PDF Junction
Reference Guide Version 3.4

Standard and Professional Editions

Version 3.41
May 2017

© Aquaforest Limited 2001-2017
Web: www.aquaforest.com
E-mail: info@aquaforest.com

Aquaforest

CONTENTS

1 INTRODUCTION 3

1.1 INSTALLATION 3
1.2 SYSTEM REQUIREMENTS 3
1.3 CONFIGURATION FOR CONVERSION TO PDF 3
1.3.1 Native Application 3
1.3.2 Service Configuration 4
1.4 LICENCING AND PURCHASING 4
1.5 CONVENTIONS 4

2 PDF JUNCTION WINDOWS INTERFACE 5

2.1 MAIN WINDOW 5
2.1.1 Job Definition Section 6
2.1.2 Any File to PDF Section 7
2.1.3 PDF to TIFF Section 8
2.1.4 Split Options Section 8
2.1.5 Merge PDF Section 9
2.1.6 Extract Text from PDF Section 10
2.1.7 Web Optimize Section 10
2.1.8 PDF Barcode 10
2.2 DOCUMENT OPTIONS 12
2.2.1 Security Settings 12
2.2.2 Properties 13
2.2.3 Msoffice 14
2.2.4 Conversions 16
2.2.5 Script 18

3 USING PDF JUNCTION FROM THE COMMAND LINE 19

3.1 GENERATING COMMAND LINE 19
3.2 COMMAND LINE SYNTAX 19
3.3 EXAMPLE COMMANDS 21
3.3.1 Example jobid Commands 21
3.3.2 Example Split Commands 21
3.3.3 Example Merge Commands 22
3.3.4 Example Set Commands 22
3.3.5 Example Get Command 23
3.4 /JOBOPTION SPECIFICATIONS 24
3.4.1 Job Options for Split Operations 24
3.4.2 Job Options for Merge Operations 26
3.4.3 Job Options for Set Operations 26
3.4.4 Job Options for Get Operations 27
3.5 /METADATA SPECIFICATION 28
3.6 /SECURITY SPECIFICATION 28
3.7 /DOCOPTIONS SPECIFICATION 29

4 CONVERSION TO PDF –COMMAND LINES INTERFACE (PROFESSIONAL EDITION) 30

4.1 COMMAND LINE INTERFACE 30
4.2 CONFIGURATION FILE 34
4.2.1 Overview of the Different Conversions Methods 34

4.2.2 Comparison of the features provided by the different methods 35
4.2.3 In-depth Look at the Configuration file..... 36

5 PDF TO TIFF CONVERSION COMMAND LINE INTERFACE 38

5.1 COMMAND LINE INTERFACE 38

6 PDF TEXT EXTRACTION COMMAND LINE INTERFACE 39

6.1 COMMAND LINE INTERFACE 39

7 WEB OPTIMIZATION (LINEARIZATION) CLI 39

7.1 COMMAND LINE INTERFACE 39

8 SET XMP PROPERTIES (PROFESSIONAL EDITION ONLY)..... 40

8.1 SETTING METADATA 40
8.2 COMMAND LINE INTERFACE 42

9 PDF BARCODE COMMANDLINE INTERFACE 43

10 XML JOB TICKET FILES 46

11 CUSTOM SCRIPTS..... 48

12 XML "INFORMATION" FILES 50

13 PDF JUNCTION FILES AND DIRECTORIES..... 51

14 SUPPORT 51

15 ACKNOWLEDGEMENTS 51

1 INTRODUCTION

PDF Junction provides a set of processing tools for PDF files, including splitting, merging, setting of security and other attributes, and generation of XML metadata files. In addition, the Professional version of the product also supports conversion of Microsoft Office and other documents to PDF, and the setting of XMP properties.

The product has both a GUI and command-line interface. Batch processes can be defined using the GUI to create an XML Job Ticket file which can be run using the GUI, or using the command line. In addition, all options can be used from the command line without having to use the GUI.

1.1 Installation

The installation MSI will guide you through the process of setting up PDF Junction.

1.2 System Requirements

- Server 2016, Server 2012 (R2), Server 2008 (R2).
- Windows 7 (64 bit), Windows 8 (64 bit), Windows 10 (64 bit).
- Microsoft .Net Framework 4.5.2
- Visual C++ Runtime Redistributable x64 2012 and 2010
- Native Applications (such as Microsoft Office) are required for PDF Conversion.
- PDF Junction may be configured to allow Open Office for MS Office document conversions
- 50Mb of disk space is required.

1.3 Configuration for Conversion to PDF

The “Convert any file to PDF” process makes use of a print driver to generate the PDF which has the following pre-requisites to successfully perform a conversion:

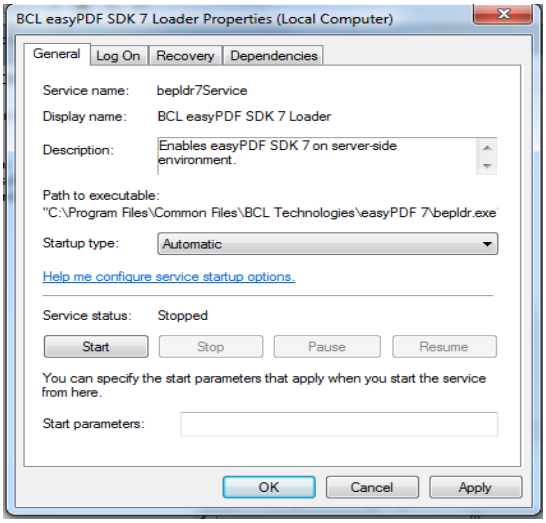
1.3.1 Native Application

An application must be installed which is capable of printing the file type to be converted to PDF. Most commonly this will be applications or office suites such as Microsoft Office or Open Office. In addition, the application must be able to be automated as the print driver process occurs as a background automated task via a service.

Section 4.3 of the reference guide details how PDF Junction can be configured to make use of specific applications for certain file types – for example how to use Open Office to assist converting Microsoft Word documents and therefore eliminate the need for Microsoft Office to be installed.

1.3.2 Service Configuration

The conversion print driver is called via the BCL easyPDF SDK 7 Loader service which can be configured via Control Panel | Administrative Tools | Services.



As the conversion is automated it is important that the Log On user for the service has permission to access the required native applications without any pop-up dialog boxes. It may be convenient to configure a specific user – eg "bcluser" and configure the service to run as this user - having first logged in as the user and launch each of the relevant applications to ensure that there are no popup dialog boxes that would remain when accessed in batch mode.

1.4 Licencing and Purchasing

Unlicensed versions are fully functional and not time-restricted, but all generated PDF files have an alert message which indicates that the document was generated with an unlicensed versions of the product.

When a license is purchased (see http://www.aquaforest.com/en/tif_pdf_buy.asp#collapseFour) a permanent license key will be issued for use with the product which should be entered using the Help | License option of the GUI. There is no need to download another version of the software.

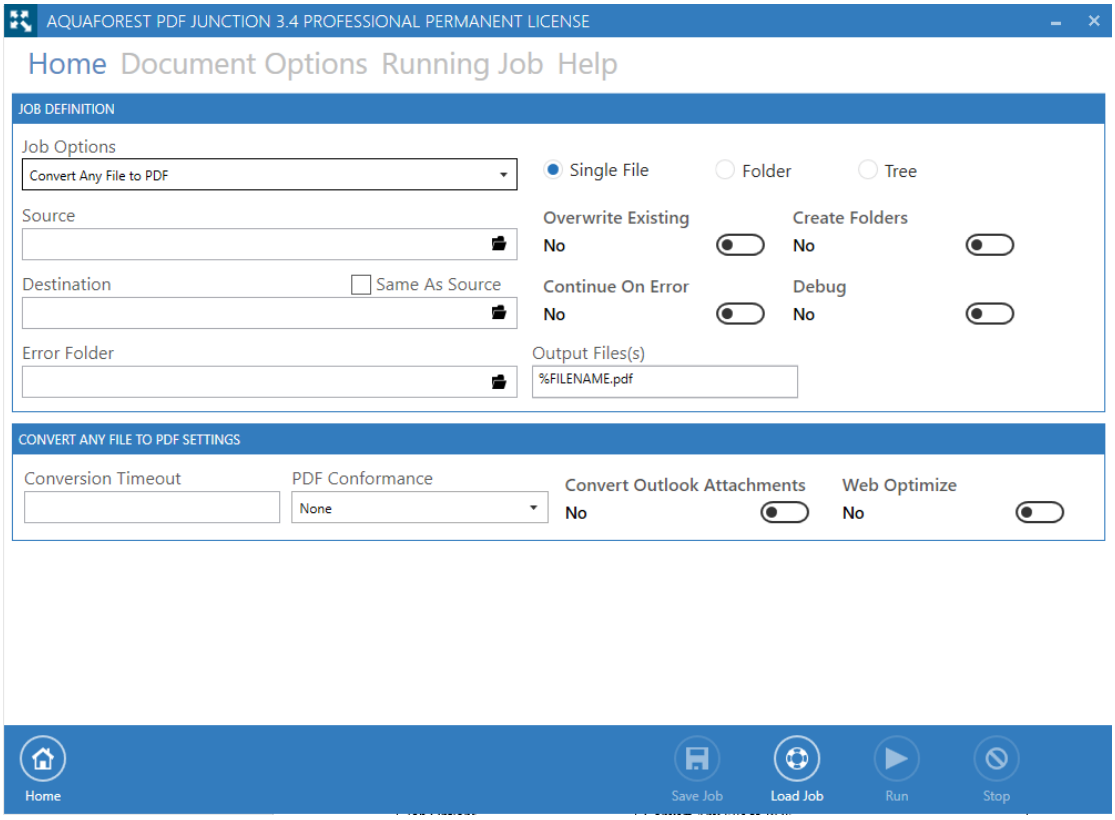
1.5 Conventions

Throughout this guide type in **`bold courier`** is used to represent operating system commands and *PDFJUNCTIONDIR* represents the directory in which PDF Junction was installed, by default C:\Aquaforest\PDF Junction.

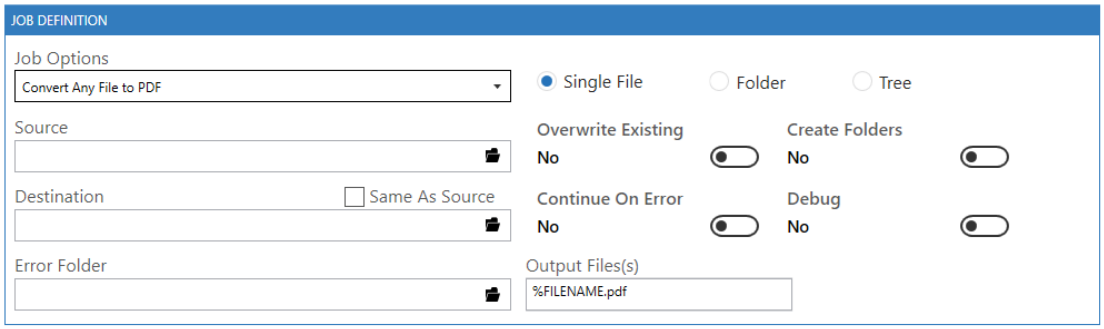
2 PDF JUNCTION WINDOWS INTERFACE

2.1 Main Window

When PDF Junction is launched, the main window is displayed as shown below. This allows the definition and running of jobs, as well as saving and loading job definitions to and from files.



2.1.1 Job Definition Section

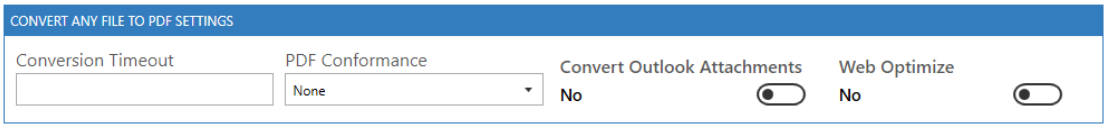


Screen Field / Button	Description
Job Options	Convert Any File to PDF Convert PDF to TIFF Split PDF Merge PDF Extract Text Web Optimize Create Property XML File Set PDF Properties Set XMP Properties PDF Barcode
Run	Runs the current job. Job output will appear in the "Running Job" Tab.
Create Folders	If this is checked, TIFF Junction will create the target directories if they do not exist.
Overwrite Existing Files	If checked, generated PDF files will overwrite existing files of the same name (but not source files).
Continue On Error	When checked, folder and tree processing jobs will continue even if there is an error processing individual files.
Explore Destination	Launches Windows Explorer using the destination as the starting point.
Save Job	This buttons allows a user to save an already configured job to an xml file.
Load Job	This button allows a user to load a job from an xml job file.
Process File	When this is selected only a single file is processed in a run.
Process Folder	When this is selected all the files in the top most directory are processed in a run.
Process Tree	When this is selected all the files in the whole folder tree are processed in a run.
Source	The source PDF file or directory.
Destination	The location where the generated PDF file(s) will be placed. This can be set to be the same as the source location. Note that PDF Junction cannot overwrite source files, a new file is always made.

<table>
<tr><td>Output File(s)</td><td>This defines the template output file name.

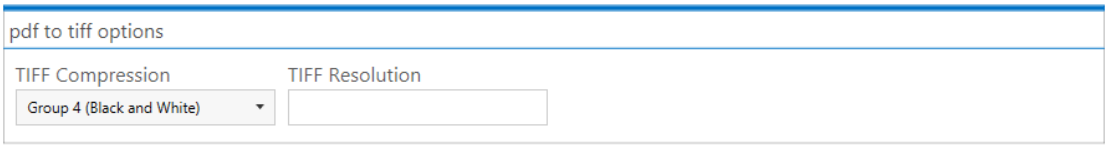
<table>
<tr><td>%BOOKMARK</td><td>Bookmark title</td></tr>
<tr><td>%PAGEn</td><td>The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used.</td></tr>
<tr><td>%FILENAME</td><td>Source file name without .pdf</td></tr>
<tr><td>%UNIQUEn</td><td>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>String</td><td>(Any string)</td></tr>
<tr><td>%TIMESTAMP</td><td>Uses Date and time as the output file name.</td></tr>
<tr><td>%GUID</td><td>Uses GUID value as the output name.</td></tr>
</table></td></tr>
</table>

2.1.2 Any File to PDF Section



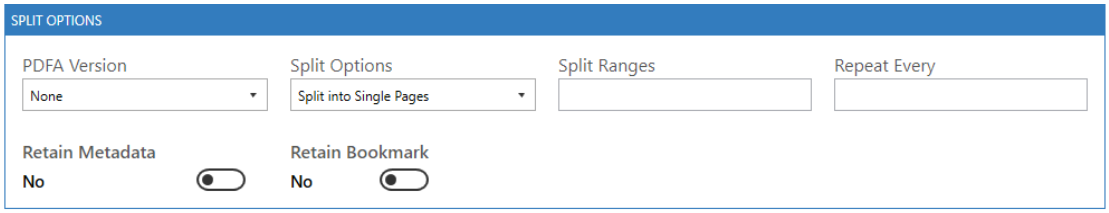
Screen Field / Button	Description
PDF/A-1b	The generated PDF file will be PDF/A-1b compliant.
PDF/X-1a	The generated PDF file will be PDF/X-1a compliant.
PDF/X-3	The generated PDF file will be PDF/X-3 compliant.
Conversion Timeout	Limits the amount of time in seconds that can be spent on a conversion. A value of zero means wait indefinitely.
Conversion Error Folder	Define the folder where documents are placed if their conversion raised an error.
Web Optimize	The generated PDF file will be Web Optimized (Linearized)
Convert Outlook Attachments	If selected (when converting .msg files) the generated PDF file will comprise the email plus each page of each attachment.

2.1.3 PDF to TIFF Section



Screen Field / Button	Description
Compression	This specifies the compression that will be used in the TIFF version of the PDF file. The choices are Group 4 (black and white only) or LZW (which also supports color). The default is Group 4.
Resolution	This specifies the resolution in DPI of the result TIFF file. The default is 200. A higher figure of 300 can give better results in some cases, at the cost of a larger file size.

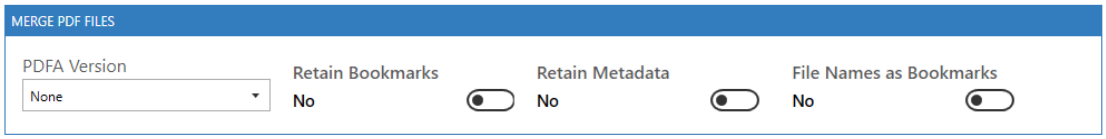
2.1.4 Split Options Section



Screen Field / Button	Description
Split Type	*One of :* Split into single pages Split by top-level bookmarks Split by page ranges (See below) Split by repeating page ranges (See below)
Page Ranges	Set of page ranges separated by commas that defines which pages from the original should be extracted. The following types of page ranges are allowed : 1 — Specifies a single page 1-3 — Specifies a range of pages 10-20even — Specifies even pages within the range 10-20odd — Specifies odd pages within the range 20-lastpage — Note that *lastpage* can be used to

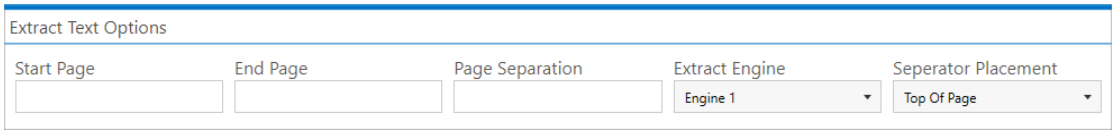
	represent the last page in the document.
Repeating Range	Apply the page range to each set of Page Ranges pages within the document. For example if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages. Hence the file is split such that the first output file contains pages 2-4 from the original document, the second contains pages 6-8 and so on.
Retain Bookmarks	If checked, generated files will include bookmarks from the original file.
Retain Metadata	If checked, generated files will include metadata (such as Author and Title) from the original file.
PDF/A Version	This option allows the user to choose a particular PDF/A conformance level as the PDF file type.

2.1.5 Merge PDF Section



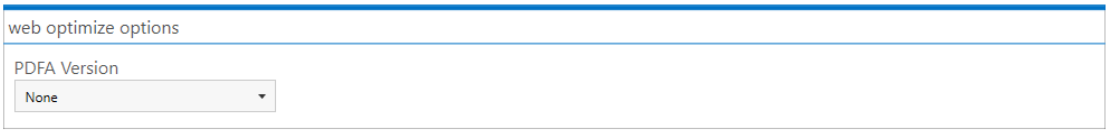
Screen Field / Button	Description
Use File Names As Bookmarks	If checked a new bookmark with the file name will be placed on the first page of each file.
Retain Bookmarks	If checked, generated files will include bookmarks from the original file.
Retain Metadata	If checked, generated files will include metadata (such as Author and Title) from the original file.
PDF/A Version	This option allows the user to choose a particular PDF/A conformance level as the PDF file type.

2.1.6 Extract Text from PDF Section



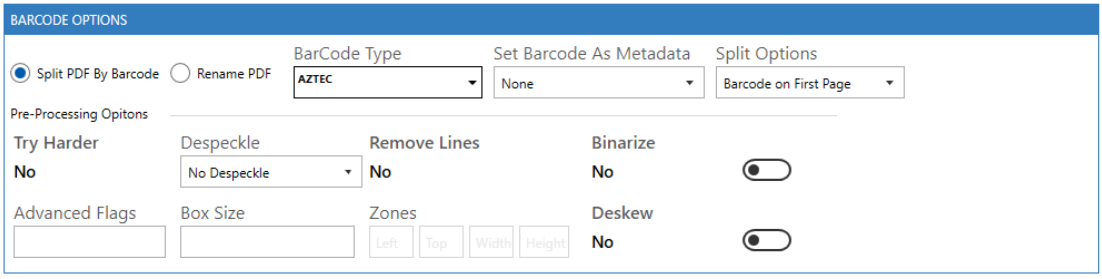
Screen Field / Button	Description
From, To	This defines the range of pages that will have text extracted. By default all pages are processed.
Page Separation	This allows the definition of an optional page separator string in the output text file, it will replace any occurrence of n with the current page number
Extract Engine	You can switch between three extract engines; they take a different approach to text extraction. • Engine 1= PDFBox with formatting • Engine 2 = BCL Engine • Engine 3 = PDFBox
Separator Placement	You can choose if you want to place the Page separator at the top or bottom of the page.

2.1.7 Web Optimize Section



Screen Field / Button	Description
PDF/A Version	This option allows the user to choose a particular PDF/A conformance level as the PDF file type.

2.1.8 PDF Barcode

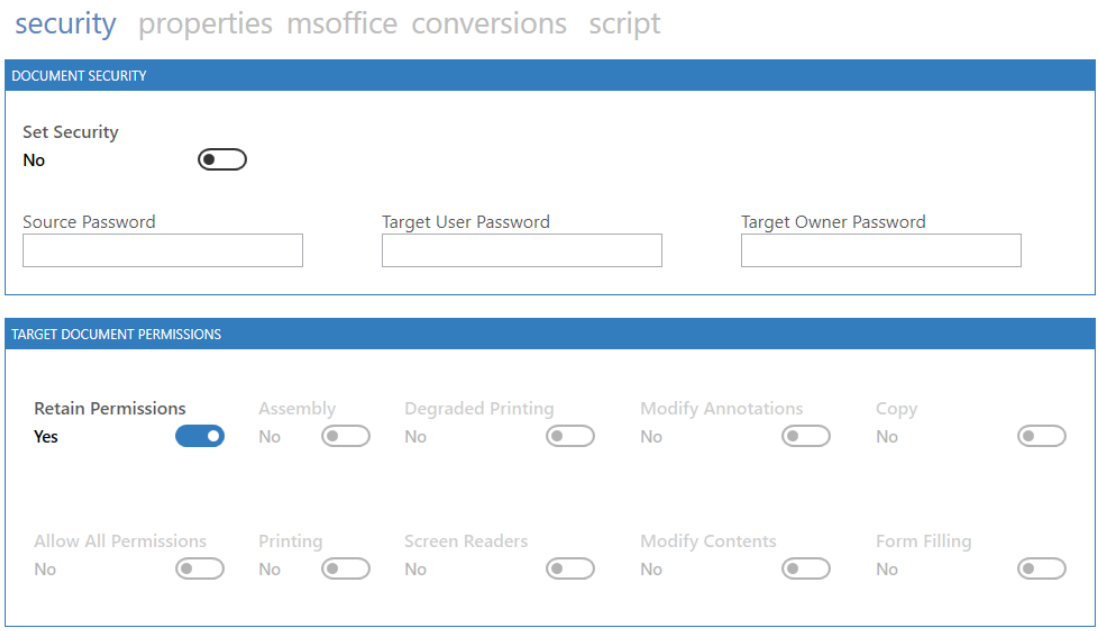


Screen Field / Button	Description
Split PDF By Barcode	Choose this option to split the PDF file by Barcode
Rename PDF	Choose this option to rename the PDF file based on

	Barcode.
Barcode Type	Barcode formats supported.
Split options	Various Options for splitting Files by Barcode
Set Barcode as Metadata	We now provide the ability to set the barcode extracted as a PDF metadata, just choose the Metadata you want.
Metadata Value	The Value of the Metadata selected above. - %VALUE% : Replaced by the barcode value found.
Try Harder	Spend more time to try to find a barcode; optimize for accuracy, not speed. Default is true.
Output File(s) With Barcode	The output file path template where the split files will be saved. - %VALUE% : Replaced by the barcode value found. - %INDEX% : Replaced by the current split index. - %FILENAME : Replaced by the file name
Output File(s) No Barcode	The renaming template to use for page ranges where no barcodes were identified. Allowed templates: - %INDEX% ; Replaced by the current split index. - %FILENAME : Replaced by the filename of the source file.
Advanced Flags	Additional PDF Junction advanced command line flags may be entered here (see section 3)
Binarize	Switch this on if your input files are coloured, it first converts it to black and white before conversion.
Remove Lines	Remove lines from images.
Zones	Only examine the region specified for barcode(s).
Box Size	This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the barcode reader. Technically, this option removes connected elements with a minimum area (in pixels and defined by this property). This option is currently only applied for bitonal images.
Deskew	Deskew (straighten) the image. The default value is false

2.2 Document options

2.2.1 Security Settings



2.2.1.1 Source Document Security

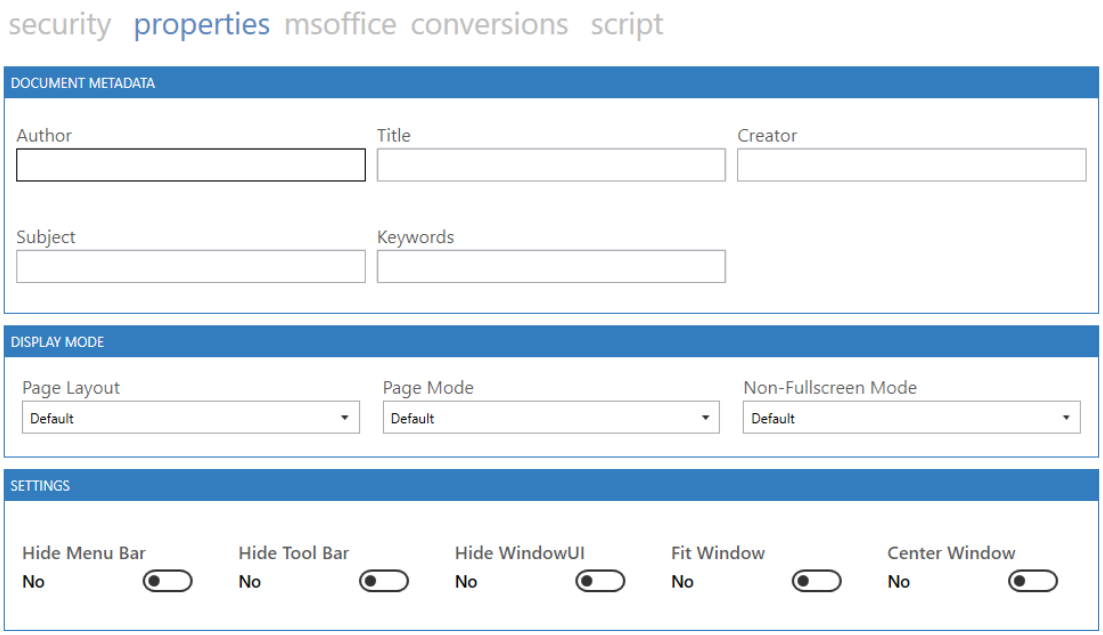
Screen Field / Button	Description
Source Password	If the source document requires a password it should be entered here.

2.2.1.2 Target Document Security

Option	Description
User Password	A password that will be required to open the document.
Owner Password	A password that will be required to change the document permissions.
Permissions	

Permission	Description
Allow ALL Permissions	All the permissions below.
Printing	Allow high-quality printing
Modify Contents	Allow assembly (see below) and other document medications
Copy	Allow text and graphic copying and extraction
Modify Annotations	Allow modification of annotations
Form Filling	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

2.2.2 Properties



This page allows setting of the generated PDF Document Properties. Custom and XMP metadata may also be set – see section 7 for further details.

The **Document Metadata** Section allows the user to change the metadata to the target PDF file.

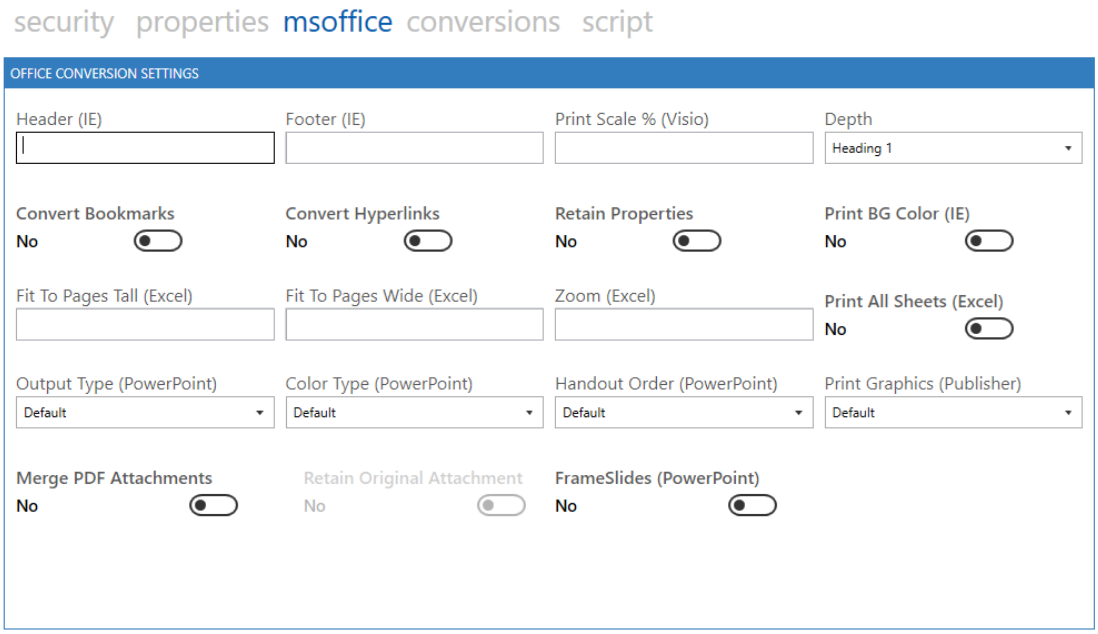
2.2.2.1 Display Modes

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

2.2.2.2 Settings

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

2.2.3 Msoffice

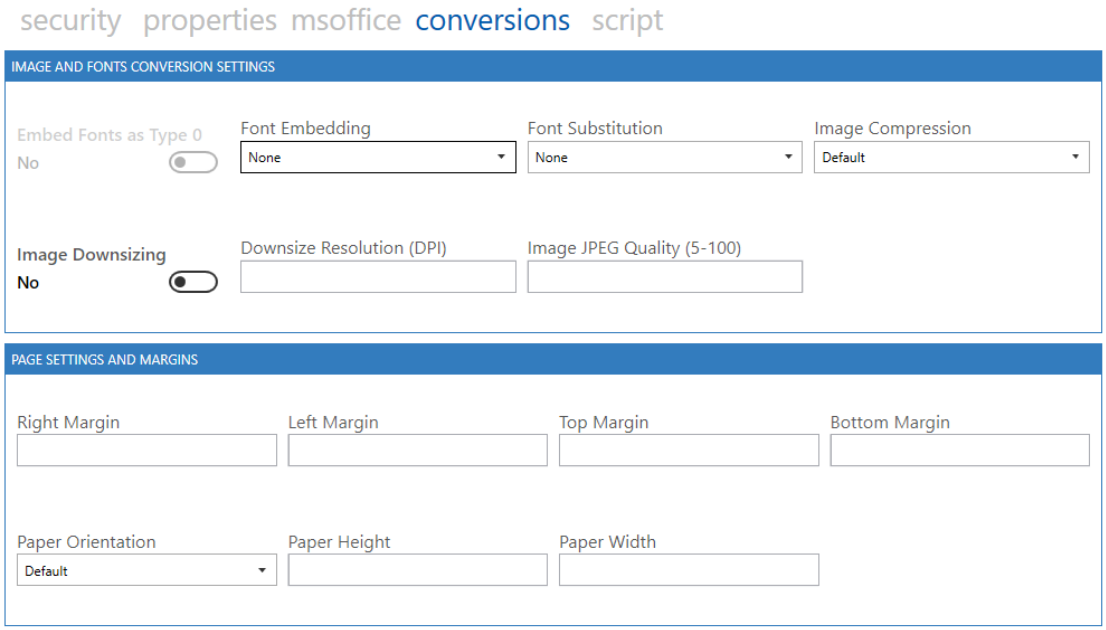


Option	Description
Convert Bookmarks	For MS Word, convert bookmarks
Depth	This property will take effect only when the ConvertBookmarks property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, empty string is functionally equivalent to: Heading 1|1|Heading 2|2|Heading 3|3|Heading 4|4|Heading 5|5|Heading 6|6|Heading 7|7|Heading 8|8|Heading 9|9 **Note:** If you use non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.

Convert Hyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
Print All Sheets	The flag that indicates whether to print all Excel worksheets or not.
Print BG Color	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
Print Scale	For Visio files, sets the print scale
Header	This property modifies Internet Explorer's header setting.
Footer	This property modifies Internet Explorer's footer setting.
Zoom	Sets printing zoom of the worksheet. The allowed value range is from 10 to 400.
FitToPagesWide	Sets number of pages wide the worksheet will be scaled to. This property is ignored if the Zoom property is set.
FitToPagesTall	Sets number of pages tall the worksheet will be scaled to. This property is ignored if the Zoom property is set.
Merge PDF Attachments	Set this flag to true if you want if you want converted pdf attachments to merged to the output pdf file. Otherwise the converted files will be merged back to the pdf.
Retain Original Attachment	Set this flag to true if you want the original attachments to be retained in the pdf file after they are merged. This flag only gets picked up if the MergeAttachments flag above is used.
Retain Properties	Set this flag if you want the MS Office properties to be transferred to the target pdf document.
Handout Order	Sets the handout order, this flag only applies to Power Point jobs. The possible values are: • Vertical First • Horizontal First
Output Type	Sets the output type, it only works with the PowerPoint files. The possible values are: • Slides • Build slides • Two slides handouts • Three slides handouts • Four slides handouts • Six slides handouts • Nine slides handouts • Notes

	• Outline
Print Graphics	Sets the graphics setting for printing. • Print Full Resolution • Print Low Resolution • Print Graphics

2.2.4 Conversions

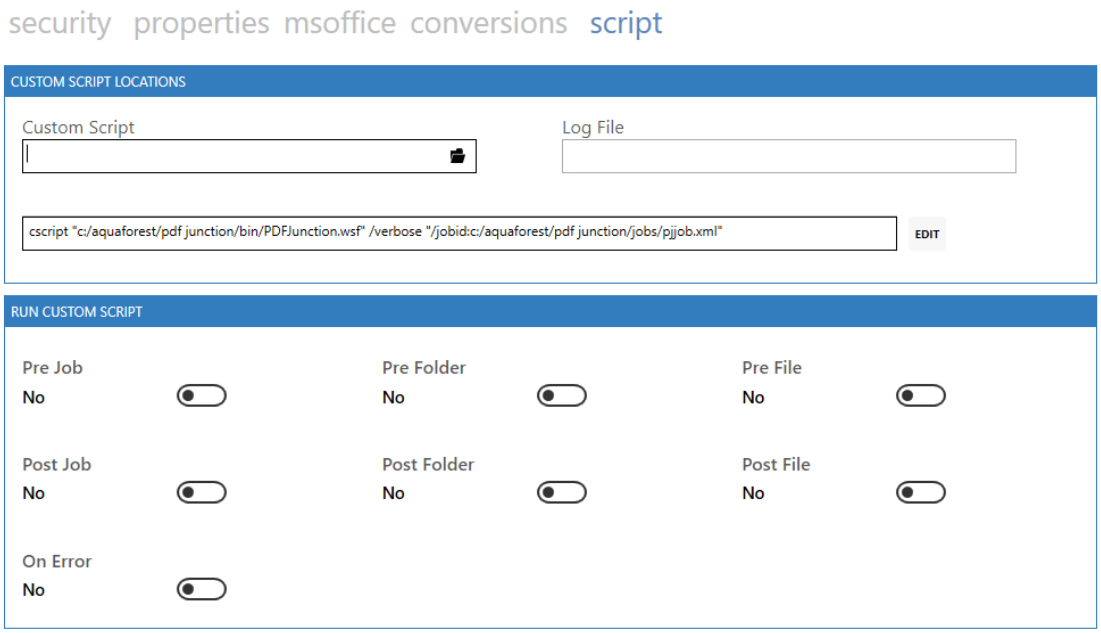


Option	Description
Font Embedding	The option PRN_FONT_EMBED_FULLSET (embedding full set of font) will cause significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
Font Substitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from Control Panel).
Embed Fonts as Type 0	This option is recommended if you have non-standard fonts like barcode font.
Top Margin	Sets top margin.
Bottom	Sets bottom margin.
Left	Sets left margin.
Right	Sets right margin.
Width	Sets a custom page width.

Height	Sets a custom page height.
Paper Orientation	Sets paper orientation e.g. PRN_PAPER_ORIENT_LANDSCAPE
Image Compression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
Image Downsizing	If this property is set to True, then the resolution of images are reduced to the DPI value specified in the ImageDownsizeResolution property.
Image JPEG Quality	The allowed value range is from 5 to 100 with 100 being the highest quality.
Downsize Resolution	If the ImageDownsizing property is set to True, then the resolution of images are reduced to the DPI value specified in this property.

2.2.5 Script

See section 5 for full details of running Custom Scripts.



2.2.5.1 Custom Script and Log Files

Option	Description
Custom Script	The path of a custom script to be used with the Job.
Log File	The name of a log file (which will be placed in *PDFJUNCTIONDIR*\logs). The log file will contain the same output as is written to the output panel.

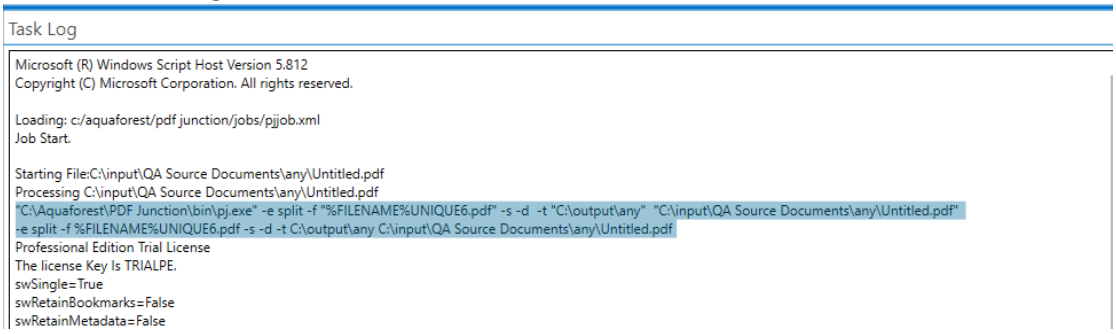
2.2.5.2 Run Custom Script

Option	Description
Pre Job	If checked, the custom script will be called at the start of the job
Post Job	If checked, the custom script will be called at the end of the job
Pre Folder	If checked, the custom script will be called at the start of processing each folder
Post Folder	If checked, the custom script will be called at the end of processing each folder
Pre File	If checked, the custom script will be called at the start of processing each file
Post File	If checked, the custom script will be called at the end of processing each file
On Error	If checked, the custom script will be called when a processing error occurs.

3 USING PDF JUNCTION FROM THE COMMAND LINE

3.1 Generating Command line

The easiest way to generate PDF Junction command line arguments is to create the Job on the GUI and run it in debug mode, the command line will be displayed in the task log as shown in the image below.



Copy the highlighted line and run it in the command prompt.

3.2 Command Line Syntax

All operations supported by PDF Junction can be accessed via the command line as covered in this section and sections 4, 5 and 6.

Usage:

cscript pdfjunction.wsf [/jobid:value] /operation:value [/sourcetype:value] [/target:value] [/joboptions:value] [/metadata:value] [/security:value] [/docoptions:value] sourcefiles1 [sourcefiles2...] [/version[+|-]] [/verbose[+|-]] [/debug[+|-]] [/onerrorcontinue[+|-]] [/silent[+|-]]

Parameter	Notes
/jobid:value	If used, specifies the full path of an XML Job Ticket file saved from the GUI or generated by some other means. Job files in the PDF Junction jobs directory may be referred to by a shorthand method of *jobname* where the Job Ticket file is *jobname*.xml.
/operation:value	Must be split, merge, set or get
/sourcetype:value	Source type (file, folder, tree). The default value is file.
/target:value	Target directory for result files. If not specified, the source directory will be used as the target.
/joboptions:value	Set of options for the specified operation. Described in detail in section 3.3.
/metadata:value	Metadata settings for result PDF files. By default metadata fields are

	left blank. Described in detail in section 3.4.
/security:value	Security settings for result PDF files. By default files are not secured. Described in detail in section 3.5.
/docoptions:value	Document option settings for result PDF files. Described in detail in section 3.6.
Sourcefiles	Source PDF file, folder or tree. May be multiple files for merge operations.
/version	Provide version and license information
/verbose	Verbose progress messages
/debug	Debug progress messages
/silent	No progress messages
/onerrorcontinue	Allow job to continue after errors

Notes

1. cscript.exe is required to run the command. This is usually included in the standard PATH, but if not you will need to be specified explicitly e.g.

```
C:\WINDOWS\SYSTEM32\cscript.exe ….
```

2. Options that contain spaces will need to be in quotes. For example
"/target:c:\output files"

3.3 Example Commands

3.3.1 Example jobid Commands

Run the job specified in the job file C:\jobs\job001.xml:

```
cscript pdfjunction.wsf /jobid:C:\Jobs\job001.xml
```

Run the job specified in the job file C:\Program Files\PDF Junction\jobs\job001.xml:

```
cscript pdfjunction.wsf /jobid:job001
```

3.3.2 Example Split Commands

Split a PDF file, c:\demo\in\demo.pdf into single page PDF files placed in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:split /joboptions:-s
/target:c:\demo\out c:\demo\in\demo.pdf
```

Split a PDF file, c:\demo\in\\demo.pdf into single pages, each named demo_split999.pdf where demo is the page number. The files will be placed in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:split "/joboptions:-s -
f%F%_split%U3%" /target:c:\demo\out c:\demo\in\demo.pdf
```

Split a PDF file, c:\demo\in\demo.pdf by bookmarks. The files will be placed in the directory c:\demo\out and named according to the bookmark.

```
cscript pdfjunction.wsf /operation:split "/joboptions:-k -b -f
%B%" /target:c:\demo\out c:\demo\in\demo.pdf
```

Split all PDF files in c:\demo\in into single page PDF files placed in the directory c:\demo\out\name where name is the name of the original pdf file, creating the directories where they don't exist.

```
cscript pdfjunction.wsf /operation:split /sourcetype:folder
"/joboptions:-c -s -f %F%/%U%" /target:c:\demo\out c:\demo\in
```

Split all PDF files in c:\demo\in and each subdirectory into single page PDF files placed in the directory c:\demo\out\sub\name where sub is the subdirectory name, and name is the name of the original pdf file, creating the directories where they don't exist.

```
cscript pdfjunction.wsf /operation:split /sourcetype:tree
"/joboptions:-c -s -f %F%/%U%" /target:c:\demo\out c:\demo\in
```

3.3.3 Example Merge Commands

Merge PDF files demo1.pdf, demo2.pdf, demo3.pdf in c:\demo\in\ into a single PDF file named demomerge.pdf in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:merge /target:c:\demo\out "/joboptions:-f demomerge c:\demo\in\demo1.pdf*c:\demo\in\demo2.pdf*c:\demo\in\demo3.pdf
```

Merge all the PDF files in c:\demo\in\ into a single PDF file named demomerge.pdf in the directory c:\demo\out, overwriting the file if it already exists.

```
cscript pdfjunction.wsf /operation:merge /sourcetype:folder /target:c:\demo\out "/joboptions:-o -f demomerge" c:\demo\in
```

Merge all the PDF files in c:\demo\in\ and each subdirectory into a single PDF file per source directory. The result file will be named dir.pdf where dir is the source directory name, and placed in the corresponding directory in c:\demo\out\

```
cscript pdfjunction.wsf /operation:merge /sourcetype:tree /target:c:\demo\out "/joboptions:-c -o -f %D%" c:\demo\in
```

3.3.4 Example Set Commands

Note that the set parameters (/metadata /...) can be used with the split and merge commands to set characteristics of the result files.

Creates a copy of c:\demo\in\demo3.pdf with the metadata settings shown, in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:set /target:c:\demo\out c:\demo\in\demo3.pdf "/metadata:Author=Tony Bush;Title=A Special Document"
```

Creates a copy of each PDF file in c:\demo\in\ with a user password of "secret", with printing and modification of annotations allowed. The result PDF files are placed in in the directory c:\demo\out.

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o /target:c:\demo\out c:\demo\in "/security:ownerpassword=Secret; permissions=allowprinting,allowmodifyannotations"
```

Creates a copy of each PDF file in c:\demo\in\ with viewer preferences set to hide toolbar and to open the document showing bookmarks. The result PDF files are placed in in the directory c:\demo\out.

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o /target:c:\demo\out c:\demo\in "/docoptions:hidetoolbar,pagemodeuseoutlines"
```

3.3.5 Example Get Command

Generates an XML file with metadata from c:\demo\in\demo3.pdf with the metadata settings shown, in the directory c:\demo\out

```
cscript pdfjunction.wsf /operation:get /target:c:\demo\out
c:\demo\in\demo3.pdf "/joboptions:-f %F%_metadata"
```

3.4 /joboption specifications

3.4.1 Job Options for Split Operations

<table>
<tr><th>Parameter</th><th>Notes</th></tr>
<tr><td>-b</td><td>Include bookmarks</td></tr>
<tr><td>-p</td><td>Password for opening document</td></tr>
<tr><td>-pdfa</td><td>The output documents will conform to the PDF/A standards.</td></tr>
<tr><td>-version</td><td>The PDF/A version
<table>
<tr><td>1</td><td>PDF/A-1b</td></tr>
<tr><td>2</td><td>PDF/A-2b</td></tr>
<tr><td>3</td><td>PDF/A-3b</td></tr>
</table></td></tr>
<tr><td>-m</td><td>Retain metadata</td></tr>
<tr><td>-c</td><td>Create result directories if required</td></tr>
<tr><td>-o</td><td>Overwrite result files if they already exist</td></tr>
<tr><td>-r range{,range}</td><td>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>
<tr><td>1</td><td>Specifies a single page</td></tr>
<tr><td>1-3</td><td>Specifies a range of pages</td></tr>
<tr><td>10-20even</td><td>Specifies even pages within the range</td></tr>
<tr><td>10-20odd</td><td>Specifies odd pages within the range</td></tr>
<tr><td>20-lastpage</td><td>Note that lastpage can be used to represent the last page in the document.</td></tr>
</table></td></tr>
<tr><td>-g grouppages</td><td>Apply the page range to each set of grouppages pages within the document. For example if –r 2-4 and –e 5 is specified, 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.</td></tr>
<tr><td>-k</td><td>Split by top-level bookmark</td></tr>
<tr><td>-s</td><td>Split by single pages (The Default)</td></tr>
<tr><td>-f string</td><td>Output file name specifier. If –f is not specifed, the default parmeter setting is %FILENAME%UNIQUE6% which would give (for example) filename.pdf split into filename000001.pdf filename000002.pdf etc.
<table>
<tr><td>%BOOKMARK</td><td>Bookmark title</td></tr>
<tr><td>%PAGEn</td><td>The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used</td></tr>
<tr><td>%FILENAME</td><td>Source file name without .pdf</td></tr>
<tr><td>%UNIQUEn</td><td>Unique number starting at 1. If n is supplied, then zero padding to n digits is used. If n is not</td></tr>
</table></td></tr>
</table>

		supplied or is zero, then zero padding is not used
	%DIRNAME	Source directory name of the directory currently being processed.
	String	(Any string)
	The .pdf extension is automatically added if not specified.	

3.4.2 Job Options for Merge Operations

<table>
<tr><th>Parameter</th><th>Notes</th></tr>
<tr><td>-b</td><td>Include bookmarks</td></tr>
<tr><td>-p</td><td>Password for document reading</td></tr>
<tr><td>-m</td><td>Retain metadata (taken from the first file in the merge list).</td></tr>
<tr><td>-c</td><td>Create result directories if required</td></tr>
<tr><td>-o</td><td>Overwrite result files if they already exist</td></tr>
<tr><td>-f string</td><td>Output file name specifier. If –f is not specified, the merged file will be called mergeoutput.pdf.

<table><tr><td>%DIRNAME</td><td>Source directory name of the directory currently being processed.</td></tr><tr><td>String</td><td>(Any string)</td></tr></table>
The .pdf extension is automatically added if not specified.</td></tr>
<tr><td>-pdfa</td><td>The output documents will conform to the PDF/A standards.</td></tr>
<tr><td>-version</td><td>The PDF/A version
<table><tr><td>1</td><td>PDF/A-1b</td></tr><tr><td>2</td><td>PDF/A-2b</td></tr><tr><td>3</td><td>PDF/A-3b</td></tr></table></td></tr>
</table>

3.4.3 Job Options for Set Operations

<table>
<tr><th>Parameter</th><th>Notes</th></tr>
<tr><td>-c</td><td>Create result directories if required</td></tr>
<tr><td>-o</td><td>Overwrite result files if they already exist</td></tr>
<tr><td>-f string</td><td>Output file name specifier.

<table><tr><td>%FILENAME</td><td>Source file name without .pdf</td></tr><tr><td>%DIRNAME</td><td>Source directory name of the directory currently being processed.</td></tr><tr><td>String</td><td>(Any string)</td></tr></table>
The .pdf extension is automatically added if not specified.</td></tr>
<tr><td>-p</td><td>Password for document reading</td></tr>
</table>

3.4.4 Job Options for Get Operations

<table>
<tr><th>Parameter</th><th>Notes</th></tr>
<tr><td>-c</td><td>Create result directories if required</td></tr>
<tr><td>-o</td><td>Overwrite result files if they already exist</td></tr>
<tr><td>-f string</td><td>Output file name specifier.

<table><tr><td>%BOOKMARK</td><td>Bookmark title</td></tr><tr><td>%PAGEn</td><td>The first page number from source file, zero padded to n digits. If n is not supplied or is zero, then zero padding is not used.</td></tr><tr><td>%FILENAME</td><td>Source file name without .pdf</td></tr><tr><td>%UNIQUEn</td><td>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>String</td><td>(Any string)</td></tr></table>
The .xml extension is automatically added</td></tr>
<tr><td>-p</td><td>Password for document reading</td></tr>
</table>

3.5 /metadata specification

This may be used to specify metadata properties as in the following example:

```
cscript pdfjunction.wsf /operation:set /target:c:\demo\out c:\demo\in\demo3.pdf "/metadata:Author=Tony Bush;Title=A Special Document"
```

The attributes that may be set are:

- Author
- Title
- Subject
- Keywords
- Creator

3.6 /security specification

This may be used to specify security properties as in the following example:

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o c:\demo\in
/target:c:\demo\out
"/security:userpassword=Secret;ownerpassword=secret
;permissions=allowprinting,allowmodifyannotations"
```

<table>
<tr><th>Parameter</th><th>Description</th></tr>
<tr><td>Userpassword</td><td>A password that will be required to open the document.</td></tr>
<tr><td>Ownerpassword</td><td>A password that will be required to change the document permissions.</td></tr>
<tr><td>Permissions</td><td>A comma separated list of document permissions
<table>
<tr><th>Permission</th><th>Description</th></tr>
<tr><td>allowprinting</td><td>Allow high-quality printing</td></tr>
<tr><td>allowmodifycontents</td><td>Allow assembly (see below) and other document medications</td></tr>
<tr><td>allowcopy</td><td>Allow text and graphic copying and extraction</td></tr>
<tr><td>allowmodifyannotations</td><td>Allow modification of annotations</td></tr>
<tr><td>allowfillin</td><td>Allow filling of form fields</td></tr>
<tr><td>allowscreenreaders</td><td>Allow extraction of text and 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>
</td></tr>
<tr><td>strength</td><td>Unless this is set to some other value (such as 40), the default of 128 is used.</td></tr>
</table>

3.7 /docoptions specification

This may be used to specify document viewing preferences as in the following example:

```
cscript pdfjunction.wsf /operation:set /sourcetype:folder /joboptions:-o
/target:c:\demo\out c:\demo\in "/docoptions:hidetoolbar,pagemodeuseoutlines"
```

Propery	Description
pagelayoutsinglepage	Single Page (SinglePage)
pagelayoutonecolumn	Continuous(OneColumn)
pagelayouttwocolumnleft	ContinuousFacing [odd pages left] (TwoColumnLeft)
pagelayouttwocolumnright	ContinuousFacing [odd pages right] (TwoColumnRight)
pagemodeusenone	No Thumbnails or Bookmarks Visible (UseNone)
pagemodeuseoutlines	Bookmarks Visible (Use Outlines)
pagemodeusethumbs	Thumbnails Visible (UseThumbs)
pagemodefullscreen	Full Screen Mode (FullScreen)
hidetoolbar	The viewer's tool bar will be hidden
hidemenubar	The viewer's menu bar will be hidden
hidewindowui	The viewer's UI elements (scrollbars etc) will be hidden
fitwindow	The viewer will resize the document's window to fit the size of the first displayed page.
centerwindow	The document window will be positioned in the center of the screen.
nonfullscreenpagemodeusenone	On exiting full screen mode, neither outline nor thumbnail images are shown.
nonfullscreenpagemodeusenoutline	On exiting full screen mode outlines are visible.
nonfullscreenpagemodeusethumbs	On exiting full screen mode thumbnails are visible.

4 Conversion to PDF –Command Lines Interface (Professional Edition)

4.1 Command Line Interface

A command line interface to PDF generation is provided topdf.exe in the PDF Junction bin directory, this exe can also be used to convert a pdf file to PDF/A.

topdf.exe /source=*sourcefiledefinition* {/*option*=*value*}

Options can be any of the options listed in section 4.1 in addition to the following:

Parameter	Notes
/Source	Source file definition. Can include wildcards.
/TargetFolder	Target folder for converted file.
/Output	This is the name that the output file will have, there are some variables that can help you have more meaning full output file names. • **%FILENAME**: Replaces any occurrence of the variable with the original filename without extensions. • **%DIRNAME**: Replaces the variable with the directory name during a merge operation. • **%GUID**: Replaces the variable with a guid, this is used for unique filenames. • **%TIMESTAMP**: Replaces the variable with the current time, this is used for unique filenames.
/CreateFolders	Create output and error folders where required.
/fileconversiontimeout	Limits the amount of time in seconds that can be spent on a conversion. A value of zero means wait indefinitely.
/Zoom	Sets printing zoom of the worksheet. The allowed value range is from 10 to 400.
/FitToPagesWide	Sets number of pages wide the worksheet will be scaled to. This property is ignored if the Zoom property is set.
/FitToPagesTall	Sets number of pages tall the worksheet will be scaled to. This property is ignored if the Zoom property is set.
/MergeAttachments	Set this flag to true if you want if you want converted pdf attachments to merged to the output pdf file. Otherwise the converted files will be merged back to the pdf.
/RetainPDFAttachment	Set this flag to true if you want the original attachments to be

	retained in the pdf file after they are merged. This flag only gets picked up if the **MergeAttachments** flag above is used.
/IncludeDocProps	Set this flag if you want the MS Office properties to be transferred to the target pdf document.
/HandoutOrder	Sets the handout order, this flag only applies to Power Point jobs. The possible values are: PRN_PP_HANDOUT_VERTICAL_FIRST, PRN_PP_HANDOUT_HORIZONTAL_FIRST
/PPOutPutType	Sets the output type, it only works with the PowerPoint files. The possible values are: Slides = PRN_PP_OUTTYPE_SLIDES Build slides = PRN_PP_OUTTYPE_BUILDSLIDES Two slides handouts = PRN_PP_OUTTYPE_TWOSLIDEHANDOUTS Three slides handouts = PRN_PP_OUTTYPE_THREESLIDEHANDOUTS Four slides handouts = PRN_PP_OUTTYPE_FOURSLIDEHANDOUTS Six slides handouts = PRN_PP_OUTTYPE_SIXSLIDEHANDOUTS Nine slides handouts = PRN_PP_OUTTYPE_NINESLIDEHANDOUTS Notes= PRN_PP_OUTTYPE_NOTESPAGES Outline = PRN_PP_OUTTYPE_OUTLINE
/PrintGraphics	Sets the graphics setting for printing. Print Full Resolution = PRN_PB_PRINT_GRAPHICS_HIGHRES Print Low Resolution = PRN_PB_PRINT_GRAPHICS_LOWRES Print Graphics = PRN_PB_PRINT_GRAPHICS_NONE
/convertbookmarks	For MS Word, convert bookmarks
/bookmarkdepth	This property will take effect only when the ConvertBookmarks property is set to True. Numbers defining bookmark levels must be equal to or larger than one. Word style names must not repeat in the string. The string must not start or end with the delimiter. When this property is empty, the default style mapping (Heading one through nine will be mapped to level one through nine) will be used. Therefore, empty string is functionally equivalent to: Heading 1|1|Heading 2|2|Heading 3|3|Heading 4|4|Heading 5|5|Heading 6|6|Heading 7|7|Heading 8|8|Heading 9|9 **Note:** If you use non-English version of Microsoft Word, then you may need to replace the word "Heading" with its localized version.
/converthyperlinks	Sets the flag to indicate whether to convert Word hyperlinks to PDF hyperlinks.
/printallsheets	The flag that indicates whether to print all Excel worksheets or not.
/printbgcolor	For files printed via IE Sets the flag that indicates whether to print background color or not when printing.
/Printscale	For Visio files, sets the print scale
/ieheader	This property modifies Internet Explorer's header setting.

/iefooter	This property modifies Internet Explorer's footer setting.
/imagecompression	If you want a lossless image compression, use PRN_IMAGE_COMPRESS_ZIP (ZIP compression).
/imagedownsizing	If this property is set to True, then the resolution of images are reduced to the DPI value specified in the ImageDownsizeResolution property.
/imagejpegquality	The allowed value range is from 5 to 100 with 100 being the highest quality.
/imagedownsizeresolution	If the ImageDownsizing property is set to True, then the resolution of images are reduced to the DPI value specified in this property.
/fontembedding	The option PRN_FONT_EMBED_FULLSET (embedding fullset of font) will cause significant increase in PDF file size, especially for CJK font, and therefore not recommended. If you need to embed font, PRN_FONT_EMBED_SUBSET (embed subset of font) will be a better choice.
/fontsubstitution	For the PRN_FONT_SUBST_TABLE (use font substitution table) option, you need to configure the substitution table. The table is stored under the "Device Setting" section of the printer driver properties (can be accessed from Control Panel).
/fontembedastype0	This option is recommended if you have non-standard fonts like barcode font.
/margintop	Margin in Inches.
/marginbottom	Margin in Inches.
/marginleft	Margin in Inches.
/marginright	Margin in Inches.
/pagewidth	In Inches.
/pageheight	In Inches.
/paperorientation	PORTRAIT or LANDSCAPE
/papersize	Word does not use the paper size setting from printer driver. If you need to change the paper size, use this property instead. If you need to use a custom paper size, set this property to PRN_MSO_PAPER_CUSTOM, and set the PageWidth and PageHeight properties.
/ PrintColorType	With this property, you can set the PowerPoint to print either with • Color : PRN_PP_PRINT_COLOR • Grey scale: PRN_PP_PRINT_GRAYSCALE • Black and White : PRN_PP_PRINT_BLACK_AND_WHITE
/queuewaittimeout	Since PowerPoint cannot perform multiple conversions at the same time, each conversion request will be put into queue, and they are converted one

	at a time. A waiting print job will timeout if it waits for more than the value specified in QueueWaitTimeout property. Timeout value is in milliseconds. Set to zero if you want to wait indefinitely.
/disablescriptdebugger	Enables/Disables Internet Explorer's script debugger. It is highly recommended that Internet Explorer's script debugger is turned off during printing so that PDF conversion process is not interrupted.
/DisplayErrorDialogOnEveryError	Enables/Disables Internet Explorer's error dialog displayed on every error. It is highly recommended that this feature is turned off from Internet Explorer so that PDF conversion process is not interrupted.
/pdfa	The output file will be PDF/A-1b compliant.
/pdfx1a	The output file will be PDF/X-1a compliant.
/pdfx3	The output file will be PDF/X-3 compliant.
/ConvertAttachments	For msg files, email attachments are processed so that the resulting PDF is composed of the email plus each page of each attachment.
/IncludeDocumentMarkups	The "Include Document Markups" option allows Word comments and markup to be included in the generated PDF.

Example Commands:

Convert the Excel spreadsheet sales.xls to PDF. The result file will be in the same directory as the source file (as targetfolder was not specified). All the sheets in the spreadsheet will be printed, using landscape orientation.

```
topdf.exe /source=c:\demo\in\demo.xls
/paperorientation=PRN_PAPER_ORIENT_LANSCAPE
/printallsheets=true
```

Convert all the files in c:\infolder to PDF and place the output files in c:\outfolder. The output file name will be converted_FILENAME.pdf where FILENAME was the original file name prefix. e.g. sales.xls is converted to converted_sales.xls.

```
topdf.exe /source=c:\demo\in\*.* /targetfolder=c:\demo\out
/output=converted_%FILENAME.pdf
```

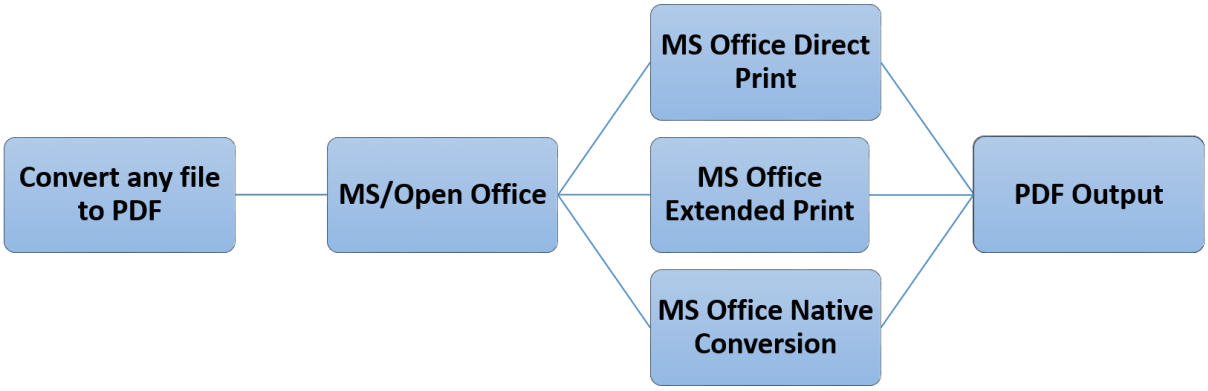
4.2 Configuration File

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

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

4.2.1 Overview of the Different Conversions Methods

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



4.2.1.1 MS Office Native Conversion

This step uses the “Save as PDF” option in MS Office to save the PDF file directly. This requires Microsoft Office 2007 or newer. Microsoft Office 2007 requires the free "Save as PDF or XPS" add-in for Office 2007 to be installed. We recommend you use the 2010 and new versions.
If you are not interested in changing properties like Image Compression, Image Downsizing, Font embedding and you are ok with the default PDF/A output from your version of office, we recommend using this approach.

4.2.1.2 MS Office Direct Print

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

4.2.1.3 MS Office Extended Print

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

4.2.2 Comparison of the features provided by the different methods

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

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

4.2.3 In-depth Look at the Configuration file

The configuration file can be found in the following locations:

- Autobahn DX: “Autobahn DX Installation directory \pj\bin\topdf.exe.config”
- PDF Junction: “PDF Junction Installation directory \bin\topdf.exe.config”

Below are the contents of the configuration file, to use a method for a file type, just write the extension in the value property of the conversion Method as shown below.
Note: Do not write an extension in more than one method

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

4.2.3.1 Mappings Between Configuration File and the Conversion Methods

The table below maps the configuration file with the Conversion Methods discussed earlier. Note: The Configuration options ending with "Ex" work with the "NativeOfficePDF" config option to select a suitable conversion method.

Configuration	Conversion Method	File Types	Comment
AutoExtension	Direct Print	All file types	This option works with any file extension that has an application that is compatible with the BCL printer
AutoExtensionEx	Extended Print	All file types	
AutoExtensionNative	Native Conversion	MS Office files	
OpenOfficeExtension	Direct Print	Open Office files	This option works with all files that can be opened in open office.
AutoExtensionOpenNative	Native Conversion	Open Office files	
WordExtension	Direct Print	.doc, .docx, .rtf… (All files that can be opened in MS Word)	
WordExtensionEx NativeOfficePDF=true	Native Conversion		
WordExtensionEx NativeOfficePDF=false	Extended Print		
ExcelExtension	Direct Print	.xls, .csv, .xlsx… (All files that open in excel)	
ExcelExtensionEx NativeOfficePDF=true	Native Conversion		
ExcelExtensionEx NativeOfficePDF=false	Extended Print		
PowerPointExtension	Direct Print	.ppt, .pptx (All power point files)	
PowerPointExtensionEx NativeOfficePDF=true	Native Conversion		
PowerPointExtensionEx NativeOfficePDF=false	Extended Print		
VisioExtension	Direct Print	.vsd	
VisioExtensionEx NativeOfficePDF=true	Native Conversion		
VisioExtensionEx NativeOfficePDF=false	Extended Print		
IEExtension	Direct Print	.html, .xml, .mht All IE files.	
IEExtendedExtension	Extended Print		
HTMLExtension	Direct Print	.htm, .html	
PublisherExtension	Direct Print	.pub, .puz	
PublisherExtensionEx NativeOfficePDF=true	Native Conversion		
PublisherExtensionEx NativeOfficePDF=false	Extended Print		
OutlookExtension	Direct Print	.msg files	
PDFExtension		PDF files	We usually convert PDF attachments if you select the option
ExcludedExtensions		others	Skips all the extensions present here.

5 PDF TO TIFF CONVERSION COMMAND LINE INTERFACE

5.1 Command Line Interface

A command line interface to PDF to TIFF conversion is provided pdf2tiff.exe in the PDF Junction bin directory

pdf2tiff.exe /source=*sourcefiledefinition* {/*option*=*value*}

Command Line / Config File Name	**Description**
/Source	Source file defintition.
/Target	Output folder. If not specified the result file will be placed in the same folder as the source file.
/Output	Output file name. If not specified the output file name will be the same as the source file with a .TIF extension instead of a .PDF one.
/TIFFCompression	GROUP4 or LZW. If not specified the default of GROUP4 will be used.
/TIFFResolution	The DPI of the output file. If not specified the default of 200 DPI will be used.

6 PDF TEXT EXTRACTION COMMAND LINE INTERFACE

6.1 Command Line Interface

A command line interface to PDF text extraction is provided pdf2txt.exe in the PDF Junction bin directory

pdf2txt.exe /source=*sourcefiledefinition* {/*option*=*value*}

Command Line / Config File Name	**Description**
/Source	Source file definition.
/Target	Output folder. If not specified the result file will be placed in the same folder as the source file.
/Output	Output file name. If not specified the output file name will be the same as the source file with a .txt extension instead of a .pdf one.
/PageFrom	First page to be processed. If not specified processing will start at page 1.
/PageTo	Last page to be processed. If not specified processing will end at the last page.
/PageSep	Page separator string which will be inserted in the text file between pages.

7 WEB OPTIMIZATION (LINEARIZATION) CLI

7.1 Command Line Interface

A command line interface to web optimization provided pdfo.exe in the PDF Junction bin directory

pdfo.exe /source=*sourcefiledefinition* {/*option*=*value*}

Command Line / Config File Name	**Description**
/Source	Source file definition.
/Target	Output folder. If not specified the result file will be placed in the same folder as the source file.
/Output	Output file name.
/pdfa	The output documents will conform to the PDF/A standards.
/version	The PDF/A version 1 PDF/A-1b 2 PDF/A-2b 3 PDF/A-3b

8 SET XMP PROPERTIES (PROFESSIONAL EDITION ONLY)

PDF Junction allows the setting of XMP Metadata and custom DocInfo metadata. XMP (Extensible Metadata Platform) facilitates the embedding of metadata in PDF files using a subset of RDF. For further XMP background information see http://www.adobe.com/products/xmp/.

Note: Any existing XMP metadata in the source file will not be copied into the new file, so the directive XML file needs to include all required XMP metadata. However, existing DocInfo metadata can be automatically mapped to XMP – see below for further details.

8.1 Setting Metadata

To set metadata, an XML directive file is required. By default the file is expected to be named *filename.xml* where *filename.pdf* is the name of the source PDF file.

Directive files should follow the model shown below. A sample PDF and XML directive file can be found in the PDF Junction xmp folder (usually C:\Program Files\Aquaforest\PDF Junction 2.5\xmp.

```
<?xml version='1.0' encoding='utf-8'?>
<rdf:RDF    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#">
<setxmp type='MAPINFO'>
  <custom name='RiskID' value='12345'></custom>
  <custom name='Department' value='AS12'></custom>
  <rdf:Description                                    rdf:about=''
xmlns:pdfaid='http://www.aiim.org/pdfa/ns/id/'>
    <pdfaid:part>1</pdfaid:part>
    <pdfaid:conformance>B</pdfaid:conformance>
  </rdf:Description>
  <rdf:Description                                        about=''
xmlns:stEvt='http://ns.adobe.com/xap/1.0/sType/ResourceEvent#'

xmlns:xmpMM='http://ns.adobe.com/xap/1.0/mm/'>
    <xmpMM:DocumentID>adobe:docid:indd:9705C6F1-8db6-11d5-
a7dd-cdf3d5439594</xmpMM:DocumentID>
    <xmpMM:History>
      <rdf:Seq>
        <rdf:li rdf:parseType='Resource'>
          <stEvt:action>rendition created</stEvt:action>
          <stEvt:parameters>some                          additonal
info</stEvt:parameters>
          <stEvt:softwareAgent>Adobe    Photoshop    6.0    for
Windows</stEvt:softwareAgent>
          <stEvt:when>2001-07-25T09:45:06-07:00</stEvt:when>
        </rdf:li>
        <rdf:li rdf:parseType='Resource'>
          <stEvt:action>version 5 created</stEvt:action>
          <stEvt:parameters>helpful info</stEvt:parameters>
          <stEvt:softwareAgent>Adobe    InDesign    2.0    for
Windows</stEvt:softwareAgent>
          <stEvt:when>2001-08-10T10:45:39Z</stEvt:when>
        </rdf:li>
      </rdf:Seq>
```

```
    </xmpMM:History>
  </rdf:Description>
</setxmp>
</rdf:RDF>
```

<table>
<tr><th>Tag</th><th>Notes</th></tr>
<tr><td>Setxmp</td><td>Required. Type should be either “MAPINFO” or “NOMAPINFO”. If set to “MAPINFO” then the standard PDF Info Dictionary items will be mapped to their XMP equivalents as described in the ISO 19005 (PDF/A) crosswalk as shown below.

<table>
<tr><th>Entry</th><th>PDF</th><th>Property</th><th>XMP</th><th>type</th></tr>
<tr><td>Title</td><td>text</td><td>dc:title</td><td>Text</td><td></td></tr>
<tr><td>Author</td><td>text</td><td>dc:creator</td><td>seq</td><td>Text</td></tr>
<tr><td>Subject</td><td>text</td><td>dc:subject</td><td>Text</td><td></td></tr>
<tr><td>Keywords</td><td>text</td><td>pdf:Keywords</td><td>Text</td><td></td></tr>
<tr><td>Creator</td><td>text</td><td>xmp:CreatorTool</td><td>Text</td><td></td></tr>
<tr><td>Producer</td><td>text</td><td>pdf:Producer</td><td>Text</td><td></td></tr>
<tr><td>CreationDate</td><td>date</td><td>xmp:CreateDate</td><td>Date</td><td></td></tr>
<tr><td>ModDate</td><td>date</td><td>xmp:ModifyDate</td><td>Date</td><td></td></tr>
</table>
</td></tr>
<tr><td>Custom</td><td>Optional. This can be used to specify name / value pairs for custom entries in the PDF Info Dictionary. The custom DocInfo metadata is visible in Adobe Acrobat via File | Document Properties | Custom.</td></tr>
<tr><td>Rdf:Description</td><td>Optional. This can be used to define the RDF for a particular XMP schema. XMP Metadata is visible in Adobe Acrobat via File | Document Properties | Description | Additional Metadata | Advanced.

The Adobe XMP reference fully describes XMP syntax.: http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf</td></tr>
</table>

8.2 Command Line Interface

A command line interface to PDF to TIFF conversion is provided topdf.exe in the PDF Junction bin directory

setxmp.exe /source=sourcefile /target=targetfile /output=outfilefile [/debug=true] [/xmlfile=xmlfile]

For example, the following command processes the file test.pdf using the directive file test.xmp in the same folder (this is the default assumption unless /xmlfile is specified). The output PDF is written to the d:\dev\out folder and is called test_xmp.pdf.

```
setxmp.exe"          /output="%FILENAME_xmp.pdf"          /Source="c:\demo\in\demo.pdf"
/Target="c:\demo\out"
```

Command Line / Config File Name	Description
/Source	Source file definition.
/Target	Output folder. If not specified the result file will be placed in the same folder as the source file.
/Output	Output file name mask. May include %FILENAME If not specified the output file name will be based on the source file name.
/xmlfile	This specifies the XML directive file. If not specified, the XML directive file is assumed to be %FILENAME.xml where the source file is %FILENAME.pdf.

9 PDF BARCODE COMMANDLINE INTERFACE

Command Line	Description
/ SplitByBarcode (bool)	Choose this option to split the PDF file by Barcode
/RenameByBarcode (bool)	Choose this option to rename the PDF file based on Barcode.
/ Source	The source file.
/BarcodeFormats	Barcode formats supported. See table below for more explanation
/ ExcludedBarcodeFormats	Barcode formats not supported.
/SplitMode (bool)	Various Options for splitting Files by Barcode
/ TryHarder (bool)	Spend more time to try to find a barcode; optimize for accuracy, not speed. Default is true.
/SplitOutputPath (string)	The output file path template where the split files will be saved. - %VALUE% : Replaced by the barcode value found. - %INDEX% : Replaced by the current split index. - %FILENAME : Replaced by the file name
/SplitNoBarcodeTemplate (string)	The renaming template to use for page ranges where no barcodes were identified. Allowed templates: - %INDEX% ; Replaced by the current split index. - %FILENAME : Replaced by the filename of the source file.
Advanced Flags	Additional PDF Junction advanced command line flags may be entered here (see section 3)
/Binarize	Switch this on if your input files are coloured, it first converts it to black and white before conversion.
/RemoveLines (float)	Remove lines from images. Value = 100.5
/Zones (string)	Only examine the region specified for barcode(s). Values =left(int), top(int), width(int), height(int) e.g. /Zones= 5,10,200,200
/BoxSize	This option is ideal for forms where sometimes boxes around text can cause an area to be identified as graphics. This option removes boxes from the temporary copy of the imaged used by the barcode reader. Technically, this option removes connected

	elements with a minimum area (in pixels and defined by this property). This option is currently only applied for bitonal images.
/PdfToImageCompression	PDF files get converted to TIFF files before the barcode are recognized, this flag represents the compression of the TIFF file. “CCITT4” or “LZW”
/PdfToImageDpi	PDF files get converted to TIFF files before the barcode are recognized, this flag represents the DPI of the TIFF file. 150,200,300
/Deskew (bool)	Deskew (straighten) the image. The default value is false
/ Despeckle (int)	Removes specks below the specified pixel size from the image. The default value is 0 and the maximum value is 9.
/ Debug (bool)	Set this to true if you want more output one the console.
/ TempFolder (string)	This flag can be used to specify a temp location where the processing will be carried out.
/PerformPreprocessing (true)	Set this to true so that the image can be pre-processed and ready for recognition, this is important for images that are not very clear and visible.
/EndPage (int)	The page you want to start recognizing the barcode.
/StartPage (int)	The page you want to start recognizing the barcode.
/ BlankPageThreshold (int)	Any page with black pixels lower than the threshold set will be ruled as a blank page.
/ReadMultipleBarcodes (bool)	This is used to read multiple barcodes in a single page
/Morph	Morphological options that will be applied to the binarized image before OCR. If set to empty none is applied. Common options include those listed below but for more options please contact support@aquaforest.com: - d2.2 – 2x2 dilation applied to all black pixel areas, useful for faint prints. - e2.2 – 2x2 erosion applied to all black pixel areas, useful for heavy prints. - c2.2 – closing process that performs a 2x2 dilation followed by a 2x2 erosion with the result that holes and gaps in the characters are filled. This option should generally only be used under guidance from technical support.
/ PInvoke (bool)	Setting this to true will use Pinvoke to create new processes.
/ ProductsOnly	Only decode the UPC and EAN families of barcodes.

/ LibTiffSavePageAsBmp	Sometimes if there is an image which is 1bpp and has LZW compression, the pre-processing can cause the colour of the image to be inverted (black to white and white to black). Set this to true to avoid this.

10 XML JOB TICKET FILES

PDF Junction uses XML Job Ticket files to store details of a particular job specification before this is run using the product either from the GUI or using the /jobid option from the command line – see 3.2.1. Note that some functionality, in particular custom scripts and logfiles can only be used in conjunction with XML Job Ticket files. Other functionality can also be used via the equivalent command line flags.

Developers may wish to create or generate XML job files using the information below.

A sample job file is shown here:

```
<pdfjunction_job>
<operation>split</operation>
<sourcetype>folder</sourcetype>
<sourcefiles>C:\qatest\pj\in\set001</sourcefiles>
<target>c:\qatest\pj\out\qa001</target>
<joboptions>-o -c -f %F%\%U%</joboptions>
<metadata></metadata>
<security></security>
<docoptions></docoptions>
<custom_script>custom.wsf</custom_script>
<custom_script_pre_job>true</custom_script_pre_job>
<custom_script_post_job></custom_script_post_job>
<custom_script_pre_folder></custom_script_pre_folder>
<custom_script_post_folder></custom_script_post_folder>
<custom_script_pre_file>true</custom_script_pre_file>
<custom_script_post_file></custom_script_post_file>
<custom_script_on_error></custom_script_on_error>
<logfile>qa001.txt</logfile>
</pdfjunction_job>
```

Attribute	Description
operation	Must be split, merge, set, get or pdf (for conversion to PDF)
sourcetype	Source type (file, folder, tree). The default value is file.
target	Target directory for result files. If not specified, the source directory will be used as the target.
Joboptions	Set of options for the specified operation. Described in detail in section 5.3.
Metadata	Metadata settings for result PDF files. By default metadata fields are left blank. Described in detail in section 5.4.
Security	Security settings for result PDF files. By default files are not secured. Described in detail in section 5.5.
docoptions	Document option settings for result PDF files. Described in detail in section 5.6.
Sourcefiles	Source PDF file, folder or tree. May be multiple files for merge operations.
custom_script	If not left blank, specifies the name of a custom windows script file in *PDFJUNCTIONDIR*\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

11 CUSTOM SCRIPTS

The functionality of PDF Junction can be extended by using custom scripts which may be called at various points during the processing of a particular job. The custom scripts are Windows Script Files which allow scripting in either VBScript or Jscript.

Custom scripts reside in *PDFJUNCTIONDIR*\custom. In order to implement custom scripts, a developer should take a copy of the script template (custom.wsf, shown below) and implement the script according to their requirements. The custom script name should be specified and required exists should be flagged as being called, either using the GUI or directly in the Job Ticket File (See section 4).

Parameters are passed as follows:

Parameter	**Description**
jobid	The current job id
currentfolder	The path of the folder currently being processed, or "null" for non-applicable exits (eg job_start)
currentfile	The path of the file currently being processed, or "null" for non-applicable exits (eg job_start)
exitpoint	One of the following string values, according to the exit point that was called. pre_job post_job pre_folder post_folder pre_file post_file on_error

```
<job>
<runtime>
  <named name="jobid" helpstring="Job ID" type="string" required="false"/>
  <named  name="currentfolder" helpstring="Current Folder" type="string" required="false" />
  <named  name="currentfile" helpstring="Current File" type="string" required="false" />
  <named  name="exitpoint" helpstring="Script Exit Point" type="string" required="false" />
</runtime>
<script language="VBScript">
'
' Use this section for VBSript
'
'jobID=WScript.Arguments.Named("jobid")
'currentFolder=WScript.Arguments.Named("currentfolder")
'currentFile=WScript.Arguments.Named("currentfile")
'exitPoint=WScript.Arguments.Named("exitpoint")
'WScript.StdOut.WriteLine("Parameter Values - "+jobID+":"+currentFolder+":"+currentFile+":"+exitPoint)
</script>
<script language="JScript">
//
// Use this section for JScript
//
var jobID=WScript.Arguments.Named("jobid");
var currentFolder=WScript.Arguments.Named("currentfolder");
var currentFile=WScript.Arguments.Named("currentfile");
```

```
var exitPoint=WScript.Arguments.Named("exitpoint");
WScript.StdOut.WriteLine("Parameter Values - "+jobID+":"+currentFolder+":"+currentFile+":"+exitPoint);
</script>
</job>
```

12 XML "INFORMATION" FILES

PDF Junction can be used to generate XML files containing metadata relating to source PDF files, as in the following example:

```
cscript pdfjunction.wsf /operation:get /target:c:\demo\out
c:\demo\in\demo3.pdf "/joboptions:-f %F%_metadata"
```

Attribute	Description
Path	Full path name of the PDF file
Pages	The number of pages in the document
Title Author Subject Keywords Creator	PDF Document Metadata
Pdfversion	Version of PDF that the document complies with
Filesize	The size in bytes of the file
Permissions	Permissions bit mask. If set to 0, there is no security set on the file.
Isencrypted	True if the file is encrypted
Xmlmetadata	The XML PDF metadata as extracted from the document. See XXX for more details.

```
<?xml version="1.0"?>
<pdfdocument>
<path>c:\demo\in\demo3.pdf</path>
<pages>3</pages>
<title>1</title>
<author>Neil Pitman</author>
<subject></subject>
<keywords></keywords>
<creator>Acrobat PDFMaker 6.0 for Word</creator>
<pdfversion>5</pdfversion>
<filesize>517811</filesize>
<permissions>0</permissions>
<isencrypted>false</isencrypted>
<xmlmetadata>
<?xpacket begin='ï»¿' id='W5M0MpCehiHzreSzNTczkc9d'?>
<?adobe-xap-filters esc="CRLF"?>
<x:xmpmeta xmlns:x='adobe:ns:meta/' x:xmptk='XMP toolkit 2.9.1-13, framework 1.6'>
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
xmlns:iX='http://ns.adobe.com/iX/1.0/'>
…..
<rdf:Description rdf:about='uuid:3a543dda-858c-4fb6-88f9-3b7358591e42'
xmlns:dc='http://purl.org/dc/elements/1.1/'
dc:format='application/pdf'><dc:title><rdf:Alt><rdf:li xml:lang='x-
default'>1</rdf:li></rdf:Alt></dc:title><dc:creator><rdf:Seq><rdf:li>Neil
Pitman</rdf:li></rdf:Seq></dc:creator><dc:subject><rdf:Seq><rdf:li></rdf:li></rdf:Seq>
</dc:subject></rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end='w'?>
</xmlmetadata>
</pdfdocument>
```

13 PDF Junction Files and Directories

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

Directory	Contents
bin	Executables and Scripts.
docs	Reference Guide
jobs	Standard directory for XML Job Ticket files.
logs	Log files
temp	Temporary storage for XML Job Ticket files.
License	License File.

14 Support

Any problems running this application should give evident error messages, but if you are unable to resolve a problem please contact support@aquaforest.com

15 Acknowledgements

This product makes use of a number of Open Source components which are included in binary form. They are listed below.

Name	Homepage
BitMiracle.LibTiff.NET	Homepage | GitHub
IKVM.NET	Homepage | Sourceforge
Libtiff	Homepage
PDFBox	Homepage