



Autobahn DX Reference Guide

Version 2.50

October 2010

© Copyright 2005-2010 Aquaforest Limited

<http://www.aquaforest.com/>

CONTENTS

1	PRODUCT OVERVIEW AND CONCEPTS.....	3
1.1	OVERVIEW	3
2	AUTOBAHN DX ARCHITECTURE AND CONCEPTS.....	4
2.1	AUTOBAHN WINDOWS SERVICE	4
2.2	XML JOB DEFINITION FILES	4
2.3	AUTOBAHN ADMINISTRATION.....	5
2.3.1	<i>Quick Job</i>	5
2.3.2	<i>Job Manager</i>	5
2.3.3	<i>Designer</i>	5
2.3.4	<i>Monitor</i>	5
2.4	DOCUMENT DIRECTORIES	5
2.5	DOCUMENT PROCESSING.....	5
2.6	JOB STATUS.....	5
2.7	AUTOBAHN DX .NET JOB API.....	5
2.8	MULTI-CORE/CPU SUPPORT	5
3	INSTALLATION AND LICENSING.....	6
3.1	SYSTEM REQUIREMENTS	6
3.2	PRODUCT LICENSING.....	6
3.3	MULTI-CPU/CORE LICENSING	6
3.4	ENTERING LICENSE KEYS.....	6
4	AUTOBAHN DX QUICK START.....	7
5	USING THE QUICK JOB TAB.....	10
5.1.1	<i>Job Definition Section</i>	12
5.1.2	<i>OCR Options Section</i>	13
5.1.3	<i>OCR PDF Options Section</i>	13
5.1.4	<i>PDF to TIFF Options Section</i>	14
5.1.5	<i>Split Options Section</i>	14
5.1.6	<i>Other Options Section</i>	15
5.1.7	<i>Task Log Section</i>	15
5.2	SECURITY SETTINGS.....	16
5.2.1	<i>Target Document Security</i>	17
5.3	DOCUMENT METADATA	17
5.4	DOCUMENT DISPLAY OPTIONS	18
5.4.1	<i>Display Modes</i>	18
5.4.2	<i>Settings</i>	19
5.5	ADVANCED FLAGS	19
6	THE AUTOBAHN DX ADMINISTRATION TOOL	20
6.1	JOB MANAGER	20
6.2	JOB DESIGNER.....	21
6.2.1	<i>Menu Items</i>	21
6.2.2	<i>Fields</i>	22
6.2.3	<i>Workflow Processing versus In-Place Processing</i>	24
6.2.4	<i>Example In-Place Job Setup</i>	24
6.2.5	<i>Step Types</i>	25
6.2.6	<i>Step Type Properties</i>	26
6.2.7	<i>Convert TIFF to PDF</i>	27
6.2.8	<i>OCR Image-Only PDF</i>	29
6.2.9	<i>Convert Any Document to PDF</i>	31
6.2.10	<i>Merge TIFFs to PDF</i>	33
6.2.11	<i>Set PDF Properties</i>	34
6.2.12	<i>Custom Script</i>	35

6.2.13	Merge PDF.....	35
6.2.14	Split PDF.....	35
6.2.15	Merge TIFFs.....	36
6.2.16	Split TIFF.....	36
6.2.17	Read Inbox.....	36
6.2.18	Send Documents.....	37
6.2.19	Convert PDF to TIFF.....	37
6.2.20	Extract Text from PDF.....	37
6.2.21	Sharepoint Download.....	38
6.2.22	Sharepoint Upload.....	38
6.2.23	Stamps.....	38
6.3	MONITOR.....	40
6.4	JOB SCHEDULING.....	41
6.4.1	Ad-Hoc.....	41
6.4.2	Watched Folder / Continuous Scheduling.....	42
6.4.3	Daily Scheduling.....	42
6.5	EMAIL OPTIONS.....	42
6.6	RUNNING JOBS VIA JOB MANAGER OR DESIGNER.....	43
7	SCRIPTING CUSTOM STEPS.....	44
7.1	CUSTOM SCRIPT EXAMPLE.....	44
7.2	CUSTOM EXAMPLE SCRIPT.....	45
7.3	FURTHER EXAMPLES.....	46
7.3.1	Postscript to PDF : custom_ps.wsf.....	46
7.3.2	MSG to PDF : custom_MSG.wsf.....	47
8	JOB DEFINITION XML FILES.....	48
8.1	OVERVIEW OF JOB DEFINITION CREATION AND PROCESSING.....	48
8.2	JOB IDS.....	48
8.3	SAMPLE JOB DEFINITION FILE (EXCLUDING STEP DETAILS).....	48
8.4	SAMPLE STEP DETAILS.....	50
9	EXTENDED STEP TYPES : STEP TYPE DEFINITION XML FILES.....	52
9.1	SAMPLE STEP TYPE DEFINITION.....	52
10	AUTOBAHN DX COMMAND LINE INTERFACE.....	54
10.1.1	General Job Options.....	54
10.1.2	Split Parameters [used with splittiff and splitpdf operations].....	55
10.1.3	Conversion Settings [used with tifftopdf and ocrimagepdf operations].....	55
10.1.4	PDF Conversion Options [used with topdf operations].....	56
10.1.5	PDF File Processing.....	58
10.1.6	OCR [used with ocrimagepdf and tiffiopdf when searchable pdf has been chosen].....	58
10.1.7	PDF Output File Settings – [used with pdfprops].....	60
10.1.8	Sharepoint Settings – [used with sharepointdownload and sharepointupload].....	61
11	AUTOBAHN DX .NET API.....	62
11.1	API EXAMPLE.....	62
12	ACKNOWLEDGEMENTS.....	64
13	PRODUCT FOLDER STRUCTURE.....	65

1 PRODUCT OVERVIEW AND CONCEPTS

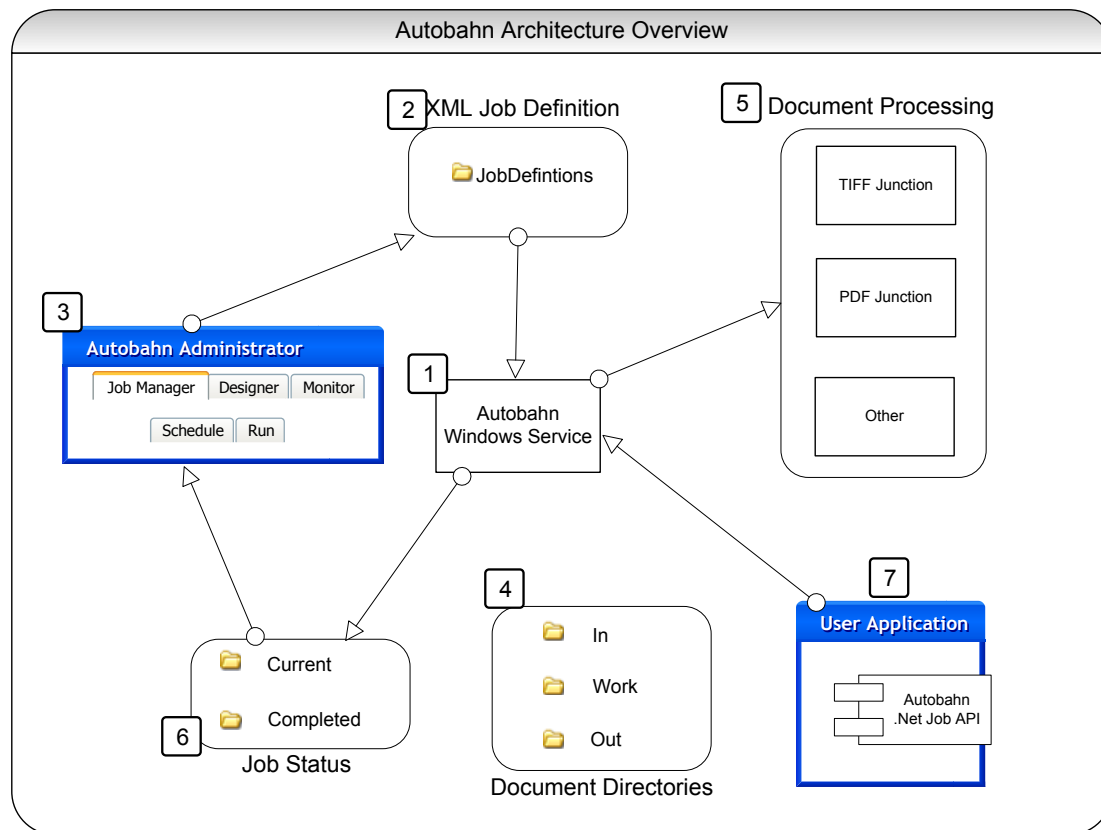
1.1 Overview

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 (Network Server)
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			•	•	•	•	•
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, XP Professional and Vista	•	•	•	•	•	•	•
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.



2.1 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 7 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 9 for further details.

2.8 Multi-Core/CPU Support

Enhanced support for multi-core/CPU architectures has been 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 introduces threaded processing on the "OCR a 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.

3 INSTALLATION AND LICENSING

3.1 System Requirements

Supported Operating Systems	Windows 2003 (32 bit) Windows 2003 (64 bit) Windows 2008 (32 bit) Windows 2008 (64 bit) Windows XP (32 bit) Windows Vista (32 bit)
Disk Space	150Mb
.Net Framework	2.0

3.2 Product Licensing

Autobahn DX has three license levels which are described below :

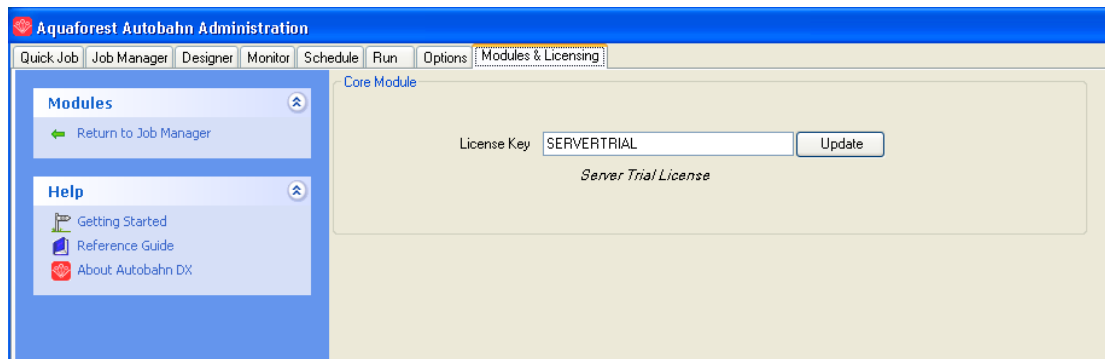
License	Description
Workstation	This license only allows use of the Quick Job tab. To test this license in trial module use the license key <i>WORKSTATIONTRIAL</i>
Server	This allows all functions with the following exceptions <ul style="list-style-type: none">• Use of the Read Mailbox & Send Documents steps.• Access to watched folders by multiple users is disallowed. To test this license in trial module use the license key <i>SERVERTRIAL</i>
Network Server	This allows all functions including <ul style="list-style-type: none">• Use of the Read Mailbox & Send Documents steps.• Access to watched folders by multiple users. To test this license in trial module use the license key <i>NETWORKSERVERTRIAL</i>

3.3 Multi-CPU/Core Licensing

The Network Server edition is licensed for use of 4 threads. Standard Server licenses allow use of 2 threads. It is possible to upgrade a Server license to a Server (4-Core) license. Please contact sales@aquaforest.com for details.

3.4 Entering License Keys

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

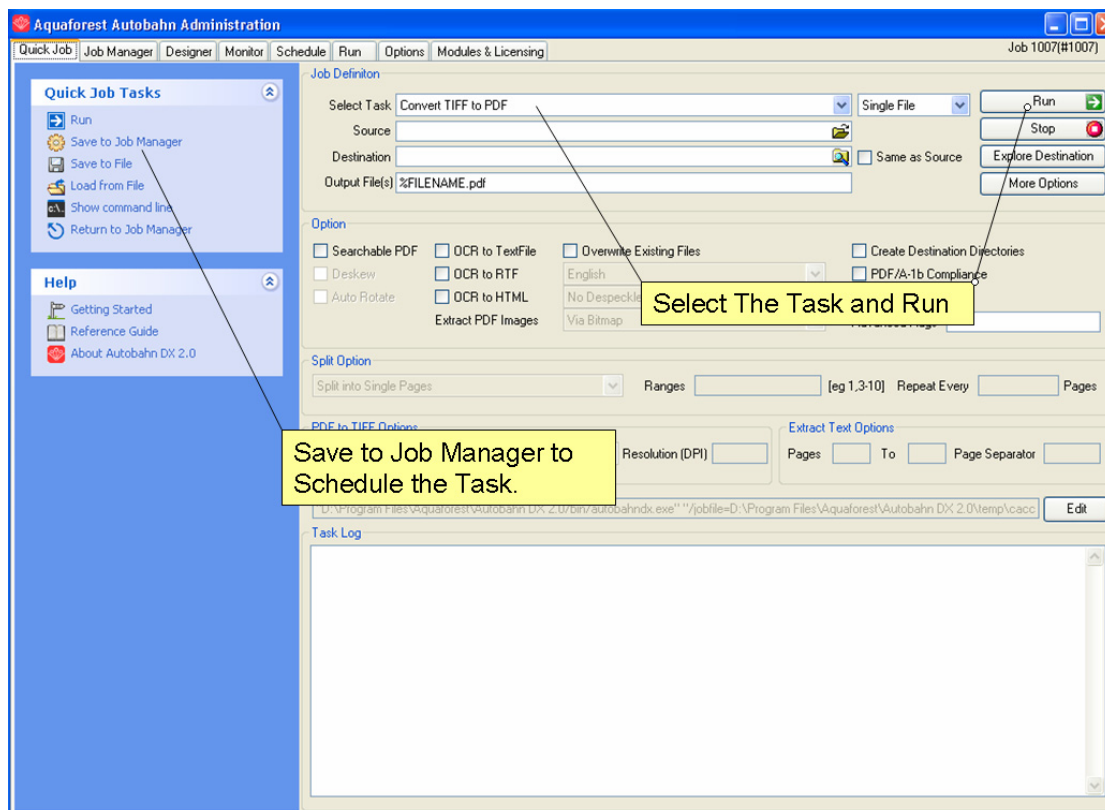


4 AUTOBAHN DX QUICK START

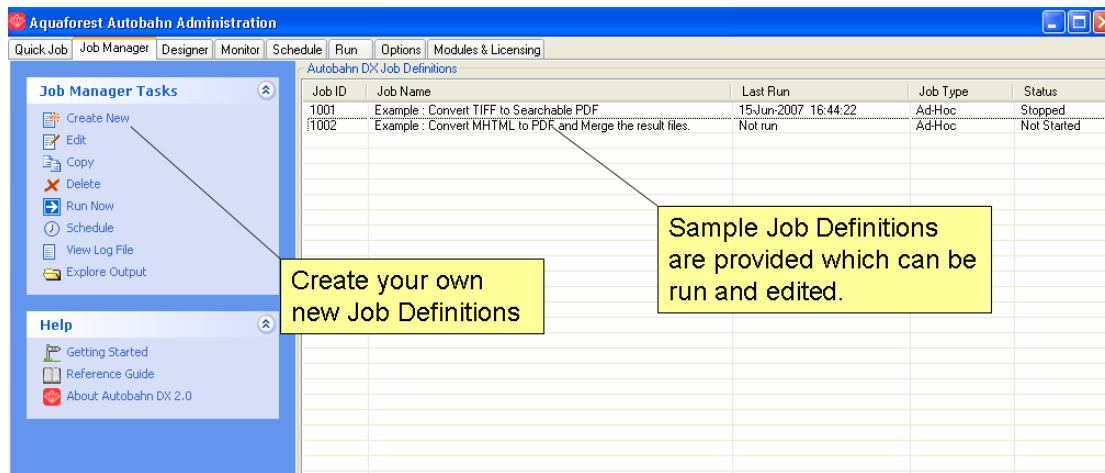
Autobahn DX is designed to allow definition of document processing jobs (“Job Definitions”) and then scheduling them to run continuously on watched folders or on a daily scheduled basis. The annotated screen shots in this section are designed to enable rapid familiarization with the product.

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.

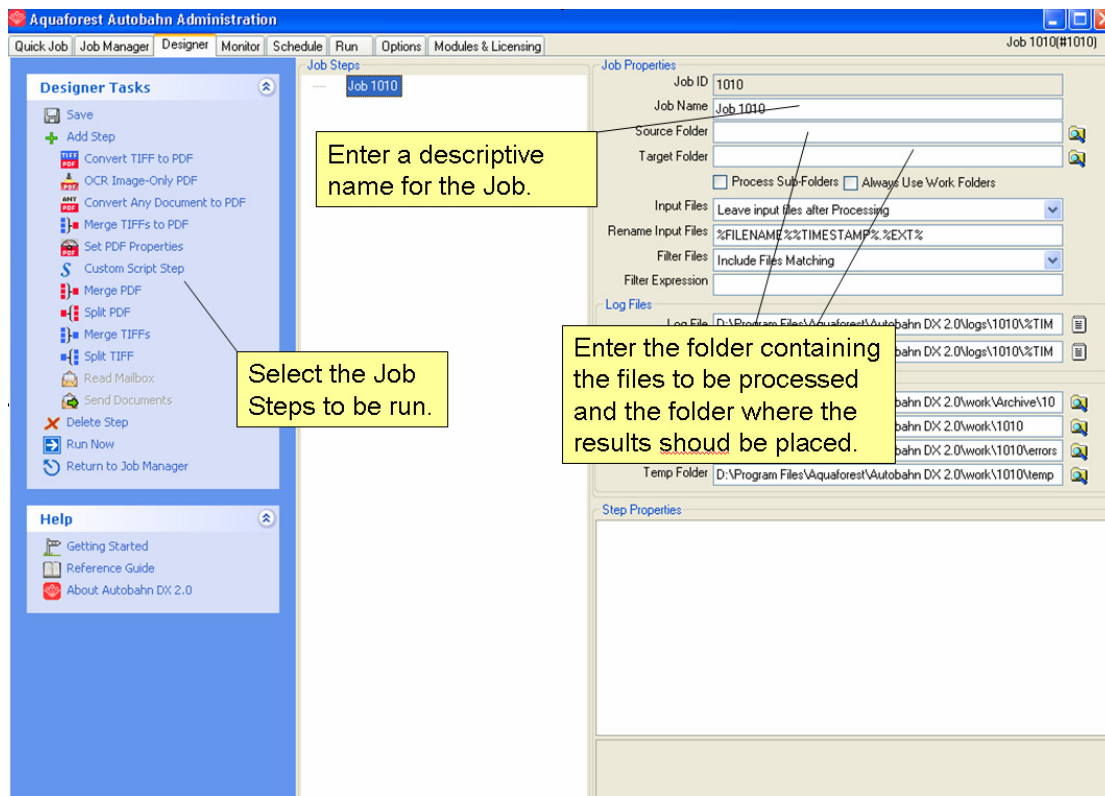
1. The Quick Job Screen



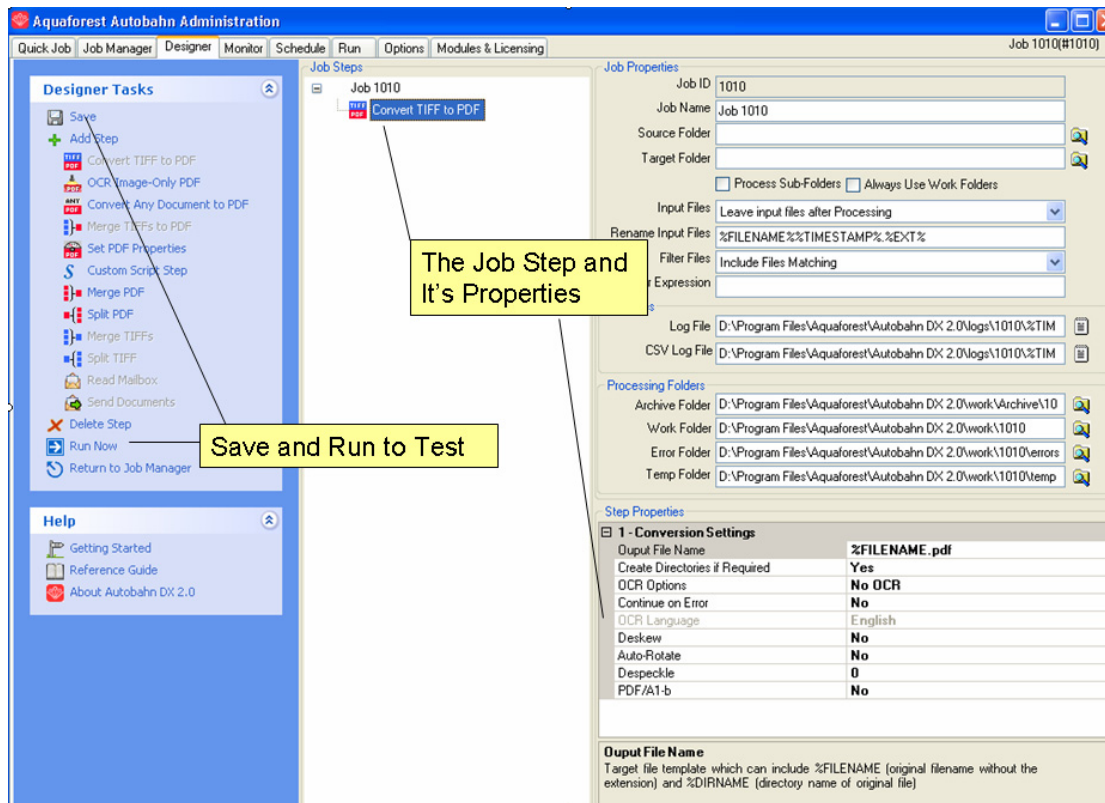
2. The Job Manager Screen



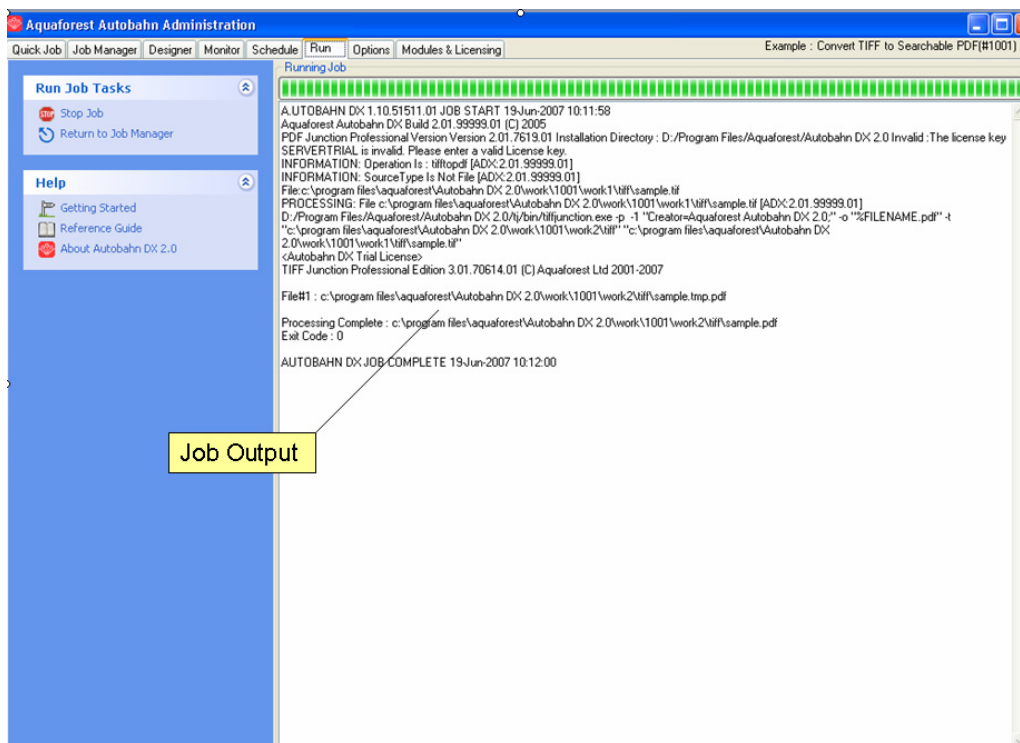
3. The Job Designer Screen



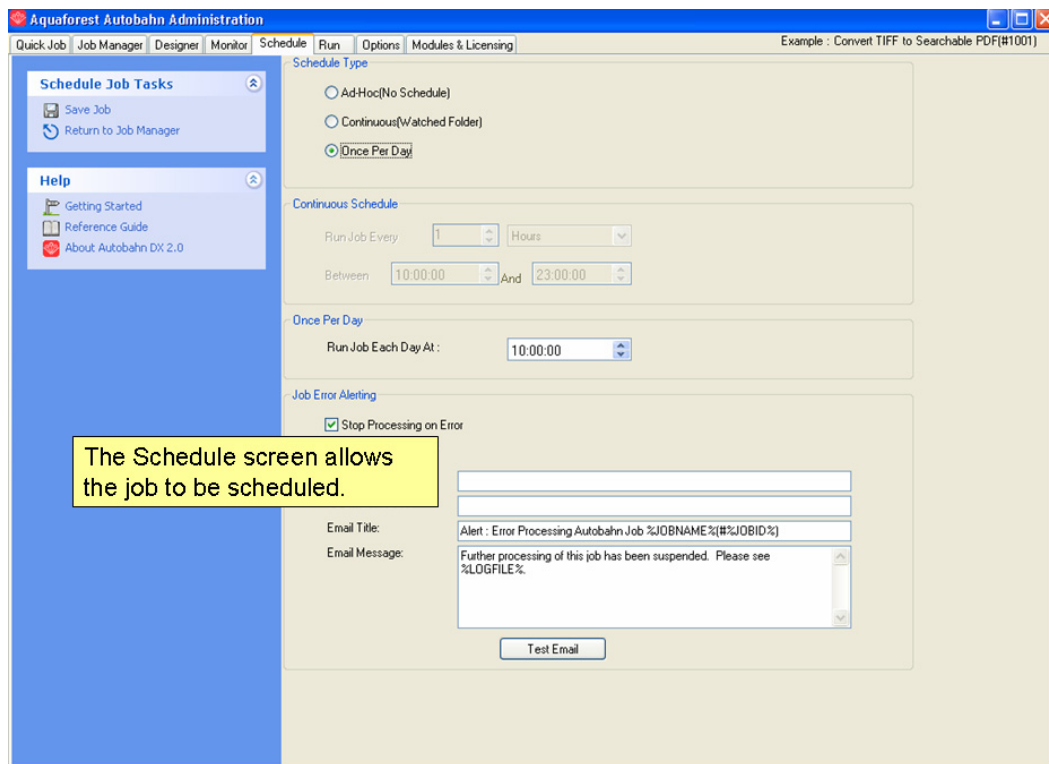
4. Defining a Job



5. Testing the Job Definition



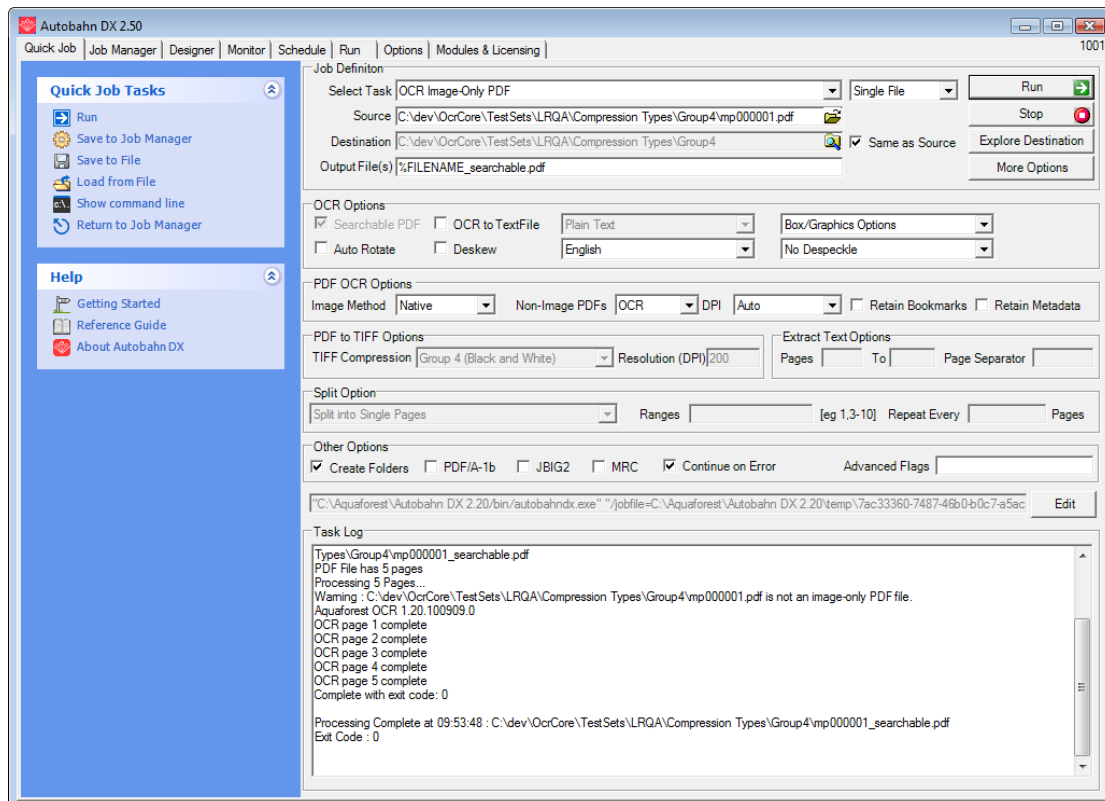
6. Scheduling the Job



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 or Network 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	<p>This defines which task is to be run and may be one of :</p> <ul style="list-style-type: none"> 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)	<p>This defines the template output file name.</p> <table border="1"> <tr> <td>%FILENAME</td><td>Source file name without .pdf</td></tr> <tr> <td>%UNIQUEn</td><td>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.</td></tr> <tr> <td>%DIRNAME</td><td>Source directory name of the directory currently being processed.</td></tr> <tr> <td><i>String</i></td><td>(Any string)</td></tr> </table> <p>For example, a split operation using %FILENAME-part%UNIQUE6.pdf would give <i>filename.pdf</i> split into <i>filename-part000001.pdf filename-part000002.pdf</i> etc.</p>	%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.	<i>String</i>	(Any string)
%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.								
<i>String</i>	(Any string)								

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	<p>This allows control over the method used to extract images from PDF files for OCR processing when using the “OCR an Image-Only PDF”. The options are :</p> <ul style="list-style-type: none">• Auto – 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 OCR'd text directly into a copy of the original PDF rather than creating an entirely new PDF. <p>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 (eg a scanned document with text bates stamps) – for these images the Bitmap or conversion to TIFF method is required.</p> <p>Note methods other than “Native” actually changes the non-image parts of the PDF to image + text.</p> <p>IMPORTANT NOTE : As the “Native” method does not make any changes to the images in the PDF file, it is not compatible with the JBIG2 or MRC compression, PDF/A compliance option 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.</p>

Non-Image PDFs	<p>This allows control over the treatment of non image-only PDFs, ie PDFs that have some text in them as well as images. The options are :</p> <ul style="list-style-type: none"> • OCR. The document will be OCR'd using the image method defined by "Image Method" • Raise Error. The task will terminate with an error. If "On Error Continue" is set this then behaves as Skip. This is the default. • Skip. The document will not be processed. • Pass Through. The file will not be processed, but a copy of the document will be made and named as if the processing had occurred.
DPI	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	<p><i>One of:</i></p> <p>Split into single pages</p> <p>Split by ranges (See below)</p> <p>Split by repeating ranges (See below)</p>				
Page Ranges	<p>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 :</p> <table border="1"> <tr> <td>1</td><td>Specifies a single page</td></tr> <tr> <td>1-3</td><td>Specifies a range of pages</td></tr> </table>	1	Specifies a single page	1-3	Specifies a range of pages
1	Specifies a single page				
1-3	Specifies a range of pages				
Repeating Range	<p>Apply the page range to each set of <i>Page Ranges</i> pages within the document. For example if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages.</p> <p>Hence the file is split such that the first output file contains pages 2-4 from the original document, the second contains</p>				

	pages 6-8 and so on.
--	----------------------

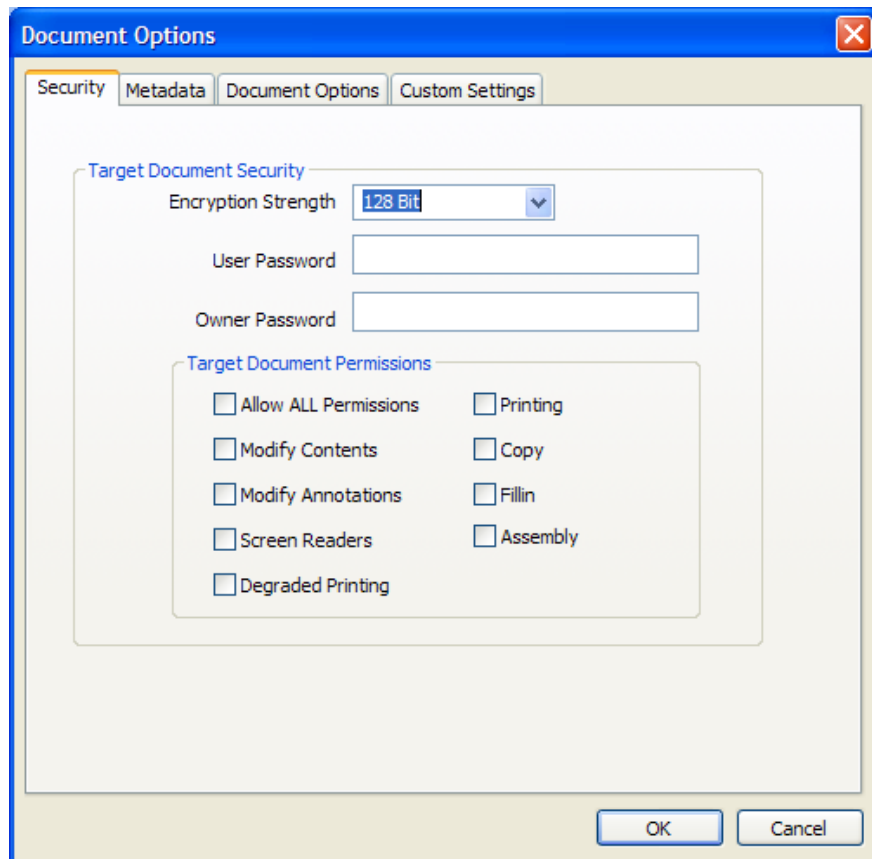
5.1.6 Other Options Section

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 Compliance	Processes the output PDF file to ensure it is compliant with ISO 19005 / PDF/A-1b. See section 6 for more details.
JBIG2	If this option is chosen, bitonal images in the output PDF will be compressed using JBIG2 compression rather than the default of Group 4. This will result in smaller files for PDFs with black and white images.
MRC	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



The image shows a 'Document Options' dialog box with a blue title bar and a close button. It contains four tabs: 'Security' (selected), 'Metadata', 'Document Options', and 'Custom Settings'. The 'Security' tab is active and displays two main sections: 'Target Document Security' and 'Target Document Permissions'. The 'Target Document Security' section includes a dropdown for 'Encryption Strength' set to '128 Bit', and text input fields for 'User Password' and 'Owner Password'. The 'Target Document Permissions' section contains two columns of checkboxes. The first column includes 'Allow ALL Permissions', 'Modify Contents', 'Modify Annotations', 'Screen Readers', and 'Degraded Printing'. The second column includes 'Printing', 'Copy', 'Fillin', and 'Assembly'. All checkboxes are currently unchecked. At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

Document Options

Security Metadata Document Options Custom Settings

Target Document Security

Encryption Strength: 128 Bit

User Password:

Owner Password:

Target Document Permissions

<input type="checkbox"/> Allow ALL Permissions	<input type="checkbox"/> Printing
<input type="checkbox"/> Modify Contents	<input type="checkbox"/> Copy
<input type="checkbox"/> Modify Annotations	<input type="checkbox"/> Fillin
<input type="checkbox"/> Screen Readers	<input type="checkbox"/> Assembly
<input type="checkbox"/> Degraded Printing	

OK Cancel

5.2.1 Target Document Security

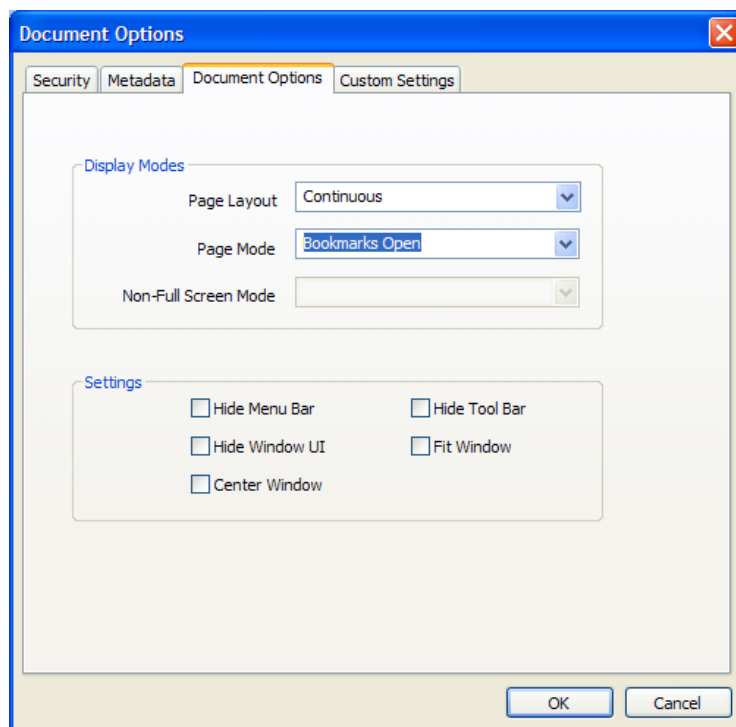
Option	Description	
User Password	A password that will be required to open the document.	
Owner Password	A password that will be required to change the document permissions.	
Permissions		
	Permission	Description
	Allow ALL Permissions	All all the permissions below.
	Printing	Allow high-quality printing
	Modify Contents	Allow assembly (see below) and other document medications
	Copy	Allow text and graphic copying and extraction
	Modify Annotations	Allow modification of annotations
	Fillin	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

The screenshot shows a Windows-style dialog box titled "Document Options". It has four tabs: "Security", "Metadata" (which is selected and highlighted in orange), "Document Options", and "Custom Settings". Inside the "Metadata" tab, there is a section titled "Document Metadata" containing five text input fields with labels: "Author", "Title", "Subject", "Keywords", and "Creator". At the bottom right of the dialog box are "OK" and "Cancel" buttons.

This page allows setting of the generated PDF Document Properties.

5.4 Document Display Options



5.4.1 Display Modes

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

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.
Center 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.

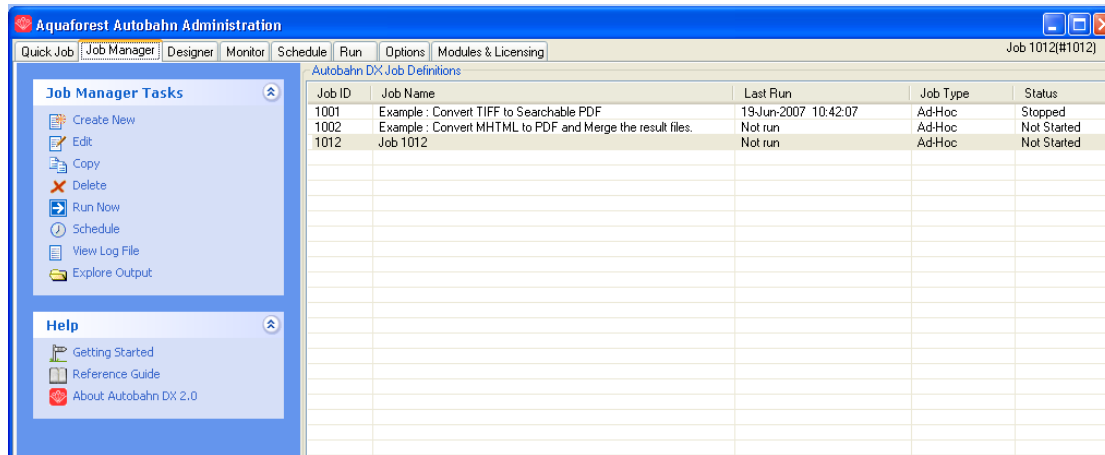
Option	Description
-B 100	Blank Page Removal This option can be used when converting TIFF files to Searchable PDFs. A value should be provided which specifies the pixel threshold to be used to determine whether a page is blank or not. If a page is deemed to be blank then it is omitted from the output file. A suggested value is 100 ie using the new -B advanced flag : -B 100
-D	Dot Matrix Processing T his 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 despeckling the page for OCR processing, but the image retained in the final file is not despeckled.
-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

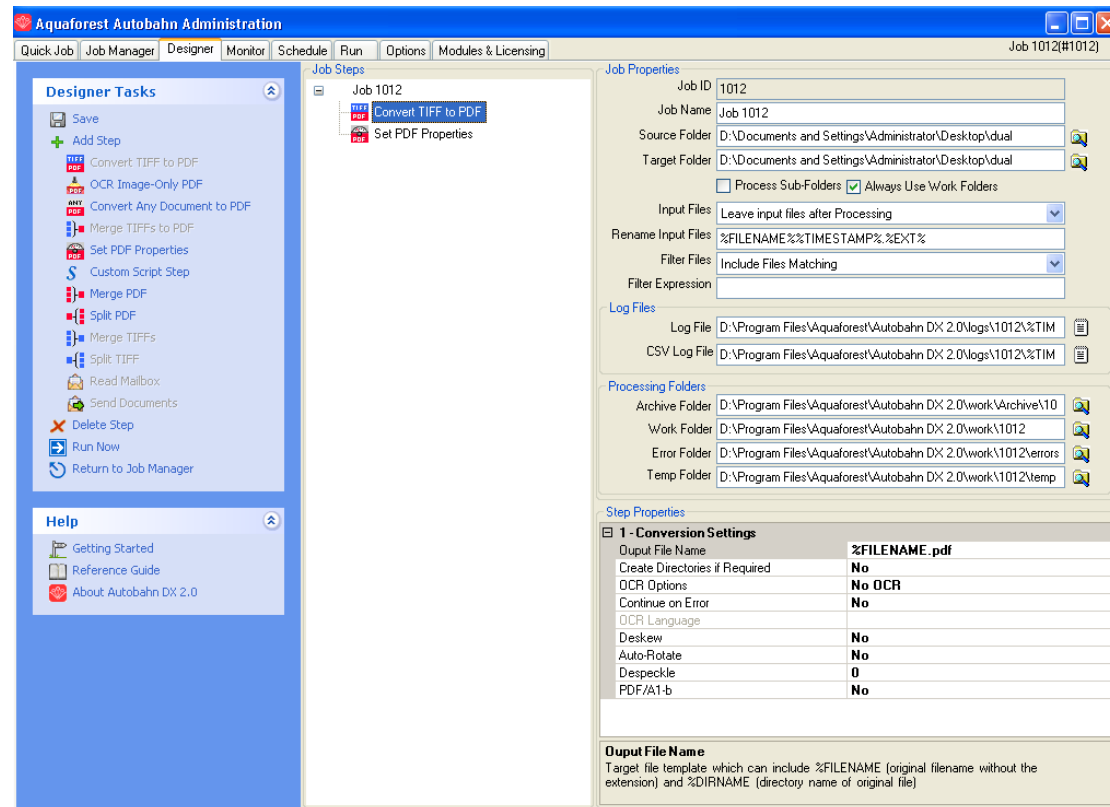
This section of the application displays a list of all current Job Definitions and the set of operations to be performed shown in the menu table below.



Menu Item	Action
Create New	Creates a new, skeleton job definition from jobtmpl.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 (eg correctly processing files that are added to the source folder after a job has started) but can slow down large jobs.
Input files	<p>This option determines what happens to the input files once processing has been completed. The options are :</p> <p><i>Leave input files after processing:</i> Files are left in the Source Folder.</p> <p><i>Move to archive after processing:</i> Files are moved to the Archive Folder.</p> <p><i>Copy to archive after processing:</i> Files are copied to the Archive Folder.</p> <p><i>Move input files to target folder after processing:</i> Input files are placed in the same folder as the output files.</p> <p><i>Delete input files after successful processing:</i> Input files are deleted.</p>
Input Files renaming	<p>This determines how input files will be renamed when moved to the Target or Archive folder. The default is :</p> <p>%FILENAME%%TIMESTAMP%.%EXT%</p>

Filter Files	<table><tr><th>Filter Files Option</th><th>Description</th></tr><tr><td>Include Files Matching</td><td>Only files matching the Filter Expression are included.</td></tr><tr><td>Exclude</td><td>Files matching the Filter Expression are excluded.</td></tr><tr><td>Include with Document Count Limit</td><td>For example “*.pdf; 3000” would limit the job to 3000 PDF files.</td></tr><tr><td>Include Unprocessed PDFs Only</td><td><p>This would limit files selected to PDFs that have not been OCR'd.</p><p>A file is deemed to have been OCR'd if either</p><ul style="list-style-type: none">(a) It has a custom metadata tag “AQUAFORSTOCR”(b) It has one image per page and only has “invisible” text.<p>This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</p></td></tr><tr><td>Include Unprocessed PDFs Only – with Document Count Limit</td><td>As above, but limited to the number of files specified in the filter.</td></tr></table>	Filter Files Option	Description	Include Files Matching	Only files matching the Filter Expression are included.	Exclude	Files matching the Filter Expression are excluded.	Include with Document Count Limit	For example “*.pdf; 3000” would limit the job to 3000 PDF files.	Include Unprocessed PDFs Only	<p>This would limit files selected to PDFs that have not been OCR'd.</p> <p>A file is deemed to have been OCR'd if either</p> <ul style="list-style-type: none">(a) It has a custom metadata tag “AQUAFORSTOCR”(b) It has one image per page and only has “invisible” text. <p>This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</p>	Include Unprocessed PDFs Only – with Document Count Limit	As above, but limited to the number of files specified in the filter.
	Filter Files Option	Description											
	Include Files Matching	Only files matching the Filter Expression are included.											
	Exclude	Files matching the Filter Expression are excluded.											
	Include with Document Count Limit	For example “*.pdf; 3000” would limit the job to 3000 PDF files.											
	Include Unprocessed PDFs Only	<p>This would limit files selected to PDFs that have not been OCR'd.</p> <p>A file is deemed to have been OCR'd if either</p> <ul style="list-style-type: none">(a) It has a custom metadata tag “AQUAFORSTOCR”(b) It has one image per page and only has “invisible” text. <p>This should be used in conjunction with a “Non-Image PDF” setting of “Rasterize and OCR” to ensure that all PDF files are processed.</p>											
	Include Unprocessed PDFs Only – with Document Count Limit	As above, but limited to the number of files specified in the filter.											
NB : Work Folders must be used to enable use of filters.													
Filter Expression	<p>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.</p> <p>Examples :</p> <ul style="list-style-type: none">*.pdf*.doc *.ppt *.xls*.pdf; 3000												
Log File	Path of the job log file. This can include %TIMESTAMP%, which is the timestamp of the job start time, and will therefore give a new log file for each run.												
CSV Log File	<p>Path of the job CSV log file. This can include %TIMESTAMP%, which is the timestamp of the job start time, and will therefore give a new log file for each run. The columns in the CSV file are :</p> <ul style="list-style-type: none">• Job Start – Time Job Started• Source Files – Full path to the source file• Target File – Full path to the target file• Job Stopped – Time Job Finished• Success – True or False; Files that could not be processed will have a value of False.												

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 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:\qat\221008\in 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:\qat\221008\arch.

The screenshot displays the 'Job Properties' window for a job named 'In-Place Conversion of PDF to Searchable PDF'. The interface is divided into several sections:

- Job Steps:** A tree view on the left shows the job structure, including 'In-Place Conversion of PDF to Searchable PDF' and a sub-step 'OCR Image-Only PDF'.
- Job Properties:**
 - Job ID:** 1001
 - Job Name:** In-Place Conversion of PDF to Searchable PDF
 - Source Folder:** C:\qat\221008\in
 - Target Folder:** C:\qat\221008\in
 - Process Sub-Folders:** ☒ **Always Use Work Folders:** ☒
 - Input Files:** Copy to Archive after Processing
 - Rename Input Files:** %FILENAME%.%EXT%
 - Filter Files:** Include Unprocessed PDFs Only - with Document Count Limit
 - Filter Expression:** *.pdf;1000
 - Files in Error:** Move to Error Folder
 - Retain Folder Structure:** ☐
- Log Files:**
 - Log File:** C:\Program Files\Aquaforest\Autobahn DX 2.0\logs\1012\%TIME
 - CSV Log File:** C:\Program Files\Aquaforest\Autobahn DX 2.0\logs\1012\%TIME
- Processing Folders:**
 - Archive Folder:** C:\qat\221008\arch
 - Work Folder:** C:\Program Files\Aquaforest\Autobahn DX 2.0\work\1001
 - Error Folder:** C:\Program Files\Aquaforest\Autobahn DX 2.0\work\1012\errors
 - Temp Folder:** C:\Program Files\Aquaforest\Autobahn DX 2.0\work\1012\temp
- Step Properties:**
 - OCR Engine:** Balanced OCR
 - OCR Language:** English
 - Deskew:** No
 - Auto-Rotate:** No
 - Despeckle:** 0
 - OCR to TextFile:** No
 - Output File:** Plain Text
 - Extract PDF Images Via:** Auto
 - Non-Image PDFs:** Rasterize and OCR
 - Retain Bookmarks:** No

6.2.5 Step Types

Step Type	Description	Underlying Functionality Provided by
Convert TIFF to PDF	<p>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).</p> <p>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.</p>	TIFF Junction
OCR Image-Only PDF	<p>Creates a searchable PDF file from the set of images from an image-only PDF file.</p> <p>Depending upon the Step Type Properties chosen, a separate text or HTML file may be produced from the OCR process.</p>	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 6 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 (eg 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 Document to PDF”. Each property has a description associated with it which is displayed when the property is highlighted.

Step Properties

1 - Conversion Settings

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

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

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	Select Yes if the result file must be PDF/A1-b Compliant
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 :

	<p><i>“Treat all Graphics Areas as Text”</i>. This option will ensure the entire document is processed as text.</p> <p><i>“Remove Box Lines in OCR Processing”</i>. This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the OCR engine. It does not remove boxes from the final image. Technically, this option removes connected elements with a minimum area (by default 100 pixels).</p>
Threads	Defines the number of processing threads for conversion processes involving OCR. See section 2.8 and 3.3 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. <ul style="list-style-type: none">• Auto (default)• Via Bitmap• Extract TIFF• Convert to TIFF
PDF/A1-b	Select Yes if the result file must be PDF/A1-b Compliant
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	<p>This defines how non-image PDFs should be treated. The options are :</p> <ul style="list-style-type: none"> • Raise Error (default) • Rasterize and OCR • Skip • Pass Through
DPI	<p>When OCRing a PDF, the PDF is rasterized to produce a TIFF file which is then OCR'd. By default the TIFF image resolution is determined from the images embedded in the source PDF but this flag can be used to override default processing and specify the DPI of the TIFF that will be generated.</p>
Box/Graphics Processing	<p>By default, if an area of the document is identified as a graphic area then no OCR processing is run on that area. However, certain documents may include areas or boxes that are identified as “graphic” or “picture” areas but that actually do contain useful text.</p> <p>To ensure that the OCR engine can be forced to process such areas there are two options :</p> <p><i>“Treat all Graphics Areas as Text”</i>. This option will ensure the entire document is processed as text.</p> <p><i>“Remove Box Lines in OCR Processing”</i>. This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the OCR engine. It does not remove boxes from the final image. Technically, this option removes connected elements with a minimum area (by default 100 pixels).</p>
Threads	<p>Defines the number of processing threads. See section 2.8 and 3.3 for further details.</p>
Line Removal in OCR Processing	<p>This removes lines and boxes during OCR processing to improve recognition – particularly in cases where characters “touch” lines.</p>
MRC	<p>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.</p>
Stamps	<p>See section 6.2.23 for full details</p>

6.2.9 Convert Any Document 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 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.

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

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 (eg %UNIQUE4 for 4 digits), %BOOKMARK and %PAGE n (eg %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 Filling	Allow filling of form fields
Allow Screen Readers	Allow extraction of text and graphics in support of accessibility.
Allow Assembly	Allow rotation, insertion or deletion of pages.
Allow Degraded Printing	Allow low-quality printing
Author	Sets the Author property
Title	Sets the Title property
Subject	Sets the Subject property
Keywords	Sets the Keywords property
Creator	Sets the Creator property
Page Layout	The setting for 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.

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 (eg 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 (eg 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 or NETWORKSERVER License.

Parameter	Notes
imap_server	The IMAP server address eg <code>imap.company.co.uk</code>
email_account	The email account to be read eg <code>pdf@company.com</code>
email_password	Password for the account. This is held encrypted.
in_mailbox	Mailbox to read eg Inbox
processed_mailbox	Mailbox to move processed email to eg 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. <i>Files are named <code>name@timestamp@email@.ext</code> where <code>name.ext</code></i>

	<i>is the original filename, email is the user's email address and timestamp is an integer timestamp to ensure uniqueness.</i>
Include	Regular expression. If specified, only files matching the expression will be processed. Eg *.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. Eg *.pdf

6.2.18 Send Documents

Use of this step type requires a SERVER or NETWORKSERVER License.

Parameter	Notes
Domain	The sending domain Eg aquaforest.com
SMTP Server	SMTP Server address Eg smtp.aquaforest.com
Send Logon	Sending user logon eg admin.
Send User Password	Password for the account.
Sending User	Sending user eg 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. Note the need to use escaped quotes for prefixes and suffixes with spaces.

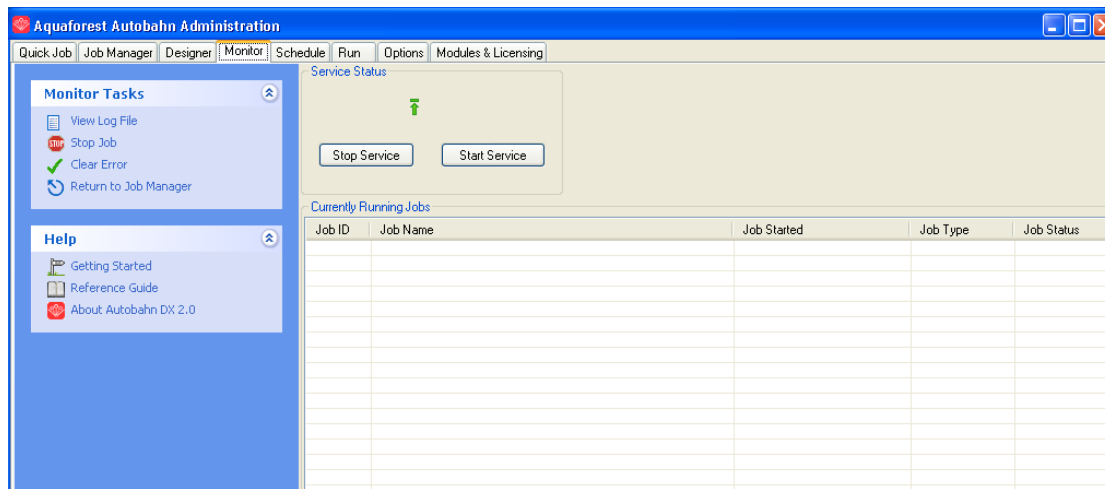
```
/stamppref=\"Page Number\" /stampsuff=Final /stampstart=123  
/stampdigits=6 /stamppos=0 /stampstype=0
```

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

6.3 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

Aquaforest Autobahn Administration

Quick Job | Job Manager | Designer | Monitor | **Schedule** | Run | Options | Modules & Licensing

Example : Convert TIFF to Searchable PDF(#1001)

Schedule Job Tasks

- Save Job
- Return to Job Manager

Help

- Getting Started
- Reference Guide
- About Autobahn DX 2.0

Schedule Type

☒ Ad-Hoc (No Schedule)

☐ Continuous (Watched Folder)

☐ Once Per Day

Continuous Schedule

Run Job Every: 1 Hours

Between: 15:00:00 And 23:00:00

Once Per Day

Run Job Each Day At: 15:00:00

Job Error Alerting

☒ Stop Processing on Error

☐ Send Email on Error

From Email Address:

To Email Address:

Email Title: Alert : Error Processing Autobahn Job %JOBNAME%(##%JOBID%)

Email Message: Further processing of this job has been suspended. Please see %LOGFILE%.

Test Email

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 5.5 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.

The screenshot shows the 'Aquaforest Autobahn Administration' window with the 'Options' tab selected. The 'SMTP Settings' section is active, displaying the following fields and values:

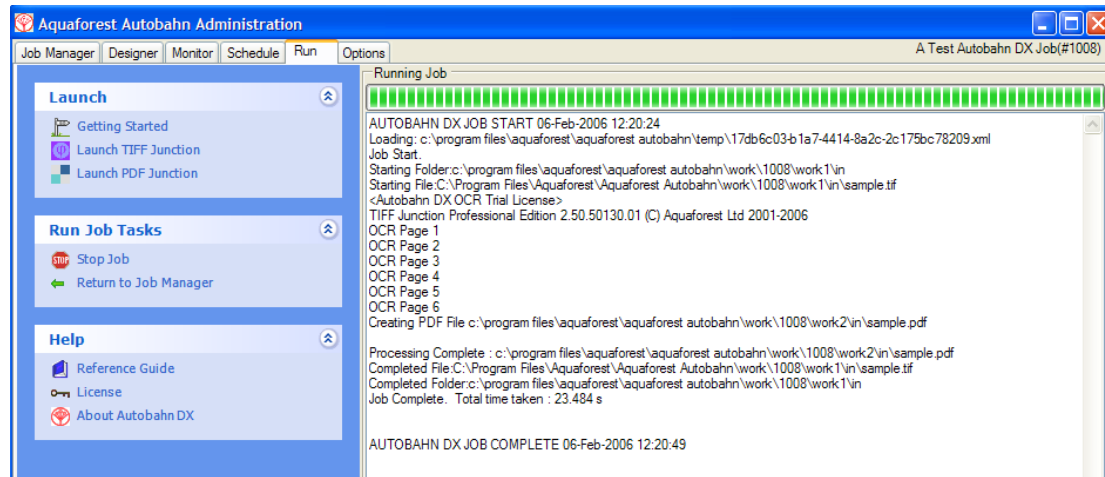
- Server: mail.system.com
- SMTP Port: 25
- Username: john.anite@system.com
- Password: [encrypted]
- Need Authentication: [checked]

On the left, there is a sidebar with 'Options' and 'Help' sections. The 'Options' section includes 'Save Options' and 'Return to Job Manager'. The 'Help' section includes 'Getting Started', 'Reference Guide', and 'About Autobahn DX'.

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.



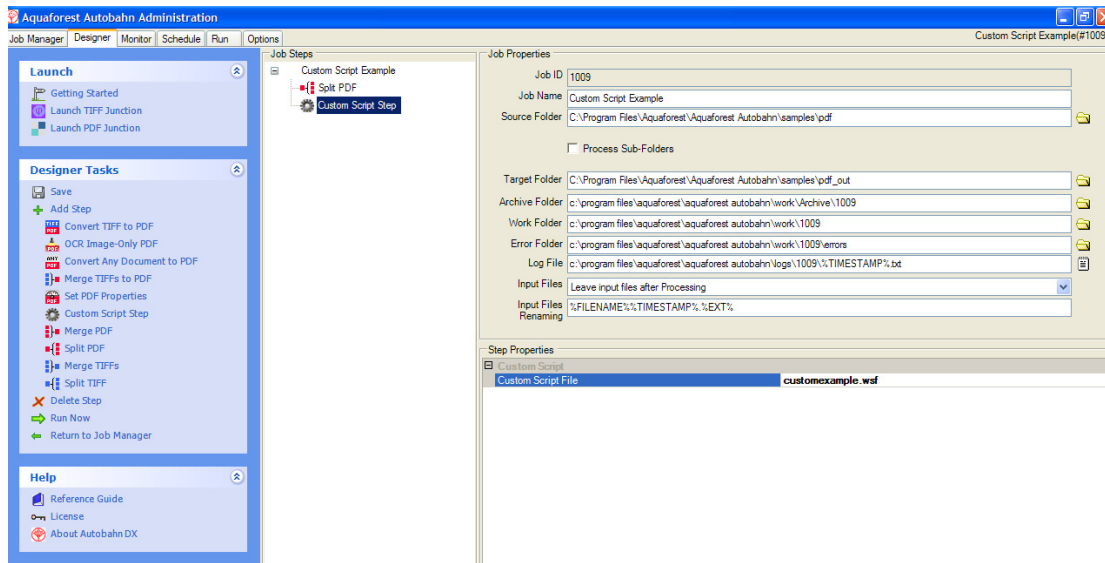
Menu Item	Action
Stop	Stops the currently running job.
Return to Job Manager	Returns to the job manager with the current job selected.

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.

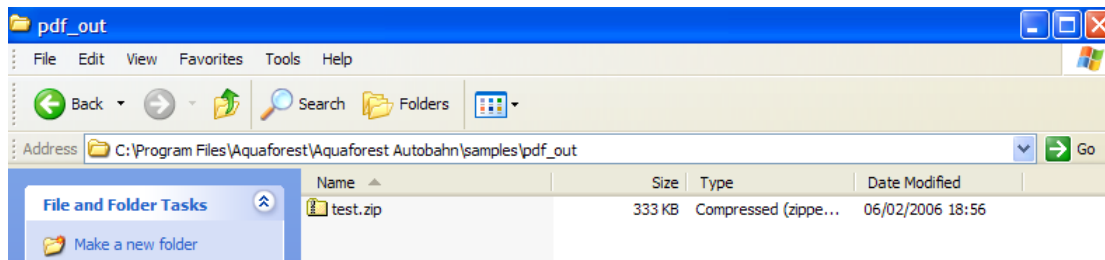
7.1 Custom Script Example

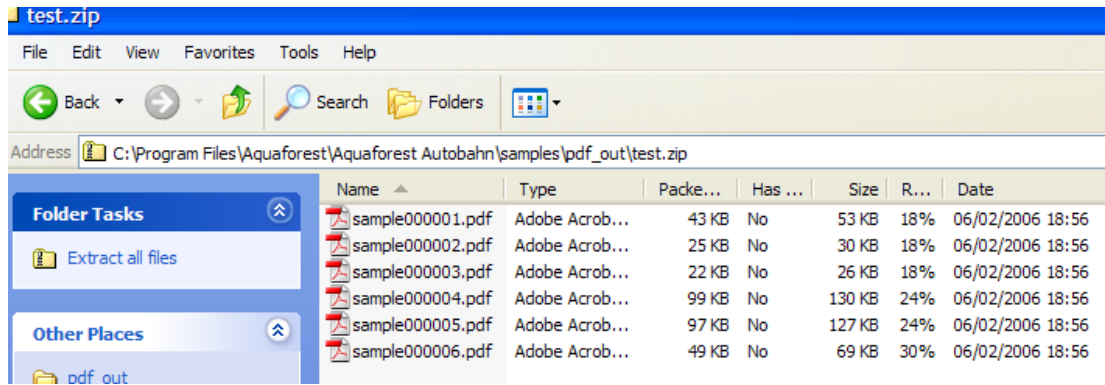
Included with the product is an example custom script (in the Autobahn DX custom folder) that will split a PDF file into single pages and then create a single ZIP file that contains the pages. The example makes use of the Freebyte Zip application which you can download from here <http://www.freebyte.com/download/fbzipack.exe>, run and unpack into C:\Aquaforest\Autobahn DX 2.1\custom.



Custom script files must be placed in C:\Aquaforest\Autobahn DX 2.1\custom.. In this case the custom script file is customexample.wsf.

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 `outFolder` 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 2.1\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.

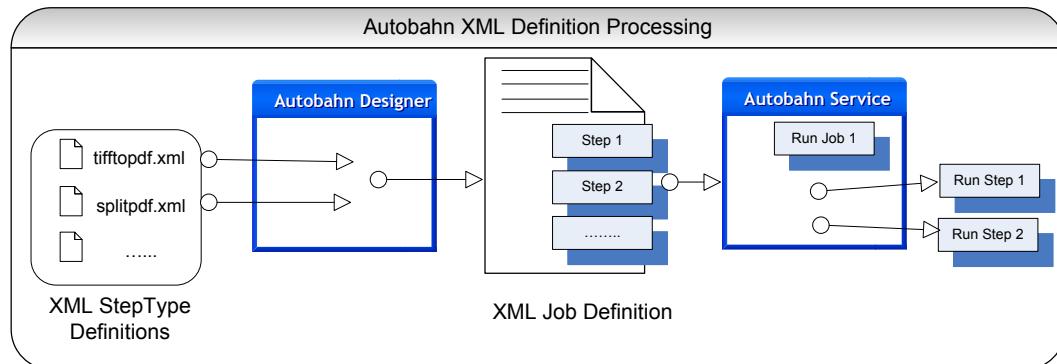
The script should be placed in the Autobahn DX custom folder (by default C:\Aquaforest\Autobahn DX 2.1\custom) and can be used by creating a job with a custom script step and entering *custom_msg.wsf* as the value of Custom Script File.

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 5). 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>
  <sceduleeveryunits>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 7.4 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 (eg 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 (eg TIFF, Word, PDF)
- Supported Source Types (eg Folder, Email..)
- Type of Document Output (eg 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>Output File Name</attributename>
    <attributetype>value</attributetype>
    <defaultvalue>%FILENAME.pdf</defaultvalue>
    < attributejoboptions >-o%ATTRIBUTEVALUE%</attributejoboptions
  </attribute>
</attributes>
```

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

	may be added to by attribute settings.
wrapper	Wrapper element to be used when creating temp step job file.
multithread	Yes or no – determines whether multiple instances of the steptype may run concurrently.
validinputtype	Type(s) of files accepted by the step. (NB this is NOT the same as a file extension). * means all types. Multiple values may be separated by commas.
outputtype	Type of file output by the step. (NB this is NOT the same as a file extension)
iconfile	Bitmap file with the steptype icon. The file is assumed to be in the %STEPTYPEDEFDIR% directory.
attribute	The definition of an attribute which the user may set a value for
attributeid	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 or Network 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 2.50.

```
autobahndx.exe /source=value [/option=value]
```

10.1.1 General Job Options

Parameter	Notes
/jobfile	This is the path to an XML Job Definition file saved from the Quick Job tab. When this parameter is specified, no other parameters are required.
/operation	This defines the type of processing to be run and is one of the following : <ul style="list-style-type: none">• Tifftopdf [Convert TIFF to PDF]• Ocrimagepdf [OCR Image PDF]• Topdf [Convert any file to PDF]• Mergetifftopdf [Merge TIFF to PDF]• Pdfprops [Set PDF Properties]• Mergepdf [Merge PDF Files]• Splitpdf [Split PDF]• Mergetiff Merge TIFF Files]• Splittiff [Split TIFF Files]• Pdftotiff [Convert PDF to TIFF]• Tiffinfo [Show Information about a TIFF file]• Version [Shows Autobahn DX version information]
/source	Source file or folder
/sourcetype	One of : file, folder, tree
/target	The Target folder
/output	The output filename mask (eg %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)				
/splitrage	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 : <table border="1" data-bbox="592 976 1275 1043"> <tr> <td>1</td><td>Specifies a single page</td></tr> <tr> <td>1-3</td><td>Specifies a range of pages</td></tr> </table>	1	Specifies a single page	1-3	Specifies a range of pages
1	Specifies a single page				
1-3	Specifies a range of pages				
/splitrepeatingrange	Apply the page range to each set of <i>Page Ranges</i> pages within the document. For example if 2-4 is specified for page ranges, 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	<p>This property will take effect only when the <u>ConvertBookmarks</u> property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, empty string is functionally equivalent to:</p> <p>Heading 1 1 Heading 2 2 Heading 3 3 Heading 4 4 Heading 5 5 Heading 6 6 Heading 7 7 Heading 8 8 Heading 9 9</p> <p>Note: If you use non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.</p>
/converthyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
/printallsheets	The flag that indicates whether to print all Excel worksheets or not.
/printbgcolor	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
/Printscale	For Visio files, sets the print scale
/ieheader	This property modifies Internet Explorer's header setting.
/iefooter	This property modifies Internet Explorer's footer setting.
/imagecompression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
/imagedownscaling	If the this property is set to True, then the resolution of images are reduced to the DPI value specified in the <u>ImageDownsizeResolution</u> property.
/imagejpegquality	The allowed value range is from 5 to 100 with 100 being the highest quality.
/imagedownsizeresolution	If the <u>ImageDownscaling</u> 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.
/marginotop	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 <u>PageWidth</u> and <u>PageHeight</u> properties.
/printcolortype	With this property, you can set the PowerPoint to print with either color, gray scale, or black and white.
/queuewaittimeout	Since PowerPoint cannot perform multiple conversion 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.
/disablescripdebugger	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.

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. <table border="1"> <thead> <tr> <th>LANGUAGE</th><th>/language Flag Value</th></tr> </thead> <tbody> <tr><td>English</td><td>0</td></tr> <tr><td>German</td><td>1</td></tr> <tr><td>French</td><td>2</td></tr> <tr><td>Russian</td><td>3</td></tr> <tr><td>Swedish</td><td>4</td></tr> <tr><td>Spanish</td><td>5</td></tr> <tr><td>Italian</td><td>6</td></tr> <tr><td>Russian English</td><td>7</td></tr> <tr><td>Ukrainian</td><td>8</td></tr> <tr><td>Serbian</td><td>9</td></tr> <tr><td>Croatian</td><td>10</td></tr> <tr><td>Polish</td><td>11</td></tr> <tr><td>Danish</td><td>12</td></tr> <tr><td>Portuguese</td><td>13</td></tr> <tr><td>Dutch</td><td>14</td></tr> <tr><td>Czech</td><td>15</td></tr> <tr><td>Roman</td><td>16</td></tr> <tr><td>Hungar</td><td>17</td></tr> <tr><td>Bulgar</td><td>18</td></tr> <tr><td>Slovenian</td><td>19</td></tr> </tbody> </table>	LANGUAGE	/language Flag Value	English	0	German	1	French	2	Russian	3	Swedish	4	Spanish	5	Italian	6	Russian English	7	Ukrainian	8	Serbian	9	Croatian	10	Polish	11	Danish	12	Portuguese	13	Dutch	14	Czech	15	Roman	16	Hungar	17	Bulgar	18	Slovenian	19
LANGUAGE	/language Flag Value																																										
English	0																																										
German	1																																										
French	2																																										
Russian	3																																										
Swedish	4																																										
Spanish	5																																										
Italian	6																																										
Russian English	7																																										
Ukrainian	8																																										
Serbian	9																																										
Croatian	10																																										
Polish	11																																										
Danish	12																																										
Portuguese	13																																										
Dutch	14																																										
Czech	15																																										
Roman	16																																										
Hungar	17																																										
Bulgar	18																																										
Slovenian	19																																										

	Latvian	20
	Lithuanian	21
	Estonian	22
	Turkish	23
/keepimages	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.	
/graphicsproc	“all” : Treat all Graphics Areas as Text. “box” : Remove Box Lines in OCR Processing.	
/extractimages	One of the following : <ul style="list-style-type: none"> • 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 	
/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.	

10.1.7 PDF Output File Settings – [used with pdfprops]

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

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 2.50.

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) <i>Create Job object using an existing jobId.</i>
	public Job(string jobdeffile) <i>Create Job object using a temporary job definition file.</i>
Methods	public Start() <i>Starts the Job.</i>
	public void WaitForExit(int ms) <i>Waits for the job to exit, for up to ms milliseconds. The job is stopped if it has not completed.</i>
	public void WaitForExit() <i>Waits indefinitely for the job to compete.</i>
	public virtual void Dispose() <i>Disposes of the resources associated with the Job.</i>
	Public string Jobstatus() <i>Returns the status of the job which may be either Stopped or Error. This should be checked in conjunction with the value of JobFinished() to determine whether the job completed before being stopped.</i>
	Public string JobFinished() <i>Indicates whether the job has completed (return value True) or has been stopped (return value False) as a result of timeout being exceeded.</i>
	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 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 al. 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.

13 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%/%%TIMESTAMP%.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.
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