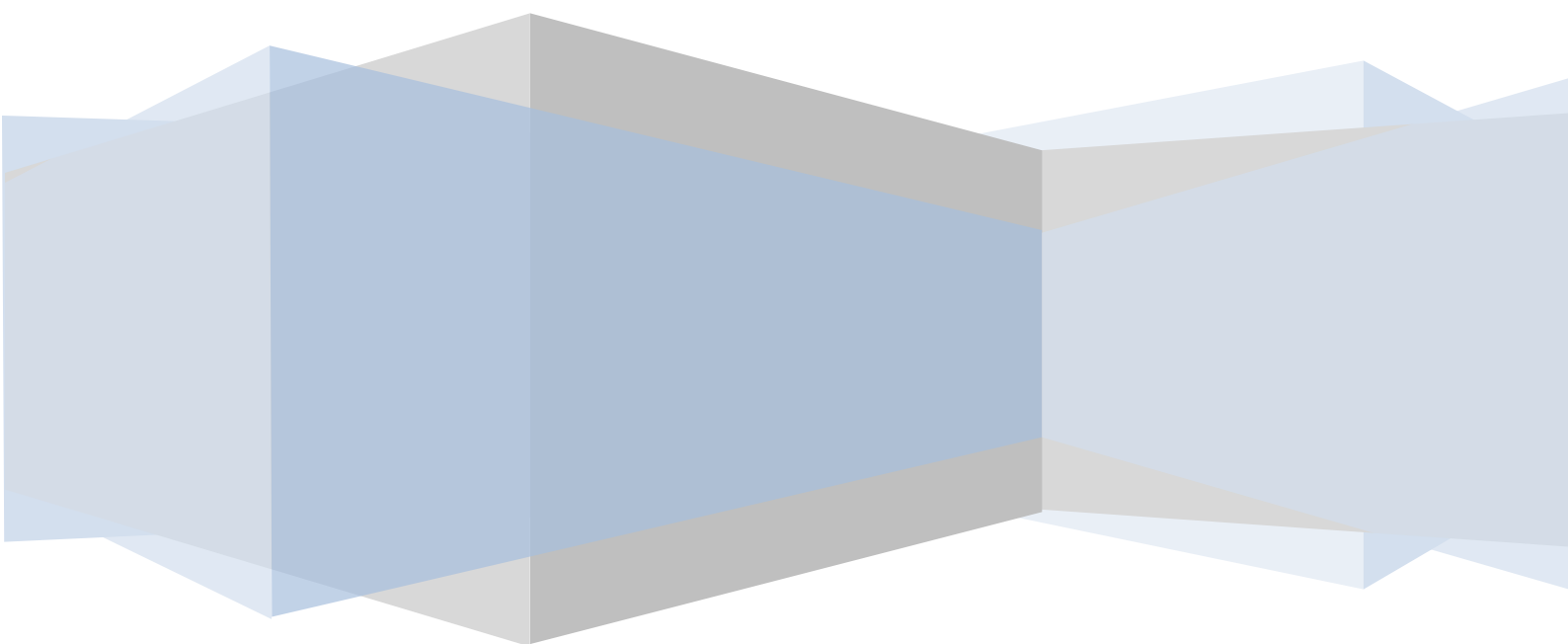




Proven Product, Proven Company

OCR Zone Data Extraction

Setup Guide



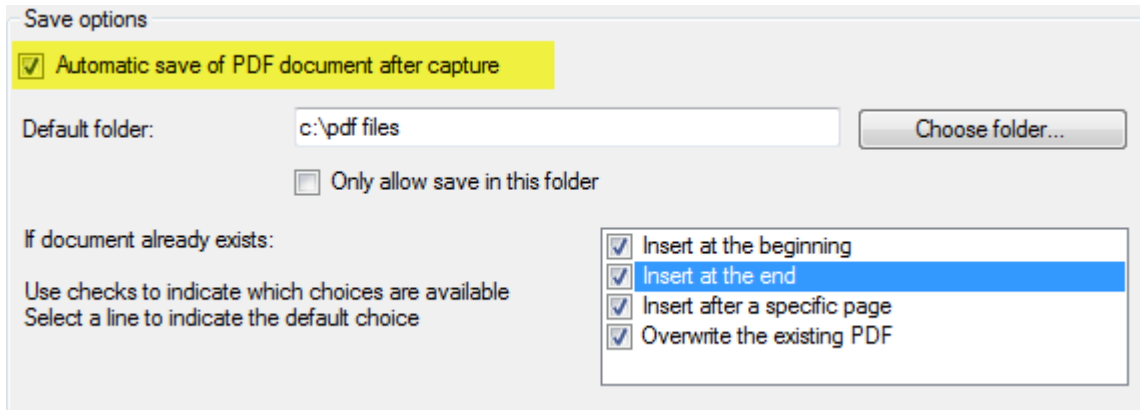
CONFIGURING SCANTOPDF SETTINGS

Start ScanToPDF and Click the Edit Menu, then select the “Options” choice from the menu



Save

Click the Save Icon

A screenshot of the 'Save options' dialog box. The 'Automatic save of PDF document after capture' checkbox is checked and highlighted in yellow. The 'Default folder' is set to 'c:\pdf files'. A list of options for 'If document already exists' is shown, with 'Insert at the end' selected. Other options include 'Insert at the beginning', 'Insert after a specific page', and 'Overwrite the existing PDF'.

Save options

Automatic save of PDF document after capture

Default folder: Choose folder...

Only allow save in this folder

If document already exists:

Use checks to indicate which choices are available
Select a line to indicate the default choice

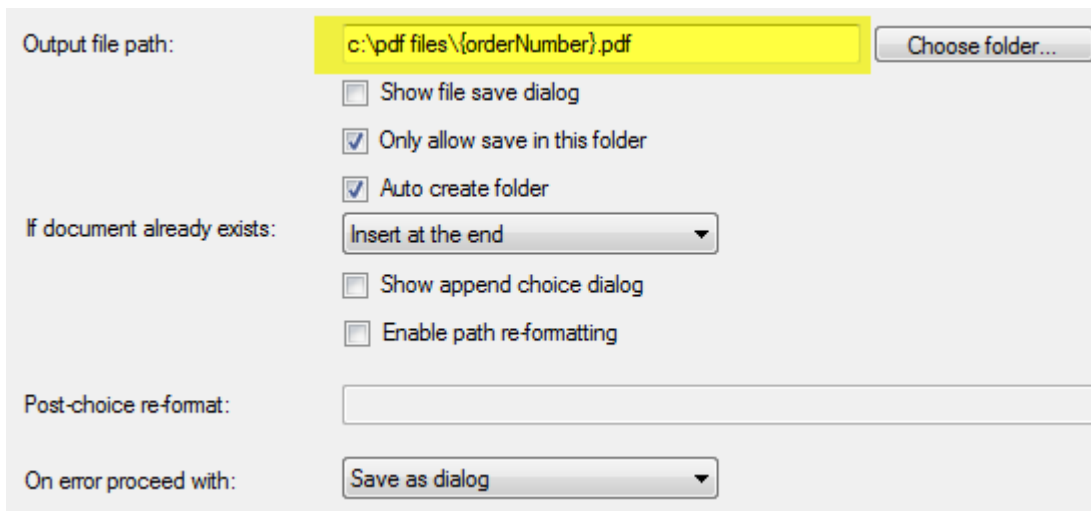
- Insert at the beginning
- Insert at the end
- Insert after a specific page
- Overwrite the existing PDF

Make sure the option “Automatic save of PDF document after capture” (*highlighted above in yellow*) is checked



File namer

Click the Filenamer Icon

A screenshot of the 'File namer' dialog box. The 'Output file path' is highlighted in yellow and set to 'c:\pdf files\{orderNumber}.pdf'. Other options include 'Show file save dialog', 'Only allow save in this folder', 'Auto create folder', 'If document already exists' (set to 'Insert at the end'), 'Show append choice dialog', 'Enable path re-formatting', 'Post-choice re-format', and 'On error proceed with' (set to 'Save as dialog').

Output file path: Choose folder...

Show file save dialog

Only allow save in this folder

Auto create folder

If document already exists:

Show append choice dialog

Enable path re-formatting

Post-choice re-format:

On error proceed with:

Output file path – Set the output file path (the location and file name for your PDF files). The example setting shows the file being named using the orderNumber value read from the document (ScanToPDF variables are always enclosed in curly braces so we enter {orderNumber}.pdf

On error proceed with – This will depend on the error process required. Choose from one of the three options

OiC – ScanToPDF OCR Zone Data Extraction Configuration Guide

1. **Error case** – Name the file using the settings configured in the “Error case file name generation” TAB
2. **Save as dialog** – Display a Save dialog to enable the user to enter a file name
3. **Cancel save** – Do nothing, processing will be suspended

ALL OTHER OPTIONS CAN BE SET AS ABOVE



Data Gatherer

Click the Data Gatherer Icon

The data gatherer is used to define the zones and create ScanToPDF variables from the data read from those zones. In this example we are extracting 5 zones (Order Number, Make, Model, Year and Colour) from an example Motor Vehicle bill of sale. To Edit Data Gatherer settings, click the Edit button to open up the template editor. The editor has TWO panes; the template image is shown on the left hand side, the settings for the selected zone on the right. The zones are shown on the template window as light blue boxes and the settings for each zone are shown when the zone is clicked. The zone settings for order number are shown below.

<input type="checkbox"/> Data	Assignment variable name	orderNumber
	Data type	String
	Zone name	Order Number
<input type="checkbox"/> Extraction		
<input checked="" type="checkbox"/> Coordinates		X: 1.370856 Y:1.370856 Width1.199499
	Data provenance	OCR
	Default to empty	False
	Exclusion list	(Collection)
	Multi-value delimiter	
	Multi-value grab	False
	Regex capture	(?<orderNumber>[0-9]{5})
	Regex grab syntax	
	Regular expression case insensitivity	True
<input type="checkbox"/> Trigger		
	Condition script	

SETTINGS NOT EXPLAINED HERE ARE NOT USED IN THIS EXAMPLE AND CAN BE DEFAULTED

Assignment variable name – The name of the variable in ScanToPDF

Data type – String, Integer, Decimal or Date (choose the one that matches your data)

Zone Name – The meaningful name for the zone

Regex Capture – A Regular Expression used to capture the variable. In this case a named group called orderNumber is used, which is 5 numerical digits long. It is beyond the scope of this document to explain all the features of Regular Expressions but the Internet has many sites dedicated to Regular Expressions (or Regex as it is widely known)

The other zones are configured in a similar way, these can be selected by clicking the zone in the template or choosing the zone from the drop down list at the top of the settings pane.

All templates must be text searchable PDF documents. To create a template you can simply scan a page and save it manually in a template folder of your choose. You can then use the “Load Template” button to load the new template and change the settings using the process above.



Data Exporter

Click the Data Exporter Icon

Data Exporter Settings

Exported file path:

Auto create folder

File type:

If file exist then:

Append

OverWrite

File encoding:

Data block expression:

Default null values to empty

SETTINGS NOT EXPLAINED HERE ARE NOT USED IN THIS EXAMPLE AND CAN BE DEFAULTED

Exported file path – Set this to a path to the log file

Data block expression – These are the fields which are exported to the log file (ScanToPDF variables). They are exported in this order and will produce a line as follows :

03-03-2011,54321,VOLVO,SILVER,2001,740,c:\pdf files\54321.pdf,1



Click the Batch separation icon

In our example, each time the order number changes we want to create a separate PDF. This is done by using the “Changing Expression” setting and setting the variable {orderNumber} as the expression, as shown below.

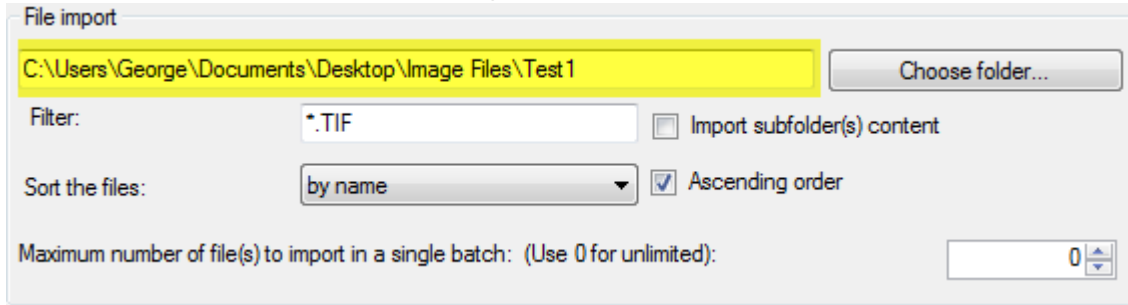
The screenshot shows the 'Separation' configuration window. It has two tabs: 'Separation' and 'Visualisation'. The 'Separation mode' is set to 'Multiple PDF files per batch'. There are two radio button options: 'Fixed number of pages' (unselected) and 'Variable number of pages' (selected). Under 'Fixed number of pages', there is a 'Page count' spinner set to 0. Under 'Variable number of pages', there is a 'Separator page is:' list with four options: 'Barcoded' (unselected), 'Black' (unselected), 'Blank' (unselected), and 'Changing Expression' (selected). Below this list are three checkboxes: 'Include separator page in document' (checked), 'Separator marks start of document' (checked), and 'Display error on attempt at producing empty document' (unchecked). There is a 'Value (leave blank for any):' text box. Below that is an 'Expression Syntax:' text box containing '{orderNumber}'. At the bottom, there is a checkbox for 'Carry expression values to subsequent pages' which is unchecked.

OCR ZONE DATA EXTRACTION SOLUTION OPERATION



In the SampleImages folder you will find 3 tiff images which can be used for testing. To use these images as a test

1. Copy the images to a new folder on your system
2. Start ScanToPDF and click Edit-Options-Import
3. Enter the folder you created into step 1 into the folder name in the Import settings



4. Click OK to close the settings screen
5. Click Import from folder on the ScanToPDF toolbar, or use File-Import-Import from folder option
6. The documents will be imported and three PDF files will be created in the output folder, named using the order numbers on the document (12345.pdf, 98765.pdf and 54321.pdf)