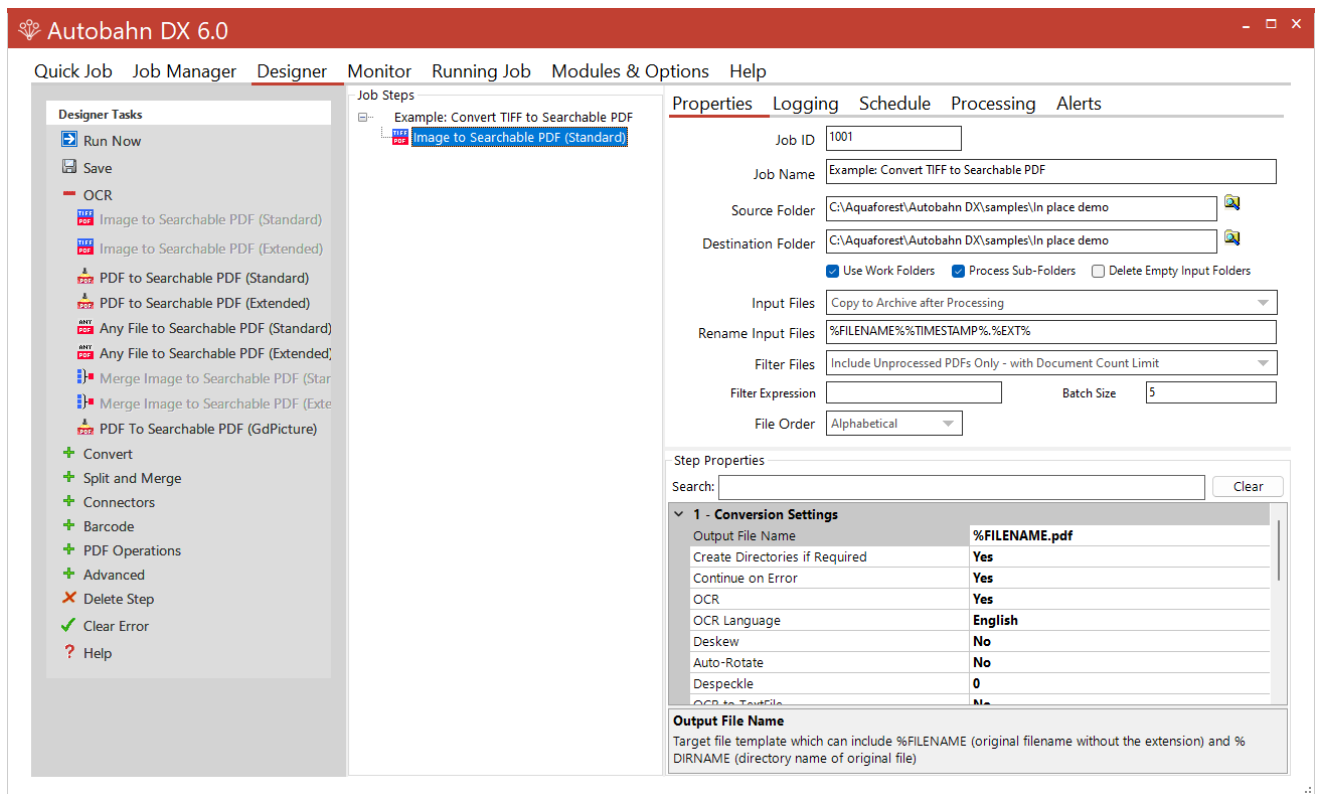
Autobahn DX is designed as a Workflow product where there is an input folder and an output folder. At the end of the process, there are options to copy, delete or move the input files that have been successfully processed.

With “in-place” processing, the input documents are turned into searchable PDFs and returned to the same location. It is possible to replace the existing file if the output file format produces the same file name. The input files can be copied to an archive location if they need to be kept (this is recommended during the development process and during testing – if this is not set, the original file cannot be recovered).

Autobahn DX can be used for in-place processing, but we have an OCR product named [Aquaforest Searchlight](#) that is designed specifically for in-place conversions to searchable PDFs, it may handle this Use-Case more effectively. Searchlight records all the files it processes, so is more efficient when there are a lot of files, as they do not need to be opened to be identified as previously processed.

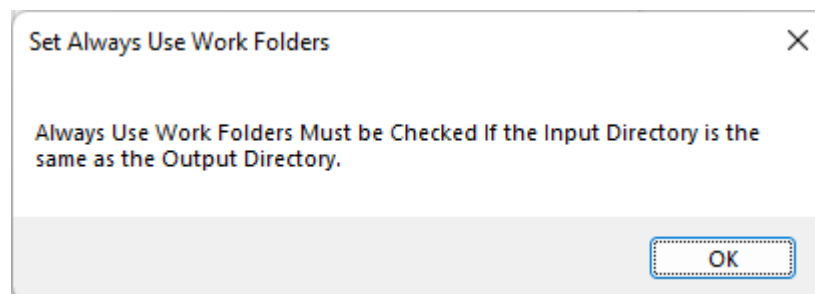
5.2.5.1 Example In-Place Job Setup

The job shown below will convert PDFs under the tree “C:\ADX Demo\Documents” to searchable PDFs, processing up to 5 files each time the job is run.



The **Source Folder** and the **Target Folder** must be the same.

The **Use Work folders** check box must be checked when processing in place. A message will be displayed when the folders are set to the same location in the UI and the check box set automatically.



Select the **Process Sub-Folders** check box.

For Audit Purposes, the **Input Files** option should be set to **Copy to archive after Processing**.

To avoid re-processing files, select the **Include Unprocessed PDFs Only - with Document Count Limit** option in the **Filter Files** combo box.

Because the **Filter Files** option selected includes the Document Count Limit, the **Batch Size** of the job can be set to 5 files per run (You can increase this to a suitable batch size).

The **Output file Name** is set in the Conversion Settings for the step and should be configured to **%FILENAME.pdf** so that it will replace the input file

5.2.6 Step Types

This section explains each of the step types.

Autobahn DX Server edition is licensed to use Standard and GdPicture steps. The Extended edition adds the Extended OCR steps

Step group	Step name
OCR	Image to Searchable PDF (Standard)
OCR	Image to Searchable PDF (Extended)
OCR	PDF to Searchable PDF (Standard)
OCR	PDF to Searchable PDF (Extended)
OCR	Any File to Searchable PDF (Standard)
OCR	Any File to Searchable PDF (Extended)
OCR	Merge Image to Searchable PDF (Standard)
OCR	Merge Image to Searchable PDF (Extended)
OCR	PDF To Searchable PDF (GdPicture)
Convert	Convert PDF to TIFF
Convert	Convert Any File to PDF
Convert	Convert PDF to PDF/A
Convert	Convert Any File To PDF (GdPicture)
Convert	Combine Any File To PDF
Convert	PDF To JPEG
Convert	PDF To PNG
Convert	PDF To TIFF
Convert	PDF To Text
Convert	Convert PDF To Office
Convert	Convert Any File To Office
Split and Merge	Merge PDF
Split and Merge	Split PDF
Split and Merge	Merge TIFF, JPEG, BMP, PNG, GIF
Split and Merge	Split TIFF
Split and Merge	Combine PDFs
Split and Merge	Split PDF (GdPicture)
Connectors	Read Mailbox
Connectors	Send Documents
Connectors	SharePoint Download
Connectors	SharePoint Upload
Connectors	Azure Storage Download
Connectors	Azure Storage Upload
Barcode	Barcode TIFF/PDF
Barcode	Split by Barcode
PDF Operations	Set PDF Properties
PDF Operations	Create XML Property File
PDF Operations	Extract Text from PDF File
PDF Operations	Optimize PDF
PDF Operations	Stamp PDF Files

Step group	Step name
PDF Operations	Modern Compress PDF
PDF Operations	Validate PDF/A
PDF Operations	Linearize PDF
PDF Operations	Create Pdf Portfolio
PDF Operation	Get Document Information
Advanced	Custom Script Step
Advanced	High Availability
Advanced	Kingfisher Job
Advanced	Distributed Polling
Advanced	PDF Recognition to JSON
Advanced	Image to Searchable PDF (Microsoft Cloud OCR)
Advanced	PDF to Searchable PDF (Microsoft Cloud OCR)
Advanced	Image to Searchable PDF (Google Cloud OCR)
Advanced	PDF to Searchable PDF (Google Cloud OCR)
Advanced	Detect Signatures
Advanced	Smart Redaction
Advanced	Key Value Pair Extraction
Advanced	Pattern Redaction
Advanced	Pattern Highlighting
Advanced	Pattern Enumeration

5.2.6.1 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.

This step is not available for the GdPicture engine; however, it can be replicated by using a combination of the Convert Any File To PDF (GdPicture) and PDF To Searchable PDF (GdPicture) steps

5.2.6.1.1 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• MS Word (rtf)• HTML
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• 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.

Parameter	Notes
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 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.
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	<p>This specifies the number of parallel files you want to be processed at a given time.</p> <p>Note: This needs a multi-core license and the number of cores used will depend on the availability of cores.</p>
Debug	Set this to true to execute the step in debug mode.

5.2.6.1.2 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	<p>One or more of the following, separated by commas if more than one is required.</p> <ul style="list-style-type: none"> • CSV * • DOCX • EPUB • EXCELML * • HTM • OPENTXT • PDF • RTF • TXT • WORDML • XLSX * • XPS <p>*These output formats are suitable for table-oriented pages that can be mapped onto a spreadsheet format.</p>
Create Folders If Required	Create an output folder if it does not exist. Default true.
OCR Engine	The OCR engine to use. This must be set to use the IRIS engine.
OCR Language 1-8	You can set up to eight different languages for OCR recognition on one page, only if they are in the same character set. English is available as a language
Automatic language detection	<p>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.</p> <p>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.</p>
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 define 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

Parameter	Notes
	<p>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	<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>
Layout	<p>The layout for the docx or rtf document</p> <ul style="list-style-type: none"> • Standard • Flow
PDFVersion	<p>This determines the PDF version of the generated PDF:</p> <ul style="list-style-type: none"> • 1.4 • 1.5 • 1.6 • 1.7 • 1.7 Extension Level 3 • 1.7 Extension Level 5 • 1.7 Extension Level 8 • PDF/A-1a • PDF/A-1b • PDF/A-2a • PDF/A-2b • PDF/A-3a • PDF/A-3b
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).

Parameter	Notes
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	Level 1 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 make the result darker).
Contrast	The contrast (lower values will make the result darker).
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 grayscale 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.

Parameter	Notes
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 Convert To TIFF set to Yes
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

Parameter	Notes
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

5.2.6.2 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.

5.2.6.2.1 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 layers that already exist 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.
Convert to TIFF	Choose the method for PDF image extraction. <ul style="list-style-type: none"> • No – (Native) • Yes – (Convert to TIFF)
DPI	When OCRing 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 Convert To TIFF=Yes .
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 • 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 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.
Advanced Flags	Command line flags to be passed through to the underlying executable.
Maximum Cores	<p>This specifies the number of parallel files you want to be processed at a given time.</p> <p>Note: This needs a multi-core license and the number of cores used will depend on the availability of cores.</p>
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 step in debug mode.

5.2.6.2.2 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	<p>One or more of the following, separated by commas if more than one is required.</p> <ul style="list-style-type: none">• CSV *• DOCX• EPUB• EXCELML *• HTM• OPENTXT• PDF• RTF• TXT• WORDML• XLSX *• XPS <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 1-8	You can set up to eight different languages for OCR recognition in one page as long as they are in the same character set.
Automatic language detection	<p>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.</p> <p>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.</p>
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 define advanced deskew properties.

Parameter	Notes
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	<p>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.</p> <p>Note: This Will Only work if: "Convert to TIFF = Yes"</p>
Layout	<p>The layout for the docx or rtf document</p> <ul style="list-style-type: none"> • Standard • Flow

Parameter	Notes
PDFVersion	<p>This determines the PDF version of the generated PDF:</p> <ul style="list-style-type: none"> • 1.4 • 1.5 • 1.6 • 1.7 • 1.7 Extension Level 3 • 1.7 Extension Level 5 • 1.7 Extension Level 8 • PDF/A-1a • PDF/A-1b • PDF/A-2a • PDF/A-2b • PDF/A-3a • PDF/A-3b <p>Note: This will only work if: "Extract Images Method = Convert to TIFF"</p>
Extract Images Method	<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'd 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	Level 1 is the basic compression level while level 3 is the most advanced Intelligent High-Quality Compression Mode.

Parameter	Notes
IHQC Quality Factor	The quality Factor for IHQC
Binarization	Whether or not to perform binarization on the document.
Brightness	The brightness (higher values will make the result darker).
Contrast	The contrast (lower values will make the result darker).
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 grayscale 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.

Parameter	Notes
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 Convert To TIFF set to Yes
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

5.2.6.3 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.

5.2.6.3.1 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/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• PDF/A3-b
Validate PDF/A	Whether or not to validate the PDF/A document after conversion

5.2.6.4 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).

Parameter	Notes
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 fonts) 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 fonts) 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	Sets paper orientation to <ul style="list-style-type: none"> • Default (Maintain Source Orientation) • Landscape • Portrait
PDF Compliance	Allows the User To choose PDF/A or PDF/X Compliant 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 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.

Parameter	Notes
	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 WordExtensionEX 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 merge 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 • 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.

Parameter	Notes
	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	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 step in debug mode.

5.2.6.5 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 %PAGEn (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

Parameter	Notes
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.

5.2.6.6 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.
Job ID	(optional) Will send an additional flag with the jobdef file location. e.g. A value of 1024 will give the flag <code>"/jobdef:C:\Aquaforest\Autobahn DX\jobdef\1024.xml"</code> given that Autobahn is installed on the default C drive location.

5.2.6.7 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.• 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

Parameter	Notes
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. Using * or leaving it blank will process all pages.
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
Bates Suffix	Specifies the suffix of the Bates stamp
Bates 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 step in debug mode.

5.2.6.7.1 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

5.2.6.8 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 step in debug mode.

5.2.6.9 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	Single Pages – Splits the file into single pages Page Ranges – Splits the file based on the range Repeated Ranges – Splits the file based on the range and the repeated range Bookmarks – Splits the file based on the original bookmarks
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 step in debug mode.

5.2.6.10 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 step in debug mode.

5.2.6.11 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 %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.
Split Type	Single Pages – Splits the file into single pages Page Ranges – Splits the file based on the range Repeated Ranges – Splits the file based on the range and the repeated range
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 step in debug mode.

5.2.6.12 Read Inbox

This can read mailboxes and extract attachments using IMAP4 or OAuth2 (Modern) Authentication, 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 for IMAP4:

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

For OAuth2, you require an access token from the Microsoft Identity Platform, which will supply you with the credentials to use our email steps with Modern Authentication.

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

Example: *file1@20210330@john.smith@aquaforest.com@.pdf*

Parameter	Notes
Authentication Mode	Choose between IMAP and Modern Authentication
IMAP Server	The IMAP server address e.g. imap.company.co.uk
Require Authentication	If anonymous authentication is set up on your server, a username and password is not needed when setting this option to 'No'
Username	The username for the account to access the IMAP server.
Password	Password for the account. This is held encrypted.
Azure Client ID	The Client ID for OAuth2 Authentication
Azure Tenant	The Tenant for OAuth2 Authentication
Azure AD Instance	The address of the Azure AD Instance e.g. https://login.microsoftonline.com
Credential Type	Select the credential type for OAuth2 Authentication. The options are Client Secret or Certification.
Client Secret	The Client secret generated by Azure
Certificate Path	The path to the certificate generated by Azure
Certificate Password	The password of the certificate generated by Azure
Source Email Account	The email account to be read e.g. pdf@company.com
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.
Output Template	The template for the name of the output file. This can include %FILENAME% for the original filename, %TIMESTAMP% for the

Parameter	Notes
	job timestamp, and %FROMADDRESS% for the 'From' email address.
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
Subject Filter	Autobahn will only download attachments from email with the subject filter in their subject.
Debug	Set this to true to execute the step in debug mode.

5.2.6.13 Send Documents

Use of this step type requires a Server License.

Attachment limit is 50MB but email provider's limits are normally lower.

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.

Example: *file1@20210330@john.smith@aquaforest.com@.pdf*

Parameter	Notes
Authentication Mode	Choose between SMTP and Modern Authentication
Domain	The sending domain e.g., aquaforest.com
SMTP Server	SMTP Server address e.g., smtp.aquaforest.com
Require Authentication	If anonymous authentication is set up on your server, a username and password is not needed when setting this option to 'No'
Username	The username for the account to access the SMTP server.
Password	Password for the account. This is held encrypted.
Azure Client ID	The Client ID for OAuth2 Authentication
Azure Tenant	The Tenant for OAuth2 Authentication
Azure AD Instance	The address of the Azure AD Instance e.g., https://login.microsoftonline.com
Credential Type	Select the credential type for OAuth2 Authentication. The options are Client Secret or Certification.
Client Secret	The Client secret generated by Azure
Certificate Path	The path to the certificate generated by Azure
Certificate Password	The password of the certificate generated by Azure
Sender Name	Name of the sending User e.g., John
From Email Address	Sending user e.g., admin@aquaforest.com
CC Addresses	Email list of CC'd email addresses. Separate addresses with a comma. e.g., admin@aquaforest.com, admin2@aquaforest.com
BCC Addresses	Email list of Bcc'd email addresses. Separate addresses with a comma. e.g., admin@aquaforest.com, admin2@aquaforest.com
Email Title	The title of the Email
Email Body	The body of the Email
Allow Multiple Attachments	By default, Autobahn sends files as individual emails. If set to 'Yes' Autobahn will try to group files by destination and send multiple files in one email.

Parameter	Notes
Attachment Number Limit	Setting this number limits the number of files that can be attached to one email sent by Autobahn.
Attachment Total Size Limit	In MB. This value limits the total size of all the files sent in each individual email by Autobahn.
Use Original Filename	Input filenames must fit a specific format. Select true if you want the final attachment to revert to its original name.
Debug	Set this to true to execute the step in debug mode.

5.2.6.14 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 step in debug mode.

5.2.6.15 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.] Note, there is a GDPicture based step (PDF to Text).

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 step in debug mode.

5.2.6.16 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	Set to "Yes" to see more processing information on the console.
Continue on Error	Continue processing if an error occurs.

5.2.6.17 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 files that match 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.

5.2.6.18 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 storage account 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.

5.2.6.19 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 storage account 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 Local Folder	Upload documents from folders and subfolders of the local folder
Replace Invalid Characters With	A pattern to replace any invalid character Windows File Storage in the file name before downloading. Invalid characters are: " * : < > ? \ Default replacement pattern is: _
Debug	If set to "Yes" the user will see more processing information on the console.

5.2.6.20 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 step in debug mode.

5.2.6.21 Optimize PDF

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

Parameter	Notes
Linearize – Fast Web View	Set 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.

	<ul style="list-style-type: none"> • Move to Error Folder • Copy to Error Folder
Debug	Set this to true to execute the step in debug mode.

5.2.6.22 OCR Any File to PDF

This step attempts to convert all files to searchable PDFs, Autobahn DX may have the following OCR engines.

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

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

5.2.6.22.1 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	Overwrites 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 of 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 step 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 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	<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 layers that already exist 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.</p>
Convert to TIFF	<p>Choose the method for PDF image extraction.</p> <ul style="list-style-type: none"> • No - (Native) • Yes - (Convert to TIFF)
DPI	<p>When OCRing 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.</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	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 Convert To TIFF='Yes' .
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 • 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.
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 fonts) 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 fonts) 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 PDF/A or PDF/X Compliant 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 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 WordExtensionEX 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 merge 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.</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>

5.2.6.22.2 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.
Overwrite Existing	Overwrites 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 of 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 step 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> <ul style="list-style-type: none"> • CSV * • DOCX • EPUB • EXCELML * • HTM • OPENTXT • PDF • RTF • TXT • WORDML • XLSX * • XPS <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 1-8	You can set up to eight different languages for OCR recognition in one page, as long as they are in the same character set.
Automatic language detection	<p>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.</p> <p>If at least one language has been detected, recognition will be performed in the first language candidate that has been</p>

	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 define 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

	<p>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>
Layout	<p>The layout for the docx or rtf document</p> <ul style="list-style-type: none"> • Standard • Flow
PDF Version	<p>This determines the PDF version of the generated PDF:</p> <ul style="list-style-type: none"> • 1.4 • 1.5 • 1.6 • 1.7 • 1.7 Extension Level 3 • 1.7 Extension Level 5 • 1.7 Extension Level 8 • PDF/A-1a • PDF/A-1b • PDF/A-2a • PDF/A-2b • PDF/A-3a • PDF/A-3b
Remove Blank Page	<p>Set this to true to remove blank pages from Tiff or PDF documents. Value needs to be set for sensitivity (see below).</p>
Sensitivity	<p>The sensitivity, from 1 to 100. With high sensitivity, fewer blank pages are detected.</p>
Work Depth	<p>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.</p>
JPEG Quality	<p>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.</p>
JPEG2000 Compression	<p>Enable/Disable JPEG2000 Compression.</p>
JPEG2000 Compression Mode	<p>The JPEG2000 Compression Mode to use.</p>

JPEG2000 Compression Value	The Value to set for the selected Compression Mode.
IHQC Compression	Apply Intelligent High-Quality Compression
IHQC Compression Level	Level 1 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 make the result darker).
Contrast	The contrast (lower values will make the result darker).
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 grayscale 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	<p>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'.</p> <p>Note: This property only applies when processing image files or when processing PDF files with the Convert To TIFF set to Yes</p>
Keep Deskewed Image	<p>Set this to true if you want to use the deskewed image in the output document.</p> <p>Note: This property only applies when Keep Original Image is set to No</p>
Keep Despeckled Image	<p>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.</p> <p>Note: This property only applies when Keep Original Image is set to No</p>
Keep Dark Border Removal	<p>Set this to true if you want to use the image after dark borders have been removed, in the output document.</p> <p>Note: This property only applies when Keep Original Image is set to No</p>
Keep Punch Hole Removal	<p>Set this to true if you want to use the image after punch holes have been removed, in the output document.</p> <p>Note: This property only applies when Keep Original Image is set to No</p>
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	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

	<p>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 fonts) 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 fonts) 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 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 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	<p>Determines whether embedded and linked files will be preserved during conversion. Default value: False (disabled).</p> <p>Note: This will work with WordExtensionEX 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 merge 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.</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

	<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)	<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>

5.2.6.23 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.

	<ul style="list-style-type: none"> • Rename by Barcode: Choose this option to rename the TIFF/PDF file based on Barcode.
Split Mode	<p>Various Options for splitting Files by Barcode</p> <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	<p>Overwrites any file that exists with the same name in the output folder.</p> <p>Note: If you have the same barcode in different pages or files, they will be overwritten if this is set to true.</p>
Metadata Name	<p>Choose the Metadata field you want to set the 'Metadata Value' for. The named fields below will have the value added to them when set.</p> <ul style="list-style-type: none"> • Author • Creator • Keywords • Producer • Subject • Title • Trapped <p>Any other entry will be used as the name for a new custom metadata item.</p>
Metadata Value	<p>Enter a value for the Metadata Value. You can use the following 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 <p>Note: 'Trapped' metadata only accepts either 'True', 'False' or 'Unknown' as a value</p>
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 step in debug mode.
-------	---

5.2.6.24 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 (Controller Replica)
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

5.2.6.25 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 shared folder 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.

5.2.6.26 Kingfisher Job

This step allows a Kingfisher job to be integrated as an Autobahn step. See the [Kingfisher Job Step Section](#) for more details.

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

5.2.6.27 PDF To PDF/A Job

This step uses GdPicture libraries to convert a PDF document to a PDF/A format.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
PDF/A Output Type	Select the type of PDF/A to output. The selection is: PDF/A-1a, PDF/A-1b, PDF/A-2a, PDF/A-2b, PDF/A-2u, PDF/A-3a, PDF/A-3b, PDF/A-3u, PDF/A-4, PDF/A-4e, PDF/A-4f
Allow Vectorization	If set to false, the job will attempt to create the PDF/A files without Vectorization
Allow Rasterization	If set to false, the job will attempt to create the PDF/A files without Rasterization
Debug	Select true if you want to see more debug output.

5.2.6.28 PDF Recognition to JSON Job

This step extracts important data from PDF files in the form of key/value pairs. Users can define their expected keys and easily retrieve the data from those fields. No templates are needed.

Screen Field / Button	Description
Output Expected Key JSON	Creates a JSON file of expected key-values as output.
Output Expected Key Values By Page JSON	Creates a JSON file of expected key-values by page as output.
Output PDF Data Pages Text	Creates a .txt file of the pdf data by page.
Output PDF Data Page Details	Creates a .txt file of key + bounding box, Values + Bounding Boxes by page

Output PDF Data Pages As CSV	Creates a CSV containing page number, key, key bounding box, value, value bounding box, page number, page dimensions
Output PDF Data Pages As JSON	Creates a CSV containing page number, key, key bounding box, value, value bounding box, page number, page dimensions
List PDF Data Pages As JSON	If true, the results of 'Output PDF Date Pages as JSON' will be included in the logging
Date Format	Set to input date format.
Use Currency Symbols	Set to false if you want symbols and strings to be removed before returning currency values.
Page Limit	Maximum number of pages to be processed.
Page Range	A string representation of the page numbers you want to process. e.g., 1,3-4.
Current Culture	Choose the expected format of date times if ambiguous e.g., 03/07/12
Expected Keys File Paths	File paths of the text files containing expected keys. (use ' ' to separate multiple paths)
Ignore Case Expected Keys	Choose if Casing is ignored when comparing recognition values to the Expected Keys set.
Custom Keys File Paths	File path of the text files containing custom keys. (use ' ' to separate multiple paths)
Ignore Case Custom Keys	Choose if Casing is ignored when comparing recognition values to the Custom Keys set.
Custom Keys Default File Path	The default file path of the text file containing custom keys. (use ' ' to separate multiple paths)
Load Default Custom Keys	Set to true if you want custom keys to be taken from the default path.
Skip Line Width	This value will be multiplied by page width and any line with its width below this calculated value will NOT be skipped.
Skip Line Word Count	Do not skip line if the number of words in the line is less than this value.
Skip Line Word Space	Any line with an average space greater than this value will NOT be skipped.
Ignore Don't Skip Space	The only time special chunks are broken into smaller chunks is if the space between two adjacent words in the chunk is greater than this value.
Chunk Break Space	Any chunk that has two adjacent words with a space between them greater than this value will be chunked.
Chunk Break Minimum	If the average space of words in a chunk is smaller than this value, 'Chunk break space' will be used to break the chunk instead of this value.
Chunk Header Font Size	Any chunk with an average font size below this value will not be considered as a header candidate.

Chunk Break Space Header	Any header chunk that has two adjacent words with a space between them greater than this value will be chunked.
Break Words By Delimiter	Switch this to true to break words by any of the Chunk Delimiters available (wordDelimiter, chunkDelimiter and chunkSpaceDelimiter).
Word Delimiter	Enter one delimiter per index. If any series of characters match this pattern, we will break the word on that index.
Chunk Delimiter	Enter one delimiter per line. If any word ends with any of these delimiters, they will be broken into chunks.
Chunk Space Delimiter	Enter one delimiter per line.
Max Horizontal Space	Skip analyzing key/value chunks that have a horizontal space greater than this value (points) between them.
Max Vertical Space	Skip analyzing key/value chunks that have a vertical space greater than this value (points) between them.
Data Types To Split	Choose the data types that the Chunker will attempt to split into smaller chunks.
Data Types To Check	Choose the data types that will not be split once identified.
Data Types To Remove	Choose the unwanted data types that will be removed in post processing.
Error On No Expected Keys	When set to 'Yes', a file that does not contain any values for expected keys will be considered an error.
Regex Dictionary Terms File Path	File path of a text file containing regex dictionary terms. (leave blank for default)
Plain Dictionary Terms File Path	File path of a text file containing plain dictionary terms. (leave blank for default)
Debug	Select true if you want to see more debug output.

5.2.6.29 Modern Compress PDF

This step uses GDPicture libraries to compress PDF documents with various options.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Remove Annotations	Select 'Yes' if you want to remove annotations.
Remove Blank Pages	Select 'Yes' if you want to remove blank pages.
Remove Bookmarks	Select 'Yes' if you want to remove bookmarks.
Remove Embedded Files	Select 'Yes' if you want to remove embedded files.
Remove Form Fields	Select 'Yes' if you want to remove form fields.
Remove Hyperlinks	Select 'Yes' if you want to remove hyperlinks.

Remove JavaScript	Select 'Yes' if you want to remove JavaScript.
Remove Metadata	Select 'Yes' if you want to remove metadata.
Remove Page Thumbnails	Select 'Yes' if you want to remove page thumbnails.
Pack Fonts	Select 'Yes' if you want to pack fonts. This greatly optimizes output file size by focusing on fonts.
Pack Documents	Select 'Yes' if you want to pack document content before saving.
Recompress Images	Select 'Yes' if you want to recompress images.
Enable MRC	Select 'Yes' if you want to enable MRC.
Downscale Resolution MRC	Set the downscale resolution of the MRC compression. The default value is 100.
Preserve Smoothing	Select 'Yes' if you want to preserve smoothing.
Image Quality	Choose which Image Quality the output files will be. The default value is Medium.
Downscale Images	Select 'Yes' if you want to downscale images.
Downscale Resolution	Set the downscale resolution of the compression. The default value is 150.
Enable Color Detection	Select 'Yes' if you want to enable automatic color detection.
Enable Char Repair	Select 'Yes' if you want to enable character repair.
Enable JPEG2000	Select 'Yes' if you want to enable JPEG2000.
Enable JBIG2	Select 'Yes' if you want to enable JBIG2.
JBIG2 PMS Threshold	Set the threshold of the JBIG2 pattern matching and substitution. The default value is 0.85.
Debug	Select true if you want to see more debug output.

5.2.6.30 Validate PDF/A

This step uses GdPicture libraries to validate if the input PDF document conforms to the selected PDF/A version.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
PDF/A Validation Type	Choose which PDF/A version the files will be validated against.
Report Location	Target folder to save reports for files that failed to validate. The location must already exist, or the report will not save.

Debug	Select true if you want to see more debug output.
-------	---

5.2.6.31 Linearize PDF

This step uses GdPicture libraries to optimize PDFs for web-viewing, rendering the document one page at a time.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Pack Document	Select 'Yes' if you want the document to be packed before it is saved, reducing its size.
Enable Compression	Select 'Yes' if you want to enable compression on the output pdf.
Debug	Select true if you want to see more debug output.

5.2.6.32 Convert Any File To PDF (GdPicture)

This step uses GdPicture libraries to convert a large variety of file types to PDF. This step does not require an Office installation to process Office files.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Author	Set the Author metadata field in the output PDF. This can include %FILENAME% (original filename without the extension) or %DIRNAME% (directory name of original file)
Title	Set the Title metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Subject	Set the Subject metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Keywords	Set the Keywords metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Producer	Set the Producer metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%

Metadata	Set the Metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Convert Email Attachments	Select 'Yes' if you want to convert email attachments to PDF.
Attach Email Attachments To Pdf	Select 'Yes' if you want to attach the email attachments to the output PDF. If set to 'No', the files will be merged to the PDF if they have been converted to PDF, otherwise they will be removed.
Email Page Height	Specifies the page height, in points, of the resulting document when converting from the source Email file.
Email Page Width	Specifies the page width, in points, of the resulting document when converting from the source Email file.
Email Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Email file.
Email Prefer One Page	Select 'Yes' if you want the email to be converted to a single page PDF if possible.
Enable ICC	Specifies if the converter shall favor preserving the ICC profile, if present in the loaded document, during the conversion.
Html Emulation Type	Specifies a type of a media to emulate.
Html Page Height	Specifies the page height, in points, of the resulting document when converting from the source Html file.
Html Page Width	Specifies the page width, in points, of the resulting document when converting from the source Html file.
Html Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Html file.
Html Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Html file.
Html Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Html file.
Html Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Html file.
Html Prefer CSS Page Size	Give any CSS page size declared in the page priority over what is declared in Html Page Width and Html Page Height. If set to false, the renderer will scale the content to fit the paper size.

Html Prefer One Page	Specifies whether the output document should contain a single page.														
Load Only First Page	Specifies that all executed actions with the loaded document will be processed using only the first page of the document.														
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.														
Pdf Bitonal Image Compression	<p>Sets the scheme to be used to compress bitonal image data when converting/saving the currently loaded document to PDF format.</p> <table> <tr> <th>ID</th><th>Scheme</th></tr> <tr> <td>0</td><td>None</td></tr> <tr> <td>1</td><td>Flate</td></tr> <tr> <td>2</td><td>CCITT4</td></tr> <tr> <td>3</td><td>JPEG</td></tr> <tr> <td>4</td><td>JBIG2</td></tr> <tr> <td>5</td><td>JPEG2000</td></tr> </table>	ID	Scheme	0	None	1	Flate	2	CCITT4	3	JPEG	4	JBIG2	5	JPEG2000
ID	Scheme														
0	None														
1	Flate														
2	CCITT4														
3	JPEG														
4	JBIG2														
5	JPEG2000														
JBIG2 PMS Threshold	Sets the threshold of the JBIG2 pattern matching and substitution. The default value is 0.85.														
Pdf Color Image Compression	Sets the scheme to be used to compress color image data when converting/saving the currently loaded document to PDF format.														
Pdf Enable Color Detection	Enables or disables the automatic color detection feature when converting/saving the currently loaded document to PDF format.														
Pdf Image Quality	Sets the level of quality used to compress images with a lossy compression scheme, which are embedded in the newly produced PDF document when converting/saving the currently loaded document to PDF format. It must be a value from 0 to 100. 0 means the worst quality and the best compression, 100 means the best quality and the worst compression.														
Pdf Use Deflate On JPEG	Specifies if the converter shall use additional Deflate compression for JPEG images in PDF output.														
Rasterization DPI	Sets the rendering resolution to be used when converting vector content to raster content, if any is included in the currently loaded document.														
Tiff Enable Exif Rotate	Specifies whether tiff encoder is using Exif rotate flag to handle page rotations.														
Timeout Milliseconds	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.														

Txt Font Bold	Specifies whether the font used for the resulting document when converting from the source txt file must have a bold style.
Txt Font Italic	Specifies whether the font used for the resulting document when converting from the source txt file must have an italic style.
Txt Font Family	Specifies the name of the font to be used for the resulting document when converting from the source txt file.
Txt Font Size	Specifies the text size, in points, to be used for the resulting document when converting from the source txt file.
Txt Horizontal Text Alignment	Specifies the horizontal text alignment of the resulting document when converting from the source txt file.
Txt Page Height	Specifies the page height, in points, of the resulting document when converting from the source Txt file.
Txt Page Width	Specifies the page width, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Txt file.
Debug	Select true if you want to see more debug output.

5.2.6.33 Combine Any File To PDF

This step uses GdPicture libraries to convert a large variety of file types to PDF, and then merges them to create a single output PDF. This step does not require an Office installation to process Office files.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %DIRNAME% (original directory name)
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Author	Set the Author metadata field in the output PDF. This can include %FILENAME% (original filename without the extension) or %DIRNAME% (directory name of original file)
Title	Set the Title metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%

Subject	Set the Subject metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Keywords	Set the Keywords metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Producer	Set the Producer metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Metadata	Set the Metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Convert Email Attachments	Select 'Yes' if you want to convert email attachments to PDF.
Attach Email Attachments To Pdf	Select 'Yes' if you want to attach the email attachments to the output PDF. If set to 'No', the files will be merged to the PDF if they have been converted to PDF, otherwise they will be removed.
Email Page Height	Specifies the page height, in points, of the resulting document when converting from the source Email file.
Email Page Width	Specifies the page width, in points, of the resulting document when converting from the source Email file.
Email Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Email file.
Email Prefer One Page	Select 'Yes' if you want the email to be converted to a single page PDF if possible.
Enable ICC	Specifies if the converter shall favor preserving the ICC profile, if present in the loaded document, during the conversion.
Html Emulation Type	Specifies a type of a media to emulate.
Html Page Height	Specifies the page height, in points, of the resulting document when converting from the source Html file.
Html Page Width	Specifies the page width, in points, of the resulting document when converting from the source Html file.
Html Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Html file.
Html Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Html file.
Html Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Html file.

Html Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Html file.														
Html Prefer CSS Page Size	Give any CSS page size declared in the page priority over what is declared in Html Page Width and Html Page Height. If set to false, the renderer will scale the content to fit the paper size.														
Html Prefer One Page	Specifies whether the output document should contain a single page.														
Load Only First Page	Specifies that all executed actions with the loaded document will be processed using only the first page of the document.														
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.														
Pdf Bitonal Image Compression	<p>Sets the scheme to be used to compress bitonal image data when converting/saving the currently loaded document to PDF format.</p> <table border="1"> <thead> <tr> <th>ID</th><th>Scheme</th></tr> </thead> <tbody> <tr> <td>0</td><td>None</td></tr> <tr> <td>1</td><td>Flate</td></tr> <tr> <td>2</td><td>CCITT4</td></tr> <tr> <td>3</td><td>JPEG</td></tr> <tr> <td>4</td><td>JBIG2</td></tr> <tr> <td>5</td><td>JPEG2000</td></tr> </tbody> </table>	ID	Scheme	0	None	1	Flate	2	CCITT4	3	JPEG	4	JBIG2	5	JPEG2000
ID	Scheme														
0	None														
1	Flate														
2	CCITT4														
3	JPEG														
4	JBIG2														
5	JPEG2000														
JBIG2 PMS Threshold	Sets the threshold of the JBIG2 pattern matching and substitution. The default value is 0.85.														
Pdf Color Image Compression	Sets the scheme to be used to compress color image data when converting/saving the currently loaded document to PDF format.														
Pdf Enable Color Detection	Enables or disables the automatic color detection feature when converting/saving the currently loaded document to PDF format.														
Pdf Image Quality	Sets the level of quality used to compress images with a lossy compression scheme, which are embedded in the newly produced PDF document when converting/saving the currently loaded document to PDF format. It must be a value from 0 to 100. 0 means the worst quality and the best compression, 100 means the best quality and the worst compression.														
Pdf Use Deflate On JPEG	Specifies if the converter shall use additional Deflate compression for JPEG images in PDF output.														
Rasterization DPI	Sets the rendering resolution to be used when converting vector content to raster content, if any is included in the currently loaded document.														

Tiff Enable Exif Rotate	Specifies whether tiff encoder is using Exif rotate flag to handle page rotations.
Timeout Milliseconds	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.
Txt Font Bold	Specifies whether the font used for the resulting document when converting from the source txt file must have a bold style.
Txt Font Italic	Specifies whether the font used for the resulting document when converting from the source txt file must have an italic style.
Txt Font Family	Specifies the name of the font to be used for the resulting document when converting from the source txt file.
Txt Font Size	Specifies the text size, in points, to be used for the resulting document when converting from the source txt file.
Txt Horizontal Text Alignment	Specifies the horizontal text alignment of the resulting document when converting from the source txt file.
Txt Page Height	Specifies the page height, in points, of the resulting document when converting from the source Txt file.
Txt Page Width	Specifies the page width, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Txt file.
Debug	Select true if you want to see more debug output.

5.2.6.34 Combine PDFs

This step uses GDPicture libraries to convert a large variety of file types to PDF, and then merges them to create a single output PDF. This step does not require an Office installation to process Office files.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %DIRNAME (original directory name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Enable Numerical Ordering	When enabled, documents will be merged in numerical order <i>e.g. file1, file3, file11, file20, file101</i> . Otherwise it will

	be ordered lexicographically e.g. <i>file1</i> , <i>file101</i> , <i>file11</i> , <i>file20</i> , <i>file3</i>
Debug	Select true if you want to see more debug output.

5.2.6.35 PDF To JPEG/PDF To PNG/ PDF To Tiff

These steps use GDPicture libraries to convert PDF files into the JPEG, PNG or TIFF format.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Tiff Compression (PDF to TIFF only)	Specifies the TIFF compression when saving images in TIFF format.
DPI	The dpi resolution to be used for rendering. A value of 72 will give the same result as Acrobat when zoom level is 100%. Values over 300 will cause excessive memory usage.
Brightness	Adjust the Brightness of the output image. Value must be between -100 and 100.
Contrast	Adjust the Contrast of the output image. Value must be between -100 and 100.
Saturation	Adjust the Saturation of the output image. Value must be between -100 and 100.
Gamma	Adjust the Gamma of the output image. Value must be between -100 and 100.
Auto Deskew	Select 'Yes' to try to deskew the image to about 15 degrees. Deskewing an image can help a lot to do OCR, OMR, barcode detection or just improve the readability of an image.
Crop Black Borders	Detects and removes margins consisting of black color around the image.
Crop Black Borders Ex	Detects and sets to White, margins consisting of black color around the image. This does not have the same behavior as Crop Black Borders; The black borders are not removed but are set to blank. Therefore, the image dimensions are kept the same.
Crop Area Height	Specifies the page height, in pixels, of the resulting document when cropping.
Crop Area Width	Specifies the page width, in pixels, of the resulting document when cropping.
Crop Location Left	Specifies the distance, in pixels, to crop from the left of the resulting document.
Crop Location Bottom	Specifies the distance, in pixels, to crop from the bottom of the resulting document.

Despeckle	Performs a 3x3 despeckle filter. It can remove black noise pixels from white backgrounds and visa versa. It also can remove random noise from multicolored backgrounds.
Despeckle More	Performs a 5x5 despeckle filter. It can remove black noise pixels from white backgrounds and visa versa. It also can remove random noise from multicolored backgrounds.
Enable ICM	Specifies if color correction is used for images embedding an ICC profile. Enables ICM results in automatic pixel transformation while opening image including an ICC profile.
Remove Hole Punch	Removes all punch holes situated on the margins of your image.
Remove Lines	Performs line removal on the image in the direction specified.
Resize New Height	New image height in pixels, of the resulting document when resizing.
Resize New Width	New image width in pixels, of the resulting document when resizing.
Resize Interpolation Mode	The interpolation mode to use when resizing the image.
Rotate By Angle	Selects whether to rotate by an angle specified, or by a preset type of rotation.
Rotation Angle	The angle of rotation for the image.
Rotation Type	The method of rotation to apply to the image.
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.
Debug	Select true if you want to see more debug output.

5.2.6.36 PDF To Text

This step uses GdPicture libraries to extract the searchable text from the pages of a PDF file, and creates an output text file. If the page is non-searchable, there is the option to enable OCR.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.
Page Separator	A text separator that will go between the text of pages
Page Separator Placement	The placement of the Page Separator. It can go above or below each page of text
Copy Input PDF To Target Folder	Set to true to copy the input pdf to the output location after the text is extracted
Preserve Paragraph	Specifies that the text extraction engine must preserve text paragraphs.
Paragraph Separator	This property specifies the separator to be utilized for splitting paragraphs. It only takes effect when the PreserveParagraphs property is set to Yes.
Enable OCR	Enables the use of the GdPicture OCR engine if the page is non-searchable.
OCR Dictionary	Add the code of languages for OCR, separated by '+'. For example, 'eng+deu+fra' would add English, German, and French.
Debug	Select true if you want to see more debug output.

5.2.6.37 PDF To Searchable PDF (GdPicture)

This step uses GDPicture libraries to carry out Optical Character Recognition on the input PDF, creating an invisible searchable text layer over the document.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
OCR Dictionary	Add the code of any additional languages for OCR, separated by '+'. For example, 'eng+deu+fra' would add English, German and French. Codes can be found in the OCR Language Codes section.
DPI	DPI of TIFF images generated or converted from the source PDF File. These images are then OCR'd to create the searchable PDF.
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.
Thread Limit	The GdPicture OCR engine processes multiple pages concurrently for optimal performance. This can take a heavy toll on the CPU. If needed, this option allows the number of pages processed consecutively to be limited.
Debug	Select true if you want to see more debug output.

5.2.6.38 PDF Portfolio

This step uses GDPicture libraries to combine a folder of files into an integrated PDF unit. There are a wide range of file types that can be used to create the PDF Portfolio.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %DIRNAME (original directory name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Pdf Portfolio Type	The initial view mode for the PDF Portfolio. This affects the way the user views the component files after opening the PDF Portfolio file.
Debug	Select true if you want to see more debug output.

5.2.6.39 Smart Redaction

This step uses GDPicture libraries to identify and redact selected sensitive information in the input document.

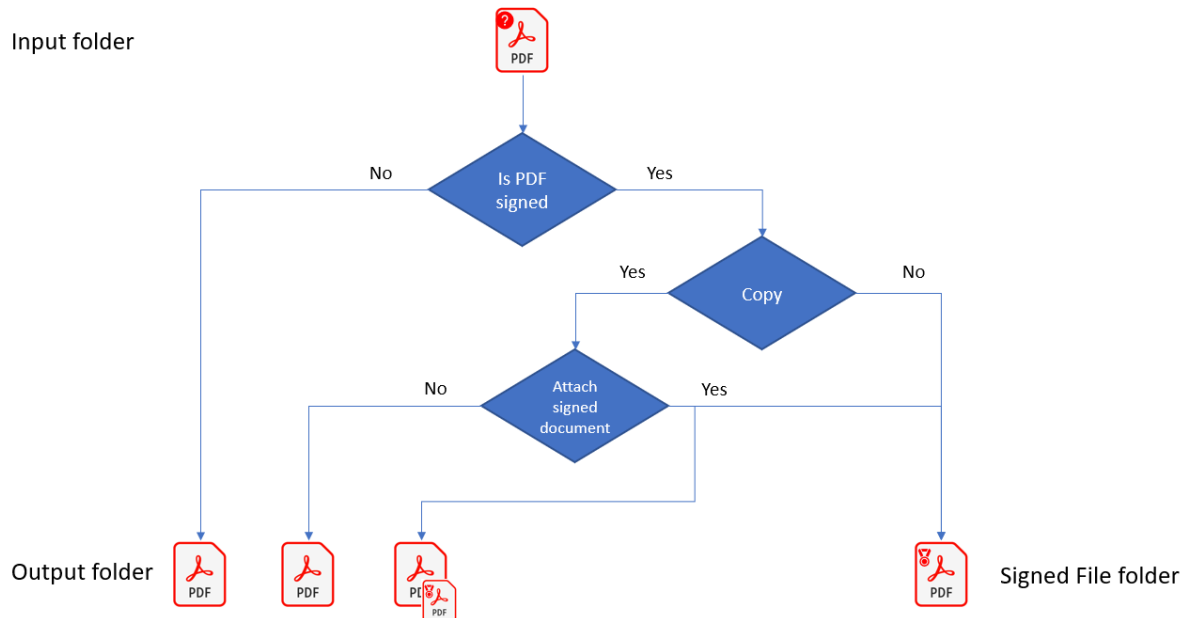
Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Redact Credit Card Numbers	Set to true if you want to redact Credit Card Numbers.
Redact Email Addresses	Set to true if you want to redact Email Addresses.
Redact IBANs	Set to true if you want to redact IBANs.
Redact Phone Numbers	Set to true if you want to redact Phone Numbers.
Redact URIs	Set to true if you want to redact URIs.
Redact VAT IDs	Set to true if you want to redact VAT IDs.
Redact Vehicle Identification Numbers	Set to true if you want to redact Vehicle Identification Numbers.
Redact Social Security Numbers	Set to true if you want to redact Social Security Numbers.
Redact Postal Addresses	Set to true if you want to redact Postal Addresses.
Redaction Color	Choose which color will be used for redacting.
OCR Dictionary	Add the code of any additional languages for OCR, separated by '+'. For example, 'eng+deu+fra' would add English, German and French. To install additional dictionaries, see the language codes .
Detect Orientation	Select 'Yes' if you want to auto detect orientation.
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.
Redaction Timeout (ms)	Limits the amount of time in milliseconds that can be spent on a redaction. A value of zero means it will wait indefinitely.
Debug	Select true if you want to see more debug output.

5.2.6.40 Detect Signatures

This step uses GDPicture libraries to identify pdf documents that contain digital signatures.

Any step that alters a digitally signed PDF will invalidate that PDF's signature. This step allows signed files to be identified, and either copied or moved to a specified folder so the signature can be preserved.

If the Copy option is selected, the original signed file can also be attached to the copy that is processed. This means that the original is attached to the file that can be subsequently processed.



Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Signed File Name	Signed file name template which can include %FILENAME (original file name).
Signed File Path	The full path (excluding file name) for the location to copy/move the signed file before processing.
Create Signed Path	Setting this to 'Yes' will create the signed file path directory if it does not exist. The file processing will fail if a signed file is processed, the signed path does not exist, and this is set to 'No'.
Overwrite Signed	Setting this to 'Yes' will automatically overwrite any file in the signed file path with the same name as the current signed file. The file processing will fail if the signed file already exists and overwrite is set to false.
Signed Action	The action to take if a signed file is detected. It can either be copied or moved to the Signed File Path.
Attach Signed Document to Output	Setting this to 'Yes' will attach a copy of the signed document to itself before being saved in the output location. This ensures a signed copy will remain with the copy that is processed.
Debug	Select true if you want to see more debug information.

5.2.6.41 Key Value Pair Extraction

This step uses the GDPicture engine to extract information about key-value pairs in pdf document. The extra information included can be the Key or Value Bounding Box, Page Number, Confidence, and Data Type.

The user can also use JSON file to declare Expected Keys. These specific keys will be added to a separate output file if a value is found. Synonyms can also be declared for each Expected Key, so that a match for any of the synonyms will be counted as a match for the Expected Key. An example is below.

For example, we have used **total** and **invoice number** as the expected keys. **grand total** is a synonym for 'total', and there are two synonyms for **invoice number** in **invoice no** and **inv no**.

```
[
  {
    "expectedKey": "total",
    "synonyms": ["grand total"]
  },
  {
    "expectedKey": "invoice number",
    "synonyms": ["invoice no", "inv no"]
  }
]
```

5.2.6.41.1 CSV Output Warning

CSV is a format commonly used by spreadsheet programs. These programs commonly transform numerical data or formula, and will save these transformations, overwriting the original data. To prevent these transformations, **we add an apostrophe** to the start of any possible transformations.

e.g. the phone number +44 115 496 0999 will appear as '+44 115 496 0999' in the CSV only.

The transformations are listed below.

- **Formula** - these are generally for values that begin with +, -, =, or @, we add an apostrophe at the beginning for the CSV output. This prevents the CSV from producing unintended formulas and functions from these values.
- **Dates/Times** – this covers many cases of date and time formats, as data can often be mistaken as a date or time, and then irreversibly transformed.
- **Long Numbers** – this covers numbers that are 11 digits or longer, as they are transformed to decimal notation

We recommend removing the apostrophes when extracting the data. This only affects CSV output, so it may be easier to extract data from the other formats if possible.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
OCR Language	Add the codes of the languages for OCR and KVP extraction, separated by '+' e.g. eng+fra Codes can be found in the OCR Language Codes section.

DPI	DPI used when performing OCR on the file as part of the KVP extraction process.
KVP Output Format	This setting determines the file output format(s). KVP data can be output in JSON, CSV and XML. e.g. json, csv, xml.
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.
Authorotate	Automatically rotate the page if the text does not have the correct orientation.
Trim Symbols	Setting this to 'Yes' will remove any symbols from the start/end of values, with the exception of the hash '#' or period '.' symbols.
Include Key Bounding Box	Setting this to 'Yes' will include the bounding box values for the key in the output.
Include Value Bounding Box	Setting this to 'Yes' will include the bounding box values for the value in the output.
Include Page Number	Setting this to 'Yes' will include the page number of the key value pair in the output.
Include Confidence	Setting this to 'Yes' will include the confidence score of the key value pair in the output. Confidence is measured between 0 (no confidence) and 100 (full confidence).
Confidence Threshold	The value of confidence (0-100) that a KVP must reach to be included in the output. Results under this confidence threshold will be discarded.
Include Type	Setting this to 'Yes' will include the data type of the key value pair in the output.
Expected Keys	The path to a JSON file for the expected keys and synonyms.
Debug	Select true if you want to see more debug information.

5.2.6.42 Pattern Redaction / Pattern Highlight

These steps use GDPicture libraries to identify and redact sensitive information (Redaction) or highlight important information (Highlight) in the input document based on a regular expression or terms list.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Pattern	A Regex pattern. The input pdf will be searched for matches to this Regex pattern, and any matches will be redacted/highlighted.
Terms Filepath	The path to a text file containing a list of terms to redact/highlight. Each line will be treated as a pattern, and any matches will be redacted/highlighted.
Case Sensitive	Determined whether or not the regex pattern matching should be case sensitive.
Red	The amount of red color to be used for the redaction/highlighted region color. Use a value between 0 and 255. Default is 0.
Green	The amount of green color to be used for the redaction/highlighted region color. Use a value between 0 and 255. Default is 0.
Blue	The amount of blue color to be used for the redaction/highlighted region color. Use a value between 0 and 255. Default is 0.
Alpha	The transparency value of the resulting region color. Use the value between 0 (full transparency) and 255 (full opacity). Default is 255.
Debug	Select true if you want to see more debug output.

5.2.6.43 Split PDF (GdPicture)

This step uses GdPicture libraries to split PDF files based on the ranges, bookmarks, or into single pages.

Screen Field / Button	Description
Output File Name	Target file template which can include %UNIQUEn (unique number starting at 1, zero padded to n digits) %FILENAME (original filename without the extension) and %PAGE n (first page of split, zero padded to n digits)
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Retain Metadata	Generated files will include metadata(such as Author and Title) from the original file.
Split Type	Sets the way that the input file will be split. <i>One of:</i> <ul style="list-style-type: none">• Split into single pages• Split by ranges (See below)• Split by repeating ranges (See below)• Split by bookmarks
Ranges	Set of page ranges separated by commas that defines which pages from the original should be extracted.
Repeat Every (Pages)	Apply the page range to each set of Page Ranges pages 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.
Remove Unused Resources	Removes unused resources from a pdf file to minimize file size.

5.2.6.44 Split by Barcode

This step uses GDPicture libraries to identify different barcode types in a PDF, and split the document at each instance of a barcode.

Screen Field / Button	Description
Output File Name	Target file template which can include %UNIQUEn or %INDEXn (unique number starting at 1, zero padded to n digits) %FILENAME (original filename without the extension) and %PAGE n (first page of split, zero padded to n digits)
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Read QRCode	Set this to true to recognize QRCode barcodes.
Read MicroQR	Set this to true to recognize MicroQR barcodes.
Read DataMatrix	Set this to true to recognize DataMatrix barcodes.
Read PDF417	Set this to true to recognize PDF417 barcodes.
Read Aztec	Set this to true to recognize Aztec barcodes.
Read MaxiCode	Set this to true to recognize MaxiCode barcodes.
Read Industrial2of5	Set this to true to recognize Industrial2of5 barcodes.
Read Inverted2of5	Set this to true to recognize Inverted2of5 barcodes.
Read Interleaved2of5	Set this to true to recognize Interleaved2of5 barcodes.
Read Iata2of5	Set this to true to recognize Iata2of5 barcodes.
Read Matrix2of5	Set this to true to recognize Matrix2of5 barcodes.
Read Code39	Set this to true to recognize Code39 barcodes.
Read Codabar	Set this to true to recognize Codabar barcodes.
Read BcdMatrix	Set this to true to recognize BcdMatrix barcodes.
Read DataLogic2of5	Set this to true to recognize DataLogic2of5 barcodes.
Read Code128	Set this to true to recognize Code128 barcodes.
Read Code93	Set this to true to recognize Code93 barcodes.
Read EAN13	Set this to true to recognize EAN13 barcodes.
Read EAN8	Set this to true to recognize EAN8 barcodes.
Read UPCA	Set this to true to recognize UPCA barcodes.
Read UPCE	Set this to true to recognize UPCE barcodes.
Read ADD5	Set this to true to recognize ADD5 barcodes.
Read ADD2	Set this to true to recognize ADD2 barcodes.
Page Range	Specifies the page range to be scanned for barcodes. A value of * will scan every page for barcodes.
Pattern	A Regex pattern. The input pdf will be searched for matches to this Regex pattern, and any matches will be redacted.

DPI	DPI of TIFF images generated or converted from the source PDF File. These images are then scanned for barcodes.
Retain Metadata	Generated files will include metadata(such as Author and Title) from the original file.
Remove Unused Resources	Removes unused resources from a pdf file to minimize file size.
Left	X coordinate of the Top Left Point of the rectangle you want to recognize the barcode.
Top	Y coordinate of the Top Left Point of the rectangle you want to recognize the barcode.
Width	Width of the rectangle you want to recognize the barcode.
Height	Height of the rectangle you want to recognize the barcode.

5.2.6.45 Pattern Enumeration

This step uses GDPicture libraries to identify terms and/or a pattern, and will produce a report based on the frequency of each term.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Pattern	A Regex pattern. The input pdf will be searched for matches to this Regex pattern, and any matches will be redacted/highlighted.
Terms Filepath	The path to a text file containing a list of terms to redact/highlight. Each line will be treated as a pattern, and any matches will be redacted/highlighted.
Case Sensitive	Determined whether or not the regex pattern matching should be case sensitive.
Pass Through	Determines whether or not the input pdf will be copied to the output folder.
Debug	Select true if you want to see more debug output.

5.2.6.46 Get Document Information

This step uses GdPicture libraries to produce a report on a the number of pdf pages that are searchable vs image. It will also calculate how many searchable pages are visible text pages vs hidden text layer.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME (original file name).
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Output Format	Choose the output format for the report. <ul style="list-style-type: none">• txt• csv• json• xml
Pass Through	Determines whether or not the input pdf will be copied to the output folder.
Debug	Select true if you want to see more debug output.

5.2.6.47 Convert PDF To Office

This step uses GdPicture libraries to convert pdf input files to various Office output formats, including docx, pptx, xlsx, and svg.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Output Format	Choose the output format for the report. <ul style="list-style-type: none">• docs• pptx• xlsx• svg
Enable ICC	Specifies if the converter shall favor preserving the ICC profile, if present in the loaded document, during the conversion.
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.
Timeout Milliseconds	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.
Debug	Select true if you want to see more debug output.

5.2.6.48 Convert Any File To Office

This step uses GDPicture libraries to convert various input file types to various Office output formats, including docx, pptx, xlsx, and svg. Not all file conversions are supported.

Screen Field / Button	Description
Output File Name	The template for the output file, which can include %FILENAME% to give the input file name without extensions.
Continue on Error	Set to true if the job should continue processing files after a file has failed.
Output Format	Choose the output format for the report. <ul style="list-style-type: none">• doc• pptx• xlsx• svg
Author	Set the Author metadata field in the output PDF. This can include %FILENAME% (original filename without the extension) or %DIRNAME% (directory name of original file)
Title	Set the Title metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Subject	Set the Subject metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Keywords	Set the Keywords metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Producer	Set the Producer metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Metadata	Set the Metadata field in the output PDF. This can include %FILENAME% or %DIRNAME%
Inject Email Header	Specifies whether the email header should be injected into the output document.
Convert Email Attachments To Office	Select 'Yes' if you want to convert email attachments to Office.
Email Attachments Filter	A regular expression that specifies the attachments that will be converted to office format. Attachments that do not match will be skipped.
Email Page Height	Specifies the page height, in points, of the resulting document when converting from the source Email file.
Email Page Width	Specifies the page width, in points, of the resulting document when converting from the source Email file.
Email Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Email file.
Email Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Email file.

Email Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Email file.						
Email Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Email file.						
Email Prefer One Page	Select 'Yes' if you want the email to be converted to a single page PDF if possible.						
Enable ICC	Specifies if the converter shall favor preserving the ICC profile, if present in the loaded document, during the conversion.						
Html Emulation Type	Specifies a type of a media to emulate.						
Html Page Height	Specifies the page height, in points, of the resulting document when converting from the source Html file.						
Html Page Width	Specifies the page width, in points, of the resulting document when converting from the source Html file.						
Html Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Html file.						
Html Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Html file.						
Html Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Html file.						
Html Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Html file.						
Html Prefer CSS Page Size	Give any CSS page size declared in the page priority over what is declared in Html Page Width and Html Page Height. If set to false, the renderer will scale the content to fit the paper size.						
Html Prefer One Page	Specifies whether the output document should contain a single page.						
Load Only First Page	Specifies that all executed actions with the loaded document will be processed using only the first page of the document.						
Page Range	Use the string of "1-5" for pages 1 to 5, or use the string of "1,5,6" to specify pages 1 and 5 and 6. You can use the string of "1,5,8-12" to specify pages 1, 5, 8 and all pages from page 8 to page 12, etc.						
Pdf Bitonal Image Compression	<p>Sets the scheme to be used to compress bitonal image data when converting/saving the currently loaded document to PDF format.</p> <table> <tr> <th>ID</th><th>Scheme</th></tr> <tr> <td>0</td><td>None</td></tr> <tr> <td>1</td><td>Flate</td></tr> </table>	ID	Scheme	0	None	1	Flate
ID	Scheme						
0	None						
1	Flate						

	2	CCITT4
	3	JPEG
	4	JBIG2
	5	JPEG2000
JBIG2 PMS Threshold	Sets the threshold of the JBIG2 pattern matching and substitution. The default value is 0.85.	
Pdf Color Image Compression	Sets the scheme to be used to compress color image data when converting/saving the currently loaded document to PDF format.	
Pdf Enable Color Detection	Enables or disables the automatic color detection feature when converting/saving the currently loaded document to PDF format.	
Pdf Image Quality	Sets the level of quality used to compress images with a lossy compression scheme, which are embedded in the newly produced PDF document when converting/saving the currently loaded document to PDF format. It must be a value from 0 to 100. 0 means the worst quality and the best compression, 100 means the best quality and the worst compression.	
Pdf Use Deflate On JPEG	Specifies if the converter shall use additional Deflate compression for JPEG images in PDF output.	
Rasterization DPI	Sets the rendering resolution to be used when converting vector content to raster content, if any is included in the currently loaded document.	
Render Sheets Headers and Footers	Specifies that the xls and xlsx headers and footers should be rendered. Affects XLSX/XLS input only.	
Split Excel Sheets Into Pages	Specifies that xls and xlsx sheets should be split into pages according to the PageSetup element of each sheet. Affects XLSX/XLS input only.	
Spreadsheet Bottom Margin Override	Specifies the spreadsheet bottom margin height in millimeters. If height is not given or negative, the margin specified in the document will be used instead. Affects XLSX/XLS input only.	
Spreadsheet Left Margin Override	Specifies the spreadsheet left margin width in millimeters. If height is not given or negative, the margin specified in the document will be used instead. Affects XLSX/XLS input only.	
Spreadsheet Maximum Content Height Per Sheet	Decimal value indicating the maximum height of the sheet content, in millimeters. Maximum content height ignores header and footer height. Affects XLSX/XLS input only.	
Spreadsheet Maximum Content Width Per Sheet	Decimal value indicating the maximum width of the sheet content, in millimeters. Maximum content width ignores margins. Affects XLSX/XLS input only.	
Spreadsheet Page Height Override	Specifies the spreadsheet page height in millimeters. If height is not given or not positive, the page height	

	specified in the document will be used instead. Affects XLSX/XLS input only.
Spreadsheet Page Width Override	Specifies the spreadsheet page width in millimeters. If width is not given or not positive, the page width specified in the document will be used instead. Affects XLSX/XLS input only.
Spreadsheet Render Only Print Area	For spreadsheet, specifies that for each sheets only the print areas must be rendered. If no print area are found the whole sheets will be rendered. Affects XLSX/XLS input only.
Spreadsheet Right Margin Override	Specifies the spreadsheet right margin width in millimeters. If width is not given or not positive, the page height specified in the document will be used instead. Affects XLSX/XLS input only.
Spreadsheet Top Margin Override	Specifies the spreadsheet top margin height in millimeters. If width is not given or not positive, the page width specified in the document will be used instead. Affects XLSX/XLS input only.
Tiff Enable Exif Rotate	Specifies whether tiff encoder is using Exif rotate flag to handle page rotations.
Timeout Milliseconds	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.
Txt Font Bold	Specifies whether the font used for the resulting document when converting from the source txt file must have a bold style.
Txt Font Italic	Specifies whether the font used for the resulting document when converting from the source txt file must have an italic style.
Txt Font Family	Specifies the name of the font to be used for the resulting document when converting from the source txt file.
Txt Font Size	Specifies the text size, in points, to be used for the resulting document when converting from the source txt file.
Txt Horizontal Text Alignment	Specifies the horizontal text alignment of the resulting document when converting from the source txt file.
Txt Page Height	Specifies the page height, in points, of the resulting document when converting from the source Txt file.
Txt Page Width	Specifies the page width, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Bottom	Specifies the bottom page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Left	Specifies the left page margin, in points, of the resulting document when converting from the source Txt file.
Txt Page Margin Right	Specifies the right page margin, in points, of the resulting document when converting from the source Txt file.

Txt Page Margin Top	Specifies the top page margin, in points, of the resulting document when converting from the source Txt file.
Debug	Select true if you want to see more debug output.

5.2.7 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

Search:

Clear

1 - Conversion Settings

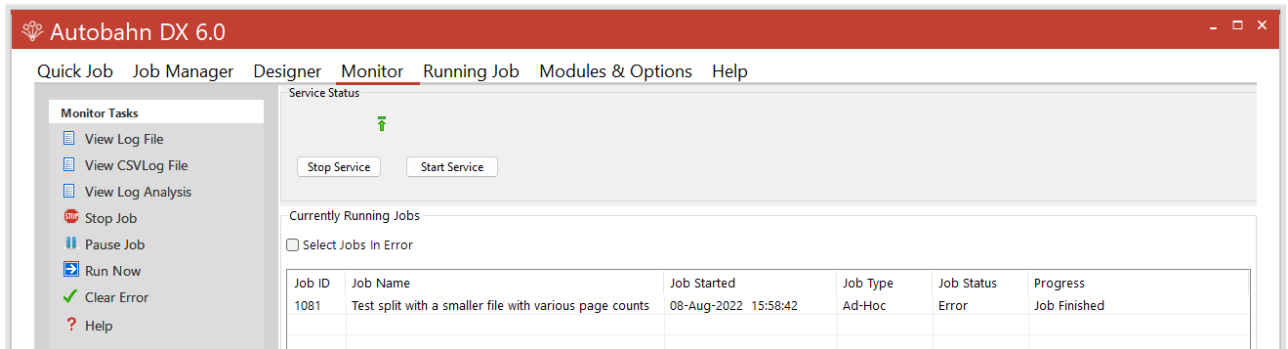
Output File Name	%FILENAME.pdf
Continue on Error	Yes
Conversion Timeout (ms)	
Convert Bookmarks	No
Bookmark Depth	
Convert Hyperlinks	No
Print All Sheets (Excel)	No
Print Background Color (IE)	No
Print Scale % (Visio)	
Header (IE)	
Footer (IE)	
Image Compression	JPEG
Image Downsizing	No
Downsize Resolution DPI	
Image JPEG Quality	
Font Embedding	SubSet
Font Substitution	None
Embed Fonts as Type 0	No
Top Margin	
Bottom Margin	

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

To look for a property, you can either use the scroll bar on the right-hand side or the search bar at the top. The search bar looks for an exact match of the text that you type but will offer suggestions that start with the text you have currently typed. Selecting a suggestion will jump you to the property and select it for editing.

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 CSV log file of the currently selected job.
View Log Analysis	View the log stats of the currently selected job.
Stop Job	Halt execution of the selected job.
Pause Job	Pauses the execution of a job, you can 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 shows the current license key, the components unlocked by the license key (including any expiry date) and the email settings.

5.4.1 License key

The license key is entered here, for more details see [section 2.3](#).

The licensed components are displayed below the key.

5.4.2 Email settings

Email settings can be entered here. This is used to support the “Send Email on Error” functionality defined in the previous section. The information entered is held in the Autobahn.config file found in the config directory. The password is held in an encrypted form.

There are two options for email authentication, SMTP or OAuth2 (Azure) authentication.

Use the drop-down box below the license properties to select the authentication method.

5.4.2.1 Email settings for SMTP:

Autobahn DX 6.0

Quick Job Job Manager Designer Monitor Running Job **Modules & Options** Help Example : Convert TIFF to Searchable PDF

Options ? Help

Core Module

License Key

License Expires On Friday, 2 Nov, 2029

Server License

- + Extended OCR License
- + Asian Languages License
- + Arabic License
- + Hebrew License
- + IHQC License
- + Multi-core License With 8 cores

SMTP

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.
Need Authentication	If checked, basic authentication will be used, and the username and password can be filled in.
Username	Username for authentication by the server.
Password	Password for the username.

5.4.2.1.1 Sending emails from Autobahn DX via Microsoft 365 (Office 365) without authentication

Although it is recommended to use (at least) Basic Authentication for sending emails, there are situations where this is not possible.

The following instructions are included as a guide (not a recommendation) to configure Autobahn DX to send emails via Microsoft 365 (Office 365).

The server field requires the MX endpoint of the email server.

This is generally of the form (where company.com is the domain):

company-com.mail.protection.outlook.com

The port number is generally 25

SMTP Settings

Server	<input type="text" value="company-com.mail.protection.outlook.com"/>	<input type="checkbox"/> Need Authentication
SMTP Port	<input type="text" value="25"/>	
Username	<input type="text"/>	
Password	<input type="password"/>	<input type="button" value="Update"/>

With these settings, the destination email address must be from the same domain as the server:

So, for this example the sender and recipient email addresses would be:

From: alice@company.com

To: bob@company.com

5.4.2.2 Email settings for OAuth2:

The screenshot shows the 'Autobahn DX 6.0' application window with the 'Modules & Options' tab selected. The window has a menu bar with 'Quick Job', 'Job Manager', 'Designer', 'Monitor', 'Running Job', 'Modules & Options', and 'Help'. Below the menu bar is a toolbar with 'Options' and 'Help' buttons. The main content area is titled 'Core Module' and contains the following elements:

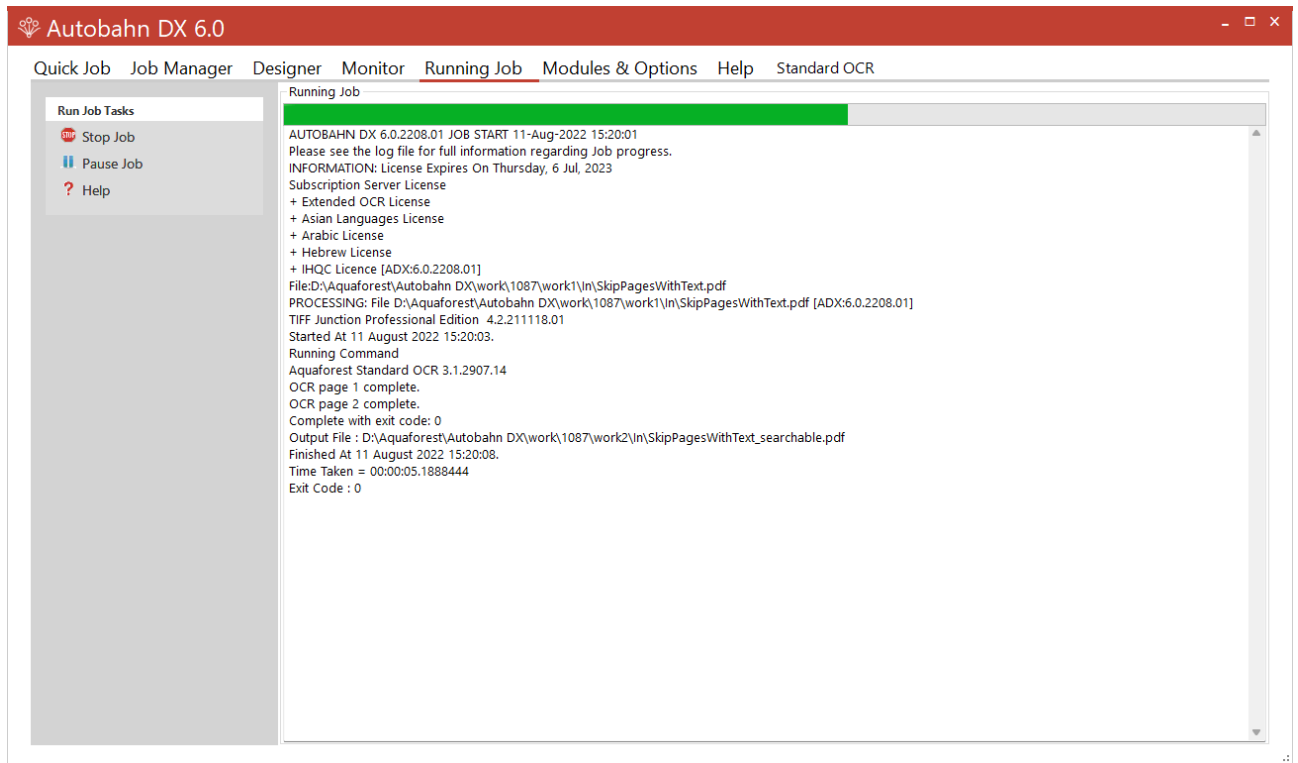
- A 'License Key' input field with an 'Update' button.
- Text indicating the license expires on Friday, 2 Nov, 2029.
- A list of licenses: 'Server License', '+ Extended OCR License', '+ Asian Languages License', '+ Arabic License', '+ Hebrew License', '+ IHQC License', and '+ Multi-core License With 8 cores'.
- A dropdown menu for 'OAuth2 (Azure)'.
- A section titled 'OAuth2 (Azure) Settings' containing:
 - 'Azure Client ID' input field.
 - 'Azure Tenant' input field.
 - 'Azure AD Instance' input field.
 - 'Credential Type' dropdown menu (set to 'Certificate').
 - 'Certificate Path' input field.
 - 'Certificate Password' input field.
 - An 'Update' button.

Field	Description
Azure Client ID	The Client ID for OAuth2 Authentication
Azure Tenant	The Tenant ID for OAuth2 Authentication
Azure AD Instance	The address of the Azure AD Instance e.g., https://login.microsoftonline.com
Credential Type	Select the credential type for OAuth2 Authentication. The options are Client Secret or Certification.
Client Secret	The Client secret generated by Azure
Certificate Path	The path to the certificate generated by Azure
Certificate Password	The password of the certificate generated by Azure

5.5 Running Jobs via Job Manager or Designer

Choosing “Run Now” from the menu on the Job Manager or Designer tabs changes the displayed tab to the Running Job tab and begins the execution of the current selected job definition.

The job's output is displayed on this tab, as well as being added to the Job log files.



Menu Item	Action
Stop	Stops the currently running job.
Pause Job	Pauses the execution of a job, you can 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	Takes you to the Help tab.

6 Scripting Custom Steps

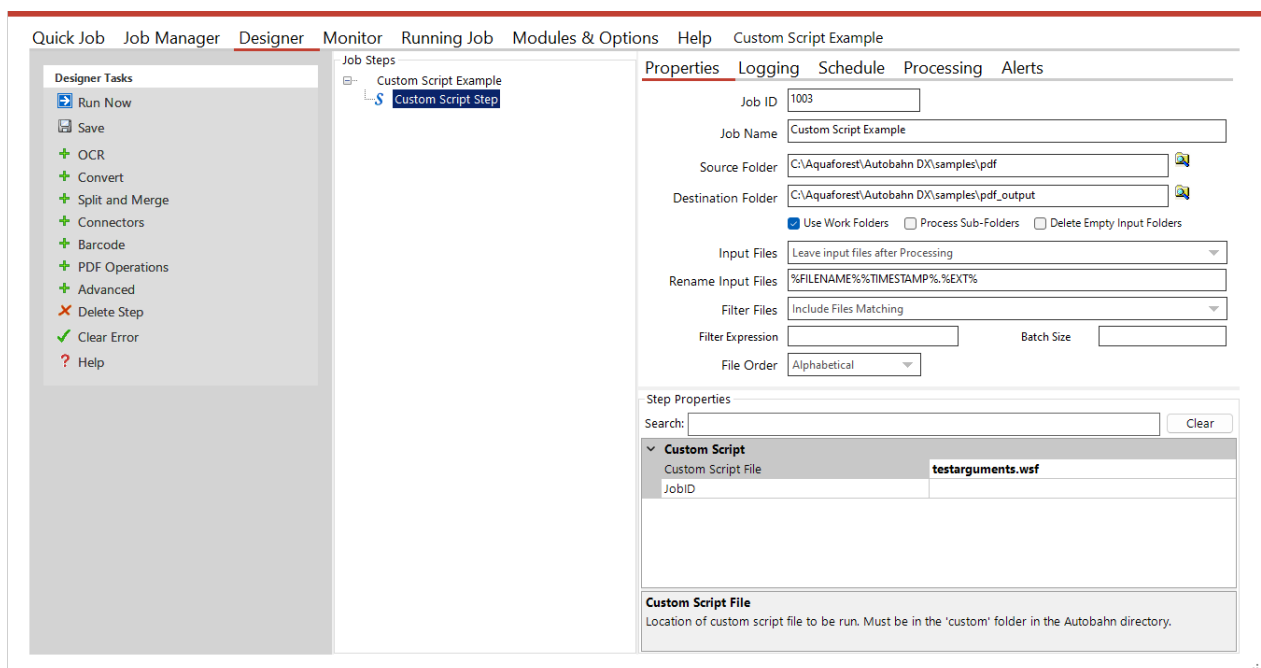
Autobahn DX allows custom job steps to be defined that can be included in a Job Definition in the same way as any other step type.

The custom step itself starts with a Windows Script File in the Custom folder of the Autobahn DX installation. This can be the complete script to be executed or the interface to any form of executable file or script.

Some template wsf scripts are included which demonstrates calling a command-line executable from within the script.

6.1 Custom Script Example

The 'Custom Script Step' is found in the 'Advanced' section of the designer.



Custom script files must be placed in "*<installation folder>\Autobahn DX\custom folder*". For this example, the custom script file used is called **testarguments.wsf**. It simply returns the arguments that autobahn sends to the script. This file can be found in the custom folder by default

The result of this custom script is shown in the logging of the running job:

Named arguments

in has value C:\Aquaforest\Autobahn DX\samples\pdf

out has value C:\Aquaforest\Autobahn DX\samples\pdf_output

type has value folder

If you require any advice for using a custom script, please contact the support team.

6.2 Building a Custom Script

6.2.1 Use case – Script only

The client has a large number of files containing scanned images. A large proportion of these images are either of poor quality (hand-written, badly scanned, or damaged originals) or contain mainly pictorial information with minimum text.

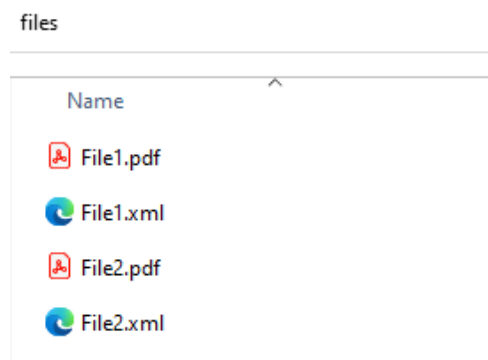
To improve the searchability of the PDFs, XML documents have been prepared (either by hand or by image recognition) to identify the contents.

These files are in pairs, named:

<filename>.pdf

<filename>.xml

For example: **File1.pdf** and **File1.xml**



The requirement is to produce searchable PDF files comprising the original PDF and the XML document of each pair.

The script is going to group each PDF-XML pair in an individual sub folder in the output folder.

The rest of the Autobahn DX job can be accomplished using normal Autobahn steps.

6.2.2 The Script

The following assumes a small amount of knowledge of the Microsoft JScript language (which is a Microsoft version of the JavaScript/ECMAScript language). The [Custom Script Example](#) above shows the information provided to the script by Autobahn DX.

6.2.2.1 WSF wrapper

All WSF scripts require the following wrapper elements.

This example is in Jscript, but other languages are available.

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

The Jscript code goes here

```
  </script>  
</job>
```

6.2.2.2 Jscript code

The full script is in the Custom sub folder of the Autobahn DX installation.

The first operation in the script is to obtain the source (In) and target (Out) folders.

```
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");
}
```

If there are values for the InFolder and OutFolder,

```
if(OutFolder!="" && InFolder!=""){
```

then create a File System Object (FSO) and use it to get the InFolder, the files in the InFolder and the OutFolder.

```
var objFSO=WScript.CreateObject("Scripting.FileSystemObject");
var objFolder = objFSO.GetFolder(InFolder);
var colFiles=objFolder.Files;
// Remove any trailing '\' in supplied Out path
OutFolder = objFSO.GetFolder(OutFolder);
```

For each file in the InFolder

```
for(var objEnum = new Enumerator(colFiles);!objEnum.atEnd();
objEnum.moveNext()) {
    sFileName = objEnum.item();
    sOutName = OutFolder + "\\";
    var rootName = objFSO.GetFileName(sFileName);
```

If the file has a suffix of "pdf", if there is no sub folder create it, then create the output file name by prefixing the original name with the letter 'a'.

```
if(objFSO.GetExtensionName(sFileName) == "pdf")
{
    rootName = rootName.substring(0,rootName.length-4)
    sOutName = sOutName + rootName;
    if(!objFSO.FolderExists(sOutName)){
        objFSO.CreateFolder(sOutName);
    }
    sOutName = sOutName + "\\a" + objFSO.GetFileName(sFileName);
}
```

If the file has a suffix of "xml", if there is no sub folder create it, then create the output file name by prefixing the original name with the letter 'b'

```
else if(objFSO.GetExtensionName(sFileName) == "xml")
{
    rootName = rootName.substring(0,rootName.length-4)
    sOutName = sOutName + rootName;
    if(!objFSO.FolderExists(sOutName)){
        objFSO.CreateFolder(sOutName);
    }
    sOutName = sOutName + "\\b" + objFSO.GetFileName(sFileName);
}
```

Any other suffix is an error so quit with an exit code of 2

```
else
{
    WSH.Echo("Unrecognised file type");
    WSH.Echo(sFileName);
    WSH.Quit(2);
}
```

Record the filename to the log and copy the file out

```
        WSH.Echo(sOutName)
        // Copy the file (overwriting any existing file)
        objFSO.CopyFile (sFileName,sOutName,true);
    }
```

Record that it has completed and return an exit code of 0.

```
        // Quit with value 0
        WSH.Echo("Done");
        WSH.Quit(0);
    }
```

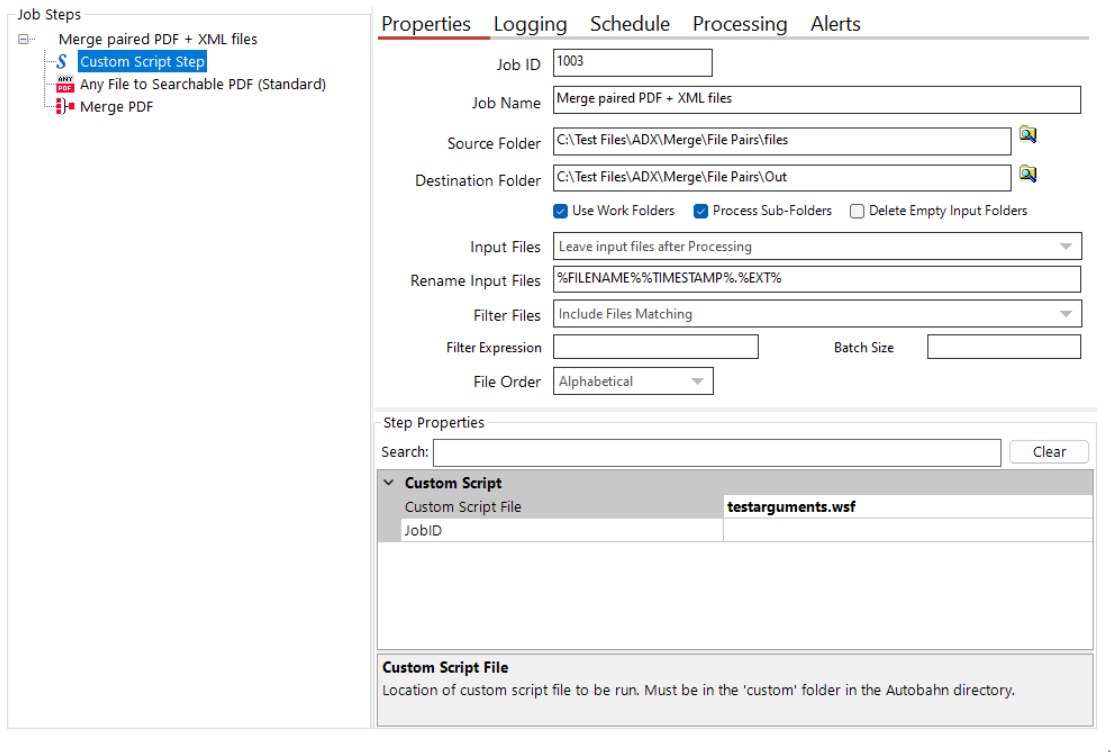
If there are missing parameters, record an error and return an exit code of 1.

```
        // Else there are one or more missing parameters
        //so quit with error
        else
        {
            WSH.Echo("Missing Named Parameter(s)");
            WSH.Quit(1);
        }
```

6.2.3 The Autobahn DX Job

The Autobahn job comprises three steps:

- The call to the custom script step
- Convert any file to searchable PDF
- Merge PDF



Job Steps

- Merge paired PDF + XML files
- Custom Script Step**
- Any File to Searchable PDF (Standard)
- Merge PDF

Properties | Logging | Schedule | Processing | Alerts

Job ID: 1003

Job Name: Merge paired PDF + XML files

Source Folder: C:\Test Files\ADX\Merge\File Pairs\files

Destination Folder: C:\Test Files\ADX\Merge\File Pairs\Out

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

Input Files: Leave input files after Processing

Rename Input Files: %FILENAME%%TIMESTAMP%%EXT%

Filter Files: Include Files Matching

Filter Expression: Batch Size: File Order: Alphabetical

Step Properties

Search: Clear

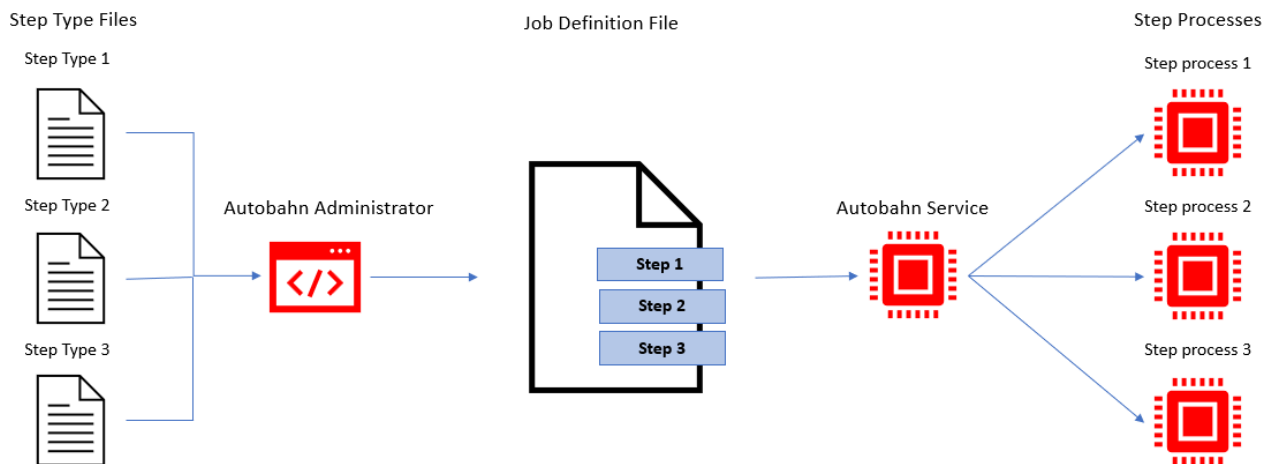
Custom Script	
Custom Script File	testarguments.wsf
JobID	

Custom Script File
Location of custom script file to be run. Must be in the 'custom' folder in the Autobahn directory.

The Merge PDF step merges the contents of each sub folder into an individual PDF.

7 Job Definition XML Files

7.1 Overview of Job Definition Creation and Processing



Autobahn Job Definitions are stored as XML files which are generally using the Autobahn Administrator (see [section 6](#)).

A job definition file contains certain standard pieces of information (source folder for example) that are common to all jobs plus "StepType" Definition sections for each step in the job.

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 <Autobahn DX Installation folder>\temp\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

Below is a simple example of a job definition which is designed to continuously (every 30 seconds) monitor a directory (c:\faxes) for new .TIFF files and convert them to searchable PDF and place the resulting files in c:\processed faxes. The step details have been removed for clarity and are covered [here](#).

```
<autobahnjob>
  <jobid>1009</jobid>
  <jobname>Monitor incoming faxes and OCR them</jobname>
  <adxversion>6.0</adxversion>
  <trigger />
  <scheduletype>continuous</scheduletype>
  <scheduleevery>30</scheduleevery>
  <scheduleeveryunits>Seconds</scheduleeveryunits>
  <schedulefrom>00:00:00</schedulefrom>
  <scheduleto>23:55:00</scheduleto>
  <scheduleat>16:00:00</scheduleat>
  <stopprocessingonerror>True</stopprocessingonerror>
  <sendemailalerts>False</sendemailalerts>
  <attachjobreport>False</attachjobreport>
  <attachlogfile>False</attachlogfile>
  <dontsendnoerror>False</dontsendnoerror>
  <dontsendnofiles>False</dontsendnofiles>
  <dontsendonsuccess>False</dontsendonsuccess>
  <SendEmailAlertsfromaddress />
  <SendEmailAlertstoaddress />
  <skipprocessedfiles>False</skipprocessedfiles>
  <skipmask>%FILENAME%.pdf</skipmask>
  <SendEmailAlertstitle>%JOBNAME% %JOBSTATUS%!</SendEmailAlertstitle>
  <SendEmailAlertsmessage>Job: '%JOBNAME%' Status: '%JOBSTATUS%'.&lt;br&gt;
Log: %LOGFILE%.&lt;br&gt;Time: %DATESTAMP% %TIMESTAMP%&lt;br&gt;
Source: &lt;a href='%JOBSOURCE%'&gt;%JOBSOURCE%&lt;/a&gt;&lt;br&gt;
Target: &lt;a href='%JOBTARGET%'&gt;%JOBTARGET%&lt;/a&gt;&lt;br&gt;
  </SendEmailAlertsmessage>
  <jobprogresscsv>C:\Aquaforest\Autobahn
DX\work\1009\JobProgress.csv</jobprogresscsv>
  <jobsteps>1</jobsteps>
  <joblogrention />
  <joblogmaxsize />
  <jobsourcetype>folder</jobsourcetype>
  <jobsource>C:\Faxes</jobsource>
  <jobtarget>C:\Processed Faxes</jobtarget>
  <jobwork>C:\Aquaforest\Autobahn DX\work\1009</jobwork>
  <jobarchive>C:\Aquaforest\Autobahn DX\work\Archive\1009</jobarchive>
  <inputfilesrename>%FILENAME%%TIMESTAMP%.%EXT%</inputfilesrename>
  <joberrors>C:\Aquaforest\Autobahn DX\work\1009\errors</joberrors>
  <joblogfile>C:\Aquaforest\Autobahn DX\logs\1009\%DATESTAMP%.txt</joblogfile>
  <jobCSVlogfile>C:\Aquaforest\Autobahn
DX\logs\1009\%DATESTAMP%.csv</jobCSVlogfile>
  <jobtemp>C:\Aquaforest\Autobahn DX\work\1009\temp</jobtemp>
  <JobUseWorkFolder>False</JobUseWorkFolder>
  <DeleteEmptyFolders>False</DeleteEmptyFolders>
  <SkipErrorFolder>False</SkipErrorFolder>
  <inputfileprocessing>leave</inputfileprocessing>
  <jobfilterfile>include</jobfilterfile>
  <jobfiltertype />
  <jobfileorder>alphabetical</jobfileorder>
  <jobreturnstructure>False</jobreturnstructure>
  <filelength>False</filelength>
  <jobinerror>Copy to Error Folder</jobinerror>
  <jobstep>
...
  </jobstep>
</autobahnjob>
```

XML Element	Description
jobid	The job ID number (see 3.2 above).
jobname	Job Description. Default is "Job %JOBID%"
adxversion	Autobahn version
trigger	Name of the trigger file. The job will not start automatically until the trigger file is in the input folder.
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
scheduleto	For onceperday, the time at which to run
scheduleat	For continuous, end time for runs
stopprocessingonerror	If true, stop processing if an error occurs
sendemailalerts	If true, send an email alert
attachjobreport	Attach the job report to the email
attachlogfile	Attach the log file to the email
dontsendnoerror	If no errors, do not send email
dontsendnofiles	If no files are processed, do not send email
dontsendonsuccess	On success, do not send email
SendEmailAlertfromaddresses	Address of sender
SendEmailAlerttoaddress	Address of recipient
skipprocessedfiles	Skip files
skipmask	Mask of processed files to skip
SendEmailAlertstitle	Title of email
SendEmailAlertsmessage	Email body (HTML elements need to be escaped)
jobprogresscsv	The location of the temporary progress file used to store status information while the job is running.
joblogfile	Location of the job log file. By default, the log file is logs/%JOBID%/%TIMESTAMP%.txt
jobsteps	The number of job steps
joblogrention	Log file retention period (days)
joblogmaxsize	Maximum log file size
jobsourcetype	File, folder or tree
jobsource	The source file or folder
jobtarget	The target folder

joberrors	Folder for error files. Files that cannot be processed will be actioned based on the jobinerror value.
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.
jobarchive	The path of the folder used as the archive location when inputfileprocessing is set to move or copy to archive.
inputfilesrename	Name template to rename input files after processing
joblogfile	The log file filepath template, default is to name the log file as the current date.
jobCSVlogfile	The CSV log file filepath template, default is to name the log file as the current date.
jobtemp	The temporary folder path.
JobUseWorkFolder	Use intermediate work folders – required if the job has multiple steps or the input and output folder are the same.
DeleteEmptyFolders	If true, delete any empty folders in the job source folder
SkipErrorFolder	If true, inaccessible folders will be skipped without throwing an error
inputfileprocessing	Action on input files: <ul style="list-style-type: none"> • Copy to Archive after processing • Move to Archive after processing • Leave input files after processing • Move input files to Target folder after processing • Delete input files
jobfilterfile	Filter files based on: <ul style="list-style-type: none"> • Include files matching • Exclude files matching • Include with document count limit • Include unprocessed PDFs only • Include unprocessed PDFs only with document count limit
jobfiltertype	File filter
jobfilterorder	Ordering of files included (mainly used in conjunction with document count limit options): <ul style="list-style-type: none"> • Alphabetical • Created date/UTC date ascending or descending • Modified date/UTC date ascending or descending
jobreturnstructure	If true, the input folder structure will be preserved in the output
filelength	If true, skip long file names.
jobinerror	Action when job is in error: <ul style="list-style-type: none"> • Move to error • Copy to error • Take no action

jobstep	Contains the definition of a job step. Multiple elements, including multiple steps. See Sample Step Details .
---------	---

7.3.1 Job Step

The following is a sample job step. Other step types may have more step details and attribute vales.

```
<jobstep>
  <stepsequence>1</stepsequence>
  <steptype>kingfisher</steptype>
  <stepdetails>
    <operation>kingfisher</operation>
    <sourcetype>folder</sourcetype>
    <source>C:\Aquaforest\Autobahn DX\work\1010\work1\In</source>
    <target>C:\Aquaforest\Autobahn DX\work\1010\work2\In</target>
    <inerror>Copy to Error Folder</inerror>
    <returnstructure>False</returnstructure>
    <errors>C:\Aquaforest\Autobahn DX\work\1010\errors</errors>
    <tempfolder>C:\Aquaforest\Autobahn DX\work\1010\temp</tempfolder>
    <joboptions />
    <advancedflags />
    <fileorder>alphabetical</fileorder>
    <kfjobid>10002</kfjobid>
    <docoptions />
    <logfile />
  </stepdetails>
  <attributevalues>
    <attribute>
      <attributeid>topdf2</attributeid>
      <currentvalue />
    </attribute>
    <attribute>
      <attributeid>topdf1</attributeid>
      <currentvalue>10002</currentvalue>
    </attribute>
  </attributevalues>
</jobstep>
```

XML Element	Description
stepsequence	The number of the step in the sequence. The steps will be processed in stepsequence order (ascending).
steptype	Identifier of the step type. This will match the files (without the ordering number) found in <installation folder>\Autobahn DX\steptype
stepdetails	Details of the step that are passed to the step processor. See below.
attributevalues	Elements that are displayed in the UI. See below.

A Job step comprises four elements.

7.3.1.1 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.

The Step Details will vary by step type.

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.
Source	Source file or folder.
Target	Target folder.
inerror	Action to be undertaken in the event of an error
returnstructure	Retain file structure when copying/moving to error folder (based on inerror setting above).
Errors	Files that cannot be processed are placed in this directory. Inherited from the Job definition.
tempfolder	The temporary folder to use for this step (defaults to job temp folder).
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.
fileorder	Order of file selection
kfjobid	Kingfisher job id. This is linked to the attribute topdf2 (see below).
Docoptions	PDF file open options derived from the options selected in the Job Designer.
Logfile	If specified, the output will be logged to a file with this name in %PDFJUNCTIONDIR%\logs or %TIFFJUNCTIONDIR%\logs

7.3.1.2 Attribute Values

The attribute values element contains the attributes that are displayed and edited in the UI.

Each attribute element comprises paired attributeid and currentvalue elements.

The attributeid can be used to link to the matching stepdetail by examining the Step Type file.

8 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.

8.1 Examples

8.1.1 Generate a searchable PDF C:\ADX Demo\Output\ US2007246939A1.pdf from a TIFF file

```
"C:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /source="C:\ADX Demo\In\Tiff\File\US2007246939A1.tif" /sourcetype=file /target="C:\ADX Demo\Output" /outputtype=pdf /operation=tifftopdf /ocr=true
```

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

```
"C:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /source="C:\ADX Demo\In\TIFF\Folder" /sourcetype=folder /target="C:\ADX Demo\Output" /output=outputfile.pdf /outputtype=pdf /ocr=true /autorotate /deskew /operation=mergetifftopdf
```

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

```
"C:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /source="C:\ADX Demo\In\PDF\Tree" /sourcetype=tree /target="C:\ADX Demo\Output" output=%FILENAME /ocr=true /outputtype=pdf /operation=ocrimagepdf /debug
```

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

```
"C:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /jobId=1001
```

8.1.5 Run an Autobahn job from an existing jobfile, e.g to run job 1002.

```
"C:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /jobfile="C:/Aquaforest/Autobahn DX/jobdef/1002.xml" /logfile="C:/Aquaforest/Autobahn DX/logs/1002"
```

8.1.6 Smart Redaction of Credit Card numbers only

```
"C:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /operation=smartredaction /ResourcePath="C:\Aquaforest\Autobahn DX\distribution\gdpicture\ocr" /Source="C:\ADX Demo\In\Redaction-Test-2.pdf" /target="C:\ADX Demo\Out" /Dictionary=eng /TempFolder="C:\ADX Demo\temp\00af66a1-ef8e-4a22-a9c3-4059be36f514" /PageRange="*" /MarkColor="black" /sourcetype=file /output=%FILENAME-No-Credit-card /createfolder=true /overwrite=true /logfile="C:\ADX Demo\Log\redaction_UnitTests.log" /RedactEmailAddresses=false /RedactPhoneNumbers=false /RedactVatIDs=false /RedactURIs=false /RedactIBANs=false /RedactVehicleIdentificationNumbers=false /RedactSocialSecurityNumbers=false /RedactPostalAddresses=false
```

8.2 General Job Options

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

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. Note: for some operations such as Merge operations the mask %FILENAME is not relevant and will be treated as a literal string
/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 maximum number of cores Autobahn DX will attempt to use 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	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 .

Parameter	Notes
	<ul style="list-style-type: none"> • tifftopdfocrimagepdf • topdf • pdftotiff • pdfprops • mergetiff • mergetifftopdf • splittiff • tiffinfo • mergepdf • splitpdf • extracttext • optimizepdf • sharepointdownload • sharepointupload • pdftopdfa • getxmlpdf • ocranyfile • ocranyfileex • barcode • highavailability • kingfisher • polling • azureupload • azuredownload • stamppdf • tifftopdfhand • ocrimagepdfhand • pdftoJson • comparerecognition • tifftopdfmicrosoft • ocrimagepdfmicrosoft • tifftopdfgoogle • ocrimagepdfgoogle • gdpicturepdfa • gdpicturecompression • senddocs • readmailbox • validatepdfa • linearizepdf • convertany • combineany • combinepdfs • pdftojpeg • pdftopng • pdftotiffgd • pdftotext • ocrpdfgd • pdfportfolio • smartredaction • detectsignatures • keyvaluepairs • patternredaction • patternhighlight • patternenumeration • getdocinformation • convertpdfotooffice • convertanytooffice

8.3 Split Parameters [used with splittiff and splitpdf operations]

Parameter	Notes				
/splitby	Pages, ranges, repeatingrange, bookmarks (PDF Only)				
/splitrange	<div>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><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></div>	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	<div>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, 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.</div>				
/RetainBookmarks	<div>Generated files will include bookmarks from the original file. Note: This works with the spit pdf operation only</div>				
/RetainMetadata	<div>Generated files will include metadata (such as Author and Title) from the original file. Note: This works with the spit pdf operation only</div>				

8.4 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 if you want converted pdf attachments to merged to the output pdf file. Otherwise, the converted files will be merged back to the pdf.
/RetainPDFAttachment	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: <ul style="list-style-type: none"> • PRN_PP_HANDOUT_VERTICAL_FIRST, • PRN_PP_HANDOUT_HORIZONTAL_FIRST
/PPOutPutType	Sets the output type, it only works with the PowerPoint files. The possible values are: Slides = PRN_PP_OUTTYPE_SLIDES Build slides = PRN_PP_OUTTYPE_BUILDSLIDES Two slide handouts = PRN_PP_OUTTYPE_TWOSLIDEHANDOUTS Three slide handouts = PRN_PP_OUTTYPE_THREESLIDEHANDOUTS Four slide handouts = PRN_PP_OUTTYPE_FOURSLIDEHANDOUTS Six slide handouts = PRN_PP_OUTTYPE_SIXSLIDEHANDOUTS Nine slide handouts = PRN_PP_OUTTYPE_NINESLIDEHANDOUTS Notes= PRN_PP_OUTTYPE_NOTESPAGES Outline = PRN_PP_OUTTYPE_OUTLINE
/PrintGraphics	Sets the graphics setting for printing. Print Full Resolution = PRN_PB_PRINT_GRAPHICS_HIGHRES Print Low Resolution = PRN_PB_PRINT_GRAPHICS_LOWRES Print Graphics = PRN_PB_PRINT_GRAPHICS_NONE
/convertbookmarks	For MS Word, convert bookmarks

Parameter	Notes
/bookmarkdepth	<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>
/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).
/imagedownscaling	If this property is set to True, then the resolution of images is reduced to the DPI value specified in the ImageDownsizeResolution property.
/imagejpegquality	The allowed value range is from 5 to 100 with 100 being the highest quality.
/imagedownsizeresolution	If the ImageDownscaling 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 fonts) 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 fonts) 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.

Parameter	Notes																																										
/marginleft	The margin in Inches.																																										
/marginright	The margin in Inches.																																										
/pagewidth	In Inches.																																										
/pageheight	In Inches.																																										
/paperorientation	PORTRAIT, LANDSCAPE or DEFAULT																																										
/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 PageWidth and PageHeight properties.</p> <p>Below is a list of Paper sizes</p> <table> <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> <tr> <td>PRN_MSO_PAPER_CUSTOM = 41</td><td>Custom Paper Size</td></tr> <tr> <td>PRN_MSO_PAPER_DSHEET = 13</td><td>D size sheet 22 x 34 in</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPE10 = 25</td><td>Envelope #10 4 1/8 x 9 1/2 in</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPE11 = 26</td><td>Envelope #11 4 1/2 x 10 3/8 in</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPE12 = 27</td><td>Envelope #12 4 3/4 x 11 in</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPE14 = 28</td><td>Envelope #14 5 x 11 1/2 in</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPE9 = 24</td><td>Envelope #9 3 7/8 x 8 7/8 in</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPEB4 = 29</td><td>Envelope B4 250 x 353 mm</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPEB5 = 30</td><td>Envelope B5 176 x 250 mm</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPEB6 = 31</td><td>Envelope B6 176 x 125 mm</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPEC3 = 32</td><td>Envelope C3 324 x 458 mm</td></tr> <tr> <td>PRN_MSO_PAPER_ENVELOPEC4 = 33</td><td>Envelope C4 229 x 324 mm</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_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_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																																										

Parameter	Notes
	<div>PRN_MSO_PAPER_ENVELOPEC5 = 34</div> <div>Envelope C5 162 x 229 mm</div> <div>PRN_MSO_PAPER_ENVELOPEC6 = 35</div> <div>Envelope C6 114 x 162 mm</div> <div>PRN_MSO_PAPER_ENVELOPEC65 = 36</div> <div>Envelope C65 114 x 229 mm</div> <div>PRN_MSO_PAPER_ENVELOPEDL = 37</div> <div>Envelope DL 110 x 220 mm</div> <div>PRN_MSO_PAPER_ENVELOPEITALY = 38</div> <div>Envelope Italy 110 x 230 mm</div> <div>PRN_MSO_PAPER_ENVELOPEMONARCH = 39</div> <div>Envelope Monarch 3.875 x 7.5 in</div> <div>PRN_MSO_PAPER_ENVELOPEPERSONAL = 40</div> <div>6 3/4 Envelope 3 5/8 x 6 1/2 in</div> <div>PRN_MSO_PAPER_ESHEET = 14</div> <div>E size sheet 34 x 44 in</div> <div>PRN_MSO_PAPER_EXECUTIVE = 5</div> <div>Executive 7 1/4 x 10 1/2 in</div> <div>PRN_MSO_PAPER_FANFOLDLEGALGERMAN = 15</div> <div>German Legal Fanfold 8 1/2 x 13 in</div> <div>PRN_MSO_PAPER_FANFOLDSTDGERMAN = 16</div> <div>German Std Fanfold 8 1/2 x 12 in</div> <div>PRN_MSO_PAPER_FANFOLDUS = 17</div> <div>US Std Fanfold 14 7/8 x 11 in</div> <div>PRN_MSO_PAPER_FOLIO = 18</div> <div>Folio 8 1/2 x 13 in</div> <div>PRN_MSO_PAPER_LEDGER = 19</div> <div>Ledger 17 x 11 in</div> <div>PRN_MSO_PAPER_LEGAL = 4</div> <div>Legal 8 1/2 x 14 in</div> <div>PRN_MSO_PAPER_LETTER = 2</div> <div>Letter 8 1/2 x 11 in</div> <div>PRN_MSO_PAPER_LETTERSMALL = 3</div> <div>Letter Small 8 1/2 x 11 in</div> <div>PRN_MSO_PAPER_NOTE = 20</div> <div>Note 8 1/2 x 11 in</div> <div>PRN_MSO_PAPER_QUARTO = 21</div> <div>Quarto 215 x 275 mm</div> <div>PRN_MSO_PAPER_STATEMENT = 22</div> <div>Statement 5 1/2 x 8 1/2 in</div> <div>PRN_MSO_PAPER_TABLOID = 23</div> <div>Tabloid 11 x 17 in</div>
/PrintColorType	<p>With this property, you can set the PowerPoint to print either with</p> <ul style="list-style-type: none"> • Color: PRN_PP_PRINT_COLOR • Grayscale: PRN_PP_PRINT_GRAYSCALE • Black and White: PRN_PP_PRINT_BLACK_AND_WHITE
/queuwaittimeout	<p>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.</p>
/disablescripdebugger	<p>Enables/Disables Internet Explorer's script debugger. It is highly recommended that Internet Explorer's script debugger is turned</p>

Parameter	Notes						
	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	<p>The default selection is PRN_DEFAULT_REVISION, meaning the current revision mode that is set inside Word will be used.</p> <p>Specify PRN_INLINE_REVISION to use inline/strikethrough revision mode and use PRN_BALLOON_REVISION to choose balloon revision mode.</p> <p>Used in conjunction with Include Document Markup</p> <p>Word revision mode constants:</p> <table border="1"> <tr> <td>0</td><td>Use the revision mode currently set inside Word</td></tr> <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>	0	Use the revision mode currently set inside Word	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)
0	Use the revision mode currently set inside Word						
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	<p>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.</p> <p>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.</p>						
/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	<p>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.</p> <p>Default is 200ms</p>						

Parameter	Notes
/HighQualityModeForGraphics	<p>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.</p> <p>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.</p>

8.5 [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 (Color)

8.6 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 layers that already exist 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	Removes dots below the specified pixel size, the possible input values range from 0-9. Note: To carry over the despeckled image to the output file, set the SavePreDespeckle flag explained below to false.
/SavePreDespeckle	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.

/autorotate	Rotates the page image if required																																																		
/language	<p>OCR Language dictionary to be used. The default is English.</p> <table> <tr> <th>LANGUAGE</th><th>/language Flag Value</th></tr> <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>Hungarian</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> </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	Hungarian	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																																																		
Hungarian	17																																																		
Bulgar	18																																																		
Slovenian	19																																																		
Latvian	20																																																		
Lithuanian	21																																																		
Estonian	22																																																		
Turkish	23																																																		
/graphicsproc	<ul style="list-style-type: none"> • "all": Treat all Graphics Areas as Text. • "box": Remove Box Lines in OCR Processing. 																																																		
/extractimages	<p>One of the following:</p> <ul style="list-style-type: none"> • Auto – TIFF Junction will select the most appropriate method. • 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. They 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. They 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	<p>See the Advanced Pre-Processing Section for a full explanation of this option.</p>										

8.7 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)																				
/pagemodefullscreen	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.

8.8 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 files that match this pattern will be included.
/ExcludePattern	Any file that matches this pattern will be excluded.

8.9 Extended OCR Steps

The Autobahn DX command line interface also supports the 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:\ADX Demo\In\PDF\File\US2007246939A1.pdf" /sourcetype=file /target="C:\ADX Demo\Output" /output=%FILENAME /outputtype=pdf,docx /operation=ocrimagepdf /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:\ADX Demo\In\TIFF\Folder" /sourcetype=folder /target="C:\ADX Demo\Output" /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:\ADX Demo\In\PDF\Tree" /target="C:\ADX Demo\Output" /sourcetype=tree /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	One or more of the following, separated by commas if more than one is required. RTF PDF DOCX CSV* SML (SpreadsheetML XML file)* HTM TXT *These output formats are suitable for table-oriented pages that can be mapped onto a spreadsheet format.
/ExtractImages	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.
/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	<p>This determines the PDF version of the generated PDF:</p> <ul style="list-style-type: none"> • 1.4 • 1.5 • 1.6 • 1.7 • 1.7 Extension Level 3 • 1.7 Extension Level 5 • 1.7 Extension Level 8 • PDF/A-1a • PDF/A-1b • PDF/A-2a • PDF/A-2b • PDF/A-3a • PDF/A-3b
/ValidatePDF/A	Set this flag to true if you want to validate the output PDF/A files.
/LanguageDetection	<p>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.</p> <p>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.</p> <p>Note: This is set to true by default.</p>
/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 exist. 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	<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. The options are:</p> <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 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	<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>
/Binarization	Whether or not to perform binarization on the document.
/Brightness	The brightness (higher values will make the result darker).
/Contrast	The contrast (lower values will make the result darker).
/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	<p>Set this to true if you want to use the deskewed image in the output document.</p> <p>Note: This property only applies when Keep Original Image is set to No</p>
/KeepDespeckle	<p>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.</p> <p>Note: This property only applies when KeepOriginalImage is set to No</p>
/KeepDarkBorderRemoval	<p>Set this to true if you want to use the image after dark borders have been removed, in the output document.</p> <p>Note: This property only applies when KeepOriginalImage is set to No</p>
/KeepPunchHoleRemoval	<p>Set this to true if you want to use the image after punch holes have been removed, in the output document.</p> <p>Note: This property only applies when KeepOriginalImage is set to No</p>
/resourcesfolder	<p>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.</p>

8.9.1 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	
Luxembourgish	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	
Korean	48	
Schinese	49	Simplified Chinese
Tchinese	50	Traditional Chinese
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	

Numeric	114	
Esperanto	115	
Maltese	116	
Zulu	117	
Afaan	118	
Asturian	119	
AzeriLatin	120	
Luba	121	
Papamiento	122	
Tatar	123	
Turkmen	124	
Welsh	125	
Arabic	126	<p>Note:</p> <ul style="list-style-type: none"> - You need to set about:blank to about:blank to use Arabic language - Arabic and English: Works only for Arabic texts with embedded English words. The result for a zone with only English will be empty.
Farsi	127	
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

8.10 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.

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

8.11.1 Standard

Parameter	Value
/operation	ocranyfile

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

8.11.2 Extended

Parameter	Value
/operation	ocranyfileex

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

8.11.3 GDPicture

There is no direct GDPicture OCR Any File to PDF, please use the GDPicture Convert Any File To PDF [convertany] followed by the OCR PDF to searchable PDF [ocrpdfgd] step.

8.12 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)	The renaming template to use for page ranges where no barcodes were identified. Allowed templates: - %INDEX%: Replaced by the current split index. - %FILENAME%sk: 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)	Remove lines from images. Value = 100.5
/Zones (string)	Only examine the region specified for barcode(s). Values =left(int), top(int), width(int), height(int) e.g. /Zones= 5,10,200,200
/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	PDF files get converted to TIFF files before the barcode is recognized, this flag represents the compression of the TIFF file. "CCITT4" or "LZW"
/PdfToImageDpi	PDF files get converted to TIFF files before the barcode is recognized, this flag represents the DPI of the TIFF file. 150,200,300
/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 end 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.
---------------	---

8.13 Stamp PDF – [used with stamppdf]

Parameter	Notes								
/stampoperation	<p>Select the way you want to stamp the PDF pages.</p> <p>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 represents 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. Using * or leaving it blank will process all pages.								
/stampstartnumber	The number that the page numbering will start with.								

Parameter	Notes
/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

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

Parameter	Notes
/storagename	The name of the Azure storage 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.
/ReplaceInvalidCharacters With (download only)	A pattern to replace any invalid character Windows File Storage in the file name before downloading. Invalid characters are: " * : < > ? \ Default replacement pattern is: _

8.15 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	The Extract Text Engine to use, below is a list of possible values. <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.

8.16 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.

8.17 GDPicture Steps

The Autobahn DX command line interface also supports the GDPicture module.

Examples:

1. OCR an image PDF to create a Searchable PDF

```
"c:\Aquaforest\Autobahn DX\bin\autobahndx.exe" /source="C:\ADX  
Demo\In\PDF\File\US2007246939A1.pdf" /sourcetype=file /target="C:\ADX Demo\Output"  
/outputtype=PDF /operation=ocrpdfgd /ocr=true /Tempfolder="C:\ADX Demo\Temp"  
/Dictionary="eng" /DictionaryPath="D:\Aquaforest\Autobahn  
DX\distribution\gdpicture\ocr" /PageRange="*" /DPI="300" /Output=pdf /Overwrite  
/logfile="C:\ADX Demo\Log\LogFile.log" /debug
```

8.17.1 Validate PDF/A – [used with validatepdfa]

Parameter	Notes																								
/operation	Use the operation validatepdfa																								
/debug	Log additional information																								
/pdfaoption	<p>PD/A option.</p> <p>Select the value from the following table</p> <table> <tr> <th>Value</th><th>Description</th></tr> <tr> <td>PDF/A-1a</td><td>The PDF conformance level is PDF/A-1a.</td></tr> <tr> <td>PDF/A-1b</td><td>The PDF conformance level is PDF/A-1b.</td></tr> <tr> <td>PDF/A-2a</td><td>The PDF conformance level is PDF/A-2a.</td></tr> <tr> <td>PDF/A-2b</td><td>The PDF conformance level is PDF/A-2b.</td></tr> <tr> <td>PDF/A-2u</td><td>The PDF conformance level is PDF/A-2u.</td></tr> <tr> <td>PDF/A-3a</td><td>The PDF conformance level is PDF/A-3a.</td></tr> <tr> <td>PDF/A-3b</td><td>The PDF conformance level is PDF/A-3b.</td></tr> <tr> <td>PDF/A-3u</td><td>The PDF conformance level is PDF/A-3u.</td></tr> <tr> <td>PDF/A-4</td><td>The PDF conformance level is PDF/A-4.</td></tr> <tr> <td>PDF/A-4e</td><td>The PDF conformance level is PDF/A-4e.</td></tr> <tr> <td>PDF/A-4f</td><td>The PDF conformance level is PDF/A-4f.</td></tr> </table>	Value	Description	PDF/A-1a	The PDF conformance level is PDF/A-1a.	PDF/A-1b	The PDF conformance level is PDF/A-1b.	PDF/A-2a	The PDF conformance level is PDF/A-2a.	PDF/A-2b	The PDF conformance level is PDF/A-2b.	PDF/A-2u	The PDF conformance level is PDF/A-2u.	PDF/A-3a	The PDF conformance level is PDF/A-3a.	PDF/A-3b	The PDF conformance level is PDF/A-3b.	PDF/A-3u	The PDF conformance level is PDF/A-3u.	PDF/A-4	The PDF conformance level is PDF/A-4.	PDF/A-4e	The PDF conformance level is PDF/A-4e.	PDF/A-4f	The PDF conformance level is PDF/A-4f.
Value	Description																								
PDF/A-1a	The PDF conformance level is PDF/A-1a.																								
PDF/A-1b	The PDF conformance level is PDF/A-1b.																								
PDF/A-2a	The PDF conformance level is PDF/A-2a.																								
PDF/A-2b	The PDF conformance level is PDF/A-2b.																								
PDF/A-2u	The PDF conformance level is PDF/A-2u.																								
PDF/A-3a	The PDF conformance level is PDF/A-3a.																								
PDF/A-3b	The PDF conformance level is PDF/A-3b.																								
PDF/A-3u	The PDF conformance level is PDF/A-3u.																								
PDF/A-4	The PDF conformance level is PDF/A-4.																								
PDF/A-4e	The PDF conformance level is PDF/A-4e.																								
PDF/A-4f	The PDF conformance level is PDF/A-4f.																								
/ErrorFolder	Path for error files																								
/Source	Source file to be validated.																								
/Target	Target path for the output file.																								
/TempFolder	Path for temporary folder (must be unique)																								

8.17.2 Linearize PDF – [used with linearizepdf]

Parameter	Notes
/operation	Use the operation linearizepdf
/debug	Log additional information
/EnableCompression	Add this parameter if you want to compress the PDF document during each save process.
/PackDocument	Add this parameter to pack the output PDF when saved to reduce its size.
/Source	Source file to be linearized.
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.3 Convert Any File to PDF – [used with convertany]

Parameter	Notes
/operation	Use the operation convertany
/debug	Log additional information
/ConvertAttachments	Convert any attachments
/AttachAttachments	Attach converted attachments rather than merge attachments
/EmailPageHeight="num"	Page height in points
/EmailPageWidth="num"	Page width in points
/EmailPageMarginBottom="num"	Page margin bottom in points
/EmailPageMarginLeft="num"	Page margin left in points
/EmailPageMarginRight="num"	Page margin right in points
/EmailPageMarginTop="num"	Page margin top in points
/EmailPreferOnePage	Add to output document as one page
/EnableICC	When present, specifies the ICC profile should be preserved during the conversion if it's present in the loaded document.
/HtmlEmulationType="num"	Specifies a type of media to emulate. 1 No emulation 2 Print 3 Screen
/HtmlPageHeight="num"	Page height in points
/HtmlPageWidth="num"	Page width in points
/HtmlPageMarginBottom="num"	Page margin bottom in points
/HtmlPageMarginLeft="num"	Page margin left in points
/HtmlPageMarginRight="num"	Page margin right in points
/HtmlPageMarginTop="num"	Page margin top in points
/HtmlPreferOnePage	Include to specify that the output document should contain a single page.
/HtmlPreferCSSPageSize	Include to give any CSS @page size declared in the page priority over what is declared in HtmlPageWidth and HtmlPageHeight. If not included, the renderer will scale the content to fit the paper size.
/LoadOnlyOnePage	Specifies that all executed actions with the loaded document will be processed using only the first page of the document.
/PageRange="range"	Range of pages to OCR, use "*" for all pages.

Parameter	Notes														
/PdfBitonalImageCompression="ID"	<p>Sets the scheme to be used to compress bitonal image data when converting/saving the currently loaded document to PDF format.</p> <table> <tr> <th>ID</th><th>Scheme</th></tr> <tr> <td>0</td><td>None</td></tr> <tr> <td>1</td><td>Flate</td></tr> <tr> <td>2</td><td>CCITT4</td></tr> <tr> <td>3</td><td>JPEG</td></tr> <tr> <td>4</td><td>JBIG2</td></tr> <tr> <td>5</td><td>JPEG2000</td></tr> </table>	ID	Scheme	0	None	1	Flate	2	CCITT4	3	JPEG	4	JBIG2	5	JPEG2000
ID	Scheme														
0	None														
1	Flate														
2	CCITT4														
3	JPEG														
4	JBIG2														
5	JPEG2000														
/JBIG2PMSThreshold	Specifies the threshold value for the JBIG2 encoder pattern matching and substitution between 0 and 1. Any number lower than 1 may lead to lossy compression. Default value is 0.85.														
/PdfColorImageCompression	Sets the scheme to be used to compress color image data when converting/saving the currently loaded document to PDF format.														
/PdfImageQuality	<p>Sets the level of quality used to compress images with a lossy compression scheme, which are embedded in the newly produced PDF document when converting/saving the currently loaded document to PDF format.</p> <p>It must be a value from 0 to 100. 0 means the worst quality and the best compression, 100 means the best quality and the worst compression.</p>														
/PDFUseDeflateOnJPEG	Include to specify if the converter will use additional Deflate compression for JPEG images in PDF output.														
/RasterizationDPI="DPI"	Sets the rendering resolution to be used when converting vector content to raster content, if any is included in the currently loaded document.														
/TiffEnableExifRotate	Include to specify that the tiff encoder will use Exif rotate flag to handle page rotations.														
/TimeoutMilliseconds="timeout in milliseconds"	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.														
/TxtFontBold	Specifies whether the font used for the resulting document when converting from the source txt file must have a bold style.														
/TxtFontItalic	Specifies whether the font used for the resulting document when converting from the source txt file must have an italic style.														
/TxtFontFamily="font family"	Specifies the name of the font to be used for the resulting document when converting from the source txt file.														

Parameter	Notes
/TxtFontSize="font size"	Specifies the text size, in points, to be used for the resulting document when converting from the source txt file.
/TxtPageHeight="num"	Page height in points
/TxtPageWidth="num"	Page width in points
/TxtPageMarginBottom="num"	Page margin bottom in points
/TxtPageMarginLeft="num"	Page margin left in points
/TxtPageMarginRight="num"	Page margin right in points
/TxtPageMarginTop="num"	Page margin top in points
/Source	Source file to be converted to PDF.
/Target	Target path for the output file. This can include the %FILENAME variable
/TempFolder	Path for temporary folder (must be unique)

8.17.4 Combine Any File- [used with combineany]

Parameter	Notes
/operation	Use the operation combineany
/debug	Log additional information
/ConvertAttachments	Convert any attachments
/AttachAttachments	Attach converted attachments rather than merge attachments
/EmailPageHeight="num"	Page height in points
/EmailPageWidth="num"	Page width in points
/EmailPageMarginBottom="num"	Page margin bottom in points
/EmailPageMarginLeft="num"	Page margin left in points
/EmailPageMarginRight="num"	Page margin right in points
/EmailPageMarginTop="num"	Page margin top in points
/EmailPreferOnePage	Add to output document as one page
/EnableICC	When present favor to preserve ICC profile, if present in the loaded document, during the conversion.
/HtmlEmulationType="num"	Specifies a type of a media to emulate.
/HtmlPageHeight="num"	Page height in points
/HtmlPageWidth="num"	Page width in points
/HtmlPageMarginBottom="num"	Page margin bottom in points
/HtmlPageMarginLeft="num"	Page margin left in points
/HtmlPageMarginRight="num"	Page margin right in points
/HtmlPageMarginTop="num"	Page margin top in points

Parameter	Notes														
/HtmlPreferOnePage	Include to specify that the output document should contain a single page.														
/HtmlPreferCSSPageSize	Include to give any CSS @page size declared in the page priority over what is declared in HtmlPageWidth and HtmlPageHeight. If not included, the renderer will scale the content to fit the paper size.														
/LoadOnlyOnePage	Specifies that all executed actions with the loaded document will be processed using only the first page of the document.														
/PageRange="range"	Range of pages to OCR, use "*" for all pages.														
/PdfBitonalImageCompression="ID"	<p>Sets the scheme to be used to compress bitonal image data when converting/saving the currently loaded document to PDF format.</p> <table border="1"> <thead> <tr> <th>ID</th><th>Scheme</th></tr> </thead> <tbody> <tr> <td>0</td><td>None</td></tr> <tr> <td>1</td><td>Flate</td></tr> <tr> <td>2</td><td>CCITT4</td></tr> <tr> <td>3</td><td>JPEG</td></tr> <tr> <td>4</td><td>JBIG2</td></tr> <tr> <td>5</td><td>JPEG2000</td></tr> </tbody> </table>	ID	Scheme	0	None	1	Flate	2	CCITT4	3	JPEG	4	JBIG2	5	JPEG2000
ID	Scheme														
0	None														
1	Flate														
2	CCITT4														
3	JPEG														
4	JBIG2														
5	JPEG2000														
/JBIG2PMSThreshold	Specifies the threshold value for the JBIG2 encoder pattern matching and substitution between 0 and 1. Any number lower than 1 may lead to lossy compression. Default value is 0.85.														
/PdfColorImageCompression	Sets the scheme to be used to compress color image data when converting/saving the currently loaded document to PDF format.														
/PdfImageQuality	<p>Sets the level of quality used to compress images with a lossy compression scheme, which are embedded in the newly produced PDF document when converting/saving the currently loaded document to PDF format.</p> <p>It must be a value from 0 to 100. 0 means the worst quality and the best compression, 100 means the best quality and the worst compression.</p>														
/PDFUseDeflateOnJPEG	Include to specify if the converter will use additional Deflate compression for JPEG images in PDF output.														
/RasterizationDPI="DPI"	Sets the rendering resolution to be used when converting vector content to raster content, if any is included in the currently loaded document.														
/TiffEnableExifRotate	Include to specify that the tiff encoder will use Exif rotate flag to handle page rotations.														

Parameter	Notes
/TimeoutMilliseconds="timeout in milliseconds"	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.
/TxtFontBold	Specifies whether the font used for the resulting document when converting from the source txt file must have a bold style.
/TxtFontItalic	Specifies whether the font used for the resulting document when converting from the source txt file must have an italic style.
/TxtFontFamily="font family"	Specifies the name of the font to be used for the resulting document when converting from the source txt file.
/TxtFontSize="font size"	Specifies the text size, in points, to be used for the resulting document when converting from the source txt file.
/TxtPageHeight="num"	Page height in points
/TxtPageWidth="num"	Page width in points
/TxtPageMarginBottom="num"	Page margin bottom in points
/TxtPageMarginLeft="num"	Page margin left in points
/TxtPageMarginRight="num"	Page margin right in points
/TxtPageMarginTop="num"	Page margin top in points
/Source	Source folder for the files to be combined.
/Target	Target path for the output file. This can include the %DIRNAME variable
/TempFolder	Path for temporary folder (must be unique)

8.17.5 Combine PDFs– [used with combinepdfs]

Parameter	Notes
/operation	Use the operation combinepdfs
/debug	Log additional information
/NumericalOrdering=true /NumericalOrdering=false	Choose whether the merging of file should happen in numerical order e.g. file1, file3, file11, file20, file101. Otherwise it will be ordered lexographically e.g. file1, file101, file11, file20, file3
/Source	Source folder for files to be combined.
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.6 Convert PDF to JPEG – [used with pdftojpeg]

Parameter	Notes
/operation	Use the operation pdftojpeg
/debug	Log additional information
/DPI="num"	The dpi resolution for rendering.
/Brightness="num"	Brightness adjustment between -100 and 100.
/Contrast="num"	Contrast adjustment between -100 and 100.
/Saturation="num"	Saturation adjustment between -100 and 100.
/Gamma="num"	Gamma adjustment between -100 and 100.
/AutoDeskew	When this flag is set, the image is skewed between ± 15 degrees to obtain the best information from the image.
/CropBlackBorders	Detects and removes margins consisting of black color around a GdPicture image.
/CropBlackBordersEx	Detects and sets to White, margins consisting of black color around a GdPicture image. This method does not have the same behavior as the CropBlackBorders. The black borders are not removed but are set to blank. Therefore, the image dimensions are kept the same.
/CropHeight="num"	Specifies the page height, in pixels, of the resulting output document after cropping
/CropWidth="num"	Specifies the page width, in pixels, of the resulting output document after cropping
/CropLeft="num"	Specifies the distance, in pixels, to crop from the top of the resulting document
/CropTop="num"	Specifies the distance, in pixels, to crop from the top of the resulting document
/Despeckle	Performs a 3x3 despeckle filter (also known as median filter)
/DespeckleMore	Performs a 5x5 despeckle filter (also known as median filter)
/EnableICM	Specifies if color correction is used for images embedding an ICC profile.
/RemoveHolePunch	Removes all punch holes situated on the margins. Be aware this can affect some barcode types.
/RemoveLines="num"	Perform line removal. Horizontal is 1 Vertical is 2
/ResizeNewHeight="num"	New image height in pixel
/ResizeNewWidth="num"	New image width in pixel

/ResizeInterpolateMode="num"	<p>The interpolation mode specifies the algorithm used when the image is stretched:</p> <table border="1"> <thead> <tr> <th>Mode</th><th>Value</th></tr> </thead> <tbody> <tr> <td>Default</td><td>0</td></tr> <tr> <td>Low</td><td>1</td></tr> <tr> <td>High</td><td>2</td></tr> <tr> <td>Bilinear</td><td>3</td></tr> <tr> <td>Bicubic</td><td>4</td></tr> <tr> <td>Nearest Neighbor</td><td>5</td></tr> <tr> <td>High Quality Bilinear</td><td>6</td></tr> <tr> <td>High Quality Bicubic</td><td>7</td></tr> </tbody> </table>	Mode	Value	Default	0	Low	1	High	2	Bilinear	3	Bicubic	4	Nearest Neighbor	5	High Quality Bilinear	6	High Quality Bicubic	7
Mode	Value																		
Default	0																		
Low	1																		
High	2																		
Bilinear	3																		
Bicubic	4																		
Nearest Neighbor	5																		
High Quality Bilinear	6																		
High Quality Bicubic	7																		
/RotateByAngle	Add parameter to rotate the image by an angle (see /RotationAngle)																		
/RotationAngle="num"	Rotate clockwise by this many degrees when /RotateByAngle parameter is present																		
/RotationType="num"	<p>Value determines rotate and/or flip action. Some actions result in the same final outcome, so the value is shared:</p> <table border="1"> <thead> <tr> <th>Action</th><th>Value</th></tr> </thead> <tbody> <tr> <td>Rotate180 Flip None</td><td>2</td></tr> <tr> <td>Rotate180 Flip X</td><td>6</td></tr> <tr> <td>Rotate180 Flip XY</td><td>0</td></tr> <tr> <td>Rotate180 Flip Y</td><td>4</td></tr> <tr> <td>Rotate270 Flip None</td><td>3</td></tr> <tr> <td>Rotate270 Flip X</td><td>7</td></tr> <tr> <td>Rotate270 Flip XY</td><td>1</td></tr> <tr> <td>Rotate 270 Flip Y</td><td>5</td></tr> </tbody> </table>	Action	Value	Rotate180 Flip None	2	Rotate180 Flip X	6	Rotate180 Flip XY	0	Rotate180 Flip Y	4	Rotate270 Flip None	3	Rotate270 Flip X	7	Rotate270 Flip XY	1	Rotate 270 Flip Y	5
Action	Value																		
Rotate180 Flip None	2																		
Rotate180 Flip X	6																		
Rotate180 Flip XY	0																		
Rotate180 Flip Y	4																		
Rotate270 Flip None	3																		
Rotate270 Flip X	7																		
Rotate270 Flip XY	1																		
Rotate 270 Flip Y	5																		

	Rotate 90 Flip None	1
	Rotate 90 Flip X	5
	Rotate 90 Flip XY	3
	Rotate 90 Flip Y	7
	Rotate None Flip None	0
	Rotate None Flip X	4
	Rotate None Flip XY	2
	Rotate None Flip Y	6
/PageRange="page range"	Range of pages to OCR, use "*" for all pages.	
/Source	Source file to be converted to JPEG.	
/Target	Target path for the output file.	
/TempFolder	Path for temporary folder (must be unique)	

8.17.7 Convert PDF to PNG – [used with pdftopng]

Parameter	Notes
/operation	Use the operation pdftopng

See [Convert PDF to JPEG](#) for the rest of the parameters

8.17.8 Convert PDF to TIFF – [used with pdftotiffgd]

Parameter	Notes																												
/operation	Use the operation pdftotiffgd																												
/TiffCompression	<table> <tr> <th>Value</th><th>Description</th></tr> <tr> <td>8</td><td>Deflate compression, as recognized by Adobe.</td></tr> <tr> <td>65536</td><td>Uses CCITT4 compression for bitonal image and LZW for others. This allows a mix of compression in a multipage tiff document.</td></tr> <tr> <td>3</td><td>CCITT Group 3 fax encoding</td></tr> <tr> <td>4</td><td>CCITT Group 4 fax encoding</td></tr> <tr> <td>32771</td><td>#1 w/ word alignment</td></tr> <tr> <td>32946</td><td>Deflate compression.</td></tr> <tr> <td>7</td><td>%JPEG DCT compression</td></tr> <tr> <td>5</td><td>Lempel-Ziv and Welch</td></tr> <tr> <td>32766</td><td>NeXT 2-bit RLE</td></tr> <tr> <td>1</td><td>No compression.</td></tr> <tr> <td>6</td><td>!6.0 JPEG</td></tr> <tr> <td>2</td><td>CCITT modified Huffman RLE</td></tr> <tr> <td>32809</td><td>ThunderScan RLE</td></tr> </table>	Value	Description	8	Deflate compression, as recognized by Adobe.	65536	Uses CCITT4 compression for bitonal image and LZW for others. This allows a mix of compression in a multipage tiff document.	3	CCITT Group 3 fax encoding	4	CCITT Group 4 fax encoding	32771	#1 w/ word alignment	32946	Deflate compression.	7	%JPEG DCT compression	5	Lempel-Ziv and Welch	32766	NeXT 2-bit RLE	1	No compression.	6	!6.0 JPEG	2	CCITT modified Huffman RLE	32809	ThunderScan RLE
Value	Description																												
8	Deflate compression, as recognized by Adobe.																												
65536	Uses CCITT4 compression for bitonal image and LZW for others. This allows a mix of compression in a multipage tiff document.																												
3	CCITT Group 3 fax encoding																												
4	CCITT Group 4 fax encoding																												
32771	#1 w/ word alignment																												
32946	Deflate compression.																												
7	%JPEG DCT compression																												
5	Lempel-Ziv and Welch																												
32766	NeXT 2-bit RLE																												
1	No compression.																												
6	!6.0 JPEG																												
2	CCITT modified Huffman RLE																												
32809	ThunderScan RLE																												

See [Convert PDF to JPEG](#) for the rest of the parameters

8.17.9 Convert PDF to Text – [used with pdftotext]

Parameter	Notes
/operation	Use the operation pdftotext
/debug	Log additional information
/PageRange	Range of pages to OCR, use "*" for all pages.
/PageSeparator="text"	Text that will separate the pages of text
/PageSeparatorPlacement=top /PageSeparatorPlacement=bottom	The placement of the text separator. The separator text can either go at the top or bottom of the page text. Default is bottom.
/PreserveParagraph	Specifies that the text extraction engine must preserve text paragraphs. Default false.
/ParagraphSeparator=	This property specifies the separator to be utilized for splitting paragraphs. It only takes effect when the PreserveParagraphs property is set to true.
/EnableOCR	Enables the use of the GdPicture OCR engine if the page is non-searchable.
/DictionaryPath=	Path to GdPicture Dictionary files
/Dictionary=	The codes of the chosen languages for OCR, separated by '+' e.g. eng+fra Codes can be found in the OCR Language Codes section.
/CopyPdf	If the flag is present, it copies the input pdf to the output location.
/Source	Source file to be converted to text.
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.10 OCR PDF to searchable PDF – [used with ocrpdfgd]

Parameter	Notes
/operation	Use the operation ocrpdfgd
/debug	Log additional information
/Dictionary	List of language codes, linked by '+' For example: /Dictionary="ENG+FRA"
/DictionaryPath	Path to GDPicture Dictionary files
/DPI	DPI used for rendering
/PageRange	Range of pages to OCR, use "*" for all pages.
/ThreadLimit	Limits that amount of pages can be processed concurrently
/IgnoreSearchablePages	Setting this flag will skip OCRing and pages that already contain visible text.
/IgnoreHiddenText	Setting this flag will skip OCRing and pages that already contain hidden text. Note: IgnoreSearchablePages must also be set
/Source	Source file to be OCRed.
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.11 Create PDF Portfolio – [used with pdfportfolio]

Parameter	Notes												
/operation	Use the operation pdfportfolio												
/debug	Log additional information												
/PdfPortfolioType	<table><tr><th>Type</th><th>Value</th><th>Description</th></tr><tr><td>Detail</td><td>0</td><td>The collection view is presented in details mode, with all information in the Schema dictionary presented in a multi-column format. This mode provides the most information to the user. The default value.</td></tr><tr><td>Hidden</td><td>2</td><td>The collection view is initially hidden, without preventing the user from obtaining a file list via explicit action.</td></tr><tr><td>Tile</td><td>1</td><td>The collection view is presented in tile mode, with each file in the collection denoted by a small icon and a subset of information from the Schema dictionary. This mode provides top-level information about the file attachments to the user.</td></tr></table>	Type	Value	Description	Detail	0	The collection view is presented in details mode, with all information in the Schema dictionary presented in a multi-column format. This mode provides the most information to the user. The default value.	Hidden	2	The collection view is initially hidden, without preventing the user from obtaining a file list via explicit action.	Tile	1	The collection view is presented in tile mode, with each file in the collection denoted by a small icon and a subset of information from the Schema dictionary. This mode provides top-level information about the file attachments to the user.
Type	Value	Description											
Detail	0	The collection view is presented in details mode, with all information in the Schema dictionary presented in a multi-column format. This mode provides the most information to the user. The default value.											
Hidden	2	The collection view is initially hidden, without preventing the user from obtaining a file list via explicit action.											
Tile	1	The collection view is presented in tile mode, with each file in the collection denoted by a small icon and a subset of information from the Schema dictionary. This mode provides top-level information about the file attachments to the user.											
/Source	Source folder.												
/Target	Target path for the output file.												
/TempFolder	Path for temporary folder (must be unique)												

8.17.12 Convert PDF to PDF/A – [used with gdpicturepdfa]

Parameter	Notes																						
/operation	Use the operation gdpicturepdfa																						
/debug	Log additional information																						
/AllowRasterization=false	Set this parameter to false, if you do not want the conversion engine to use page rasterization when the direct conversion and vectorization are not possible or allowed.																						
/AllowVectorization=false	Set this parameter to false, if you do not want the conversion engine to use page vectorization when the direct conversion is not possible. Vectorization produces vector based graphic elements where applicable, for example, fonts and paths, and combines them with image resources.																						
/Version="PDF/A Option"	<table> <tr> <th>Value</th><th>Description</th></tr> <tr> <td>PDF_A_1a</td><td>The PDF conformance level is PDF/A-1a.</td></tr> <tr> <td>PDF_A_1b</td><td>The PDF conformance level is PDF/A-1b.</td></tr> <tr> <td>PDF_A_2a</td><td>The PDF conformance level is PDF/A-2a.</td></tr> <tr> <td>PDF_A_2b</td><td>The PDF conformance level is PDF/A-2b.</td></tr> <tr> <td>PDF_A_2u</td><td>The PDF conformance level is PDF/A-2u.</td></tr> <tr> <td>PDF_A_3a</td><td>The PDF conformance level is PDF/A-3a.</td></tr> <tr> <td>PDF_A_3b</td><td>The PDF conformance level is PDF/A-3b.</td></tr> <tr> <td>PDF_A_3u</td><td>The PDF conformance level is PDF/A-3u.</td></tr> <tr> <td>PDF_A_4</td><td>The PDF conformance level is PDF/A-4.</td></tr> <tr> <td>PDF_A_4e</td><td>The PDF conformance level is PDF/A-4e.</td></tr> </table>	Value	Description	PDF_A_1a	The PDF conformance level is PDF/A-1a.	PDF_A_1b	The PDF conformance level is PDF/A-1b.	PDF_A_2a	The PDF conformance level is PDF/A-2a.	PDF_A_2b	The PDF conformance level is PDF/A-2b.	PDF_A_2u	The PDF conformance level is PDF/A-2u.	PDF_A_3a	The PDF conformance level is PDF/A-3a.	PDF_A_3b	The PDF conformance level is PDF/A-3b.	PDF_A_3u	The PDF conformance level is PDF/A-3u.	PDF_A_4	The PDF conformance level is PDF/A-4.	PDF_A_4e	The PDF conformance level is PDF/A-4e.
Value	Description																						
PDF_A_1a	The PDF conformance level is PDF/A-1a.																						
PDF_A_1b	The PDF conformance level is PDF/A-1b.																						
PDF_A_2a	The PDF conformance level is PDF/A-2a.																						
PDF_A_2b	The PDF conformance level is PDF/A-2b.																						
PDF_A_2u	The PDF conformance level is PDF/A-2u.																						
PDF_A_3a	The PDF conformance level is PDF/A-3a.																						
PDF_A_3b	The PDF conformance level is PDF/A-3b.																						
PDF_A_3u	The PDF conformance level is PDF/A-3u.																						
PDF_A_4	The PDF conformance level is PDF/A-4.																						
PDF_A_4e	The PDF conformance level is PDF/A-4e.																						

	PDF_A_4f	The PDF conformance level is PDF/A-4f.	
/Source	Source file.		
/Target	Target path for the output file.		
/TempFolder	Path for temporary folder (must be unique)		

8.17.13 Compress PDF – [used with gdpicturecompression]

Parameter	Notes
/operation	Use the operation gdpicturecompression
/debug	Log additional information
/Author="text"	Set PDF Metadata Author to supplied text
/Producer="text"	Set PDF Metadata Producer to supplied text
/Producername="text"	Set PDF Metadata Producername to supplied text
/Title="text"	Set PDF Metadata Title to supplied text
/RemoveAnnotations	When supplied, remove annotations
/RemoveBlankPages	When supplied, remove blank pages
/RemoveBookmarks	When supplied, remove bookmarks
/RemoveEmbeddedFiles	When supplied, remove embedded files
/RemoveFormFields	When supplied, remove form fields
/RemoveJavaScript	When supplied, remove JavaScript
/RemoveMetadata	When supplied, remove metadata. RemoveMetadata only removes XMP metadata in the document. PDF information (title, author, custom pdf info etc.) IF present, is not touched.
/removePageThumbnails	When supplied, remove page thumbnails
/PackFonts=false	Specifies the PDF fonts will not be packed to reduce their size.
/PackDocument=false	Specifies the PDF will not be packed to reduce its size.
/RecompressImages=false	Specifies the images from the PDF will not be recompressed.
/EnableMRC	Specifies MRC will be used for compressing the PDF contents.
/DownscaleResolutionMRC="resolution mrc"	Specifies the resolution (DPI) for downscaling the background layer by the MRC engine, if any. Default value is 100.
/PreserveSmoothing=false	Specifies that the MRC engine will not preserve smoothing between different layers.
/ImageQuality="image quality"	Specifies the quality to be used for the compression of the images from the PDF: See table below for settings.
/DownscaleImages=false	Specifies the images from the PDF will not be downscaled.
/DownscaleResolution=downscaleresolution	Specifies the resolution to be used to downscale images. Default value is 150.
/EnableColorDetection=false	Specifies whether color detection will not be performed on the images from the PDF.

Parameter	Notes
/EnableCharRepair	Specifies whether character repairing will be performed during bitonal conversion.
/EnableJPEG2000=false	Specifies that JPEG2000 compression scheme will not be used to compress the images of the PDF.
/EnableJBIG2=false	Specifies that the JBIG2 compression scheme will not be used to compress the bitonal images of the PDF.
/JBIG2PMSThreshold=threshold	Specifies the threshold value for the JBIG2 encoder pattern matching and substitution between 0 and 1. Any number lower than 1 may lead to lossy compression. Default value is 0.85.
/Source	Source folder.
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

Image quality settings

Description	Value
Image Quality Low	0
Image Quality Medium	1
Image Quality High	2
Image Quality Very High	3
Image Quality Very Very High	4
Image Quality Very Low	-1

8.17.14 Smart Redaction – [used with smartredaction]

Parameter	Notes
/operation	Use the operation smartredaction
/debug	Log additional information
/ResourcePath	Path to GDPicture Dictionary files
/Dictionary	List of language codes, linked by '+' For example: /Dictionary="ENG+FRA"
/RedactCreditCardNumbers=false	Specifies whether credit card numbers will not be redacted.
/RedactEmailAddresses=false	Specifies whether email addresses will not be redacted.
/RedactPhoneNumbers=false	Specifies whether phone numbers will not be redacted.
/RedactVatIDs=false	Specifies whether VAT IDs will not be redacted.
/RedactURIs=false	Specifies whether URIs will not be redacted.
/RedactIBANs=false	Specifies whether IBANs will not be redacted.
/RedactVehicleIdentificationNumbers=false	Specifies whether Vehicle Identification Numbers will not be redacted.
/RedactSocialSecurityNumbers=false	Specifies whether Social Security Number will not be redacted.
/RedactPostalAddresses=false	Specifies whether Postal Addresses will not be redacted.
/MarkColor	The color used to cover redacted information. The options are: Black, Transparent, Aqua, Teal, Navy, Yellow, Olive, Lime, Blue, Fuchsia, Purple, Red, Maroon, White, Gray, Silver, Green
/DetectOrientation=false	Specifies whether orientation will not be detected automatically.
/PageRange	Range of pages to redact, use "*" for all pages.
/Source	Source file to be redacted.
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.15 Detect Signatures – [used with detectsignatures]

Parameter	Notes
/operation	Use the operation detectsignatures
/debug	Log additional information
/SignedFilename=	The signed filename mask e.g., %FILENAME-signed, used for the file moved/copied to the SignedFilePath
/SignedFilePath=path	The full path (including file name) that the input file will be copy/moved to if a signature is detected
/CreateSignedPath /CreateSignedPath=false /CreateSignedPath=true	If the parameter CreateSignedPath is present or set to true, the signed file location will be created if it cannot be found. If set to false, the file will fail if the signed location does not exist.
/OverwriteSigned /OverwriteSigned=false /OverwriteSigned=true	If the parameter OverwriteSigned is present or set to true, when a signed file is moved or copied into the signed location, it will overwrite files of the same name. If set to false, the file will fail in this situation.
/SignedAction=action	Specifies if signed files will be copied or moved to the signed location. Use 'Copy' or 'Move' as values.
/AttachSignedDocToOutput /AttachSignedDocToOutput=false /AttachSignedDocToOutput=true	If parameter AttachSignedDocToOutput is present or set to true, then the signed source file will be attached to the no longer signed output file. Only applied if /SignedAction="copy" is set.
/Source	Source file to be checked for signatures
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.16 Extract Key Value Pairs – [used with keyvaluepairs]

Parameter	Notes
/operation	Use the operation keyvaluepairs
/debug	Log additional information
/DictionaryPath	Path to GDPicture Dictionary files
/Dictionary	The codes of the chosen languages for OCR, separated by '+' e.g. eng+fra Codes can be found in the OCR Language Codes section.
/DPI	DPI used for rendering
/KVPFormat	The output formats are separated is one of JSON, CSV or XML.
/Autorotate /Autorotate=false	Setting this to false will disable the ability to rotate pages if the text orientation is not correct. The default is true.
/TrimSymbols	Setting this to 'true' will remove any symbols from the start/end of values, except for the hash '#' or period '.' symbols.
/IncludeKeyBoundingBox	Setting this to 'Yes' will include the bounding box values for the key in the output.
/IncludeValueBoundingBox	Setting this to 'Yes' will include the bounding box values for the value in the output.
/IncludePageNumber	Setting this to 'Yes' will include the page number of the key value pair in the output.
/IncludeConfidence	Setting this to 'Yes' will include the confidence score of the key value pair in the output. Confidence is measured between 0 (no confidence) and 100 (full confidence).
/ConfidenceThreshold	The value of confidence (0-100) that a KVP must reach to be included in the output. Results under this confidence threshold will be discarded.
/IncludeType	Setting this to 'Yes' will include the data type of the key value pair in the output.
/ExpectedKeys	The path to a JSON file for the expected keys and synonyms. An example of how to structure this can be found in the Reference Guide.
/ExpectedKeysIsPath=true	Flag to show that the /ExpectedKeys flag is pointing to a JSON file location
/Source	Source file to be checked for key value pairs
/Target	Target path for the output file. No extension is needed, as extensions will be added based on the KVP Output Format
/TempFolder	Path for temporary folder (must be unique)

8.17.17 Pattern Redaction – [used with patternredaction]

Parameter	Notes
/operation	Use the operation patternredaction
/debug	Log additional information
/Pattern="regex"	A Regex pattern. The input pdf will be searched for matches to this Regex pattern, and any matches will be redacted.
/RedactedTermsPath=path	The path to a text file containing a list of terms to redact. Each line will be treated as a pattern, and any matches will be redacted.
/CaseSensitive	Determined whether or not the regex pattern matching should be case sensitive.
/Red=0	The amount of red color to be used for the redaction region color. Use a value between 0 and 255.
/Green=0	The amount of green color to be used for the redaction region color. Use a value between 0 and 255.
/Blue=0	The amount of blue color to be used for the redaction region color. Use a value between 0 and 255.
/Alpha=255	The transparency value of the resulting region color. Use the value between 0 (full transparency) and 255 (full opacity). The redacted text will be removed, and so will not be behind the colour layer if made transparent.
/Source	Source file to be redacted
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.18 Pattern Highlight – [used with patternhighlight]

Parameter	Notes
/operation	Use the operation patternhighlight
/Alpha=100	The transparency value of the resulting region color. Use the value between 0 (full transparency) and 255 (full opacity). The highlighted text will appear behind the highlight layer.

See [Pattern Redaction](#) for the rest of the parameters

8.17.19 Split PDF (GdPicture) – [used with splitpdfgdpicture]

Parameter	Notes
/operation	Use the operation splitpdfgdpicture
/debug	Log additional information
/RetainMetadata /RetainMetadata=false	When set to true, generated files will include metadata (such as Author and Title) from the original file.
/SplitType=Single /SplitType=Ranges /SplitType=Repeatingranges /SplitType=bookmarks	Sets the way that the input file will be split. <i>One of:</i> <ul style="list-style-type: none"> • Split into single pages • Split by ranges (See below) • Split by repeating ranges (See below) • Split by bookmarks
/SplitRange=range	Set of page ranges separated by commas that defines which pages from the original should be extracted.
/RepeatEvery=num	Apply the page range to each set of Page Ranges pages 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.
/RemoveUnusedResources /RemoveUnusedResources=false	Removes unused resources from a pdf file to minimize file size.
/Source	Source file to be redacted
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.20 Split By Barcode – [used with splitbybarcode]

Parameter	Notes
/operation	Use the operation splitbybarcode
/debug	Log additional information
/ReadQRCode	Set this to true to recognize QRCode barcodes.
/ReadMicroQR	Set this to true to recognize MicroQR barcodes.
/ReadDataMatrix	Set this to true to recognize DataMatrix barcodes.
/ReadPDF417	Set this to true to recognize PDF417 barcodes.
/ReadAztec	Set this to true to recognize Aztec barcodes.
/ReadMaxiCode	Set this to true to recognize MaxiCode barcodes.
/ReadIndustrial2of5	Set this to true to recognize Industrial2of5 barcodes.
/ReadInverted2of5	Set this to true to recognize Inverted2of5 barcodes.
/ReadInterleaved2of5	Set this to true to recognize Interleaved2of5 barcodes.
/Readlata2of5	Set this to true to recognize lata2of5 barcodes.
/ReadMatrix2of5	Set this to true to recognize Matrix2of5 barcodes.
/ReadCodabar	Set this to true to recognize Codabar barcodes.
/ReadBcdMatrix	Set this to true to recognize BcdMatrix barcodes.
/ReadDataLogic2of5	Set this to true to recognize DataLogic2of5 barcodes.
/ReadCode128	Set this to true to recognize Code128 barcodes.
/ReadCode93	Set this to true to recognize Code93 barcodes.
/ReadCode39	Set this to true to recognize Code39 barcodes.
/ReadEAN13	Set this to true to recognize EAN13 barcodes.
/ReadEAN8	Set this to true to recognize EAN8 barcodes.
/ReadUPCA	Set this to true to recognize UPCA barcodes.
/ReadUPCE	Set this to true to recognize UPCE barcodes.
/ReadADD5	Set this to true to recognize ADD5 barcodes.
/ReadADD2	Set this to true to recognize ADD2 barcodes.
/SplitRange=range	Specifies the page range to be scanned for barcodes. A value of * will scan every page for barcodes.
/Pattern="regex"	A regular expression that each barcode will be compared against. Only barcodes that match the pattern will be returned.
/DPI=int	DPI of TIFF images generated or converted from the source PDF File. These images are then scanned for barcodes.

/RetainMetadata /RetainMetadata=false	When set to true, generated files will include metadata (such as Author and Title) from the original file.
/RemoveUnusedResources /RemoveUnusedResources=false	Removes unused resources from a pdf file to minimize file size.
/Left=int	X coordinate of the Top Left Point of the rectangle you want to recognize the barcode.
/Top=int	Y coordinate of the Top Left Point of the rectangle you want to recognize the barcode.
/Width=int	Width of the rectangle you want to recognize the barcode.
/Height=int	Height of the rectangle you want to recognize the barcode.
/Source	Source file to be redacted
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.21 Pattern Enumeration

Parameter	Notes
/operation	Use the operation patternenumeration
/debug	Log additional information
/Pattern="regex"	A Regex pattern. The input pdf will be searched for matches to this Regex pattern, and any matches will be enumerated.
/RedactedTermsPath=path	The path to a text file containing a list of terms to enumerate. Each line will be treated as a pattern, and any matches will be enumerated.
/CaseSensitive	Determined whether or not the regex pattern matching should be case sensitive.
/PassThrough	Determines if the source file should be copied to the output folder alongside the report
/Source	Source file to be enumerated
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.22 Get Document Information

Parameter	Notes
/operation	Use the operation getdocinformation
/debug	Log additional information
/OutputFormat	The output format for the report <ul style="list-style-type: none">• txt• csv• json• xml
/PassThrough	Determines if the source file should be copied to the output folder alongside the report
/Source	Source pdf file
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.23 Convert PDF To Office

Parameter	Notes
/operation	Use the operation convertpdfotooffice
/debug	Log additional information
/OfficeFormat	The output format for the report <ul style="list-style-type: none">• docx• pptx• xlsx• svg
/EnableICC	When present, specifies the ICC profile should be preserved during the conversion if it's present in the loaded document.
/PageRange	Range of pages to convert, use "*" for all pages.
/TimeoutMilliseconds	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.
/Source	Source file to be converted
/Target	Target path for the output file.
/TempFolder	Path for temporary folder (must be unique)

8.17.24 Convert Any File To Office

Parameter	Notes
/operation	Use the operation convertanytooffice
/debug	Log additional information
/EmailInjectHeader	Specifies whether the email header should be injected in the output document.
/ConvertEmailAttachmentsToOffice	Convert any attachments to office format
/EmailAttachmentsFilter	A regular expression that specifies the email attachments that will be picked up and converted. Files that don't match will be ignored.
/EmailPageHeight="num"	Page height in points
/EmailPageWidth="num"	Page width in points
/EmailPageMarginBottom="num"	Page margin bottom in points
/EmailPageMarginLeft="num"	Page margin left in points
/EmailPageMarginRight="num"	Page margin right in points
/EmailPageMarginTop="num"	Page margin top in points
/EmailPreferOnePage	Add to output document as one page
/EnableICC	When present, specifies the ICC profile should be preserved during the conversion if it's present in the loaded document.
/HtmlEmulationType="num"	Specifies a type of media to emulate. 1 No emulation 2 Print 3 Screen
/HtmlPageHeight="num"	Page height in points
/HtmlPageWidth="num"	Page width in points
/HtmlPageMarginBottom="num"	Page margin bottom in points
/HtmlPageMarginLeft="num"	Page margin left in points
/HtmlPageMarginRight="num"	Page margin right in points
/HtmlPageMarginTop="num"	Page margin top in points
/HtmlPreferOnePage	Include to specify that the output document should contain a single page.
/HtmlPreferCSSPageSize	Include to give any CSS @page size declared in the page priority over what is declared in HtmlPageWidth and HtmlPageHeight. If not included, the renderer will scale the content to fit the paper size.
/LoadOnlyOnePage	Specifies that all executed actions with the loaded document will be processed using only the first page of the document.
/PageRange="range"	Range of pages to OCR, use "*" for all pages.

Parameter	Notes														
/PdfBitonallImageCompression="ID"	<p>Sets the scheme to be used to compress bitonal image data when converting/saving the currently loaded document to PDF format.</p> <table> <tr> <th>ID</th><th>Scheme</th></tr> <tr> <td>0</td><td>None</td></tr> <tr> <td>1</td><td>Flate</td></tr> <tr> <td>2</td><td>CCITT4</td></tr> <tr> <td>3</td><td>JPEG</td></tr> <tr> <td>4</td><td>JBIG2</td></tr> <tr> <td>5</td><td>JPEG2000</td></tr> </table>	ID	Scheme	0	None	1	Flate	2	CCITT4	3	JPEG	4	JBIG2	5	JPEG2000
ID	Scheme														
0	None														
1	Flate														
2	CCITT4														
3	JPEG														
4	JBIG2														
5	JPEG2000														
/JBIG2PMSThreshold	Specifies the threshold value for the JBIG2 encoder pattern matching and substitution between 0 and 1. Any number lower than 1 may lead to lossy compression. Default value is 0.85.														
/PdfColorImageCompression	Sets the scheme to be used to compress color image data when converting/saving the currently loaded document to PDF format.														
/PdfImageQuality	<p>Sets the level of quality used to compress images with a lossy compression scheme, which are embedded in the newly produced PDF document when converting/saving the currently loaded document to PDF format.</p> <p>It must be a value from 0 to 100. 0 means the worst quality and the best compression, 100 means the best quality and the worst compression.</p>														
/PDFUseDeflateOnJPEG	Include to specify if the converter will use additional Deflate compression for JPEG images in PDF output.														
/RasterizationDPI="DPI"	Sets the rendering resolution to be used when converting vector content to raster content, if any is included in the currently loaded document.														
/RenderSheetHeadersAndFooters	Specifies that the xls and xlsx headers and footers should be rendered. Affects XLSX/XLS input only.														
/SplitExcelSheetsIntoPages	Specifies that xls and xlsx sheets should be split into pages according to the PageSetup element of each sheet. Affects XLSX/XLS input only.														
/SpreadsheetBottomMarginOverride	Specifies the spreadsheet bottom margin height in millimeters. If height is not given or negative, the margin specified in the document will be used instead. Affects XLSX/XLS input only.														
/SpreadsheetLeftMarginOverride	Specifies the spreadsheet left margin width in millimeters. If height is not given or negative, the margin specified in the document will be used instead. Affects XLSX/XLS input only.														

Parameter	Notes
/SpreadsheetMaximumContentHeightPerSheet	Decimal value indicating the maximum height of the sheet content, in millimeters. Maximum content height ignores header and footer height. Affects XLSX/XLS input only.
/SpreadsheetMaximumContentWidthPerSheet	Decimal value indicating the maximum width of the sheet content, in millimeters. Maximum content width ignores margins. Affects XLSX/XLS input only.
/SpreadsheetPageHeightOverride	Specifies the spreadsheet page height in millimeters. If height is not given or not positive, the page height specified in the document will be used instead. Affects XLSX/XLS input only.
/SpreadsheetPageWidthOverride	Specifies the spreadsheet page width in millimeters. If width is not given or not positive, the page width specified in the document will be used instead. Affects XLSX/XLS input only.
/SpreadsheetRenderOnlyPrintArea	For spreadsheet, specifies that for each sheets only the print areas must be rendered. If no print area are found the whole sheets will be rendered. Affects XLSX/XLS input only.
/SpreadsheetRightMarginOverride	Specifies the spreadsheet right margin width in millimeters. If width is not given or not positive, the page height specified in the document will be used instead. Affects XLSX/XLS input only.
/SpreadsheetTopMarginOverride	Specifies the spreadsheet top margin height in millimeters. If width is not given or not positive, the page width specified in the document will be used instead. Affects XLSX/XLS input only.
/TiffEnableExifRotate	Include to specify that the tiff encoder will use Exif rotate flag to handle page rotations.
/TimeoutMilliseconds="timeout in milliseconds"	Specifies the timeout of the subsequent conversion process, in milliseconds. Default value is -1, which means no timeout.
/TxtFontBold	Specifies whether the font used for the resulting document when converting from the source txt file must have a bold style.
/TxtFontItalic	Specifies whether the font used for the resulting document when converting from the source txt file must have an italic style.
/TxtFontFamily="font family"	Specifies the name of the font to be used for the resulting document when converting from the source txt file.
/TxtFontSize="font size"	Specifies the text size, in points, to be used for the resulting document when converting from the source txt file.
/TxtPageHeight="num"	Page height in points
/TxtPageWidth="num"	Page width in points
/TxtPageMarginBottom="num"	Page margin bottom in points
/TxtPageMarginLeft="num"	Page margin left in points
/TxtPageMarginRight="num"	Page margin right in points

Parameter	Notes
/TxtPageMarginTop="num"	Page margin top in points
/Source	Source file to be converted to PDF.
/Target	Target path for the output file. This can include the %FILENAME variable
/TempFolder	Path for temporary folder (must be unique)

9 Autobahn DX .NET API

The Autobahn DX .NET API allows a .NET application to execute a selected job definition.

A sample Visual Studio project is provided in the <Autobahn DX installation folder>\code examples\AutobahnAPIExample folder.

9.1 API Example (Local Job)

```
/// <summary>
/// Call Autobahn and run specified JobID
/// </summary>
/// <param name="jobID">The Autobahn JobID</param>
/// <param name="configFile">Full file path to the Autobahn config file</param>
/// <param name="timeOut">Time out in ms. Default is zero which is no time out</param>
/// <returns>Exit Code - zero is success</returns>
public int Call(int jobID, string configFile, int timeOut = 0)
{
    int exitCode;
    Job job = null;
    try
    {
        // Create new Job instance for the supplied JobID and the Autobahn config file
        job = new Job(jobID, configFile);
        // Start the job
        job.Start();
        // If a time out value is supplied, wait until time out complete
        if (timeOut > 0)
        {
            job.WaitForExit(timeOut);
        }
        else
        {
            string started = job.JobFinished();
            job.WaitForExit();
            Console.WriteLine("Job execution finished:" + job.JobFinished());
            Console.WriteLine("Exit code is:" + job.Jobstatus());
        }
        // Job status analysis
        if (job.Jobstatus().ToUpper().Equals("ERROR"))
        {
            exitCode = -1;
        }
        else if (job.Jobstatus().ToUpper().Equals("STOPPED"))
        {
            if (job.JobFinished().ToUpper().Equals("TRUE"))
            {
                exitCode = 0;
            }
            else
            {
                exitCode = -1;
            }
        }
        else
        {
            exitCode = -2;
        }
    }
    catch (Exception ex)
    {
        exitCode = -1;
        Console.WriteLine(ex.Message);
    }
    finally
    {
        if (job != null)
        {
            job.Dispose();
        }
    }
    return exitCode;
}
```

9.2 API details

Constructors	<code>public Job(int jobid)</code> <i>Create a Job object using an existing jobid.</i>
	<code>public Job(string jobdeffile)</code> <i>Create a Job object using a temporary job definition file.</i>
	<code>public Job(int jobid, string config)</code> <i>Create a Job object using an existing jobid and config file with the configuration of the remote machine</i>
	<code>public Job(string jobdeffile, string config)</code> <i>Create a Job object using a temporary job definition file and config file with the configuration of the remote machine</i>
Methods	<code>public Start()</code> <i>Starts the Job.</i>
	<code>public void WaitForExit(int ms)</code> <i>Waits for the job to exit, for up to ms milliseconds. The job is stopped if it has not completed.</i>
	<code>public void WaitForExit()</code> <i>Waits indefinitely for the job to complete.</i>
	<code>public virtual void Dispose()</code> <i>Disposes of the resources associated with the Job.</i>
	<code>Public string Jobstatus()</code> <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>
	<code>Public string JobFinished()</code> <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>
	<code>Public void ClearError(string JobStatusPath)</code> When a job is error, <code>ClearError(string jobStatusPath)</code> will clear the error job and set the job status to stopped.
	<code>Public string GetLastRunDate()</code> Returns the last Date and Time the job executed.
	<code>Public List<DocumentResults> GetDocumentsStatus()</code> Returns a list of files processed and their status.

10 OCR Properties File and The Advanced Pre-Processing Option

10.1 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 - Hungarian
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 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 by setting it to false.

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.

10.2 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.

```
<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>
```

11 Upgrade Job Definitions

As new versions of Autobahn are released, the way that specific jobs are defined may change to incorporate new options or remove obsolete properties. This may require any job definitions files that were created on previous Autobahn versions to be upgraded to fit the new definitions that current Autobahn can recognize.

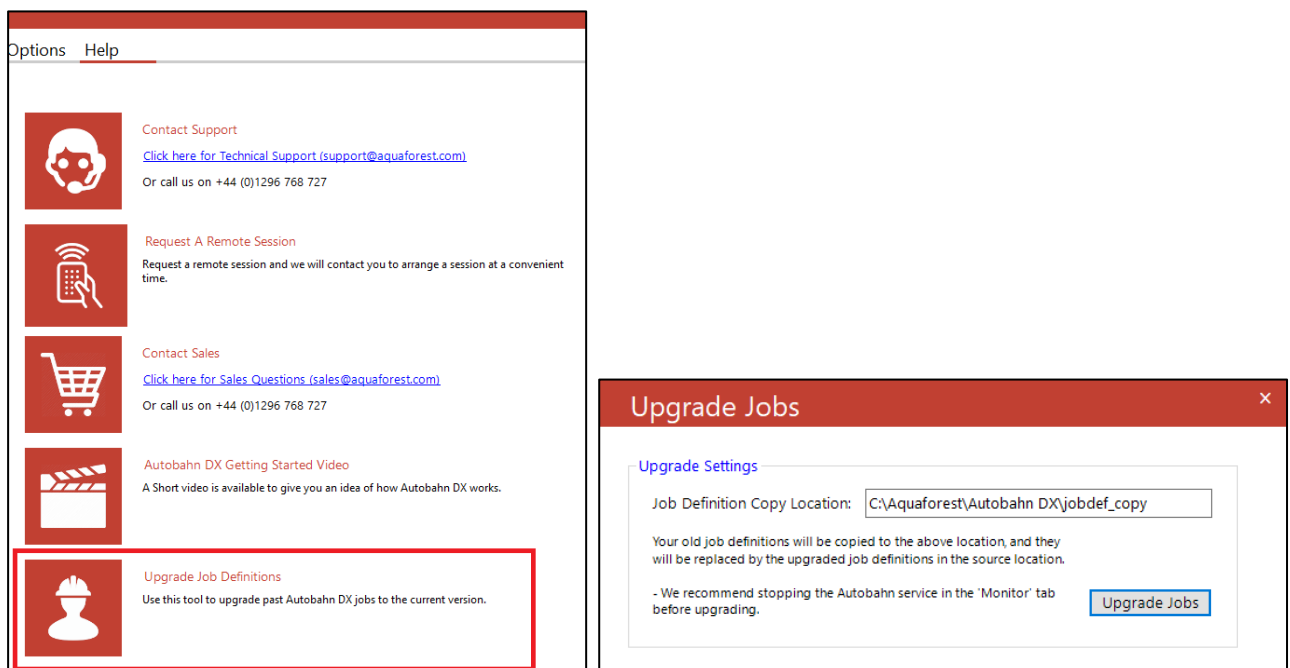
11.1 Automatic Upgrades

In the latest version of Autobahn, jobs are upgraded automatically upon opening the UI. A window will show the progress of the upgrades, but it will be very quick if there are no jobs that need upgrading. The previous job definitions are archived in the "...Aquaforest\Autobahn DX\jobdef_copy" location, and are upgraded in place. The upgrade only occurs on jobs with an internal **adxversion** value that does not match the current version of Autobahn running. Jobs matching this version will be ignored.

11.2 Upgrading Manually

In the **'Help'** tab, there is a button displayed with the text **"Upgrade Job Definitions"**. Clicking this opens a tool to carry out the job definition upgrades. This tool will create a copy of the old job definitions, and then upgrade the files into up-to-date job definitions that match the current Autobahn version's steps. We recommend checking that the settings of the job match your requirements after upgrading, as some fields may have been set to their default value.

The picture below shows the location of the **Job Upgrader**, and the window of the application. Files in the "...Aquaforest\Autobahn DX\jobdef" folder will be the jobdef file to be upgraded. You can manually set the location where the original job definition files are copied to.



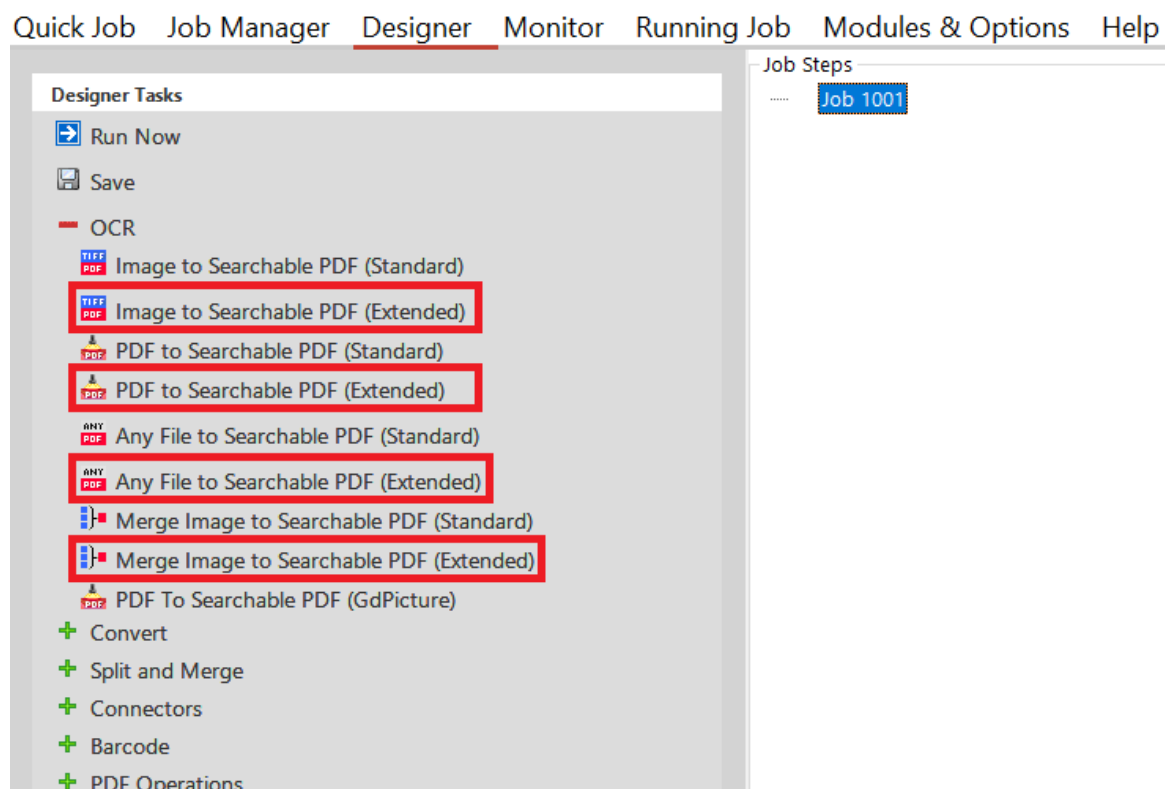
12 Extended OCR Module

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

Search: Clear

1 - Conversion Settings	
Output File Name	%FILENAME_searchable.pdf
Output File Type	PDF,RTF,DOCX
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	
OCR Language 6	
OCR Language 7	
OCR Language 8	
Automatic language detection	No
Autorotate	No
Deskew	No
Despeckle	0
Convert to TIFF	Yes
Retain Bookmarks	No
Retain Metadata	No

Output File Type
Choose the type of output file. One or more of the following, separated by commas:
CSV,DOCX,EPUB,EXCELML,HTM,OPENTXT,PDF,RTF,TXT,WORDML,XLSX,XPS

Annotations:

- Different output file (points to Output File Type)
- New IRIS OCR Engine (points to OCR Engine)
- You can select up to 8 different languages (points to OCR Language 1-8)

12.2 IHQC Module

The IHQC Module is a module included with the 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

Search: Clear

JPEG2000 Compression	No
JPEG2000 Compression Mode	Use quality factor
JPEG2000 Compression Value	
iHQC Compression	Yes
iHQC Compression Level	1
iHQC Quality Factor	Medium
No OCR	No
Advanced Despeckle	
Remove White Pixels	No
Dilate	
Advanced Deskew	No

iHQC Compression
Apply intelligent High Quality Compression.

Annotations:

- IHQC (points to iHQC Compression)

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] /ihqclevel=[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 /ihqclevel=4 /ihqcqf=5 /ocrengine=1

The following parameters are needed to use IHQC:

Parameter	Notes										
/mrc	Apply intelligent High-Quality Compression. True or False										
/IHQCLLevel	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></table>	Quality	Code	Maximal quality	9	Very high quality	8	High quality	7	Favor quality over size	6
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	

12.3 Image Requirements

Extended OCR has the following image limits:

- Max Height = 32,768 pixels
- Max Width = 32,768 pixels
- Max Size = 75,000,000 pixels

These limitations are also valid when working with image preprocessing: image resize, rotate, etc. The following table displays the image sizes at maximum resolutions:

Paper standard	Size	Maximum resolution	Image at maximum resolution
A0	33.11 x 46.81 in 841 x 1189 mm	219	7251 x 10251
A1	23.39 x 33.11 in 594 x 841 mm	311	7274 x 10297
A2	16.54 x 23.39 in 420 x 594 mm	440	7277 x 10291
A3	11.69 x 16.54 in 297 x 420 mm	622	7271 x 10287
A4	8.27 x 11.69 in 210 x 297 mm	880	7277 x 10287
A5	5.83 x 8.27 in 148 x 210 mm	1200	7270 x 10312
A6	4.13 x 5.83 in 105 x 148 mm	1200	7285 x 10284
Letter	8.5 x 11 in 216 x 279 mm	895	7607 x 9845
Legal	8.5 x 14 in 216 x 356 mm	793	6740 x 11102
Junior legal	8.0 x 5.0 in 203 x 127 mm	1200	10952 x 6845
Ledger	17 x 11 in 432 x 279 mm	633	10761 x 6963
Tabloid	11 x 17 in 279 x 432 mm	633	6963 x 10761

There are various workarounds that you can try if you experience an image size issue.

- 1) Reduce the DPI
- 2) Try using the Standard OCR engine instead
- 3) Consider exporting to PDF from the CAD application - these exported PDF files are more and more searchable, even for polygonised and SHX text (depending on CAD application). Try the following applications:
 - a. [How to create selectable and searchable text in a PDF from AutoCAD](#)
 - b. [Searchable Text \[AutoCAD to PDF\]](#)

13 Barcode Support

Autobahn DX can Split and rename PDF/TIFF files based on barcodes found within the file.

13.1 Barcode Formats

The following Barcode Formats are supported by Autobahn DX

Barcode Format	Code	Group
CODABAR	2	1D
CODE_39	4	1D
CODE_93	8	1D
CODE_128	16	1D
EAN_8	64	1D
EAN_13	128	1D
QR_CODE	2048	2D
UPC_A	16384	1D

14 High Availability

14.1 Background information

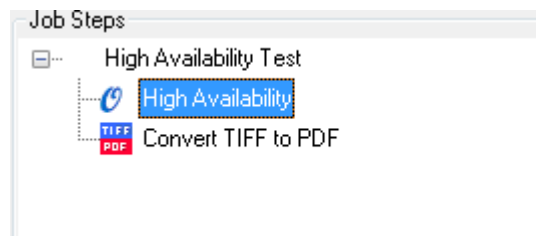
The high availability step in Autobahn DX is designed to utilize two instances of the product running on separate hosts, each instance will be configured with a job which will be scheduled to run at given intervals using the Autobahn scheduler.

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

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

14.2 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.

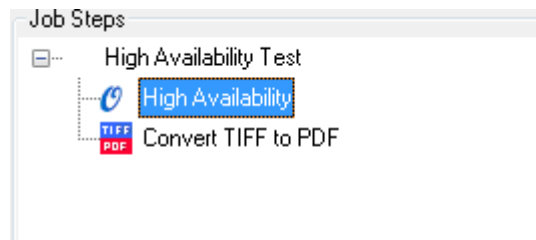
Job Properties	
Job ID	1012
Job Name	High Availability Test
Source Folder	\\HMNHD-T12RX1\qa\adx_high_availability_in
Target Folder	\\HMNHD-T12RX1\qa\adx_high_availability_out

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 (Controller | Replica)
 - 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	Controller
Shared Status File	\\HMNHD-T12RX1\Backups\Shrevin\qa\adx_high_availability_shared\shared.txt
2 - Pair Settings	
Hostname	aquaforestqa
ADX Install Path	C:\Aquaforest\Autobahn DX
Job ID	1007

14.3 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-T12RX1\Backups\qa\adx_high_availability_in
Target Folder	\\HMNHD-T12RX1\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 (Controller | Replica)
 - 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	Replica
Shared Status File	\\HMNHD-T12RX1\Backups\Shrevin\qa\adx_high_availability_shared\shared.txt
2 - Pair Settings	
Hostname	filesaver
ADX Install Path	D:\Aquaforest\Autobahn DX
Job ID	1012

14.4 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 controller host the scheduler will attempt to start the job after another minute.

When the scheduler attempts to start the job on the replica host, the replica will check to see if the controller 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.

15 ToPDF (BCL easyPDF)

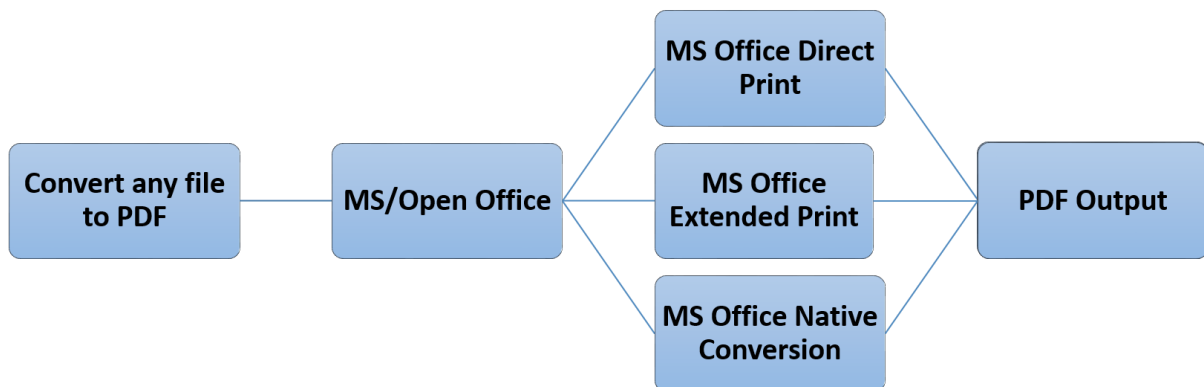
This converts any file type to PDF, the searchability status of the output PDF depends on the source document, E.G MS Office Input file.

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.

15.1.1 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.



15.1.1.1 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.

15.1.1.2 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.

15.1.1.3 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.

15.1.2 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)	✓	✗	✗

15.1.2.1 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" />
```

15.1.3 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.

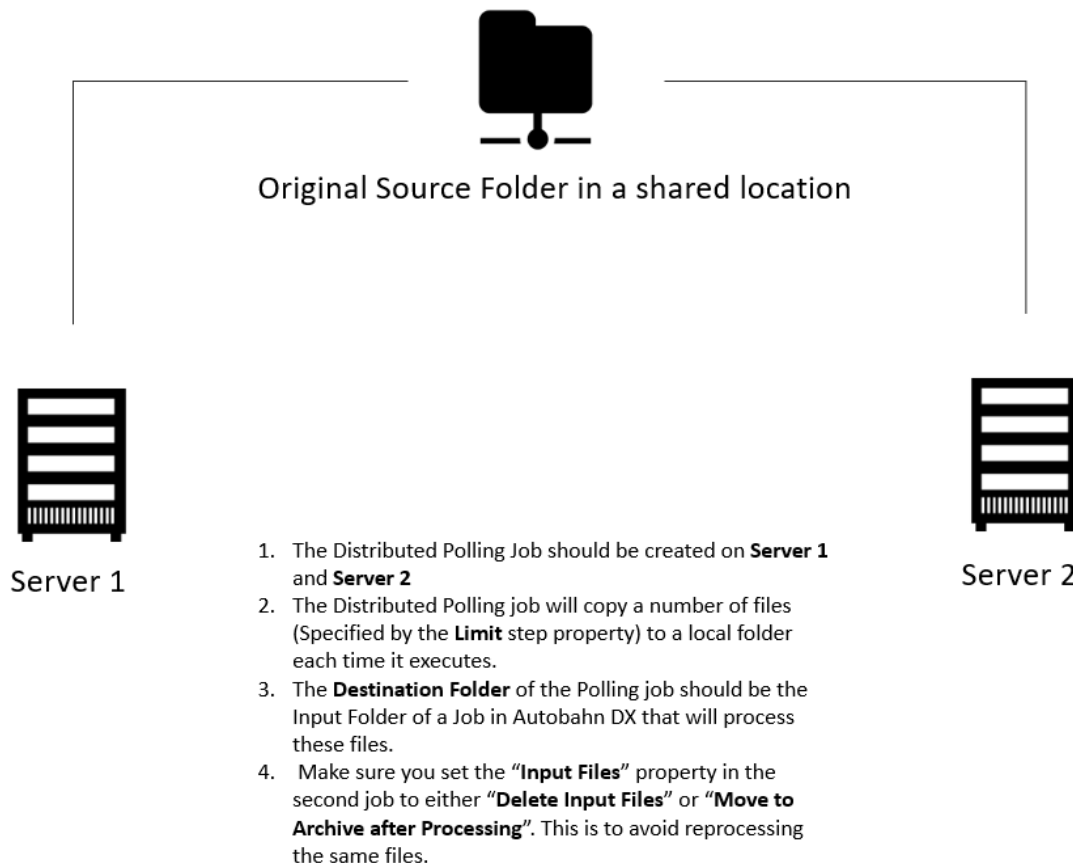
Configuration	Conversion Method	File Types	Comment
AutoExtension	Direct Print	All file types	This option works with any file extension that has an application that is compatible with the BCL printer
AutoExtensionEx	Extended Print	All file types	
AutoExtensionNative	Native Conversion	MS Office files	
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...	
WordExtensionEx NativeOfficePDF=true	Native Conversion	(All files that can be opened in MS Word)	
WordExtensionEx NativeOfficePDF=false	Extended Print		
ExcelExtension	Direct Print	.xls, .csv, .xlsx...	
ExcelExtensionEx NativeOfficePDF=true	Native Conversion	(All files that can be opened in Excel)	
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	
IEExtendedExtension	Extended Print	All IE files.	
HTMLExtension	Direct Print	.htm, .html, eml	

Configuration	Conversion Method	File Types	Comment
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.

16 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.

16.1 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.

16.1.1 Distributed Polling Job

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

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

Monitor Running Job Modules & Options Help Job 1008

Job Steps

- Job 1008
 - Distributed Polling

Properties Logging **Schedule** Processing Alerts

Job ID 1008

Job Name Job 1008

a Source Folder \\share\QA Folder\QA FILES

b Destination Folder J:\QA Folder\QA FILES

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

Input Files Leave input files after Processing

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

Filter Files Include with Document Count Limit

Filter Expression Batch Size 2

- 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 shared location 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.

16.1.2 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 is used in the **Distributed Polling** 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\Output 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

17 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 Handwriting 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 four step types to the **Advanced** section of the **Job Designer** tab of Autobahn DX, the steps are named:

- Image to Searchable PDF (Microsoft Cloud OCR)
- PDF to Searchable PDF (Microsoft Cloud OCR)
- Image to Searchable PDF (Google Cloud OCR)
- PDF to Searchable PDF (Google 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.
End Point (Microsoft Only)	The URL to the cognitive services endpoint where the OCR will be performed
Subscription Key (Microsoft Only)	The subscription key to the above endpoint if you are using Microsoft. See section 18.1 for more.
Google Key File Path (Google Only)	The path to the JSON subscription key file if you are using Google. See section 18.2 for more.
Text Recognition Mode (Microsoft only)	Types of text to recognize <ul style="list-style-type: none">• Handwritten• Printed Default is Printed

Handwritten Results Retries (Microsoft only)	The number of times to wait for the handwritten OCR results
Handwritten Results Wait (Microsoft only)	The amount of time (in milliseconds) to wait between each retry
OCR Language	<p>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.</p> <p>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</p>
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	<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.
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.										
Advanced Flags	Command line flags to be passed through to the underlying executable.										
Maximum Cores	This specifies the number of 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 step 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 border="1"> <thead> <tr> <th>Member name</th><th>Description</th></tr> </thead> <tbody> <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> <tr> <td>Skip</td><td>The document will not be processed.</td></tr> <tr> <td>Pass Through</td><td>The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.</td></tr> </tbody> </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.	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.
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 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 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>										
Convert to TIFF	<p>This allows control over the method used to extract images from PDF files for OCR processing. The default value is 'No' for native processing.</p> <ul style="list-style-type: none"> No – (Native) Yes – (Convert to TIFF) 										
DPI	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.										

	The default value for this property is taken from each page in the source PDF file.
TIFF Compression	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. The default value for this property is taken from each page in the source PDF file. Valid values are CCITT4 or LZW
Retain Bookmark*	Retains any bookmarks from the source file in the output.
Retain Metadata*	Retains any metadata from the source file in the output when using.
Retain Viewer Preferences*	Retains any PDF Viewer Preferences, Page Mode and Page Layout from source file in the output when using.

* Note: Convert To TIFF must be set to 'Yes' for this to work.

17.1 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:

- 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.

[Home](#) > [Create](#)

Create

Computer Vision

*

 Name

*

 Subscription

Microsoft Azure (Enterprise)

▼

*

 Location

Australia East

▼

*

 Pricing tier [\(View full pricing details\)](#)

▼

*

 Resource group

AzureBackupRG_eastus_1

▼

[Create new](#)

Create

Automation options

17.1.1 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+

17.2 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.

17.2.1 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

18 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>

Warning: The job in Kingfisher must be set to run as a single core job.

You can check here for more details on how to set up a Kingfisher job in Autobahn DX.

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

19 GdPicture Steps

The addition of the GdPicture SDK library has allowed the development of new, valuable steps that will present users more choice in the way documents are processed. The new steps added are below:

- **Validate PDF/A**
- **Linearize PDF**
- **Convert Any File to PDF (GdPicture)**
- **Combine Any File to PDF**
- **Combine PDFs**
- **PDF to JPEG**
- **PDF to PNG**
- **PDF to TIFF (GdPicture)**
- **PDF to Text**
- **PDF to Searchable PDF (GdPicture)**
- **Create PDF Portfolio**
- **Convert PDF to PDF/A**
- **Compress PDF**

The GdPicture library has many capabilities beyond the new steps that have been created, so users can expect more steps utilizing this library to be added in the future.

Useful information for the new steps can be found in the sections below.

19.1 PDF/A Validation

When archiving pdf files, if the files successfully conform to an ISO standard as PDF/A files, it will ensure that the document will be able to be rendered in the future and appear as expected. Setting a file to a PDF/A version ensures it's preservation, a necessity in certain industries when archiving for extended periods.

The **PDF/A Validation** step ensures that files in a directory fit all the requirements of the selected PDF/A version:

- If a file is in valid PDF/A format for the selected version, it will be copied to the output folder
- If the file does not fit the selected format, the file will go the selected error folder for the job

Users can run the files in the error folder through the **Convert PDF To PDF/A** step to create valid PDF/A files that conform to your selected PDF/A version.

19.2 Linearize PDF

This step optimizes PDFs by enabling Fast Web View mode for web-viewing, allowing the rendering the document one page at a time. This enhances the user experience when viewing larger PDFs on the web.

19.3 Convert Any File To PDF (GdPicture)

This step can convert a large variety of file types to PDF.

Description	Suffix
Windows bitmap format	BMP
Microsoft Word (.doc) binary file format	DOC
Microsoft Word OpenXML	DOCX
Microsoft Word Macro-Enabled OpenXML format	DOCM
Enhanced Windows Meta-format	EMF
Graphics Interchange Format	GIF
HTML format	HTML
Icon and cursor format (single or multi page)	ICO
Joint Photographic Expert Group	JPEG
Portable Gray-map File	PGM
Portable Network Graphics Format	PNG
Portable Pix-map File	PPM
Microsoft Powerpoint Presentation format	PPTX
Microsoft PowerPoint Macro-Enabled Presentation format	PPTM
Rich Text File Format	RTF
Tagged Image Format	TIFF
Plain text file	TXT
Standard Windows Meta-format	WMF
Microsoft Excel (.xls) binary file format	XLS
Microsoft Excel Spreadsheet format	XLSX
Electronic Mail format	EML
Outlook Item File Format	MSG
Scalable Vector Graphics File	SVG
Device Independent Bitmap format	DIB
24-bit compressed JPEG Graphic format	JPE
MIME HTML format	MHTML
OpenDocument Text file format	ODT
Portable Bitmap Image file format	PBM
Picture Exchange image file format	PCX
Target raster graphics format	TGA

This step uses the GdPicture engine to render the file and thus does not require an Office installation to process Office files.

19.4 Combine Any File To PDF

Converts a folder of files into PDF and then merges them, to create a single output PDF.

See [Convert Any File To PDF \(GdPicture\)](#) for the file formats.

This step uses the GdPicture engine to render the file and thus does not require an Office installation to process Office files.

19.5 Combine PDFs

Merges a folder of PDF files to create a single output PDF.

19.6 PDF To JPEG

Converts an input PDF page by page into a set of JPEG files using the GdPicture toolkit.

19.7 PDF To PNG

Converts an input PDF page by page into a set of PNG files using the GdPicture toolkit.

19.8 PDF To TIFF (GdPicture)

Converts an input PDF into a multipage TIFF file using the GdPicture toolkit.

19.9 PDF To Text

Extracts the searchable text from the pages of a PDF file and creates an output text file.

19.10 PDF To Searchable PDF (GdPicture)

Carries out Optical Character Recognition on the input PDF using the GdPicture toolkit, creating an invisible searchable text layer over the document.

19.10.1 OCR Language Codes

For the new GdPicture OCR step, a user can specify from over 100 languages from the table below by adding their code to the **Additional Dictionary** field. You can also specify multiple languages in this field by separating their code with a '+' symbol. e.g. Using 'deu+fra+spa' will include all three dictionaries in the OCR process.

New language files need to be added to the "...\\Autobahn DX\\distribution\\gdpicture\\ocr" folder. Please download the OCR languages pack, including over 100 languages from:

http://www.gdpicture.com/download/tesseract_ocr_4x_language_pack.zip

Language	Code	Language	Code	Language	Code
Afrikaans	afr	German - Fraktur	deu_frak	Portuguese	por
Albanian	sqi	Greek	ell	Pushto	pus
Amharic	amh	GreekAncient	grc	Quechua	que
Arabic	ara	Gujarati	guj	Romanian	ron
Armenian	hye	HaitianCreole	hat	Russian	rus

Assamese	asm	Hebrew	heb	Sanskrit	san
Azerbaijani	aze	Hindi	hin	Scottish Gaelic	gla
AzerbaijaniCyrillic	aze_cyrl	Hungarian	hun	Serbian	srp
Basque	eus	Icelandic	isl	SerbianLatin	srp-latn
Belarusian	bel	Indonesian	ind	Sindhi	snd
Bengali	ben	Inuktitut	iku	Sinhala	sin
Bosnian	bos	Irish	gle	Slovak	slk
Breton	bre	Italian	ita	Slovak (Fraktur)	slk_frak
Bulgarian	bul	Italian_Old	ita_old	Slovenian	slv
Burmese	mya	Japanese	jpn	Spanish	spa
CatalanValencian	cat	Javanese	jav	Spanish_Old	spa_old
Cebuano	ceb	Kannada	kan	Sundanese	sun
CentralKhmer	khm	Kazakh	kaz	Swahili	swa
Cherokee	chr	Kirghiz	kir	Swedish	swe
ChineseSimplified	chi_sim	Korean	kor	Syriac	syr
ChineseTraditional	chi_tra	Kurdish	kur	Tagalog	tgl
Corsican	cos	Kurmanji	kmr	Tajik	tgk
Croatian	hrv	Lao	lao	Tamil	tam
Czech	ces	Latin	lat	Tatar	tat
Danish	dan	Latvian	lav	Telugu	tel
Danish – Fraktur	dan_frak	Lithuanian	lit	Thai	tha
Dutch	nld	Luxembourgish	ltz	Tibetan	bod
Dzongkha	dzo	Macedonian	mkd	Tigrinya	tir
English	eng	Malay	msa	Tonga	ton
English (Middle)	enm	Malayalam	mal	Turkish	tur
Esperanto	epo	Maltese	mlt	Uighur	uig
Estonian	est	Maori	mri	Ukrainian	ukr
Faroese	fao	Marathi	mar	Urdu	urd
Filipino	fil	Maths	equ	Uzbek	uzb
Finnish	fin	Mongolian	mon	UzbekCyrillic	uzb-cyrl
Frankish	frk	Nepali	nep	Vietnamese	vie
French	fra	Norwegian	nor	Welsh	cym
French (Middle)	frm	Occitan	oci	Western Frisian	fry
Galician	glg	Oriya	ori	Yiddish	yid
Georgian	kat	Panjabi	pan	Yoruba	yor
Georgian_Old	kat_old	Persian	fas		
German	deu	Polish	pol		

19.11 PDF Portfolio

Creates a PDF Portfolio file by embedding files from various file types. On opening the PDF Portfolio, these files will be displayed on selection.

19.12 Convert to PDF/A

Convert a PDF file to PDF/A format file.

19.13 Compression

Compress a PDF file to reduce the output file size.

19.14 Detect Signatures

Detect if a PDF file contains digital signatures.

19.15 Smart Redaction

Redact text in a PDF file based on common categories for sensitive information.

19.16 Key Value Pair Extraction

Extract important data pairs from PDF or supported image files.

19.17 Pattern Redaction

Redact text in a PDF file based on regex patterns or a terms list.

19.18 Pattern Highlighting

Highlights text in a PDF file based on regex patterns or a terms list.

19.19 Split PDF (GdPicture)

Split PDF files based on page ranges, bookmarks, or into single pages.

19.20 Split by Barcode

Split PDF pages based on barcodes found in the document.

19.21 Pattern Enumeration

Create a report based on text in a PDF file found with regex patterns or a terms list.

19.22 Get Document Information

Split PDF pages based on barcodes found in the document.

19.23 Convert PDF To Office

Convert PDF files to the selected office format (docx, xlsx, pptx, svg).

19.24 Convert Any File To Office

Convert other file formats to the selected office format (docx, xlsx, pptx, svg).

20 PDF Recognition to JSON Job Step

The new PDF Recognition to JSON step automatically extracts important data from searchable PDF files in the form of Key/Value pairs. The output is given as a JSON file that contains each expected key along with its value pair.

A UI program exists to test pdf files and show what data pairs will be extracted from the file. This program can be found at:

"<Autobahn DX Installation directory>\distribution\recognition\AquaforestDataExtractorUI.exe"

You must use an 'Expected Key' file to tell Autobahn which keys to extract from the input files. You can also specify synonyms for your keys, so that values paired with any synonym will also be extracted with the key. This is very useful when processing files with varying formats and different ways of framing the same data. Our example 'Expected Key' file below highlights how this file can be used to cover multiple naming for the same key.

You can get more information about each step property in [Section 5.2.7.29](#)

Example of an "Expected Key" file:

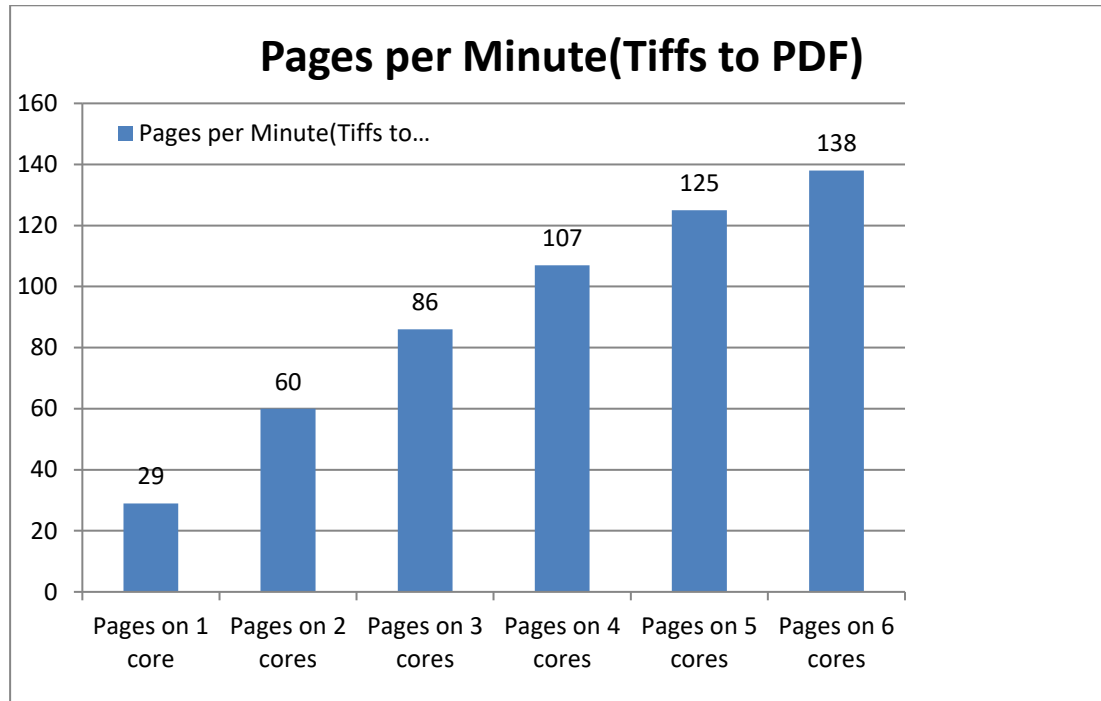
```
{
  "expectedKeys": [
    {
      "expectedKey": "Invoice No",
      "synonyms": [
        "Invoice Number",
        "Invoice No.",
        "Invoice Num"
      ]
    },
    {
      "expectedKey": "Inv Date",
      "synonyms": [
        "Invoice Date",
        "Inv. Date",
        "Inv date"
      ]
    },
    {
      "expectedKey": "Reference",
      "synonyms": [ ]
    },
  ],
}
```

21 Multicore Support

The Autobahn DX Server is available as single core, four cores or multiples of four cores.

This allows Autobahn DX to take full advantage of the number of cores available on the computer.

The current version of Autobahn DX can be licensed for up to 64 cores. This allows the simultaneous processing of 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.



21.1 Using multiple cores in a job

More than one core is available on most job step types, but the improvement may not be significant on fast job steps like Merge TIFFs, Split TIFFs, Merge PDFs, and Split PDFs.

Where the use of more than one core would be significant, the number of cores to be used in the step can be set via the Maximum Cores property in the step properties section as shown below.

Step Properties	
Search: <input type="text"/> Clear	
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	1 ▼
Password Files	take no action
Debug	No

21.2 Maximum number of cores allowed

The maximum number of cores that can be used by Autobahn DX is limited to the smaller of:

- The number of cores licensed
- The number of logical processors (not physical cores) in the CPU.

It is worth bearing in mind that setting this value to the number cores on your system may cause 100% usage of CPU resources by the process. This will have an effect on any other process running on the system.

The Max Available option allows .NET to decide the best number of processors to be used.

21.3 Multicore Throttling

Your license key will support a specific number of CPU cores. Autobahn DX 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"/>
```

21.4 Command Line Interface

The Autobahn DX command line interface can also set the number of cores to use. The parameter **/maxcores=x** can be used to set the number of cores to use to x.

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

To set the number of cores to use, add the parameter listed below with the number of cores.

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

22 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.

22.1.1 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](#)

22.1.2 File Name Length – Autobahn DX Working Folder Paths

For integrity reasons, Autobahn DX will (by default) copy source files and write 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*

22.1.3 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.

22.1.3.1 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" />`

22.1.3.2 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

23 Autobahn DX Configuration Files

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

23.1 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 advise that users should not change any undocumented settings.

These settings are important for Autobahn DX to function properly and should never be 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>
```

23.1.1 The Input Delay

This is 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.

23.1.2 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.

23.1.3 Output Trigger Files

This feature makes Autobahn DX 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" />
```

23.1.4 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.

23.1.5 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 22](#) 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.

23.1.6 Allow users to disable SSL in email steps

SSL is a security feature that helps establish secure connections when sending emails. However, sometimes emails are only sent internally, an SSL certificate is may not be required (or available).

Though we recommend keeping SSL enabled in most cases, it can now be disabled by setting the config option "secureoption" to **None**. It is set to **Auto** by default.

```
<add key="secureoption" value="Auto" />
```

Setting Name	Values	
secureoption	None	Disable SSL certificate requirement when sending emails
	Auto	Use default settings
	StartTls	StartTLS informs the email server that the email client wants to upgrade from an insecure connection to a secure one using TLS or SSL.
	SslOnConnect	The email connection should use SSL or TLS encryption immediately.

23.2 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 each setting.

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 these settings when advised from Aquaforest support.

23.3 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.

23.4 Other configuration files

Filename	Location	Comment
topdf.exe.config	distribution\pj\bin	Check section 16.2.2.1 for more details
AnyFileToSearchablePDF.exe.config	bin	Chooses the operation for anyfiletopdf job steps based on the file extension.

24 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.
jobdef_copy	When Autobahn starts, it will check all the job definitions and upgrade them to the current version. This folder will contain the previous versions of the job definitions. A new, clean installation will upgrade the two example files that are included in the installer.
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

25 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