



Data Gatherer Quick Setup Guide

Introduction

Thank you for your interest in the Data Gatherer. The Data Gatherer plug-in extracts information from zones on pages as they are captured and then populates that information into ScanToPDF variables for use in automatic file naming or data export.

This guide provides instructions for configuration of the Data Gatherer plugin using OCR zone extraction. The Data Gatherer uses a template to draw zones onto a page and then each of the zones has parameters to control the extraction of data from that zone.

Please read and follow these instructions to configure zones using the Data Gatherer.

Allow Template Setting

- Start ScanToPDF
- On the menu bar, click **Edit** then choose **Options**
- Click the **UI** icon
- Check the show **Save as template** option on the page contextual menu in the **Optional items on UI** section
- Click **Save** on the lower right of the settings window
- On the toolbar, click the **Trigger an automatic document save after capture** option so **it is not ticked**

OCR Settings Required

- Click the OCR icon
- Check the **Advanced termination mode** setting
- Click **Save** on the lower right of the settings window

Import or Scan A File To Create A Template

- Scan the page (or import a file) so the page(s) are displayed in the thumbnail pane
- Right mouse-click on the page required for the template and choose **Save as template**
- A **Save as** dialog will open. Enter a name for the template file (any filename is allowed)
- Click **Save** (THE FILE MUST BE SAVED AS A PDF FILE)
- The **Use newly created template** dialog will open asking to confirm use of template for Data Gatherer, answer **Yes** and click Save

Setting Up Data Gatherer Zones

Creating A New Zone

- Click **Edit** and choose **Options**
- Click the **Data Gatherer** icon
- Click **Edit**
- The Data Gather zone edit window will open
- On the menu bar click **Zones** and choose **Add**
- Use the mouse to draw a zone representing the area that data will be extracted from the template

Move the mouse to the property pages on the right and enter the details for the zone extraction

Data Gatherer Property Settings

Capture Settings

Coordinates – these are the coordinates used for the zone and represent the area drawn by the mouse when the zone was added

Data provenance - For extracting data using OCR (text values for example invoice number) choose **OCR**, for extracting barcode data choose **Barcode**

Default to empty - Set the value to a valid "" (empty) instead of null which would trigger an error in any expression (like in Filenamer) as null does not resolve to a valid expression in most syntax when empty will

Exclusion list – Enter a list of values that should not be matched or captured. For example, if the exclusion list is 12345 and that value is captured, the field will return empty (or null)

Multi-value delimiter – The delimiter character used to delimit a multi-value field.

Multi-value grab – A Boolean indicating that the field is a multi-value field.

Regex capture – Enter a valid regular expression for a pattern match used to capture the data, for example to capture a 6 digit variable called poNumber, the regex (?<poNumber>[0-9]{6}) could be used in this field. OIC can help in Regex settings if required or there are many website dedicated to Regex (for example www.regexlib.com)

Regex is match boolean – When this field is set to true, the zone will return a Boolean rather than the value of the zone when the regular expression is matched. A value of True returned if the data in the zone matches the Regex, and a value of false if the Regex is NOT matched.

Regular expression case insensitivity (true/false) – A Boolean to indicate if the regex should consider the case of the matched value. For example, if the Regex should be case sensitive set the value to false

Data Settings

Assignment variable name – the name of the variable to be captured and populated, for example poNumber

Data type – the type representing the data. Integer, date or String

Date format – If the data type is Date, enter a date format for the variable enclosed in double quotes, for example "dd/mm/yyyy"

Zone name – The name of the zone. Use this field to enter a meaningful name for the zone, for example PO Number

Edit Settings

Manual entry – Boolean indicating whether the field can be corrected using manual intervention

Manual entry validation regular expression – Enter a Regular Expression used to validate the manual entry of the field

Post-Capture Settings

Action Script – In some zones where text of varying sizes is close to the text being extracted it may be useful to add the following syntax, which cleans up the extracted data further to improve results.

```
{OCRZone(ZoneField.Coordinates,"","","")}
```

Cleanup Script – Use scripting syntax to change the output (for example to remove spaces or replace characters), for example if the captured data was S12345 and the "S" character was not needed a **RemoveChars** function could be used, for example **{RemoveChars(poNumber,"S")}**

Reformatting Expression – This field can be used to change the format of the variable after capture such as pre-pending or appending. For example, if our assignment variable poNumber required pre-pending with "PO#" we would use "PO#" + poNumber. If the extracted data for the poNumber zone was 12345, the ACTUAL variable would then be populated as PO#12345

Trigger Settings

Condition script – A condition equating to true or false that will trigger the extraction of this zone

Trigger – Use **BeforeBatchSeparation** ONLY

When a zone is configured, click the **Test Extraction** button to test.

PLEASE NOTE : TEST EXTRACTION ONLY WORKS WHEN OCR DATA IS EXTRACTED, IT DOES NOT WORK FOR BARCODES

When all the zones have been configured, click the **OK** button to close the settings window.

THIS PROCESS CAN THEN BE REPEATED FOR ADDITIONAL FIELDS AS REQUIRED

Using Data Captured From Zones

To use captured data, the **assignment variable name** used in the zone must be enclosed in curly braces and can be used in any syntax in ScanToPDF, for example in the **Output file path setting** of the automatic file namer plugin to use 2 zones with variables called orderNumber and orderDate

```
C:\PDF\{orderDate}-{orderNumber}.pdf
```