Autobahn DX Reference Guide

Version 3.5

January 2015

http://www.aquaforest.com/

CONTENTS

1 PRODUCT OVERVIEW AND CONCEPTS

1.1 Overview

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

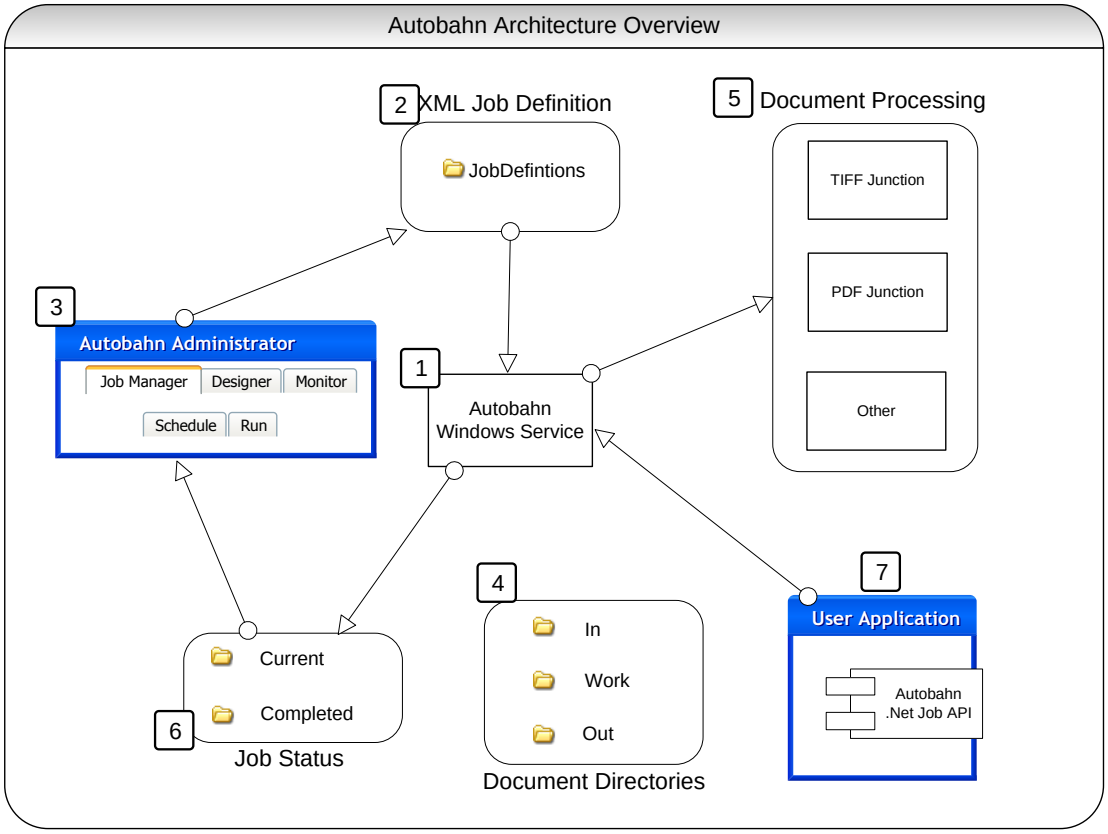
Autobahn includes PDF Junction (Professional Edition) and TIFF Junction (Professional Edition). For reference, this matrix highlights the functionality of Aquaforest's document conversion and processing products:

	TIFF Junction Standard	TIFF Junction Professional	PDF Junction Standard	PDF Junction Professional	Autobahn DX (Workstation)	Autobahn DX (Server)	Autobahn DX (Extended OCR)	Autobahn DX (Extended OCR Asian)	Autobahn DX (Extended OCR - IHQC)	Autobahn DX (Extended OCR Asian – IHQC)	Autobahn DX (Multicore)
Convert TIFF to PDF	•				•	•	•	•	•	•	•
Split & Merge TIFFs	•				•	•	•	•	•	•	•
Convert PDF to TIFF	•	•	•	•	•	•	•	•	•	•	•
Extract Raw Text from PDF			•	•	•	•	•	•	•	•	•
OCR & Convert TIFF to Searchable PDF		•			•	•	•	•	•	•	•
OCR & Convert Image-Only PDF to Searchable PDF		•			•	•	•	•	•	•	•
Split & Merge PDFs			•	•	•	•	•	•	•	•	•
Set Security on PDFs			•	•	•	•	•	•	•	•	•
Multi Core Processing											•
Extended OCR							•	•	•	•	
Asian Extended OCR								•		•	
Intelligent High Quality Compression									•	•	
Convert Microsoft Office Documents to PDF*				•	•	•	•	•	•	•	•
Convert HTML, Text and Other Documents to PDF*				•	•	•	•	•	•	•	•
Process Files, Folders or Entire Trees	•	•	•	•	•	•	•	•	•	•	•
Command Line Interface	•	•	•	•		•	•	•	•	•	•
Support for Windows 2003, 2008, XP Professional, Vista and 7	•	•	•	•	•	•	•	•	•	•	•
Graphically Define Document Processing					•	•	•	•	•	•	•
Windows Service						•	•	•	•	•	•

Scheduled & Ad-hoc Job Support via XML Job Tickets						•	•	•	•	•	•
Watched Folders						•	•	•	•	•	•
.Net API Interface						•	•	•	•	•	•
Customizable Script Support						•	•	•	•	•	•
Email Alerts						•	•	•	•	•	•
Read Mailbox & Send Converted Documents by Email						•	•	•	•	•	•
Multi-User access to watched folders						•	•	•	•	•	•

*Requires native application present on server.

2 AUTOBAHN DX ARCHITECTURE AND CONCEPTS



2.1 Autobahn Windows Service

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

2.2 XML Job Definition Files

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

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

Section 8 of this document describes the XML Job Definition Files in full detail.

2.3 Autobahn Administration

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

2.3.1 Quick Job

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

2.3.2 Job Manager

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

2.3.3 Designer

This provides a graphical interface to allow 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.

2.3.4 Monitor

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

2.4 Document Directories

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

2.5 Document Processing

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

2.6 Job Status

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

2.7 Autobahn DX .Net Job API

A .Net API is provided to allow user application to create and execute ad-hoc jobs. See section 11 for further details.

2.8 Multi-Core/CPU Support

Enhanced support for multi-core/CPU architectures was added in version 2.1. Whilst it has always been possible to run multiple concurrent jobs to exploit multiples cores/CPUs, processing of single large TIFF or image PDF files to searchable PDF has been single threaded and only been able to exploit a single core/CPU.

Version 2.1 introduced threaded processing on the "OCR an Image-Only PDF" and "Convert TIFF to PDF" (when using OCR). These step properties include a "Threads" option which may be set to 1, 2, 3 or 4. This option can provide a performance increase of up to 3.5 for multi-page files on a quad core/CPU machine.

The threads setting is used to determine how many processing threads are used. This is implemented by splitting the input TIFF or PDF into an appropriate number of parts (based on page ranges) and processing each part in parallel.

2.9 File Access Permissions

Administrators of Autobahn DX should be aware that "Quick Job" operations are run in the context of the current logged on user so rely on the permissions granted to that user and may to process files on remote file systems may make use of either UNCs or mapped drives visible to that user. Conversely,

Jobs in the Job Manager are run by the Autobahn DX windows service, so when accessing remote file systems UNC paths should be used rather than mapped drives. The job will run as the user specified in the "Log On" property page of the service so it may be necessary to change the user ID if required for security purposes.

2.10 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.

2.10.1 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:\doocuments\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$:\doocuments\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:\doocuments\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$:\doocuments\major accounts payable 2010*

2.10.2 File Name Length – Autobahn DX Working Folder Paths

For integrity reasons, Autobahn DX will (by default) copy source files and writes processed files to working folders before copying to the target output folders. The location of the working folders is defined in job manager – for example *C:\Aquaforest\Autobahn DX 3.03\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 3.03\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 3.03\work\1008*

3 INSTALLATION AND LICENSING

3.1 System Requirements

Supported Operating Systems	Windows 7 (32bit | 64bit) Windows 8 (32bit | 64bit) Windows Server 2003 (32bit | 64bit) Windows Server 2008 (32bit | 64bit) Windows Server 2012
Disk Space	500 Mb
.Net Framework	3.5 and 4.0

3.2 Product Licensing

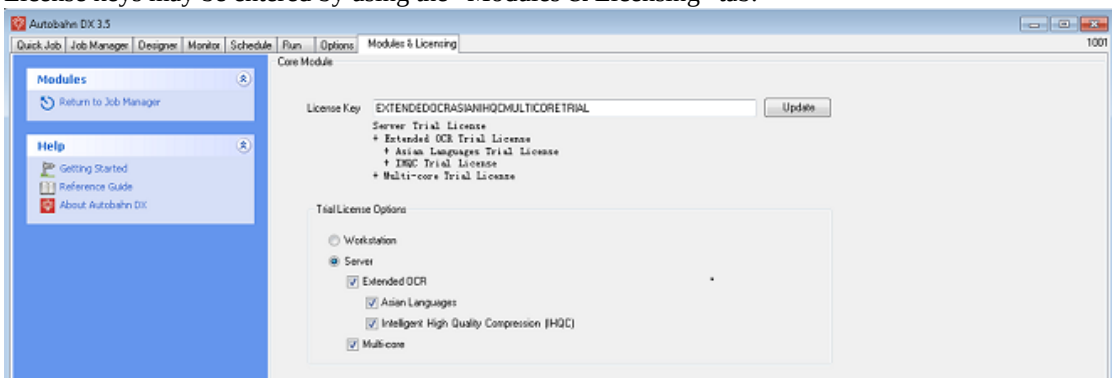
Autobahn DX has four license levels which are described below:

License	Description
Workstation	This license only allows use of the Quick Job tab. To test this license in trial mode select the WORKSTATION radio button from the "Modules & Licensing" tab.
Server	This license allows use of all of the tabs. To test this license in trial mode select the SERVER radio button from the "Modules & Licensing" tab.
Extended OCR	This module is applied on-top of the SERVER license and enables access to the IRIS OCR engine for enhanced recognition. To test this license in trial mode select the SERVER | Extended OCR radio button from the "Modules & Licensing" tab. See section 15 for more details.
IHQC	This module is applied to the SERVER + EXTENDED OCR license and enables access to IRIS COMPRESSION. To test this license in trial mode select the SERVER | Extended OCR | IHQC radio button from the "Modules & Licensing" tab. See section 16 for more details.
Asian Language Support	This module is applied to the SERVER + EXTENDED OCR license and enables Asian Language support. To test this license in trial mode select the SERVER | Extended OCR | Asian Language radio button from the "Modules & Licensing" tab. See section 16 for more details.
Multicore	This module is applied on-top of the SERVER license and enables use of multiple cores when running jobs. To test this license in trial mode select the SERVER | Multi-Core radio button from the "Modules & Licensing" tab. See section 17 for more details.

Note: The Network Server Licence has been retired from the Autobahn DX 3.03 release; all its functionalities can be handled by the Server license.

3.3 Entering License Keys

License keys may be entered by using the "Modules & Licensing" tab.

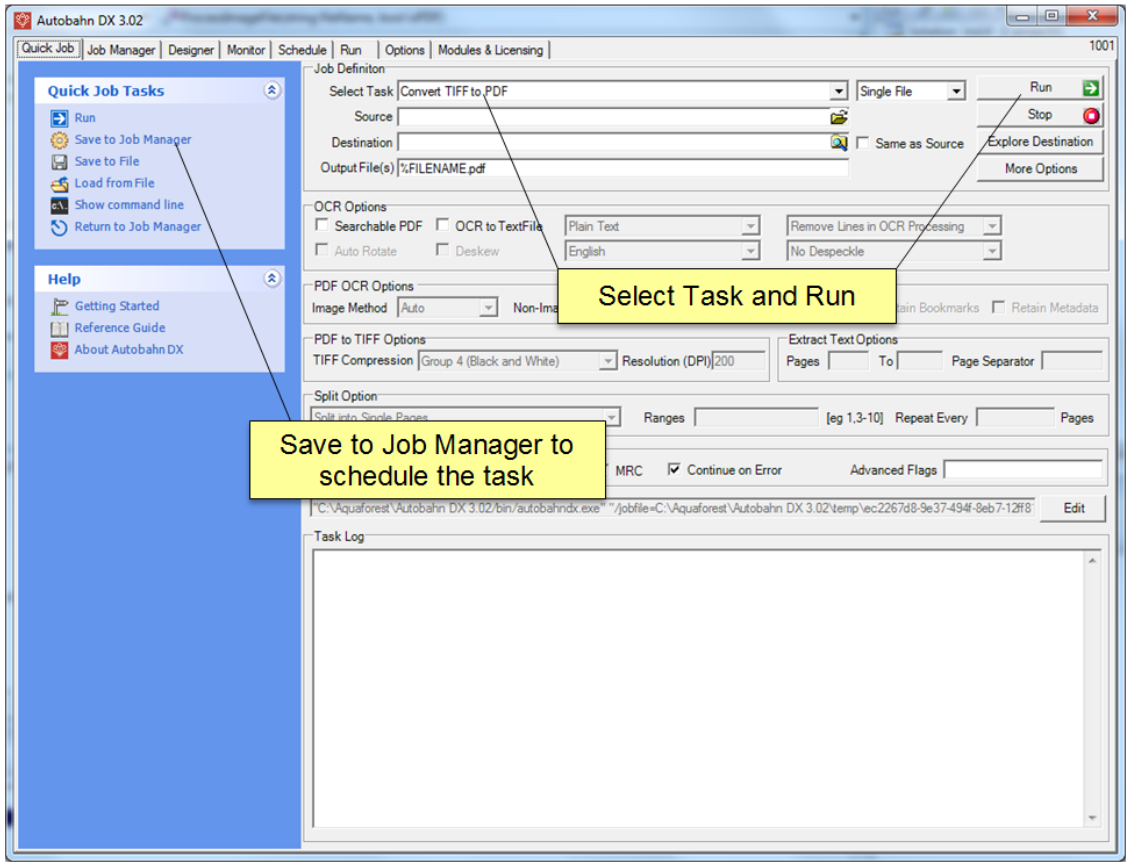


4 AUTOBAHN DX QUICK START

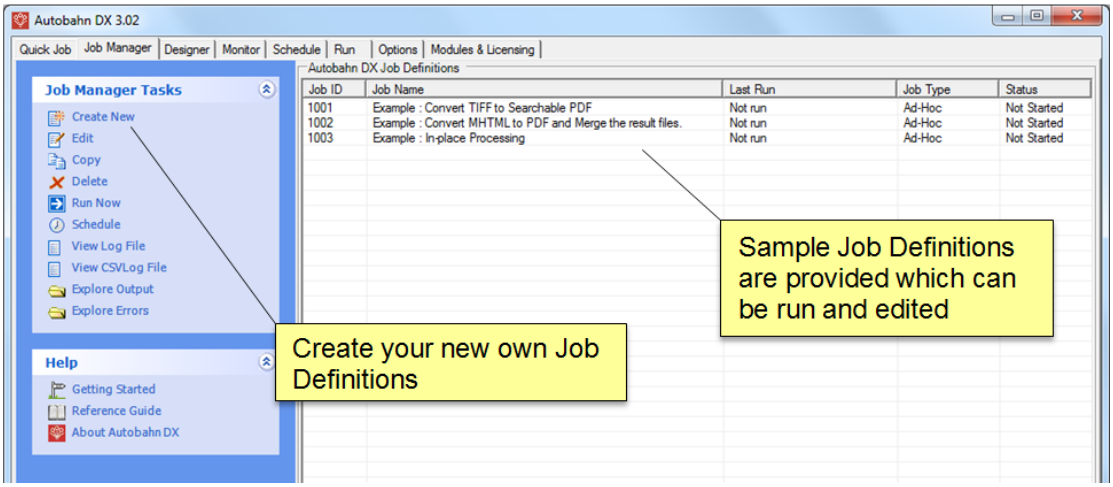
Autobahn DX is designed to allow definition of document processing jobs ("Job Definitions") and then scheduling them to run continuously on watched folders or on a daily scheduled basis. The annotated screen shots in this section are designed to enable rapid familiarization with the product.

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.

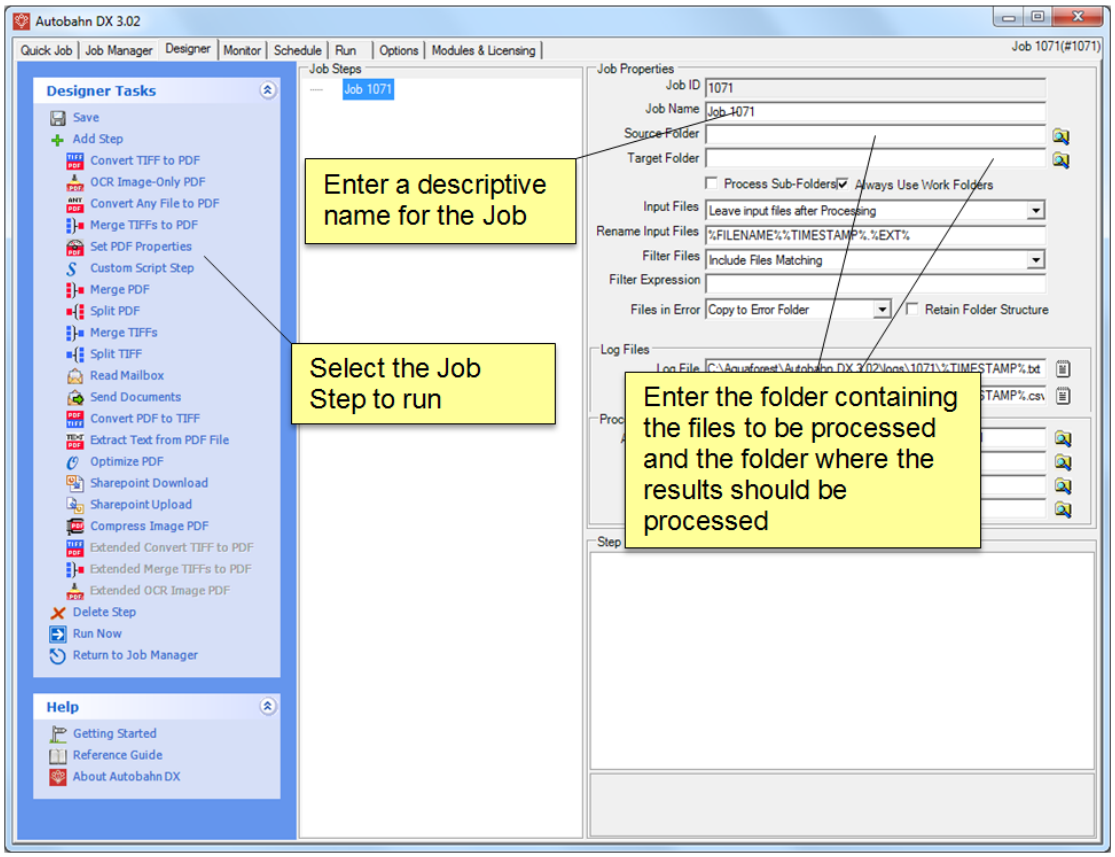
1. The Quick Job Screen



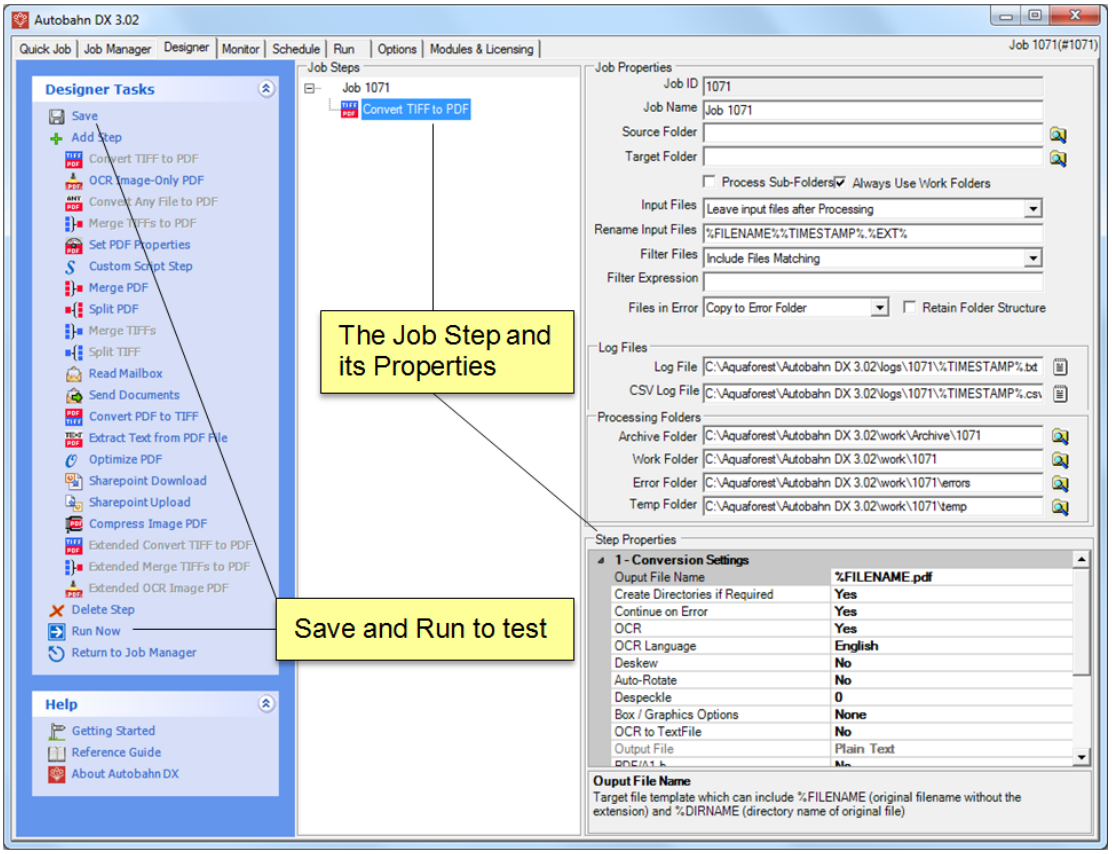
2. The Job Manager Screen



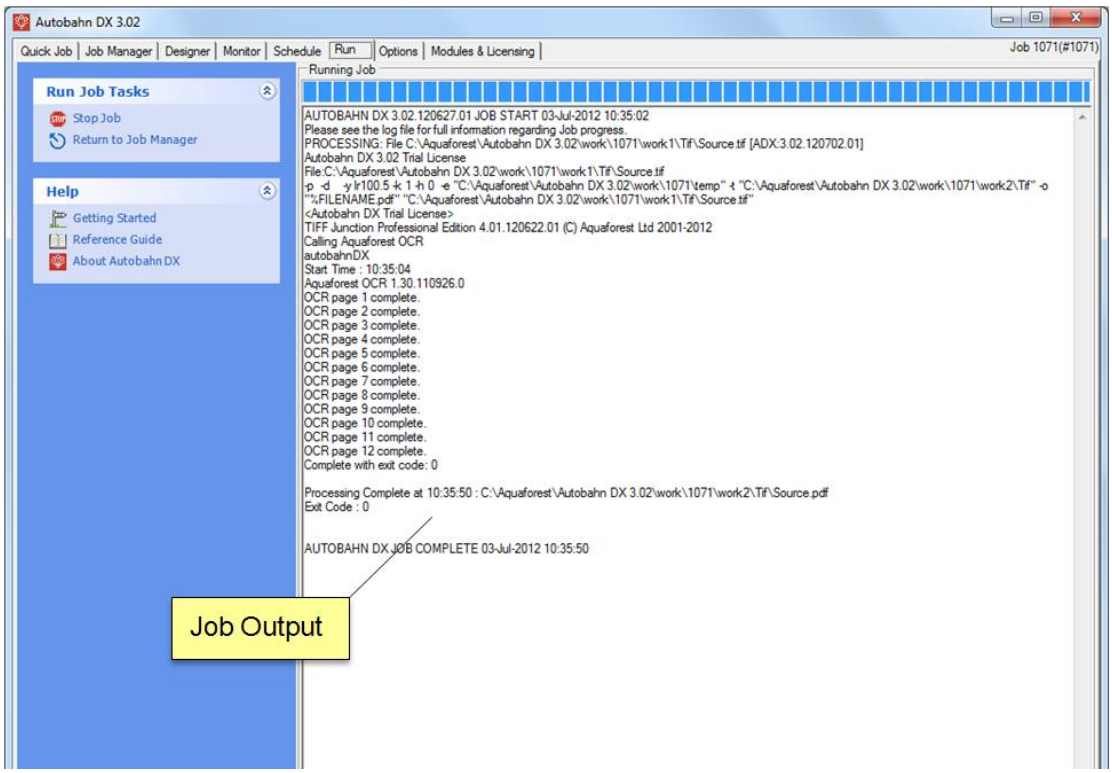
3. The Job Designer Screen



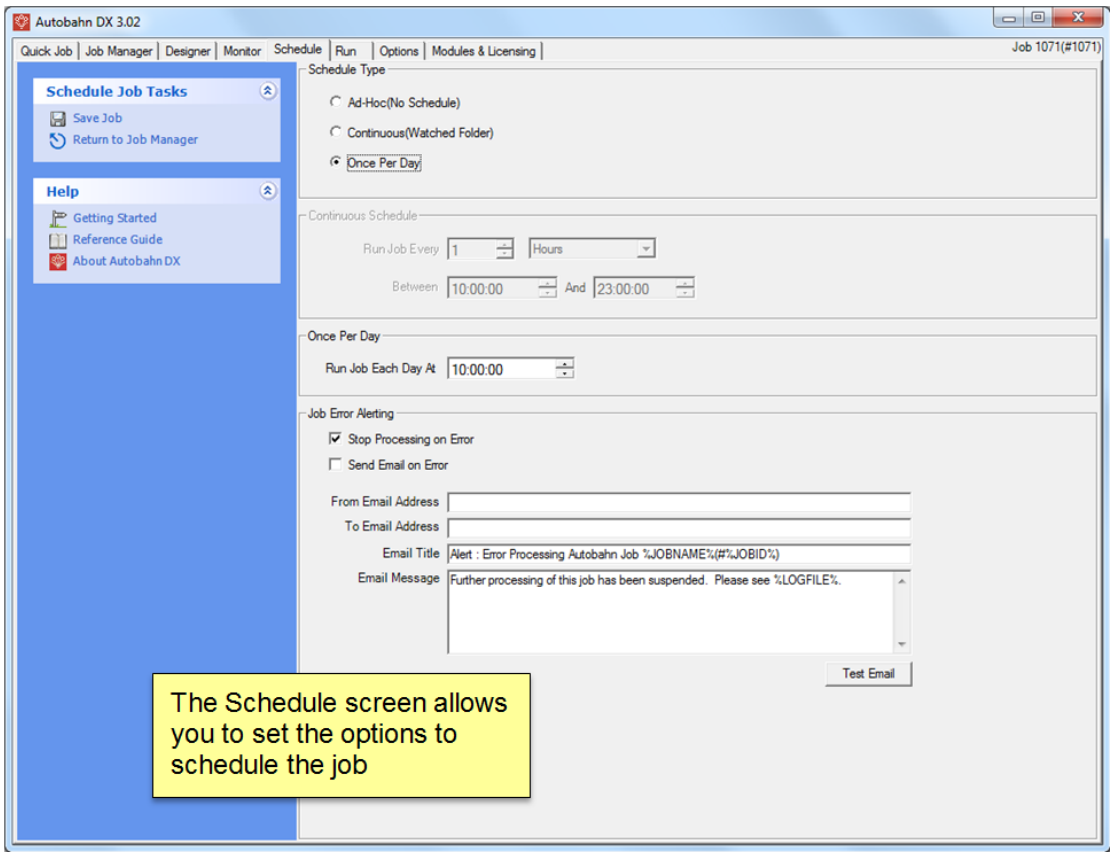
4. Defining a Job



5. Testing the Job Definition



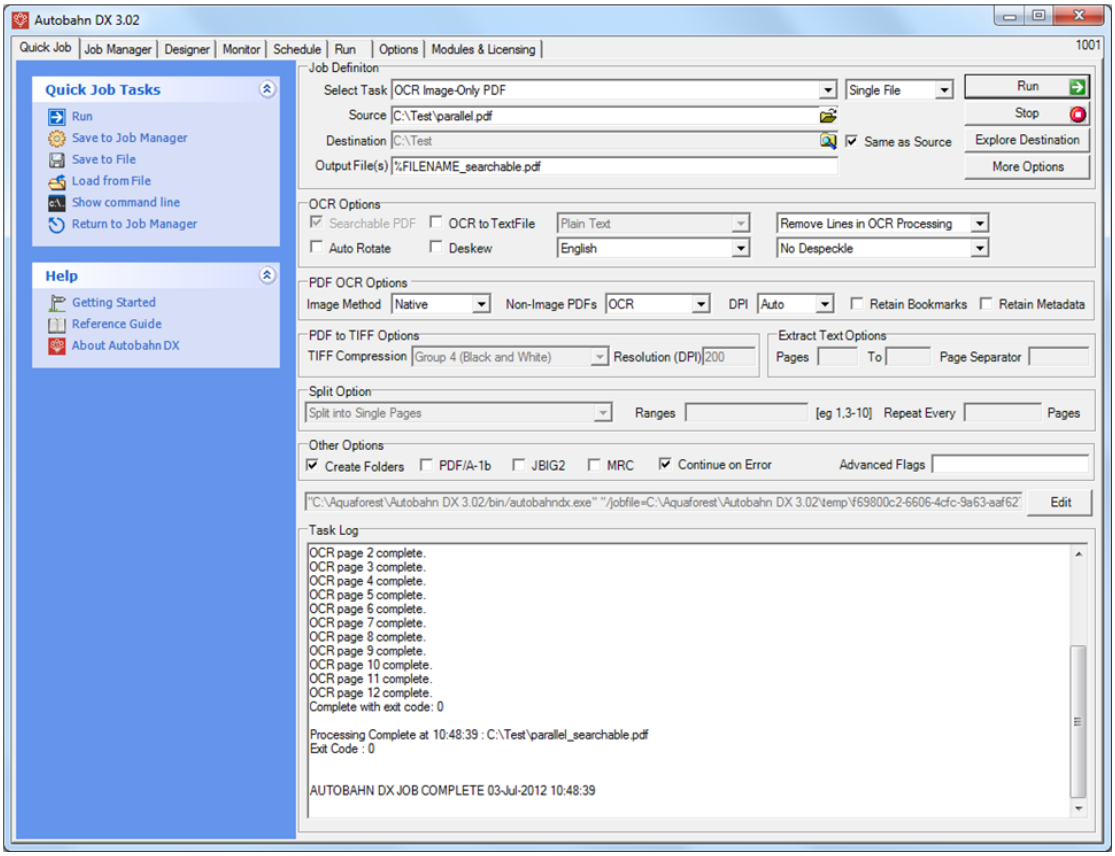
6. Scheduling the Job



5 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.

The Quick Job Screen



Menu Item	Action
Run	Runs the job using the current Quick Job definition.
Save to Job Manager	Creates a new job in the Autobahn DX job manager using the current Quick Job definition.
Save to File	Save the current Quick Job definition to an XML file.
Load from File	Loads a Quick Job definition from an XML file.
Show Command Line	Shows the command line equivalent of the current Quick Job definition. [Requires the Server license]

5.1.1 Job Definition Section

Screen Field / Button	Description
Run	Runs the current job. Job output will appear in the "task log section of the screen".
Stop	Halts processing of the currently running job.
Explore Destination	Launches Windows Explorer using the destination as the starting point.

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.

Screen Field / Button	**Description**
Select Task	This defines which task is to be run and may be one of : Convert a TIFF file to PDF OCR an Image-Only PDF File Convert any file to PDF Set PDF Properties Merge TIFF Files Merge TIFF Files to PDF Split a TIFF File Show Information for a TIFF File Split PDF Merge PDF Convert a PDF file to TIFF Extract Text from PDF
Source	The source TIFF file or directory.
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.
Output File(s)	This defines the template output file name. %FILENAME — Source file name without .pdf %UNIQUE*n* — For use with split operations. Unique number starting at 1. If *n* is supplied, then zero padding to n digits is used. If n is not supplied or is zero, then zero padding is not used. %DIRNAME — Source directory name of the directory currently being processed. *String* — (Any string) For example, a split operation using %FILENAME-part%UNIQUE6.pdf would give *filename*.pdf split into *filename*-part000001.pdf *filename*-part000002.pdf etc.

5.1.2 OCR Options Section

Screen Field / Button	**Description**
Searchable PDF	PDFs generated from TIFF files include searchable hidden text recognised by Autobahn DX's OCR Engine.
OCR to Text File (Professional Edition Only)	Create a separate text file of text recognised by Autobahn DX's OCR Engine.

Other Options (Deskew, Autorotate, Despeckle)	See section 6.2.8 for further explanation of the OCR options.

5.1.3 OCR PDF Options Section

Screen Field / Button	Description
Image Method	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 : • Auto – TIFF Junction will select the most appropriate method. This will be "Convert to TIFF" for image-only PDFs and "Native" otherwise. • Via Bitmap – The PDF is rasterized using bitmap conversion • Extract TIFF – The embedded TIFF images are extracted directly. • Convert to TIFF – The PDF is rasterized using conversion to PDF. • Native - This method places the OCRed text directly into a copy of the original PDF rather than creating an entirely new PDF. Note - Extracting PDF images via TIFF is well suited to PDFs with one scanned image per page, but certain documents that have multiple images on a page, or a mixture of image and text (e.g. a scanned document with text bates stamps) – for these images the Bitmap or conversion to TIFF method is required. Note methods other than "Native" actually change the non-image parts of the PDF to image + text. **IMPORTANT NOTE**: As the "Native" method does not make any changes to the images in the PDF file, it is not compatible with the JBIG2 or MRC compression, PDF/A compliance option or multi-threading. If these features are required then it is recommended that the "Convert to TIFF" setting for the PDF OCR image extraction method.
Non-Image PDFs	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 : • OCR. The document will 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	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.
Retain Bookmarks	Bookmarks from the original PDF are copied to the searchable result PDF.
Retain Metadata	Metadata from the original PDF are copied to the searchable result PDF.

5.1.4 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.

5.1.5 Split Options Section

Screen Field / Button	Description
Split Type	*One of :* Split into single pages Split by ranges (See below) Split by repeating ranges (See below)
Page Ranges	Set of page ranges separated by commas that defines which pages from the original should be extracted. The following types of page ranges are allowed : 1 — Specifies a single page 1-3 — Specifies a range of pages
Repeating Range	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. 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.

5.1.6 Other Options Section

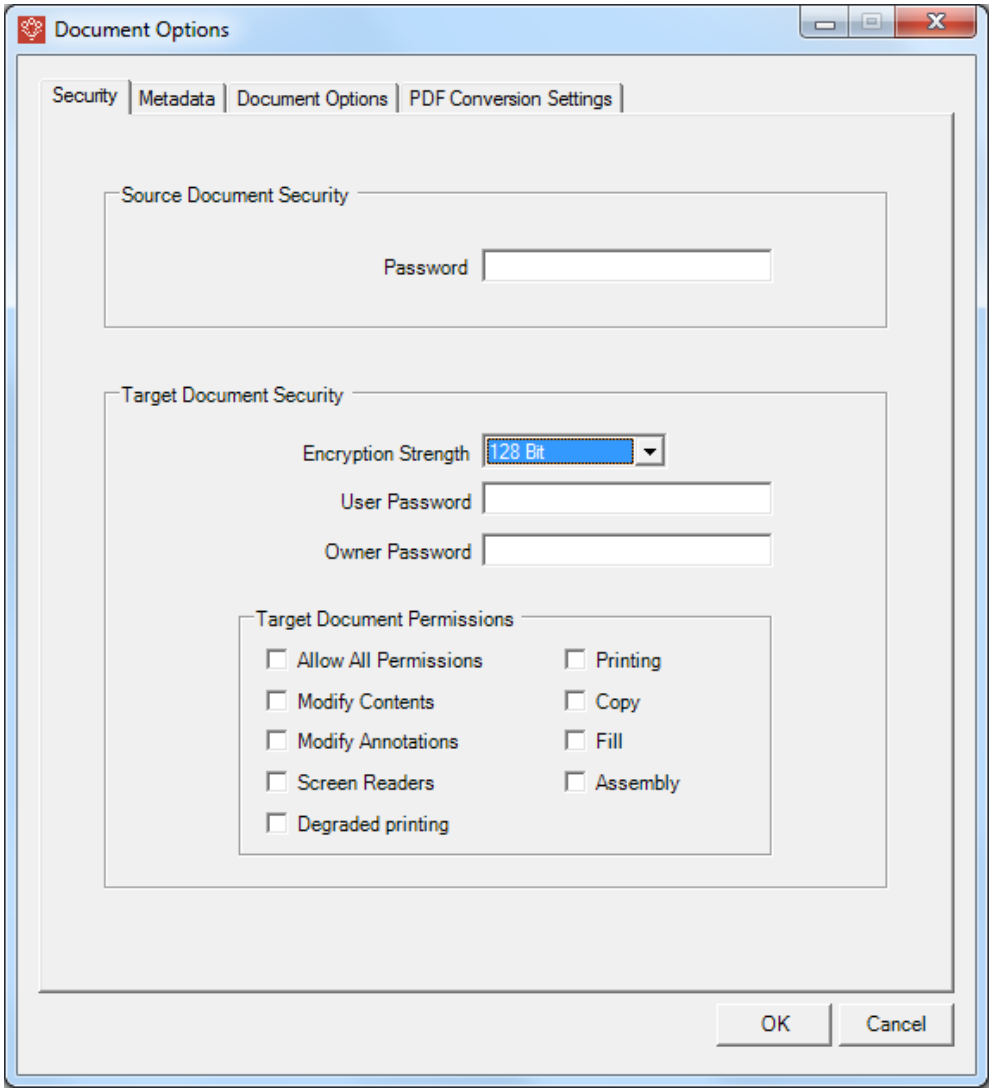
Screen Field / Button	Description
Create Directories where required	If checked, when processing a tree, subdirectories will be created in the target directory tree when required.
Continue on Error	When checked, folder and tree processing jobs will continue

	even if there is an error processing individual files.
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	If this option is chosen, bitonal images in the output PDF will be compressed used JBIG2 compression rather than the default of Group 4. This will result in smaller files for PDFs with black and white images.
MRC	This enables Mixed Raster Compression which can dramatically reduce the output size of PDFs comprising color scans. Note that this option cannot be used in conjunction with the "Native" Image Method for OCRing PDFs as the Native method does not adjust the images within the PDF.
Advanced Flags	Additional TIFF Junction advanced command line flags may be entered here (see section 3 of the TIFF Junction reference guide in the tj/docs folder)

5.1.7 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.

5.2 Security Settings

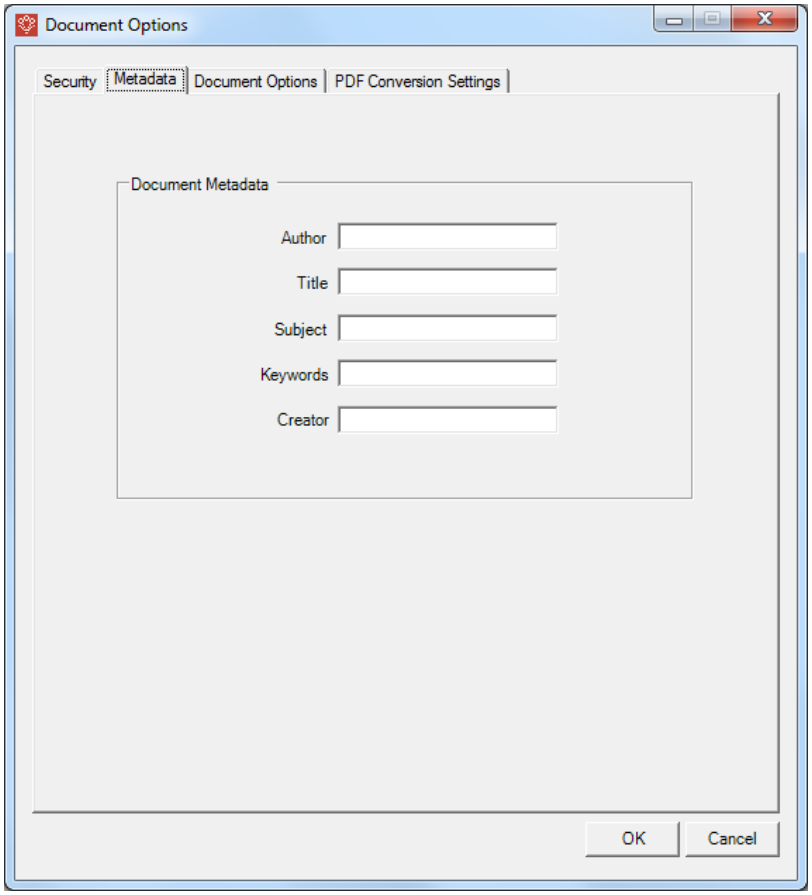


5.2.1 Target Document Security

Option	Description
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.
Permissions	

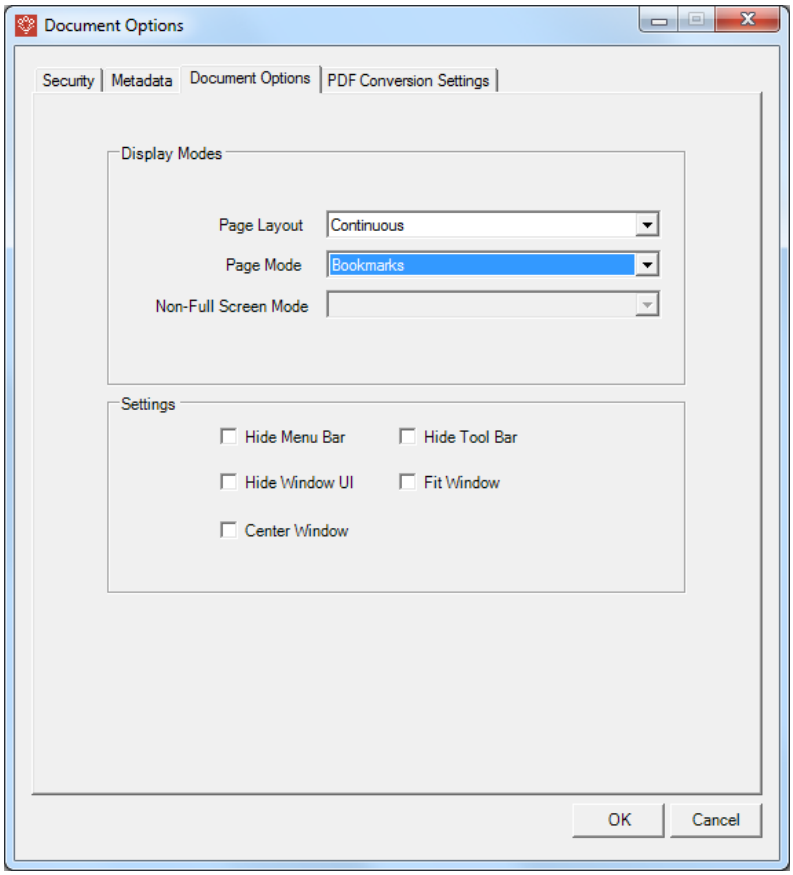
Permission	Description
Allow ALL Permissions	All the permissions below.
Printing	Allow high-quality printing
Modify Contents	Allow assembly (see below) and other document medications
Copy	Allow text and graphic copying and extraction
Modify Annotations	Allow modification of annotations
Fill	Allow filling of form fields
Screen Readers	Allow extraction of text and graphics in support of accessibility.
Assembly	Allow rotation, insertion or deletion of pages.
Degraded Printing	Allow low-quality printing

5.3 Document Metadata



This screen allows setting metadata to the generated PDF.

5.4 Document Display Options



5.4.1 Display Modes

Option	Description
Page Layout	The setting for initial document page display. One of : • Single Page • Continuous • Continuous Facing (odd pages left) • Continuous Facing (odd pages right)
Page Mode	The setting for initial viewer mode. One of : • Neither Bookmarks nor Thumbnails Open • Bookmarks Open • Thumbnails Open • Bookmarks & Thumbnails Open • Full Screen
Non-Full Screen Mode	Only applicable where Page Mode=Full Screen. The setting for document page display when exiting Full Screen mode. One of : • Neither Bookmarks nor Thumbnails Open • Bookmarks Open • Thumbnails Open

5.4.2 Settings

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

5.5 Advanced Flags

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

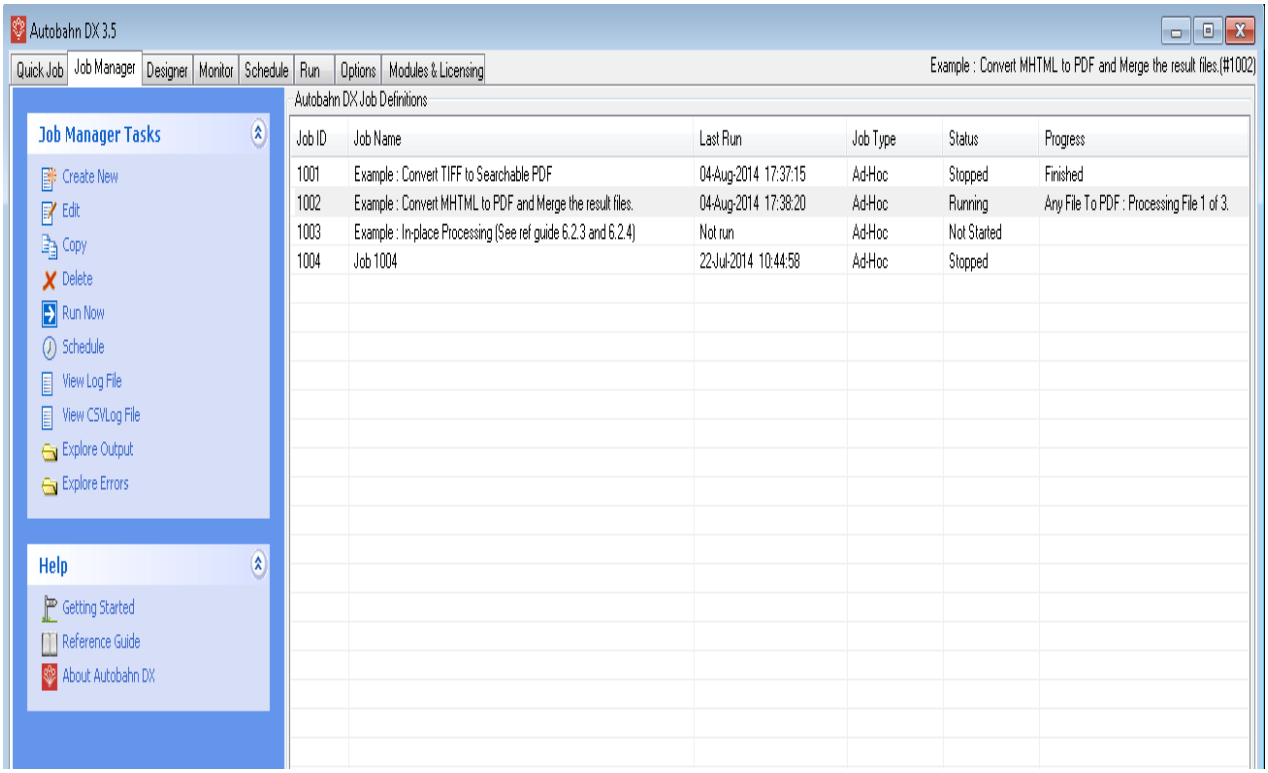
Option	Description
-B 100	Blank Page Removal This option can be used when converting TIFF files to Searchable PDFs. A value should be provided which specifies the pixel threshold to be used to determine whether a page is blank or not. If a page is deemed to be blank then it is omitted from the output file. A suggested value is 100 i.e. using the new –B advanced flag : -B 100
-D	Dot Matrix Processing 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.
-0 [zero]	Custom Despeckle 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.
-q	Custom Quantization 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).

6 THE AUTOBAHN DX ADMINISTRATION TOOL

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

6.1 Job Manager

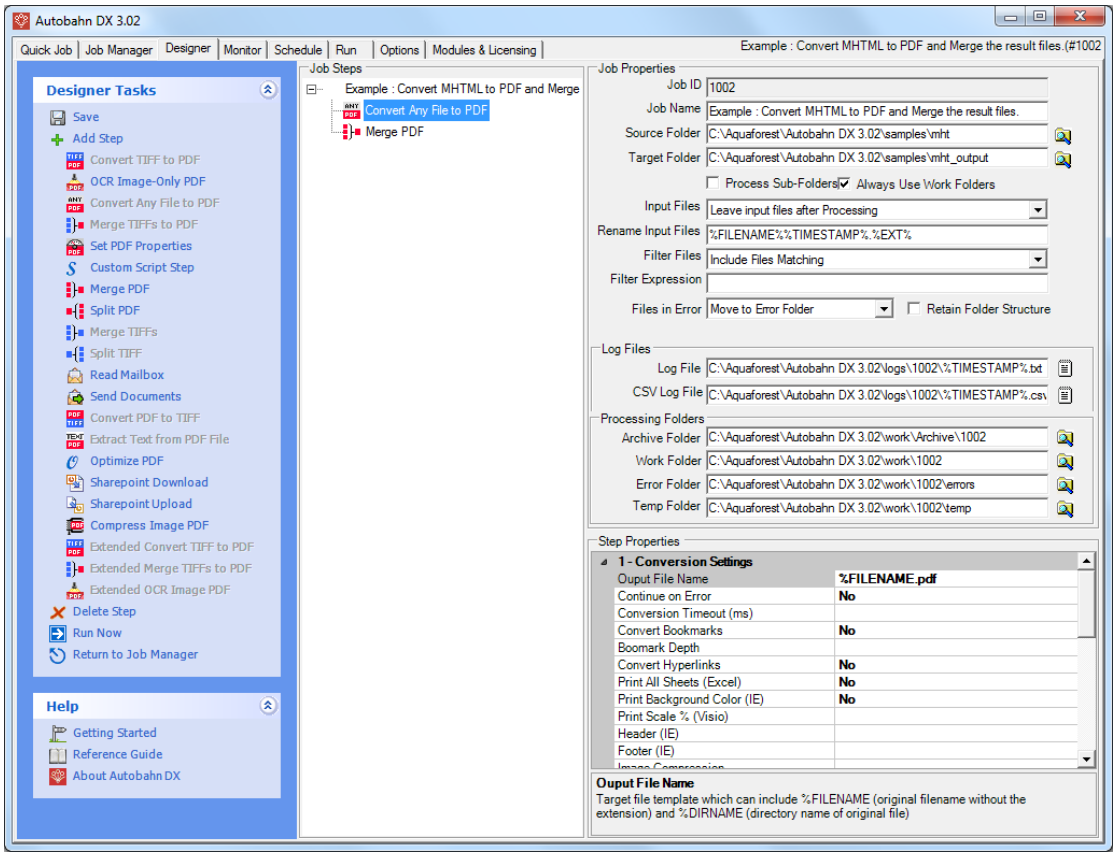
This section of the application displays a list of all current Job Definitions and the set of operations to be performed shown in the menu table below.



Menu Item	Action
Create New	Creates a new, skeleton job definition from jobtmplt.xml and then switches 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.
Schedule	Switches the view to the schedule window for the selected job.
View Log File	Displays the most recent log file for the job.
Explore Output	Launches windows explorer on the job output folder.

6.2 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. Each step type will have its own icon. Drag and drop can be used to allow reordering of steps.



6.2.1 Menu Items

Menu Item	Action
Save	Validates the current job and if valid, saves the current job definition to %JOBID%.xml in the %JOBDEFDIR% directory.
Add Step (Expands)	Expands to show a list of steps that are valid to be placed after the currently selected tree node (or as the first node if none is currently highlighted). Which steps are valid is determined by matching the outputtype of the previous step with the validinputtype of the possible steps.
Delete Step	Deletes the currently selected step node.
Run Now	Performs a save (as above) and then executes the job. The output is displayed in the Run tab screen.
Return to Job Manager	Returns to the job manager with the current job selected.

6.2.2 Fields

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

<table>
<tr><td>Filter Files</td><td>
<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>This would limit files selected to PDFs that have not been OCRed.

A file is deemed to have been OCRed if either
(a) It has a custom metadata tag “AQUAFORESTOCR”
(b) It has one image per page and only has “invisible” text.

This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</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>
N.B.: Work Folders must be used to enable use of filters.
</td></tr>
<tr><td>Filter Expression</td><td>One or more regular expressions used to determine the files in the source folder that should be processed. Multiple expressions may be used, separated by spaces.

Examples :
*.pdf
*.doc *.ppt *.xls
*.pdf; 3000</td></tr>
<tr><td>Log File</td><td>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.</td></tr>
<tr><td>CSV Log File</td><td>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 :

• 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.</td></tr>
</table>

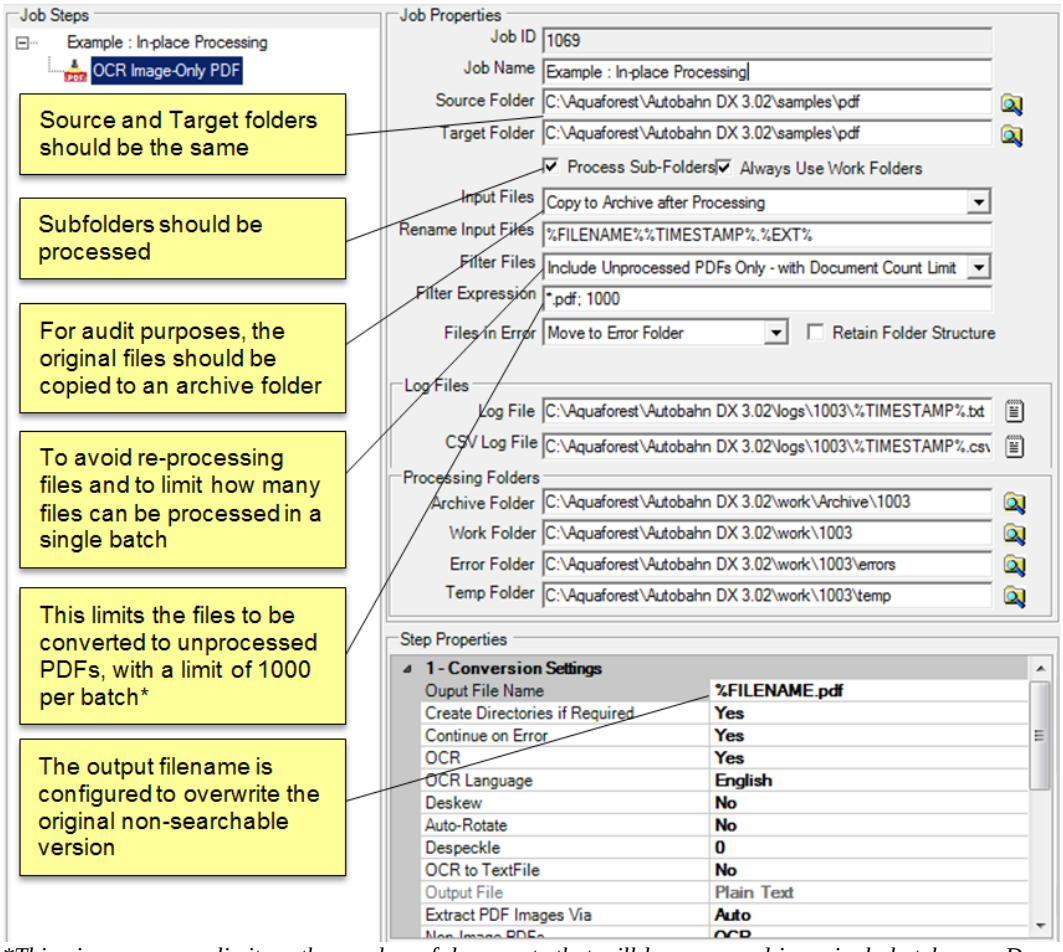
Archive Folder	The folder where the processed files will be placed if "Move to archive after processing." is chosen.
Work Folder	The folder where files will be temporarily stored during conversion and processing.
Error Folder	Source documents that have errors during processing will be placed in the specified folder.
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.

6.2.3 Workflow Processing versus In-Place Processing

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

6.2.4 Example In-Place Job Setup

The job shown below will convert PDFs under the tree "C:\Aquaforest\Autobahn DX 3.02\samples\pdf" to searchable PDFs, processing up to 1,000 files each time the job is run. Archive copies of the original files are placed in "C:\Aquaforest\Autobahn DX 3.02\work\Archive\1003".



**This gives an upper limit on the number of documents that will be processed in a single batch run. Depending upon the document store and the schedule a larger or smaller figure may be appropriate.*

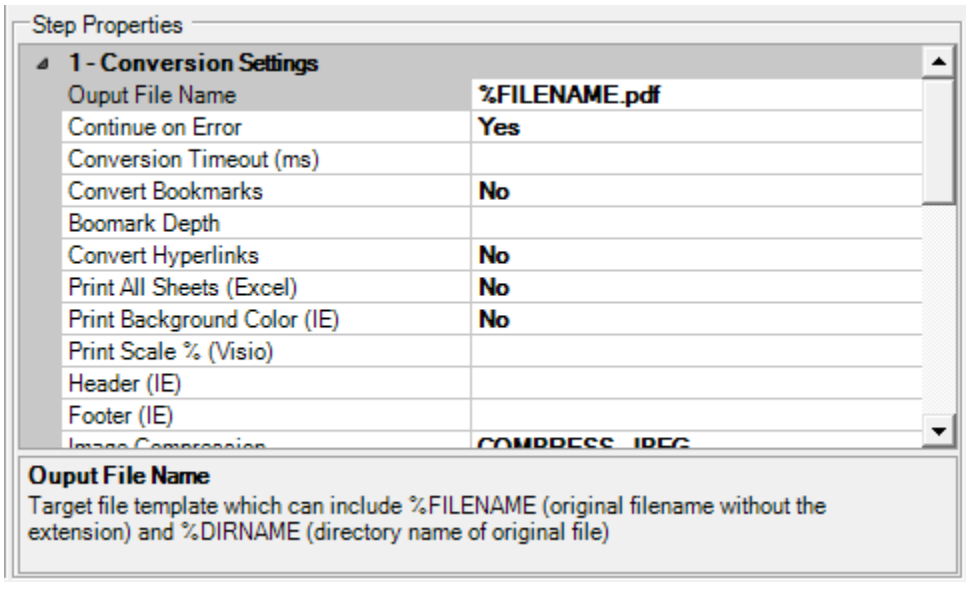
6.2.5 Step Types

Step Type	Description	Underlying Functionality Provided by
Convert TIFF to PDF	Converts a TIFF file to PDF. The TIFF compression must be one of the support formats : Group 4, Group 3, LZW, JPEG (type 6), RLE, Uncompressed (bitonal). Depending upon the Step Type Properties chosen, the PDF file may be searchable or image-only, and a separate text or HTML file may be produced from the OCR process.	TIFF Junction
OCR Image-Only 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 or HTML file may be produced from the OCR process.	TIFF Junction
Convert any document 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.	PDF Junction
Merge TIFFs to PDF	This merges a folder of TIFF files together to produce a PDF file.	TIFF Junction
Set PDF Properties	This is used to set PDF Metadata properties (such as Author, Title, etc.), Security settings and Document Display properties.	PDF Junction
Custom Step	This can be used to support a custom scripted step in the process. See section 7 for more details.	-
Merge PDF	Merges a folder of PDF files into a single file.	PDF Junction
Split PDF	Splits each input PDF file into a set of files, either a single page per file or by page ranges.	PDF Junction
Merge TIFF	Merges a folder of TIFF files into a single file.	TIFF Junction
Split TIFF	Splits each input TIFF file into a set of files, either a single page per file or by page ranges.	TIFF Junction
Convert PDF to TIFF	Rasterizes a PDF file, converting into a multi-page TIFF file.	PDF Junction
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.]	PDF Junction
Read Mailbox	This reads the specified mailbox using IMAP4 and extracts attachments for processing. This step should be followed by a regular processing step	

	(e.g. Convert Any Document to PDF) and optionally followed by Send Document.	
Send Document	This processes a folder of files resulting from conversion after a Read Mailbox step. It will email the converted files back to the users who sent the originals.	
Optimize PDF	This allows creation of Web Optimized (Linearize) PDFs.	
Sharepoint Download	This step downloads documents from the specified Sharepoint document library ready for processing	
Sharepoint Upload	This step uploads documents to the specified Sharepoint document library.	

6.2.6 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.



6.2.7 Convert TIFF to PDF

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of 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.
OCR Type	Maximum Speed Maximum Quality Balanced (the default)
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.
Output PDF	Choose "Yes" to Generate a PDF file.
Output TXT	Choose "Yes" to generate a .txt file (only applicable if OCR is specified).
Output RTF	Choose "Yes" to generate a .rtf file (only applicable if OCR is specified).
Output HTML	Choose "Yes" to generate a .htm file (only applicable if OCR is specified).
Advanced Flags	Command line flags to be passed through to the underlying executable.
PDF/A1-b PDF/A2-b PDF/A3-b	Select the output PDF/A compliant version you would like the output PDF to be.
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 increasing processing time.
Box/Graphics Options	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 "graphic" or "picture" areas but that actually do contain useful text. To ensure that the OCR engine can be forced to process such areas there are two options :

	"Treat all Graphics Areas as Text". This option will ensure the entire document is processed as text. *"Remove Box Lines in OCR Processing".* 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).
Threads	Defines the number of processing threads for conversion processes involving OCR. See section 2.8 for further details.
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.
Stamps	See section 6.2.23 for full details

6.2.8 OCR Image-Only PDF

Parameter	**Notes**
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of 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 Language	Select the language the original file is written in. This will determine the dictionary that is used.
OCR Type	Maximum Speed Maximum Quality Balanced (the default)
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.
Output PDF	Choose "Yes" to Generate a PDF file.
Output TXT	Choose "Yes" to generate a .txt file
Output RTF	Choose "Yes" to generate a .rtf file
Output HTML	Choose "Yes" to generate a .htm file
Advanced Flags	Command line flags to be passed through to the underlying executable.

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.
Extract PDF Images Via	Choose the method for PDF image extraction. • Auto (default) • Via Bitmap • Extract TIFF • Convert to TIFF
PDF/A1-b PDF/A2-b PDF/A3-b	Select the PDF/A compliant version you would like your output file to be
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 increasing processing time.
Non-Image PDFs	This defines how non-image PDFs should be treated. The options are : • Raise Error (default) • Rasterize and OCR • Skip • Pass Through •
DPI	When OCRing a PDF, the PDF is rasterized to produce a TIFF file which is then OCRed. By default the TIFF image resolution is determined from the images embedded in the source PDF but this flag can be used to override default processing and specify the DPI of the TIFF that will be generated.
Box/Graphics Processing	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 "graphic" or "picture" areas but that actually do contain useful text. To ensure that the OCR engine can be forced to process such areas there are two options : *"Treat all Graphics Areas as Text".* This option will ensure the entire document is processed as text. *"Remove Box Lines in OCR Processing".* 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).
Threads	Defines the number of processing threads. See section 2.8 for further details.
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. 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.
Stamps	See section 6.2.23 for full details

6.2.9 Convert Any File to PDF

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of 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 a conversion. A value of zero means wait indefinitely.
Convert Bookmarks	For MS Word, convert bookmarks
Bookmark Depth	This property will take effect only when the ConvertBookmarks property is set to True and defines the Word headings that will be mapped to PDF.
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 the this property is set to Yes, then the resolution of images are reduced to the DPI value specified in the ImageDownsizeResolution property.
Downsize Resolution DPI	If the ImageDownsizing property is set to True, then the resolution of images are 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 full set of font) will cause significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
Font Substitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer

	driver properties (can be accessed from Control Panel).
Embed Fonts as Type 0	This option is recommended if you have non-standard fonts like barcode font.
Top Margin	Margin in Inches.
Bottom Margin	Margin in Inches.
Bottom Left	Margin in Inches.
Bottom Right	Margin in Inches.
Page Height	In Inches.
Page Width	In Inches.
Paper Orientation	PORTRAIT or LANDSCAPE
PDF/A1-b	Select Yes if the result file must be PDF/A1-b Compliant
PDF/X3	Select Yes if the result file must be PDF/X3 Compliant
PDF/X1-a	Select Yes if the result file must be PDF/X1-a Compliant
Convert Attachments	For msg files, email attachments are processed so that the resulting PDF is composed of the email plus each page of each attachment.
Include Document Markups	The "Include Document Markups" option allows Word comments and markup to be included in the generated PDF.

6.2.10 Merge TIFFs to PDF

Parameter	**Notes**
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of 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.
OCR Type	Maximum Speed Maximum Quality Balanced (the default)

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.
Output PDF	Choose “Yes” to Generate a PDF file.
Output TXT	Choose “Yes” to generate a .txt file (only applicable if OCR is specified).
Output RTF	Choose “Yes” to generate a .rtf file (only applicable if OCR is specified).
Output HTML	Choose “Yes” to generate a .htm file (only applicable if OCR is specified).
Advanced Flags	Command line flags to be passed through to the underlying executable.
PDF/A1-b	Select Yes if the result file must be PDF/A1-b Compliant

6.2.11 Set PDF Properties

Parameter	**Notes**
Output File Name	Target file template which can include %FILENAME (original filename without the extension), %DIRNAME (directory name of 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 graphic copying and extraction
Allow Modify Annotations	Allow modification of annotations
Allow Fillin	Allow filling of form fields
Allow Screen Readers	Allow extraction of text and graphics in support of accessibility.
Allow Assembly	Allow rotation, insertion or deletion of pages.
Allow Degraded Printing	Allow low-quality printing
Author	Sets the Author property

Title	Sets the Title property
Subject	Sets the Subject property
Keywords	Sets the Keywords property
Creator	Sets the Creator property
Page Layout	The setting for 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 tool bar 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.

6.2.12 Custom Script

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

6.2.13 Merge PDF

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

6.2.14 Split PDF

Parameter	Notes
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 %DIRNAME (directory name of original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Retain Bookmarks	Generated files will include bookmarks from the original file.
Retain Metadata	Generated files will include metadata (such as Author and Title) from the original file.
Split Type	
Ranges (e.g. 1,3-10)	Set of page ranges separated by commas that 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.

6.2.15 Merge TIFFs

Parameter	Notes
Output File Name	Target file template which can include %FILENAME (original filename without the extension) and %DIRNAME (directory name of original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Advanced Flags	

6.2.16 Split TIFF

Parameter	Notes
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 %DIRNAME (directory name of original file)
Create Directories if Required	Force creation of any output directories if they do not already exist.
Split Type	
Ranges (e.g. 1,3-10)	Set of page ranges separated by commas that 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.
Advanced Flags	

6.2.17 Read Inbox

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

Check with your System Administrator and ensure the following:

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

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

Parameter	Notes
imap_server	The IMAP server address e.g. `imap.company.co.uk`
email_account	The email account to be read e.g. `pdf@company.com`
email_password	Password for the account. This is held encrypted.
in_mailbox	Mailbox to read e.g. Inbox
processed_mailbox	Mailbox to move processed email to e.g. Deleted Items. If left blank, the emails will be left in the inbox which can be useful for testing.
target	Folder in which to place the attachments extracted from the email. *Files are named name@timestamp@email@.ext where name.ext is the original filename, email is the user's email address and timestamp is an integer timestamp to ensure uniqueness.*
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

6.2.18 Send Documents

Use of this step type requires a SERVER License.

Parameter	Notes
Domain	The sending domain E.g. aquaforest.com
SMTP Server	SMTP Server address E.g. smtp.aquaforest.com
Send Logon	Sending user logon e.g. admin.
Send User Password	Password for the account.
Sending User	Sending user e.g. admin@aquaforest.com

6.2.19 Convert PDF to TIFF

Parameter	Notes
Compression	Group 4 (For bitonal images) or LZW (for color).
Resolution	The DPI of the resulting TIFF File.

6.2.20 Extract Text from PDF

Parameter	Notes
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.

6.2.21 Sharepoint Download

Parameter	Notes
Sharepoint URL	Site, the URL of the SharePoint site that you want to access, e.g. http://localhost:10480/testsite
Library	Library, the name of the library that you want to access, e.g. "Test Library"
Extensions	An optional extension mask that limits those files to manipulate, e.g. “pdf,tif”
Recurse	If set to “Yes” sub-folders of the Sharepoint Library are handled.

6.2.22 Sharepoint Upload

Parameter	Notes
Sharepoint URL	Site, the URL of the SharePoint site that you want to access, e.g. http://localhost:10480/testsite
Library	Library, the name of the library that you want to access, e.g. "Test Library"
Extensions	An optional extension mask that limits those files to manipulate, e.g. “pdf,tif”

Recurse	If set to “Yes” sub-folders of the Sharepoint Library are handled.

6.2.23 Stamps

The “Stamps” parameter for “Convert TIFF to PDF” and “OCR Image Only PDF” allows entry of a command-line style specification.

For example, the string below will produce a stamp “Page Number000123Final” on page 1, “Page Number000124Final” on page 2, etc.

-stamppref \"Page Number\" –stampsuff Final –stampstart 123 –stampdigits 6 –stamppos 0 –stamptype 0

Parameter	**Notes**
-stamppref	Prefix – a string to be added to the beginning of the stamp, before the number section.
-stampsuff	Suffix - a string to be added to the end of the stamp, after the number section.
-stampstart	Start – the value that the number portion of the stamp should start at. The number portion will be incremented by 1 each page.
-stampdigits	Digits – a value indicating the minimum length that the number portion of the stamp should be displayed as. Preceding 0’s will be used to pad any numbers less than this whilst numbers greater than this will be displayed in full.
-stamppos	Stamp Position : 0 is TopLeft, 1 is TopCenter, 2 is TopRight, 3 is CenterLeft, 4 is Center, 5 is CenterRight, 6 is BottomLeft, 7 is BottomCenter, 8 is BottomRight
-stamptype	Stamp Type : 0 stamp is added as text 1 stamp is added as an image

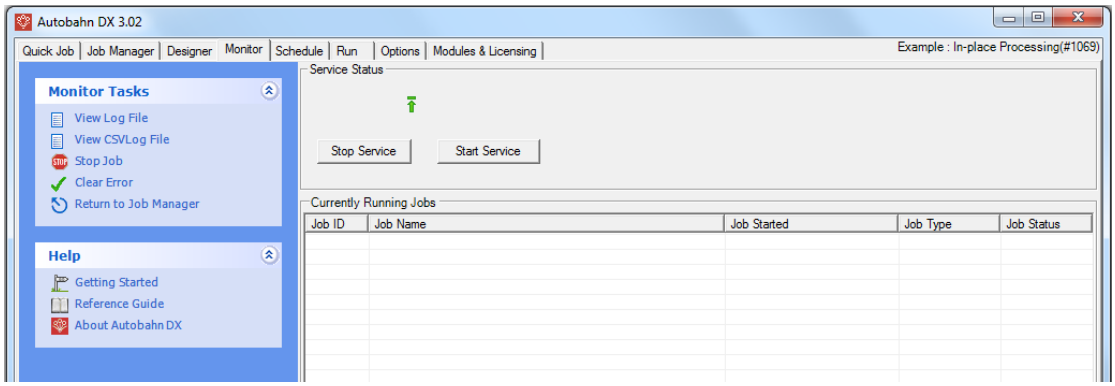
6.2.24 Compress Image PDF

Parameter	**Notes**
Output File Name	Target file template which can include %FILENAME (original filename without the extension)
MRC JPEG Quality	JPEG quality setting (percentage value 1 - 100) for use in saving the background and foreground images. Default value is 75
MRC Background Level	Sampling size for the background portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also

	quality. Default value is 3
MRC Foreground Level	Sampling size for the foreground portion of the image. The higher the number, the larger the size of the image blocks used for averaging which will result in a reduction in size but also quality. Default value is 3
Advanced Flags	A number of advanced flags are available for use under guidance from Aquaforest technical support. Contact support@aquaforest.com if needed. /pdfa –convert the compressed document to PDF/A-1b format /nojb2 – use G4 fax compression rather than JBIG2 for bitonal images /debug – generate additional informational messages /allowtext – allow PDFs containing text to be processed
Continue on Error	Continue processing files after an error occurs.

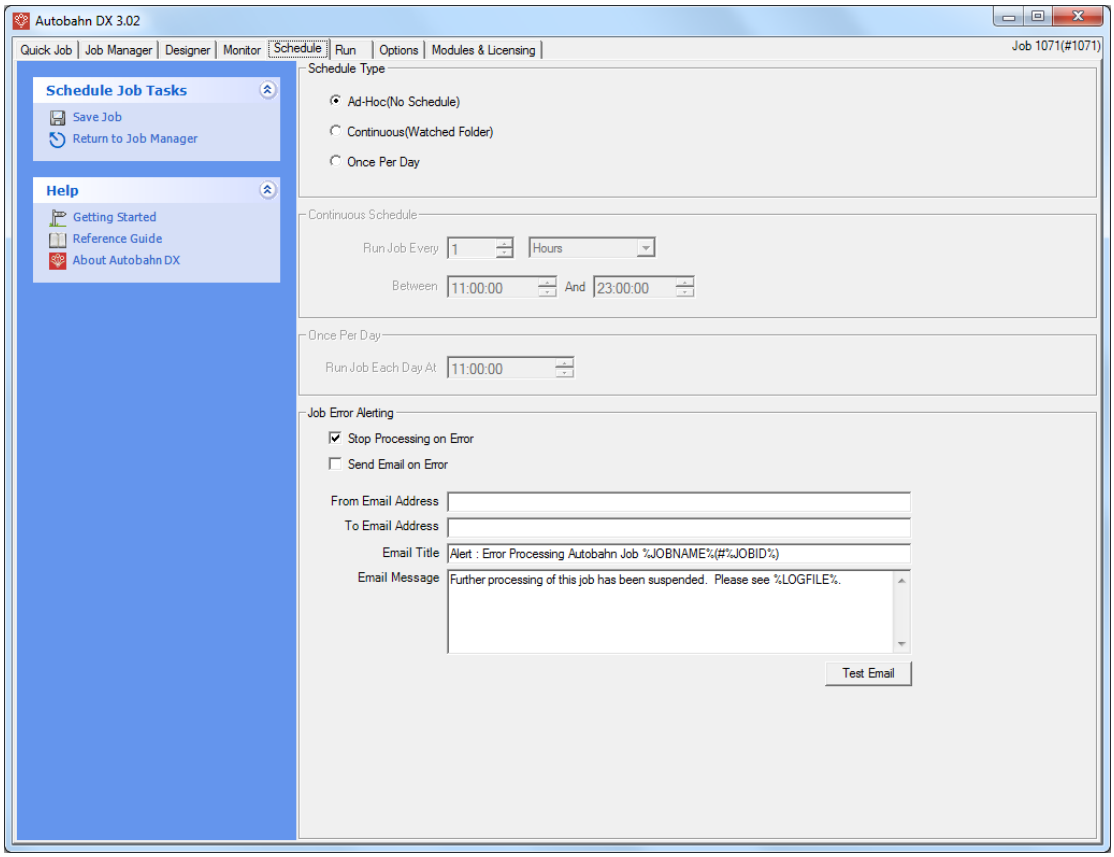
6.3 Monitor

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.
Stop Job	Halt execution of the selected job.
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.

6.4 Job Scheduling



Menu Item	Action
Save	Saves the schedule for the current job
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.
Send Email on Error	If checked, Autobahn DX will send an email if an error occurs with a job. The email server configuration should be setup on the "options" screen – see section 6.5 for more information.
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.

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

6.4.1 Ad-Hoc

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

6.4.2 Watched Folder / Continuous Scheduling

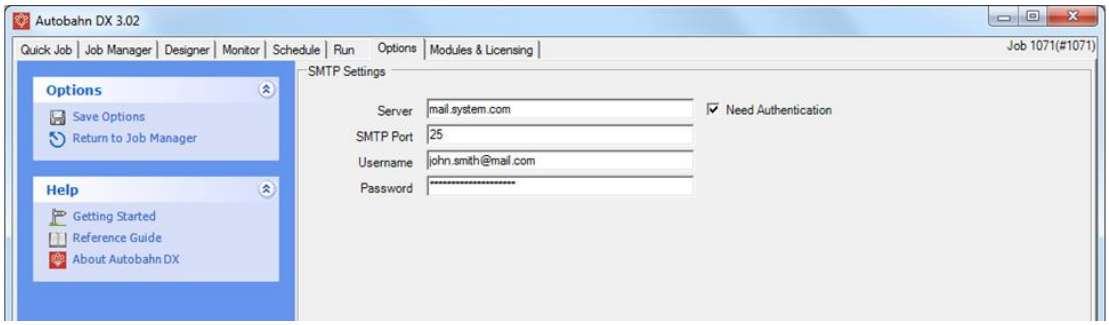
This allows the job to be scheduled to run periodically between a start time and end time each day. The periods may be seconds, minutes or hours. For example, a job may be specified to run every 30 seconds between 9:00 and 17:00.

6.4.3 Daily Scheduling

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

6.5 Email Options

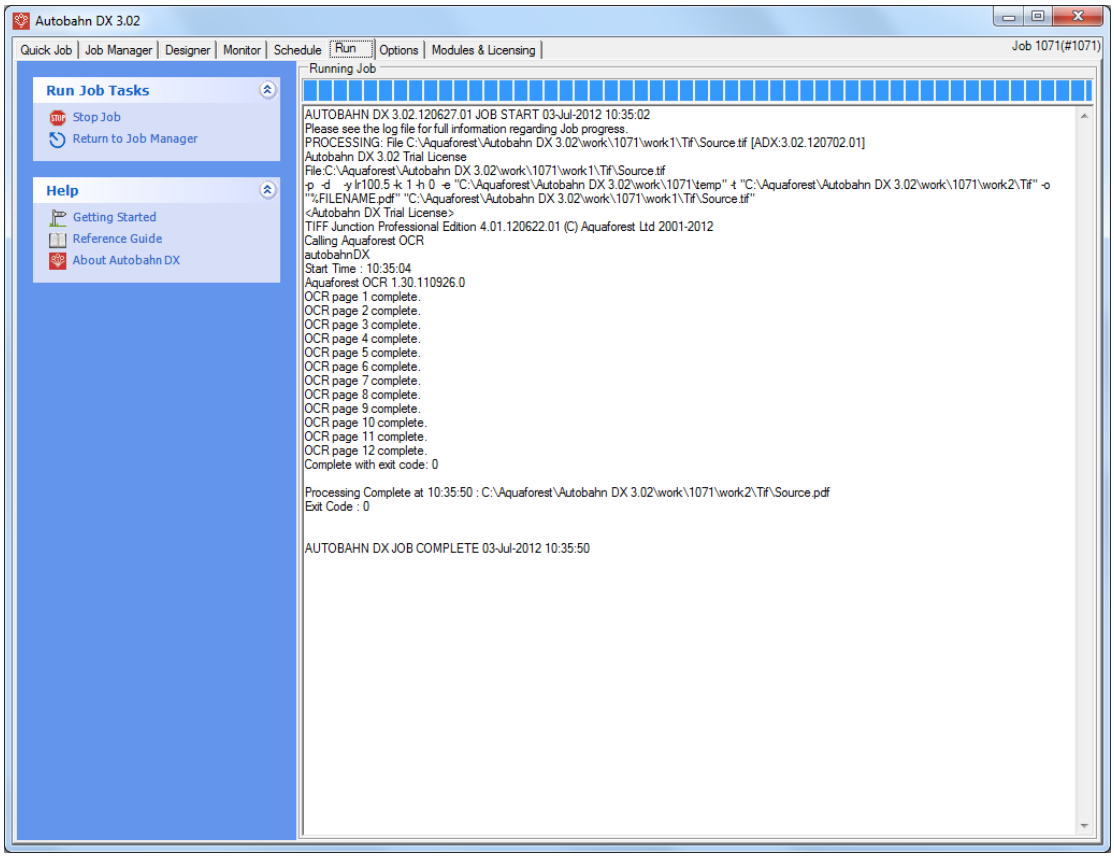
The options screen allows email server information to be defined. This is used to support the "Send Email on Error" functionality defined in the previous section. The information entered is help in Autobahn.config in the config directory. The password is held in an encrypted form.



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

6.6 Running Jobs via Job Manager or Designer

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



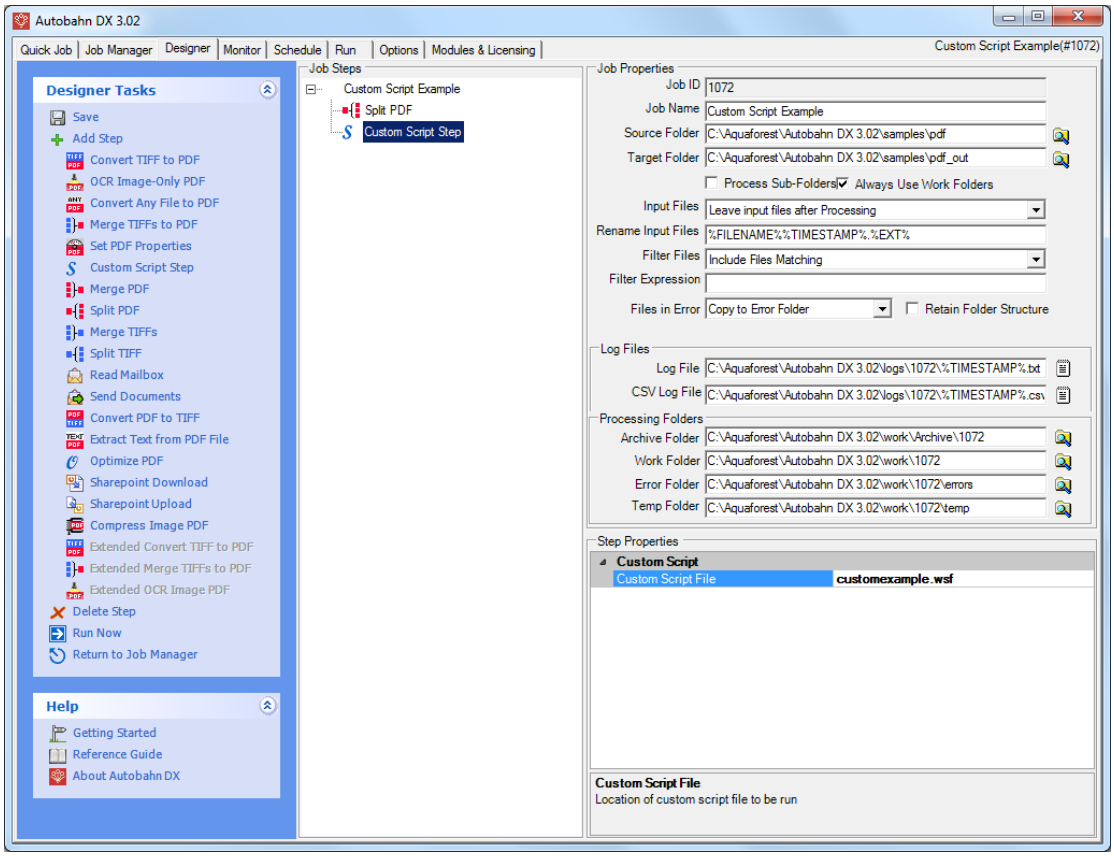
Menu Item	**Action**
Stop	Stops the currently running job.
Return to Job Manager	Returns to the job manager with the current job selected.

7 SCRIPTING CUSTOM STEPS

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

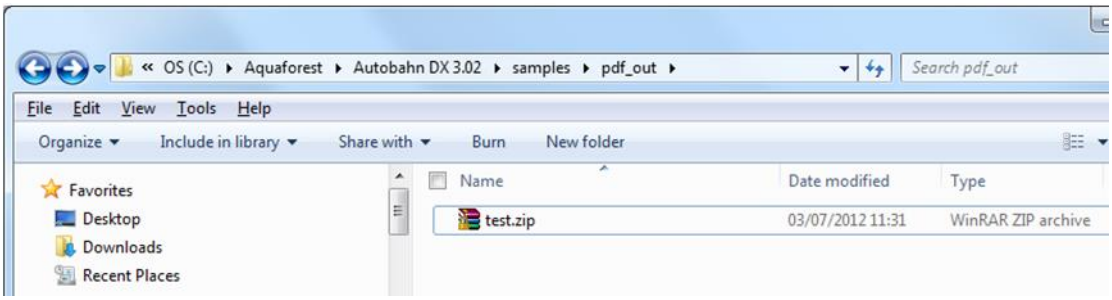
7.1 Custom Script Example

Included with the product is an example custom script (in the Autobahn DX custom folder) that will split a PDF file into single pages and then create a single ZIP file that contains the pages. The example makes use of the Freebyte Zip application which you can download from here http://www.freebyte.com/download/fbzpack.exe , run and unpack into C:\Aquaforest\Autobahn DX 3.2\custom.



Custom script files must be placed in C:\Aquaforest\Autobahn DX 3.2\custom folder. In this case the custom script file is customexample.wsh.

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



7.2 Custom Example Script

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

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

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

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

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

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

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

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

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

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

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

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

/* Write Output */

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

/* Execute Command */

function doCmd(cmd)
{
        doOut(cmd);
        var oExec    = WshShell.Exec(cmd);
        while(oExec.Status == 0)
        {
                WScript.Sleep(100);
```

```
                if(verbose)
                {
                        while(!oExec.StdOut.AtEndOfStream)
                                doOut(oExec.StdOut.ReadLine())
                }
                else
                {
                        while(!oExec.StdOut.AtEndOfStream)
                                oExec.StdOut.ReadLine()
                }
                while(!oExec.StdErr.AtEndOfStream)
                        doOut(oExec.StdErr.ReadLine())
        }

        if(verbose)
        {
                while(!oExec.StdOut.AtEndOfStream)
                        doOut(oExec.StdOut.ReadLine())
        }
        else
        {
                while(!oExec.StdOut.AtEndOfStream)
                        oExec.StdOut.ReadLine()
        }
        while(!oExec.StdErr.AtEndOfStream)
                        doOut(oExec.StdErr.ReadLine())

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

7.3 Further Examples

7.3.1 Postscript to PDF: custom_ps.wsf

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

The script should be placed in the Autobahn DX custom folder (by default C:\Aquaforest\Autobahn DX 3.2\custom) and can be used by creating a job with a custom script step and entering *custom_ps.wsf* as the value of Custom Script File.

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

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

7.3.2 MSG to PDF: custom_MSG.wsf

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

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

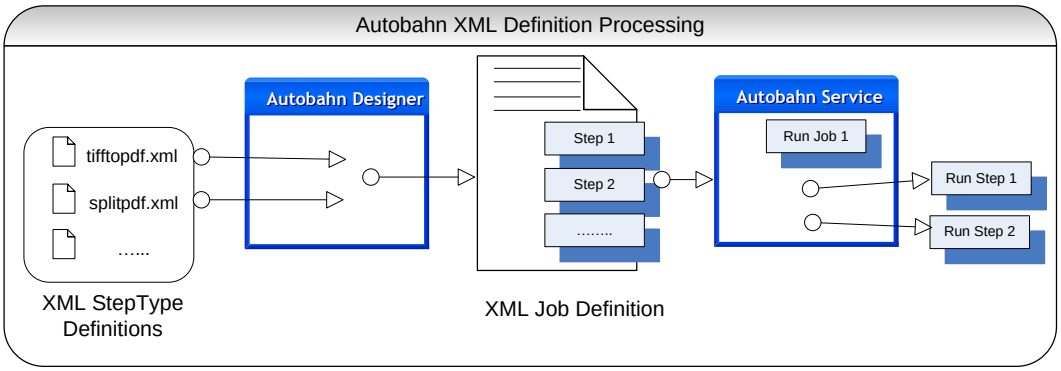
The script should be placed in the Autobahn DX custom folder (by default C:\Aquaforest\Autobahn DX 3.2\custom) and can be used by creating a job with a custom script step and entering *custom_msg.wsf* as the value of Custom Script File.

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

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

8 JOB DEFINITION XML FILES

8.1 Overview of Job Definition Creation and Processing



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

8.2 Job IDs

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

The initial contents of the file are:

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

8.3 Sample Job Definition File (Excluding Step Details)

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

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<autobahnjob>
	<jobid>1002</jobid>
	<jobname>Watch Incoming Faxes</jobname>
	<scheduletype>continuous</scheduletype>
	<scheduleevery>30<scheduleevery>
	<scehduleeveryunits>seconds<scheduleevery>
	<schedulefrom>00:00<schedulefrom>
	<scheduleto>23:59<scheduleto>
	<scheduleat></scheduleat>
	<joblogfile></joblogfile>
	<jobsteps>1</jobsteps>
	<jobsourcetype>folder</jobsourcetype>
	<jobsource>c:\faxes</jobsource>
	<jobtarget>c:\processedfaxes</jobtarget>
	<joberrors>c:\errorfaxes</joberrors>
	<jobdeleteonsuccess>yes</jobdeleteonsuccess>
	<jobwork>C:\autobahnwork\1002</jobwork>
	<jobstep>
		<stepsequence>1</stepsequence>
		<steptype>splitpdf</steptype>
		<stepdetails>
		……(See 6.5 Below)
		</stepdetails>
	</jobstep>
</autobahnjob>
```

XML Element	Description
jobid	Job ID number (see 3.2 above).
jobname	Job Description. Default is "Job %JOBID%"
scheduletype	Ad-hoc, continuous or onceperday
scheduleevery	For continuous, interval between runs
scheduleeveryunits	For continuous, unit of interval between runs
schedulefrom	For continuous, start time for runs
scheduleat	For continuous, end time for runs
scheduleto	For onceperday, the time at which to run
joblogfile	Location of the job logfile. By default the logfile is logs/%JOBID%/%TIMESTAMP%.txt
jobsteps	The number of jobsteps
jobsourcetype	File, folder or tree
jobsource	The source file or folder
jobtarget	The target folder
Joberrors	Folder for job errors. If not specified files that cannot be processed will be placed in jobwork/errors (this is also the default).
jobdeleteonsuccess	If "yes", when a job has successfully completed, all work files (hence input files) are deleted.
jobwork	The root of the temporary work directories used by the job. The work directories themselves are named work1, work2 etc.
jobstep	Contains the definition of a job step
stepsequence	Step sequence number starting at 1.
steptype	The steptype identifies which job element defines the step (see section 8.4 below)
stepdetails	Holds additional step definitions which will be used as the step definition file.

8.4 Sample Step Details

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

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

XML Element	Description
operation	The operation (e.g. split). This is defined in the step definition file for the steptype.
sourcetype	Folder, file or tree.
sourcefiles	Source file or folder.
target	Target folder.
errors	Files that cannot be processed are placed in this directory. Inherited from the Job definition.
joboptions	These are steptype-specific parameters that are derived from the options selected in the Job Designer.
advanced Flags	Additional advanced steptype-specific parameters that can be entered manually only.
metadata	PDF file metadata settings derived from the options selected in the Job Designer.
security	PDF file security settings derived from the options selected in the Job Designer.
docoptions	PDF file open options derived from the options selected in the Job Designer.
custom_script	If not left blank, specifies the name of a custom windows script file in %PDFJUNCTIONDIR%\custom or %TIFFJUNCTIONDIR%/custom that will be called in accordance with the custom_script_* settings below.
custom_script_pre_job	If =Y, the custom script will be called at the start of the job
custom_script_post_job	If =Y, the custom script will be called at the end of the job
custom_script_pre_folder	If =Y, the custom script will be called at the start of processing each folder
custom_script_post_folder	If =Y, the custom script will be called at the end of processing each folder

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

Notes

- Note that step-level custom scripts referred to above as supported by TIFF Junction and PDF Junction are different from "custom scripts" that may be used as a step.

- The step-level log files as referred to above as supported by TIFF Junction and PDF Junction are separate from the overall log file used by Autobahn, and can only be specified manually.

9 EXTENDED STEP TYPES: STEP TYPE DEFINITION XML FILES

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

A step type definition will include information such as:

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

9.1 Sample StepType Definition

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

XML Element	Description
steptype	The name that uniquely identifies the step type
command	The command line to be executed.
operation	The “internal” step type name used by the command
joboptions	Setting for the <joboptions> element in <stepdetails>. This value

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

10 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 2008 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 3.03.

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

Examples

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

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

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

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

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

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

10.1.1 General Job Options

Parameter	Notes
/source	Source file or folder
/sourcetype	One of : file, folder, tree
/target	The Target folder
/output	The output filename mask (e.g. %FILENAME)
/messagelevel	Silent,verbose,debug
/logfile	Path to a log file
/loglevel	All,erroronly

/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.
/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.
/processfilesmatching	Pattern
/ignorefilesmatching	Pattern

10.1.2 Split Parameters [used with splittiff and splitpdf operations]

Parameter	**Notes**
/splitby	Pages, ranges, repeatingrange, bookmarks (PDF Only)
/splitrange	Set of page ranges separated by commas that defines which pages from the original should be extracted. The following types of page ranges are allowed : 1 — Specifies a single page 1-3 — Specifies a range of pages
/splitrepeatingrange	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. 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.

10.1.3 Conversion Settings [used with tifftopdf and ocrimagepdf operations]

Parameter	**Notes**
/pdfa	When set to true, will produce a PDF/A-1b compliant PDF file.
/ocr	For tifftopdf, set to true if a searchable PDF is required.
/stampname	Name of the stamp (see the TIFF Junction reference guide for more details)
/stampvalue	For value stamps, this is the value to be used (see the TIFF Junction reference guide for more details)

/textpdffile	OCR, Error, Ignore, Pass through
/imageconversion	Passthrough Forceg4

10.1.4 PDF Conversion Options [used with topdf operations]

Parameter	Notes
/fileconversiontimeout	Limits the amount of time in seconds that can be spent on a conversion. A value of zero means wait indefinitely.
/continueonerror	For multiple file processing, continue if an error occurs with one file.
/errorfolder	Define the folder where documents are placed if their conversion raised an error.
/convertbookmarks	For MS Word, convert bookmarks
/bookmarkdepth	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, empty string is functionally equivalent to: 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 **Note:** If you use non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.
/converthyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
/printallsheets	The flag that indicates whether to print all Excel worksheets or not.
/printbgcolor	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
/Printscale	For Visio files, sets the print scale
/ieheader	This property modifies Internet Explorer's header setting.
/iefooter	This property modifies Internet Explorer's footer setting.
/imagecompression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
/imagedownsizing	If this property is set to True, then the resolution of images are 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 ImageDownsizing property is set to True, then the resolution of images are reduced to the DPI value specified in this property.
/fontembedding	The option PRN_FONT_EMBED_FULLSET (embedding fullset of font) will cause significant increase in PDF file size, especially for

	CJK font, and therefore not recommended. If you need to embed font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
/fontsubstitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from Control Panel).
/fontembedastype0	This option is recommended if you have non-standard fonts like barcode font.
/margintop	Margin in Inches.
/marginbottom	Margin in Inches.
/marginleft	Margin in Inches.
/marginright	Margin in Inches.
/pagewidth	In Inches.
/pageheight	In Inches.
/paperorientation	PORTRAIT or LANDSCAPE
/papersize	Word does not use the paper size setting from 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.
/printcolortype	With this property, you can set the PowerPoint to print either with color, grey scale, or black and white.
/queuewaittimeout	Since PowerPoint cannot perform multiple conversions at the same time, each conversion request will be put into queue, and they are converted one at a time. A waiting print job will timeout if it waits for more than the value specified in QueueWaitTimeout property. Timeout value is in milliseconds. Set to zero if you want to wait indefinitely.
/disablescriptdebugger	Enables/Disables Internet Explorer's script debugger. It is highly recommended that Internet Explorer's script debugger is turned off during printing so that PDF conversion process is not interrupted.
/disableerrordialogoneveryerror	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 PDF conversion process is not interrupted.
/convertwith	This can be used to specify the conversion engine to be used. It can be one of word, excel, auto or generic.
/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.

10.1.5 PDF File Processing

Parameter	Notes
/retainmetadata	Retains PDF metadata
/retainbookmarks	Retains PDF bookmarks
/password	Source PDF password
/resolution	Sets the DPI of images in the output file (used with pdftotif).
/pdfdpi	Sets the DPI of images in the output file (used with ocrimagepdf)

10.1.6 OCR [used with ocrimagepdf and tifftopdf when searchable pdf has been chosen]

Parameter	Notes
Parameter	**Notes**
/ocroutputfiles	Pdf,txt,rtf,htm
/deskew	Straightens the image
/despeckle	Removes dots below the specified pixel size
/autorotate	Rotates the page image if required
/engine	This defines the OCR type to be used and is an integer between 0 and 255. 0 represents maximum speed, 255 represents maximum quality. The default value is 100.
/language	OCR Language dictionary to be used. The default is English.

LANGUAGE	/language Flag Value
English	0
German	1
French	2
Russian	3
Swedish	4
Spanish	5
Italian	6
Russian English	7
Ukrainian	8
Serbian	9
Croatian	10
Polish	11
Danish	12
Portuguese	13
Dutch	14
Czech	15
Roman	16
Hungar	17

<table>
<tr><td></td><td>
<table>
<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>
</td></tr>
<tr><td>/keepimages</td><td>If set to true, the original page images will be used in the result file, with the despeckled or deskewed images used just for the OCR process.</td></tr>
<tr><td>/graphicsproc</td><td>“all” : Treat all Graphics Areas as Text.

“box” : Remove Box Lines in OCR Processing.</td></tr>
<tr><td>/extractimages</td><td>One of the following :

• Auto – TIFF Junction will select the most appropriate method.
• Via Bitmap – The PDF is rasterized using bitmap conversion
• Extract TIFF – The embedded TIFF images are extracted directly.
• Convert to TIFF – The PDF is rasterized using conversion to PDF.
• Native</td></tr>
<tr><td>/mrc</td><td>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.</td></tr>
<tr><td>/optimiseocr</td><td>See section 12.2 for a full explanation of this option.</td></tr>
</table>

10.1.7 PDF Output File Settings – [used with pdfprops]

<table>
<tr><th>Parameter</th><th>Notes</th></tr>
<tr><td>/userpassword</td><td>Sets the user password for the output document</td></tr>
<tr><td>/ownerpassword</td><td>Sets the owner document for the output document.</td></tr>
<tr><td>/permissions</td><td>A comma separated list of document permissions

<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 medications</td></tr>
<tr><td>allowcopy</td><td>Allow text and graphic 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</td></tr>
</table>
</td></tr>
</table>

	<table><tr><td></td><td>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>
/strength	Set to 128 to make use of security settings
/author	Sets the PDF information dictionary value.
/title	“
/subject	“
/keywords	“
/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 tool bar 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.

10.1.8 Sharepoint Settings – [used with sharepointdownload and sharepointupload]

Parameter	Notes
/sharepointurl	Site, the URL of the SharePoint site that you want to access, e.g. http://localhost:10480/testsite
/sharepointlib	Library, the name of the library that you want to access, e.g. "Test Library"
/sharepointext	An optional extension mask that limits those files to manipulate, e.g. “pdf,tif”
/sharepointrecurse	If set to “True” sub-folders of the Sharepoint Library are handled.

10.1.9 Compress PDF Settings (used with compresspdf)

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

11 AUTOBAHN DX .NET API

An API is provided which allows a .Net application to allow execution of existing job definitions and creation "on the fly" of new jobs from XML Job Ticket files. The API DLL is located in the AUTOBAHN\bin folder and a sample Visual Studio 2008 project is provided in the AUTOBAHN \code examples\AutobahnAPIExample folder where AUTOBAHN is the folder where Autobahn DX has been installed – by default this is C:\Aquaforest\Autobahn DX 3.03.

11.1 API Example

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

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

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

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

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

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

                        if(Jobstatus=="Error")
                        {
                               Console.WriteLine("Job  Failed  -  Check  Job
Log for Details");
                        }
                        else if(Jobstatus=="Stopped")
                        {
                               if(JobFinished=="True")
                                      Console.WriteLine("Job Completed");
                               else
                                      Console.WriteLine("Job            Was
Terminated due to Timeout");
                        }
                        job.Dispose();

                  }
                     catch (Exception oExp)
                     {
                        Console.WriteLine(oExp.Message);
                        Console.ReadLine();
                     }
               }
         }
}
```

Constructors	public Job(int jobid) *Create Job object using an existing jobid.*
	public Job(string jobdeffile) *Create Job object using a temporary job definition file.*
Methods	public Start() *Starts the Job.*
	public void WaitForExit(int ms) *Waits for the job to exit, for up to ms milliseconds. The job is stopped if it has not completed.*
	public void WaitForExit() *Waits indefinitely for the job to compete.*
	public virtual void Dispose() *Disposes of the resources associated with the Job.*
	Public string Jobstatus() *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.
	Public string JobFinished() *Indicates whether the job has completed (return value* True*) or has been stopped (return value* False*) as a result of timeout being exceeded.*
	Public void ClearError(string JobStatusPath) When a job is error, ClearError(string jobstatusPath) will clear the error job and set the job status to stopped.

12 OCR Properties File and The Advance Pre Processing Option

12.1 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, news print, 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. Default value is 3

DotMatrix - Set this to True to improve recognition of dot-matrix fonts. 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. 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 vales are as follows:

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

MaxDeskew – Maximum angle by which a page will be deskewed

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

d2.2 – 2x2 dilation applied to all black pixel areas, useful for faint prints.
e2.2 – 2x2 erosion applied to all black pixel areas, useful for heavy prints.
c2.2 – 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 indentified 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 "graphic" or "picture" areas but that actually do contain useful text. Setting NoPictures to True will cause it to ignore areas identified as pictures whilst setting it to False will force OCR of areas identified as pictures.

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

PdfToImage – The SDK ships with two engines for the conversion of PDF pages to images for OCR. The default engine is used when this is set to 0 but if certain PDF source documents are proving problematic then the alternate engine can be used by changing this value to 1.

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. 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 it you are experience difficulties with this value or have any questions then please contact support@aquaforest.com.

12.2 Advance Pre Processing

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

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

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

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

Sample Section of Properties.xml

```
<ImagePreProcessing ID="1">
    <Binarize>-1</Binarize>
    <Morph>c2.2</Morph>
  </ImagePreProcessing>
  <ImagePreProcessingDefaults>
    <RemoveLines></RemoveLines>
    <Binarize>200</Binarize>
    <BlackPixelLimit>0.65</BlackPixelLimit>
    <BoxSize>0</BoxSize>
    <GrayscaleQuality>0</GrayscaleQuality>
    <Jbig2EncFlags>-s</Jbig2EncFlags>
    <Language>0</Language>
    <MaxDeskew>10</MaxDeskew>
    <MinDeskewConfidence>3.0</MinDeskewConfidence>
    <Morph></Morph>
    <Mrc>
      <ForegroundFactor>3</ForegroundFactor>
      <BackgroundFactor>3</BackgroundFactor>
      <Quality>75</Quality>
    </Mrc>
    <Ocr>
      <DotMatrix>False</DotMatrix>
      <OneColumn>False</OneColumn>
      <NoPictures>False</NoPictures>
      <Tables>False</Tables>
    </Ocr>
```

```
    <PdfImageExtraction>
      <PdfToImage>0</PdfToImage>
      <PdfToImageMinRes>200</PdfToImageMinRes>
      <PdfToImageMaxRes>300</PdfToImageMaxRes>
      <PdfToImageIncludeText>False</PdfToImageIncludeText>
    </PdfImageExtraction>
    <RemoveLines></RemoveLines>
    <SavePredespeckle>False</SavePredespeckle>
    <TextLayerExtraction>
      <MaxBoxes>0</MaxBoxes>
      <FilterWidth>130</FilterWidth>
      <FilterHeight>130</FilterHeight>
      <FilterRatio>1</FilterRatio>
      <FilterPercentage>0.6</FilterPercentage>
      <FilterWidthInverted>40</FilterWidthInverted>
      <FilterHeightInverted>40</FilterHeightInverted>
      <FilterRatioInverted>0.95</FilterRatioInverted>
      <FilterPercentageInverted>0.5</FilterPercentageInverted>
    </TextLayerExtraction>
  </ImagePreProcessingDefaults>
```

13 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 3.03\config\” and 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.

14 EXTENDED OCR MODULE

The optional Extended OCR module extends Autobahn DX 3.0 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 15.3 for details)
- Optional Asian Language Support
- Support for multiple languages within a single document from the same alphabet - e.g. French + German + Italian

14.1 Trial Licensing

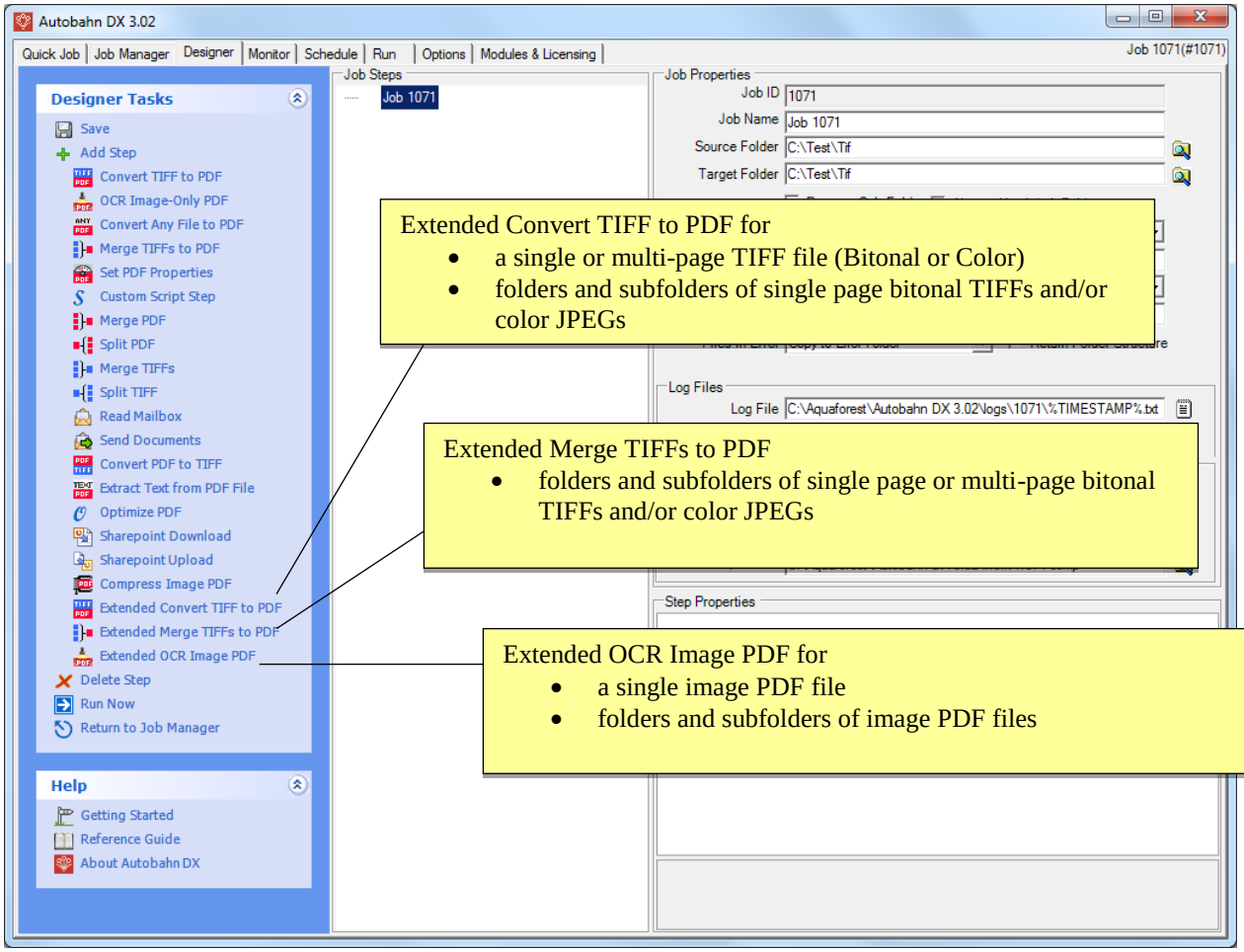
The Extended OCR Module for Autobahn DX is licensed as an optional module and is not enabled by default in the trial version of the product.

To evaluate the module you can enter EXTENDEDOCRTRIAL as the license key under the "Modules and Licensing" tab. To evaluate the Asian OCR capabilities, enter EXTENDEDOCRASIANTRIAL.

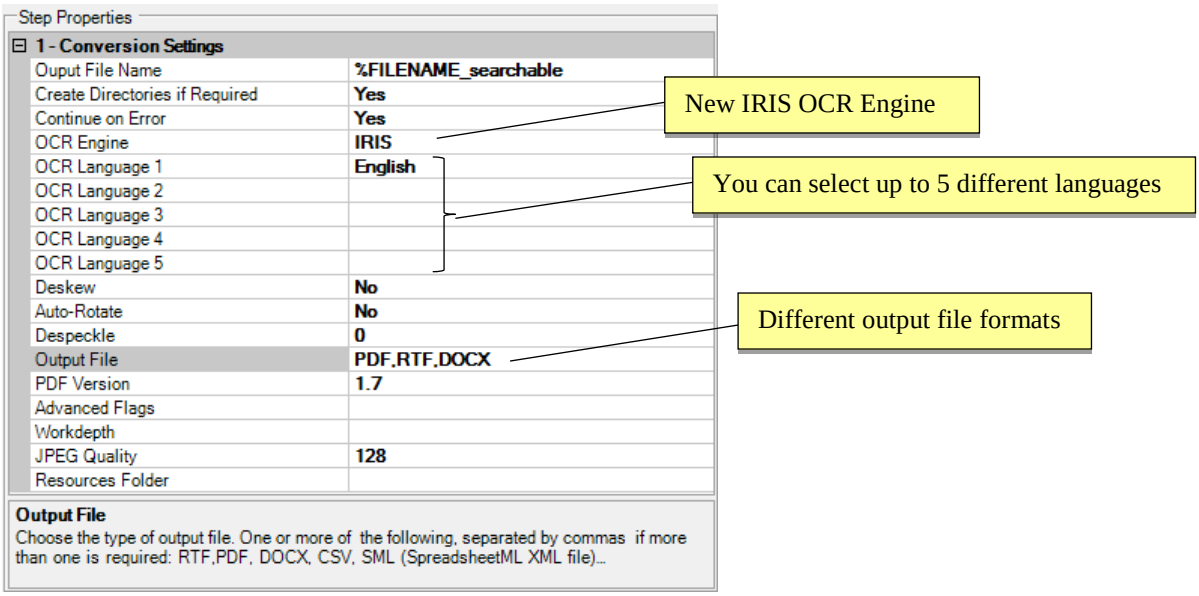
14.2 New Job Manager Steps

The Extended OCR Module provides the following three new job steps:

- Extended Convert TIFF to PDF
- Extended Merge TIFFs to PDF
- Extended OCR Image PDF



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.



14.3 Extended Command Line Interface

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

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

Examples

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

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

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

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

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

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

The three new Extended OCR options use the parameters listed in the table below.

Parameter	**Notes**
/operation	The operation that needs to be carried out: • tifftopdf • mergetifftopdf • ocrimagepdf
/ocrengine	The OCR engine to use. This must 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. • 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. A value needs to be set for sensitivity (see below).
/Sensitivity	The sensitivity, from 1 to 100. With a high sensitivity, less blank pages are detected.
/Deskew	Rotates the image to correct its skew angle.
/Despeckle	Removes all the groups of connected pixels with a number of pixels below the parameter. Suggested range: 1-20.
/Workdepth	This parameter (0 – 255) defines how deeply the OCR engine will analyze a page with 255 being the deepest. For poorer quality documents, higher values can give better recognition results.
/JPEGQuality	This parameter (0 – 255) determines the compression/quality of color JPEG images in generated PDFs. 0 gives the smallest file size whilst 255 gives the best quality. The default value is 128.
/PDFVersion	This determines the PDF version of the generated PDF : 1.4 1.5 1.6 1.7 PDFA1B *(PDF/A-1b)*
/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. **Language** **Code** ENGLISH(Default) 0 GERMAN 1 FRENCH 2 SPANISH 3 ITALIAN 4 BRITISH 5

	SWEDISH	6
	DANISH	7
	NORWEGIAN	8
	DUTCH	9
	PORTUGUESE	10
	BRAZILIAN	11
	GALICIAN	12
	ICELANDIC	13
	GREEK	14
	CZECH	15
	HUNGARIAN	16
	POLISH	17
	ROMANIAN	18
	SLOVAK	19
	CROATIAN	20
	SERBIAN	21
	SLOVENIAN	22
	LUXEMB	23
	FINNISH	24
	TURKISH	25
	RUSSIAN	26
	BYELORUSSIAN	27
	UKRAINIAN	28
	MACEDONIAN	29
	BULGARIAN	30
	ESTONIAN	31
	LITHUANIAN	32
	AFRIKAANS	33
	ALBANIAN	34
	CATALAN	35
	IRISH_GAELIC	36
	SCOTTISH_GAELIC	37
	BASQUE	38
	BRETON	39
	CORSICAN	40
	FRISIAN	41
	NYNORSK	42
	INDONESIAN	43
	MALAY	44
	SWAHILI	45
	TAGALOG	46
	JAPANESE	47 *Requires the Asian OCR option
	KOREAN	48 *Requires the Asian OCR option
	SCHINESE	49 *Requires the Asian OCR option
	TCHINESE	50 *Requires the Asian OCR option
	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 just a Latin alphabet) SERBIAN_LATIN 102 LATIN 103 LATVIAN 104
/createfolders	Create output folder if it does not exists. Default true.
/dpi	Sets the DPI of images in the output file. Set to Auto by default, alternatively can be set to 300, 200 or 150 to force a specific resolution.
/nonimagepdf	This allows control over the treatment of non image-only PDFs, i.e. PDFs that have some text in them as well as images. The options are : • OCR. The document will 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.

/noocr	Whether are not to perform OCR on the document (Yes to not perform OCR, No to perform OCR).
/AdvancedDespeckle	The size of the speckles to remove.
/RemoveWhitePixels	By default, despeckle removes black pixels. If set to true, the despeckle will remove white pixels rather than black pixels.
/Binarization	Whether or not to perform binarization on the document.
/Brightness	The brightness (higher values will darker the result).
/Contrast	The contrast (lower values will darker the result).
/SmoothingLevel	Smoothing may be useful to binarize text with a colored background in order to avoid noisy pixels (0 disables smoothing, higher values smooth more).
/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).
/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	Yes to keep the original image as it is. No to output the image generated after pre-processing is applied.
/resourcesfolder	By default the OCR resources folder is a subfolder of the extendedocr folder. This option allows the resources to be located elsewhere if required.

15 IHQC MODULE

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

15.1 Trial Licensing

The IHQC Module for Autobahn DX is licensed as an optional module and is not enabled by default in the trial version of the product.

To evaluate the module you can enter EXTENDEDOCRIHQCTRIAL as the license key under the "Modules and Licensing" tab. To evaluate IHQC with Asian OCR capabilities, enter EXTENDEDOCRASIANIHQCTRIAL.

15.2 IHQC properties

15.3 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.

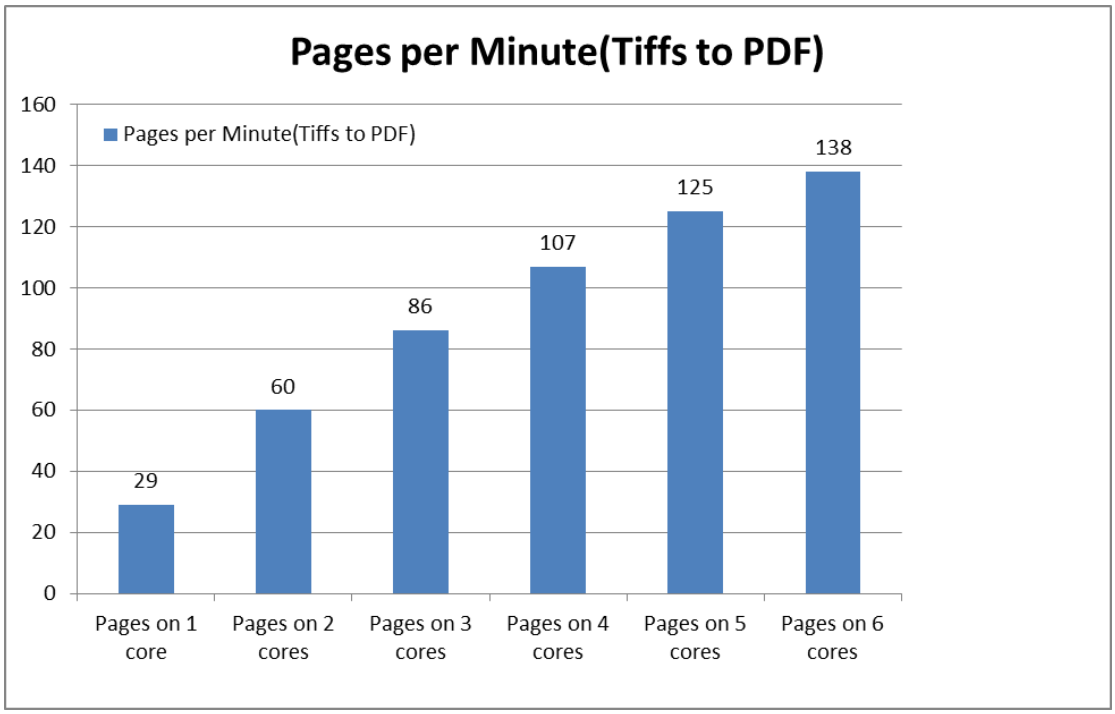
autobahndx.exe /source=c:\in\folder /sourcetype=folder /target=c:\out /output=outfilef /outputtype=pdf operation=mergetifftopdf /mrc=true /ihqclevel=4 /ihqcqf=5 /ocrengine=1

The following parameters are needed to use IHQC:

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

16 MULTICORE MODULE

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



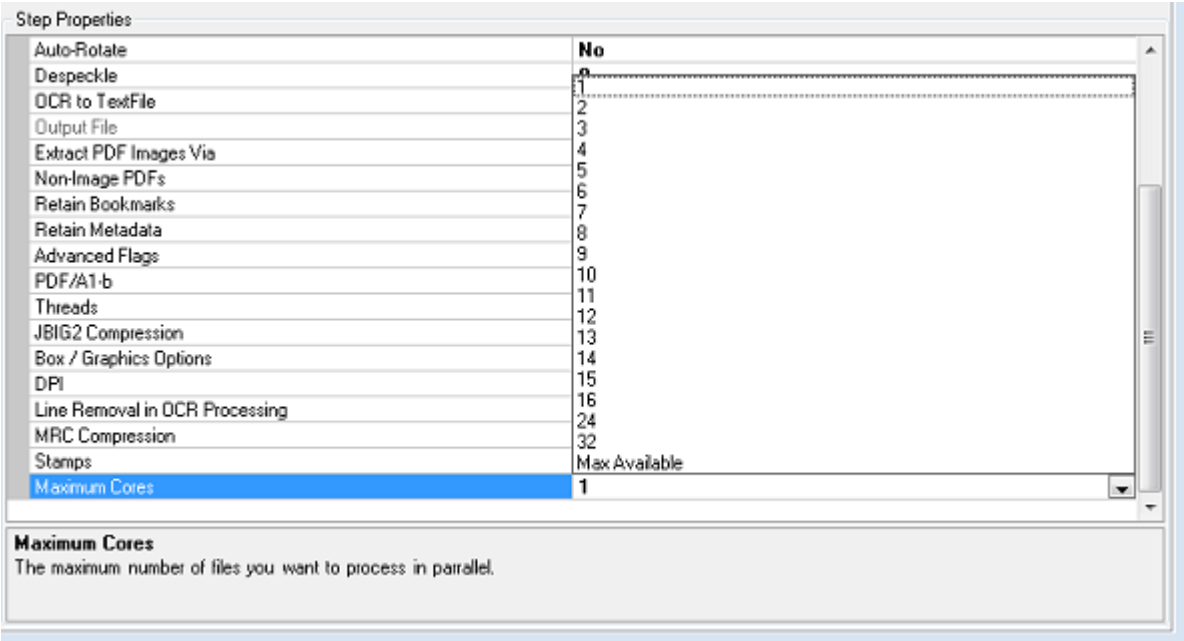
16.1 Trial Licensing

The MultiCore Module for Autobahn DX is licensed as an optional module and is not enabled by default in the trial version of the product.

To evaluate the module you can select the MULTICORETRIAL option as the license key under the "Modules and Licensing" tab.

16.2 Using the MultiCore Module

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



16.3 Command Line Interface

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

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

Example

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

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

The MultiCore module uses the parameter listed below.

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

17 ACKNOWLEDGEMENTS

This product makes use of a number of Open Source components which are included in binary form. The appropriate acknowledgements and copyright notices are given below.

LEPTONICA

Copyright (C) 2001 Leptonica. All rights reserved.

LIBJPEG

This software is based in part on the work of the Independent JPEG Group.

ZLIB

(C) 1995-2004 Jean-loup Gailly and Mark Adler.

ITEXT 4.1.6
Copyright (C) 1999-2009 by Bruno Lowagie and Paulo Soares et all. All Rights Reserved. Binaries distributed under the Mozilla Public License.

CUNEIFORM

Copyright (c) 1993-2008, Cognitive Technologies. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. Neither the name of the Cognitive Technologies nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE

LIBTIFF

Copyright (c) 1988-1997 Sam Leffler. Copyright (c) 1991-1997 Silicon Graphics, Inc.

Permission to use, copy, modify, distribute, and sell this software and its documentation for any purpose is hereby granted without fee, provided that (i) the above copyright notices and this permission notice appear in all copies of the software and related documentation, and (ii) the names of Sam Leffler and Silicon Graphics may not be used in any advertising or publicity relating to the software without the specific, prior written permission of Sam Leffler and Silicon Graphics.

THE SOFTWARE IS PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EXPRESS, IMPLIED OR OTHERWISE, INCLUDING WITHOUT LIMITATION, ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL SAM LEFFLER OR SILICON GRAPHICS BE LIABLE FOR ANY SPECIAL, INCIDENTAL, INDIRECT OR CONSEQUENTIAL DAMAGES OF ANY KIND, OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER OR NOT ADVISED OF THE POSSIBILITY OF DAMAGE, AND ON ANY THEORY OF LIABILITY, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

18 PRODUCT FOLDER STRUCTURE

Directory	Contents
bin	Executables, DLLs and WSH files.
code examples	Sample API and command line projects
config	Next_job_id.xml Jobtmplt.xml Autobahn.config
custom	Custom Script Files
logs	Default location for log files is logs/%JOBID%/%DATESTAMP%.txt
jobdef	Contains the Job Definition files.
jobstatus	Contains the Job Status files.
steptype	Contains the StepType definition files.
samples	Sample documents
work	Default work directory. Subdirectories are created as required underneath this directory and named 9999/work1, 9999/work2 etc. where 9999 is the jobid.
temp	Default directory for temporary files.
Pc	Stand-alone PDF Compression
pj	PDF Junction Install
tj	TIFF Junction Install
license	Contains key.txt which holds the license key
docs	Includes the Autobahn reference guide and release notes