



PSIGEN software inc.
the **science** of document capture

QUICKstart

02

Modified: 12.5.19
Screenshots: 7.5.0.11



PSI**capture**

Smart Zone Table Extraction

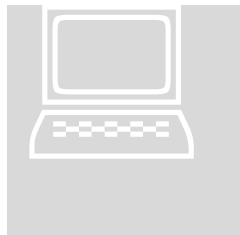
Quick Start Guide

Getting Started

Welcome to PSIsapture!

This Quick Start Guide is designed to help you through the process of creating a Capture Profile that uses Smart Zones to extract table data. This guide includes sample images that will be used to configure and test the Capture Profile. For additional information on features not covered in this guide please refer to the PSIGEN wiki at help.psigen.com. PSIGEN offers detailed hands on training with the PSIsapture software as well.

Make sure you have completed the **00 - Installation Quick Start Guide** and used it to get PSIsapture installed, licensed and running on your computer before starting with this Guide. You will need to be running version 7.6.x or newer for this guide to be fully effective.



Requirements:

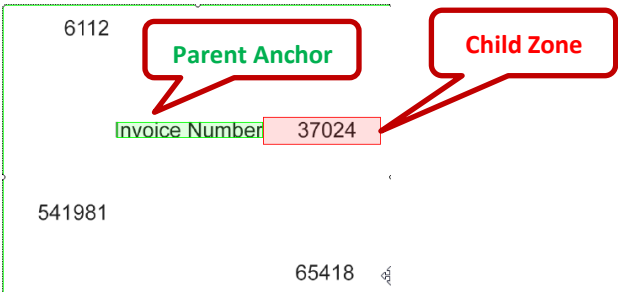
- Completed **00 - Installation Quick Start Guide**
- Ensure your computer meets the **min. requirements**
- Running version 7.6.x of PSIsapture or newer
- PSIsapture License with Table Extraction module

{This area is intentionally left blank, please continue to the next page}

What is Smart Zone Table Extraction?

Smart Zone Table Extraction allows a user to leverage the power of a single Smart Zone to return results for multiple child zones across multiple records. As with a standard Smart Zone, the anchor is found first, and then multiple child zones are established and mapped to individual index fields.

Single Smart Zone



Smart Zone Table Extraction

Date	Depart from	Depart time	Destination	Arrival time	Destination address	Phone Number	Company
1/15/09	LaGuardia	4:30 AM	Eppeley	7:30 AM	948 Elmwood Ave Omaha, NE 5701	555-543-8987	Canelux

Parent Anchor

Child 1

Child 2

Child 3

Child 4

Once the values for a single row are defined the Smart Zone anchor area can be extended to account for multiple lines.

Address	Phone Number	Company
948 Elmwood Ave Omaha, NE 5701	555-543-8987	Canelux
948 Elmwood Ave Omaha, NE 5701	555-819-0543	Singlecore
948 Elmwood Ave Omaha, NE 5701	555-245-3344	Canelux

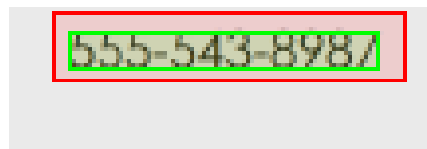
Address	Phone Number
948 Elmwood Ave Omaha, NE 5701	555-543-8987
948 Elmwood Ave Omaha, NE 5701	555-819-0543
948 Elmwood Ave Omaha, NE 5701	555-245-3344
948 Elmwood Ave Omaha, NE 5701	555-313-5544

SMART ZONE TABLE EXTRACTION
QUICK START GUIDE

Date	Depart From	Depart Time	Destination	Arrival Time	Destination address	Phone Number	Company	Comments
1/15/09	LaGuardia	4:30 AM	Eppley	7:30 AM	948 Elmwood Ave Omaha, NE 5701	555-543-8987	Canelux	Ask for Jerry
1/20/09	Eppley	6:00 PM	LaGuardia	10:00 PM				
1/23/09	LaGuardia	4:00 AM	Lambert	7:00 AM	436 Brandywine Dr St. Louis, MO 99337	555-819-0543	Singlecore	Ask for Bill
1/28/09	Lambert	5:35 PM	LaGuardia	9:35 PM				
2/1/09	LaGuardia	4:00 AM	Mitchell	7:00 AM	565 Front Street Milwaukee, WI 55678	555-245-3344	Canedrill	Ask for Mary
2/6/09	Mitchell	5:00 PM	LaGuardia	9:00 PM				
2/9/09	LaGuardia	4:30 AM	Birmingham	7:30 AM	290 Lexington Court Birmingham, AL 87745	555-314-5567	Trising	Ask for Tim
2/14/09	Birmingham	6:00 PM	LaGuardia	10:00 PM				

Smart Zone Table Extraction does **not** require a Child zone for each column in a table.

The same piece of data that is used as an Anchor can also be indexed using a Child zone.



{This area is intentionally left blank, please continue to the next page}

SMART ZONE TABLE EXTRACTION
QUICK START GUIDE

To enable Smart Zone Table Extraction, check the box in the Smart Zone Configuration window to **Create a record for each Smart Zone Anchor found on the page**.

Pre-Configured form field types

Source:

System defined

User defined

Manage

Types:

Add

Anchor expressions

Type	Expression				
Custom	[01]?[-.]?\([([2-9]]0-9]]([2-9]]0-9]](2))[-.]?[0-9]](3)[-.]?[0-9]](4)				

Locate

Child Zones will not be displayed in the viewer until the Locate button has been pressed.

Child zones

Anchor Attach Point	Child Attach Point	Width	Height	Offset X	Offset Y	Group		
---------------------	--------------------	-------	--------	----------	----------	-------	--	--

Group name is optional. Groups are used to populate multiple index fields using the same Smart Zone.

Options

☐ Anchor on matching word(s) only (use only when lack of spacing results in odd word combinations)

☒ Create a record for each smart zone anchor found on the page

☐ Combine data from this smart zone with others for complex table extraction

Zone hierarchy:

1

Determines the parent / child hierarchy when combining data

Anchor processing order:

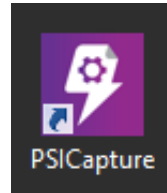
Top Down

Child zones are mapped to Index Fields using the Group column names.

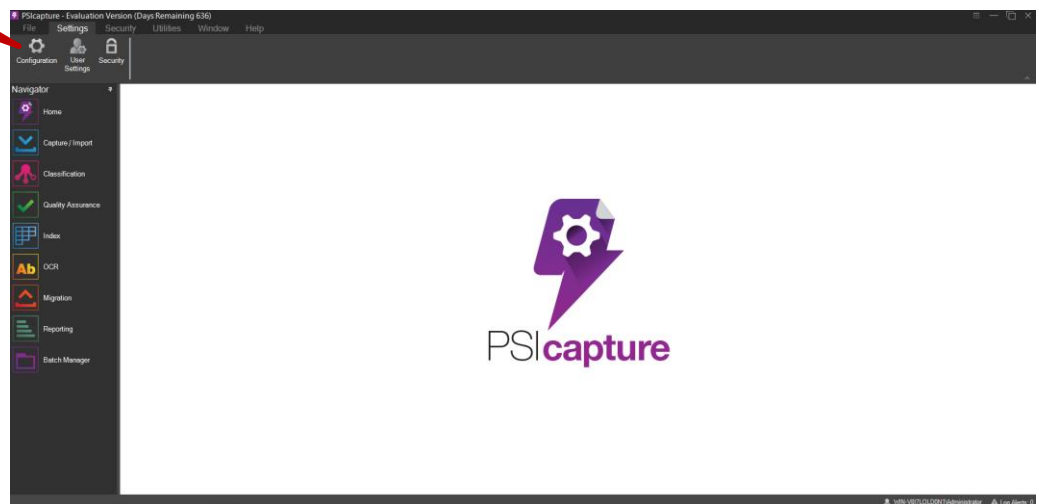
Child zones								
	Anchor Attach Point	Child Attach Point	Width	Height	Offset X	Offset Y	Group	
1	Right	Left	220	62	-2258	-2	Date	
2	Right	Left	235	62	-1996	2	Date	
3	Right	Left	299	58	-1371	-5	From	
4	Right	Left	232	59	55	4	To	
							Company	

Creating your Capture Profile

First, we'll need to get a basic Capture Profile started. Open PSICapture.



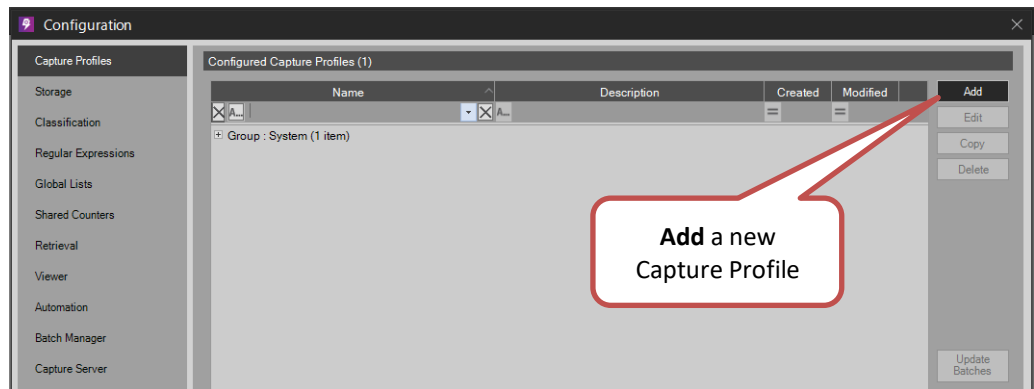
Click to open
Configuration



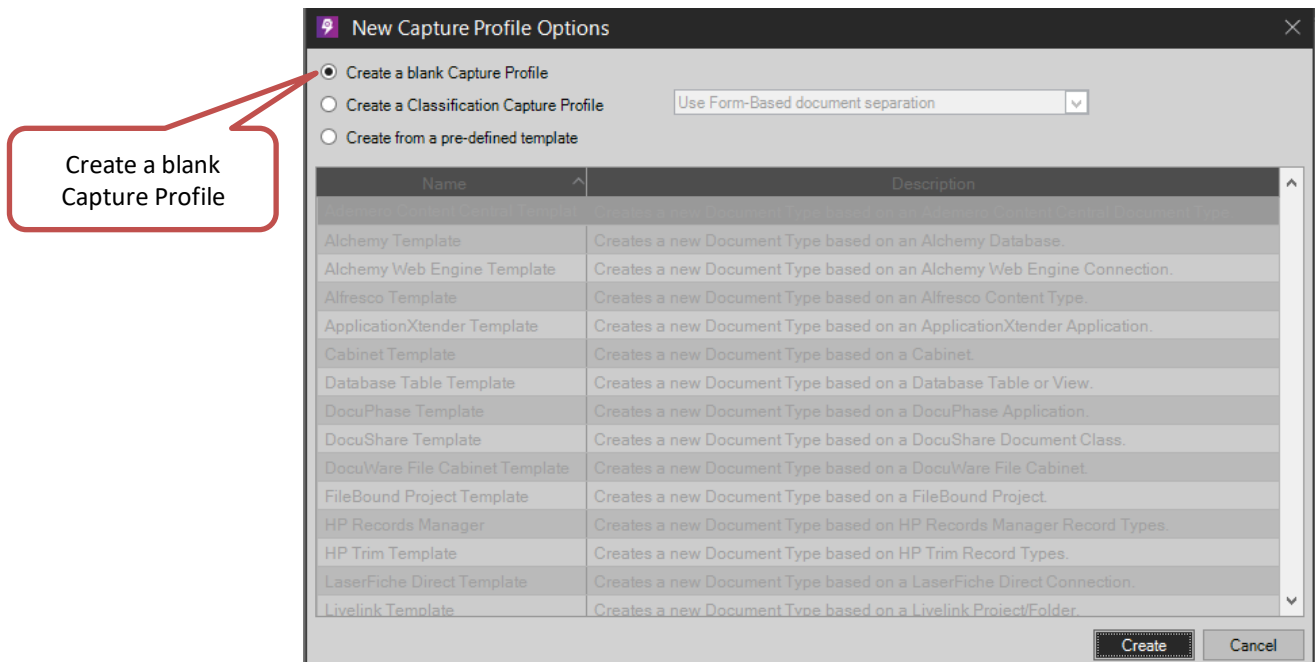
After PSICapture opens you will start in the Configuration module to create your Capture Profile.

SMART ZONE TABLE EXTRACTION QUICK START GUIDE

Once the Configuration module is open you can begin work on your Capture Profile. Please click the **Add** button on the right side of the window to proceed.



Choose to **Create a blank Capture Profile** and click **Create**.



SMART ZONE TABLE EXTRACTION QUICK START GUIDE

Capture Profile Tab

The screenshot shows the 'Capture Profile Configuration' window. The left sidebar has a 'Capture Profile' section with sub-items: General, Advanced, Image File Options, Validation And Exceptions, Auto Import, Document Records, Separation, Recognition, Indexing, Image Processing, Workflow, and Summary. The main area is titled 'Capture Profile - General Setup' and contains several sections: 'Identification, roles and security' with fields for Name (02 - Smart Zone Table Extraction), Group (QUICKstart), Description (Smart Zone Table Extraction - Quick Start Guide), and Allowed roles (All Users, Administrator); 'Batch type' with radio buttons for Local (selected) and Shared; 'Storage options' with a Storage Location dropdown (Batch Storage) and a New Location button; 'Batch naming' with radio buttons for Automatic (selected), Manual, and Custom, and a checkbox for 'When creating new batches, the batch name cannot be modified'; and 'Batch retention' with a dropdown for Batch retention type (Delete Batch), radio buttons for Immediately (selected) and After (days), and a checkbox for 'Retain retention date when re-opening Batch'. Callouts point to the Name field, the Group dropdown, the Description field, the Batch Storage dropdown, and the Batch Retention section.

Give your new Capture Profile a name, group, and description. Confirm your Batch Storage Location and set your Batch Retention.

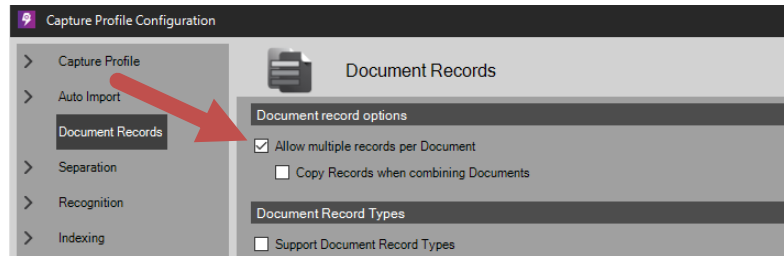
Auto Import Tab



For the purpose of this Quick Start Guide you will not be making any changes to Auto Import or its sub tabs. Many of the features within PSICapture Capture Profiles can be left at their default settings. For additional information about features available on this tab please refer to the PSIGEN [help](#).

Document Records Tab

Check the top checkbox for “Allow multiple records per document”



Separation Tab

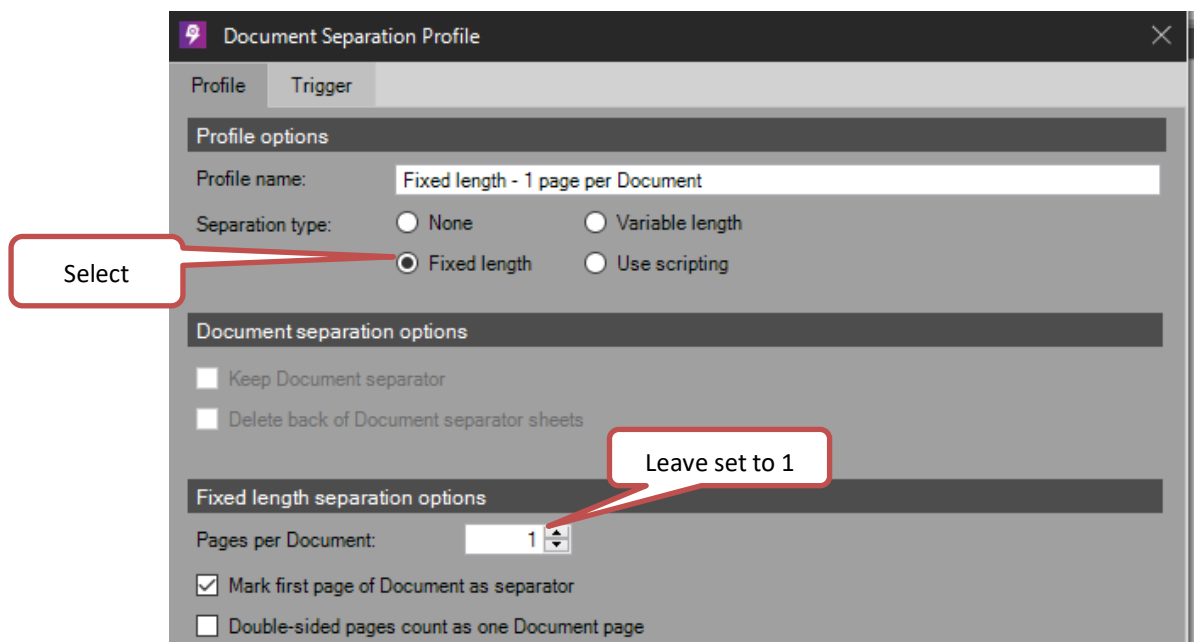
☆ Note:
Separation profiles are
unique to each Capture
Profile.

Add a Single page Fixed Length Document Separator.



Please select **Add** to create a new separation profile.

Document Separation Profile



Select Fixed Length separation. Click **Save**.

Recognition Tab

No barcode or other recognition engines are being used in this guide, so you can skip this tab.

Indexing Tab

Click the **Add** button on the middle of the window 4 times to create 4 blank index fields. Name those index fields as follows:

- Date
- Depart From
- Destination
- Company

Click **Add** 4 times (*One for each field*)

Document Index Fields

Name	Data Type
Date	Text
Depart From	Text
Destination	Text
Company	Text

General

Company

Document Index Field Definition

Default value:

Input mask:

Minimum length:

Maximum length:

Custom value: Build Custom Value

Record types:

Auto populated data options

☐ Truncate to maximum length ☐ Ignore length

☐ Trim whitespace ☐ Auto increment

Field options

☐ Required ☐ Re-key

☐ Read only ☐ Sticky

☐ Hide ☐ Auto casing

☐ Skip ☐ Regex

☐ Synchronized ☐ Requires an Advanced Data Extraction license

Import filename and path parsing

☐ Entire file / directory name ☐ Subset ☐ Split

Scripting

☐ Use custom scripts to populate and validate index fields

Notes

Apply Save Cancel

Click on the **Zone** tab and click the **Define Zones** button.

Click **Zone**

Document Index Fields

General Advanced Barcodes List Zone

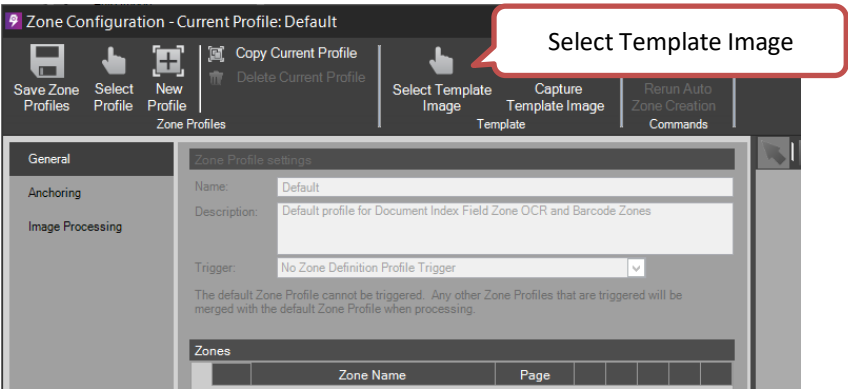
Company

Zone definition / action

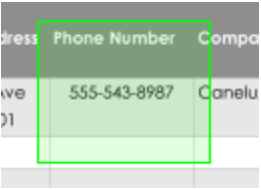
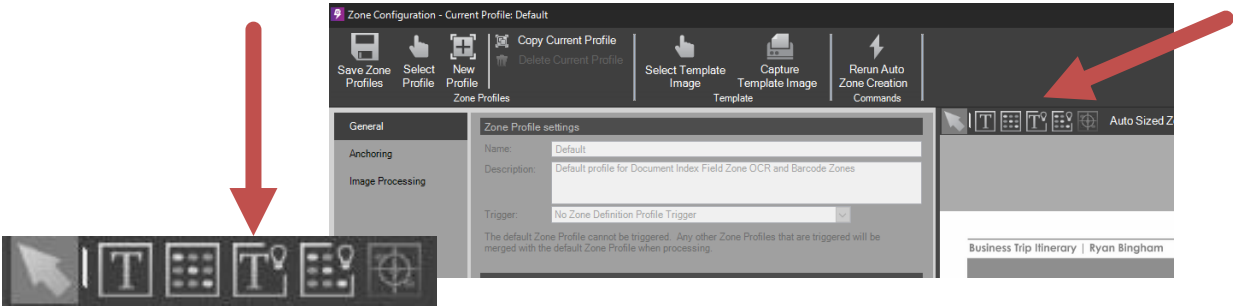
Action: Define Zones

☐ Disable page change when using the Zoom to Zone option

Load a sample of the image types we will be processing with this Capture Profile. Navigate to the folder where you saved your files and Load **02 - Smart Zones Table Extraction Sample.tif** as your template image.

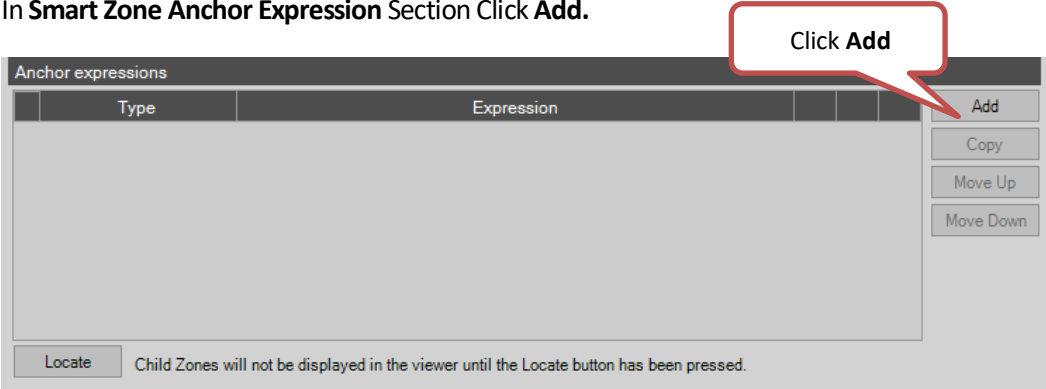


Once the template image has loaded, select the OCR Smart Zone tool.

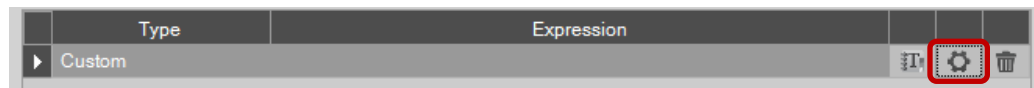


Draw an oversized OCR Smart Zone around the first Phone Number in the table. The Smart Zone Configuration will open upon releasing the mouse button.

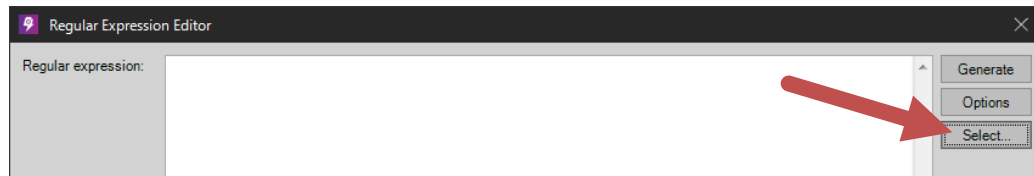
In **Smart Zone Anchor Expression** Section Click **Add**.



Click the **Edit Expression in Regular Expression editor**.



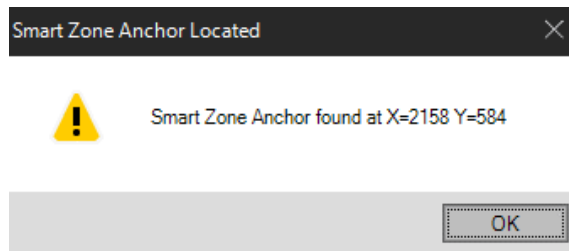
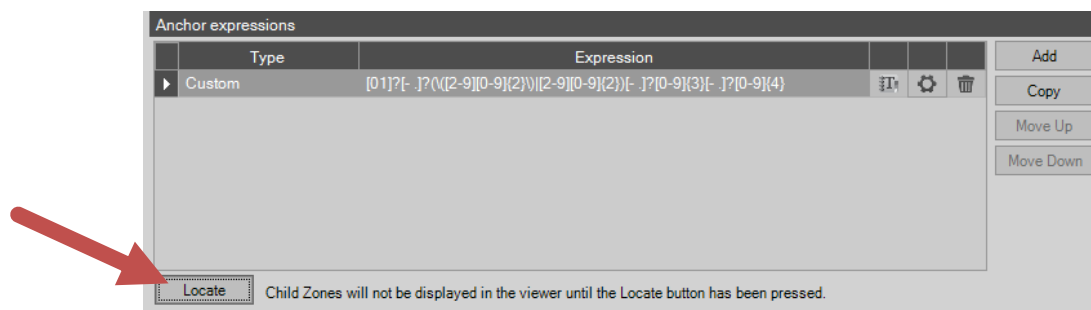
Click the **Select...** button from the right.



Select **Phone Number** and click **Select**. Then click **Save**.

Type	Description	Expression
System	Date	(1[012] 0[1-9])(<![0-9])[0-9])(/[0-9]{1,2} [/-])
System	Email Address	([0-9a-zA-Z]([-.\w]*[0-9a-zA-Z])*)@([0-9a-zA-Z]
System	Phone Number	[01]?[-]?(\([2-9][0-9](2)\) [2-9][0-9](2))[-]?[0-9]

Now that the expression is shown, press the **Locate** button. A notification message will appear stating the Anchor was located.

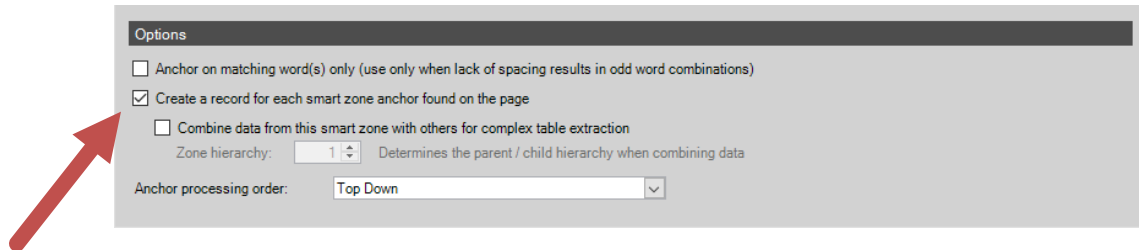


Click **OK** on the Smart Zone Anchor Located message.

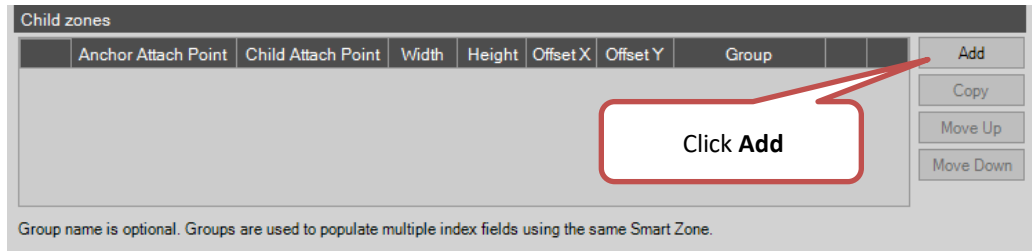
You'll now see the anchor highlighted on the image.



To account for multiple line items in the table, enable the **Create a record for each smart zone anchor found on the page** option.



Click **Add** in the lower middle of the Smart Zone Configuration screen to add the 1st Child zone.



☆ Note:
You may choose to enlarge the Smart Zone Configuration window and/or use the zoom tools to enhance precision.



A **red** Child Zone is now added to the right of our anchor in the Preview window. Click and drag the Child zone to move it so that it is positioned over the 1st Date in the sample. Using the corner selection handles, the zone can be resized to the appropriate dimensions.

Date	Depart From	Depart Time	Destination	Arrival Time	Destination address	Phone Number	Cor
1/15/09	LaGuardia	4:30 AM	Eppley	7:30 AM	948 Elmwood Ave Omaha, NE 5701	555-543-8987	Car

☆ Note:

For this Capture Profile, keep the amount of extra border space within the red child zones to a minimum. This will help ensure that no extraneous text is extracted.

Click **Add** again and repeat to add and reposition the remaining 3 Child Zones:

- Depart From
- Destination
- Company

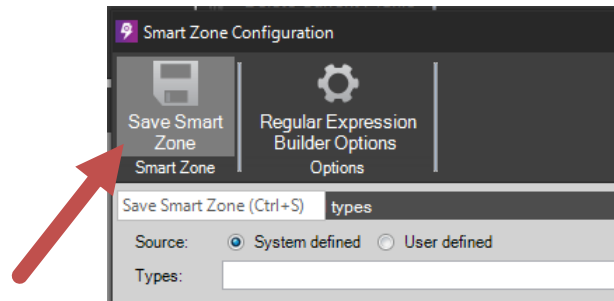
Date	Depart From	Depart Time	Destination	Arrival Time	Destination address	Phone Number	Company	Comments
1/15/09	LaGuardia	4:30 AM	Eppeley	7:30 AM	948 Elmwood Ave Omaha, NE 5701	555-561-0298	OmniLux	Ask for Jerry

Assign **Group** names to each Child zone to allow for mapping to Index Fields. Simply type in a name in the Group column to be referenced later. For this example we will use the column headers for Group names.

Child zones									
	Anchor Attach Point	Child Attach Point	Width	Height	Offset X	Offset Y	Group		
1	Right	Left	234	63	-2264	-6	Date		
2	Left	Left	359	70	-1759	-1	Depart From		
3	Right	Left	395	73	-1361	-2	Destination		
4	Right	Left	232	70	49	0	Company		

Save the Smart Zone Configuration.

Type Names



{This area is intentionally left blank, please continue to the next page}

From the Zone Configuration window, expand the Smart Zone to cover all Phone Numbers in the table.

Destination address	Phone Number	Company	Comments
948 Elmwood Ave Omaha, NE 5701	555-543-8987	Canelux	Ask for Jerry
436 Brandywine Dr St. Louis, MO 99337	555-819-0543	Singlecore	Ask for Bill
565 Front Street Milwaukee, WI 55678	555-245-3344	Cannedrill	Ask for Mary
290 Lexington Court Birmingham, AL 87745	555-313-5567	Tising	Ask for Tim

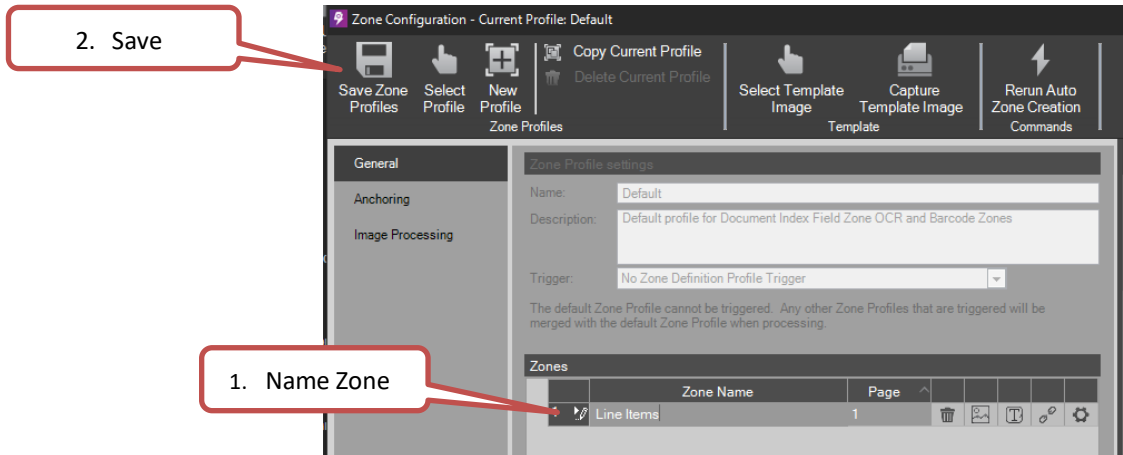
Stretch Smart Zone downward,
then Double-Click

Double-click the green Smart Zone itself to re-open the Smart Zone Configuration.

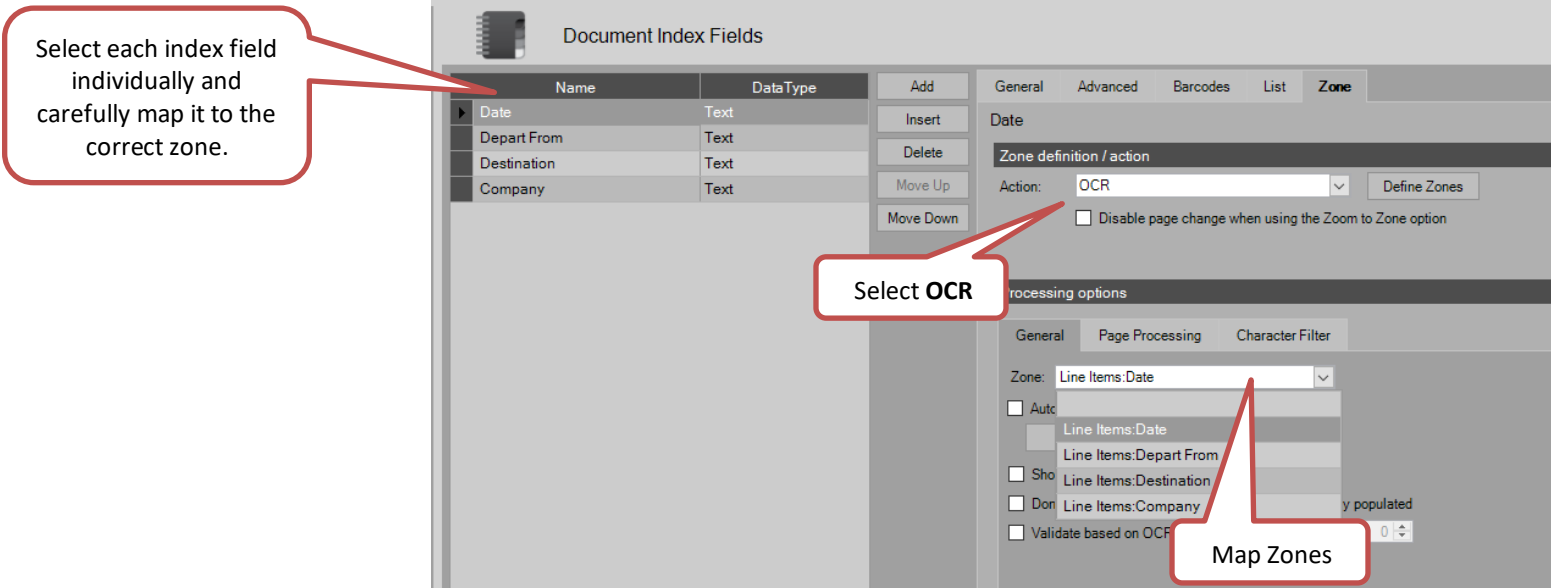
Date	Depart from	Depart Time	Destination	Arrival Time	Destination address	Phone Number	Company	Comments
1/15/09	LaGuardia	4:30 AM	Eppley	7:30 AM	948 Elmwood Ave Omaha, NE 5701	555-543-8987	Canelux	Ask for Jerry
1/20/09	Eppley	6:00 PM	LaGuardia	10:00 PM				
1/23/09	LaGuardia	4:00 AM	Lambert	7:00 AM	436 Brandywine Dr St. Louis, MO 99337	555-819-0543	Singlecore	Ask for Bill
1/28/09	Lambert	5:35 PM	LaGuardia	9:35 PM				
2/1/09	LaGuardia	4:00 AM	Mitchell	7:00 AM	565 Front Street Milwaukee, WI 55678	555-245-3344	Cannedrill	Ask for Mary
2/6/09	Mitchell	5:00 PM	LaGuardia	9:00 PM				
2/9/09	LaGuardia	4:30 AM	Birmingham	7:30 AM	290 Lexington Court Birmingham, AL 87745	555-313-5567	Tising	Ask for Tim
2/14/09	Birmingham	6:00 PM	LaGuardia	10:00 PM				

After verifying that there are now multiple Anchors, each with its own set of Child Zones, click **Save Smart Zone**.

From the Zone Configuration window, name the Smart Zone **Line Items**. Then click **Save**.



From the root of the Indexing tab, on the **Zone** Tab, set the **Action** for each of the four (4) index fields to **OCR** and map the fields to their corresponding **Line Items:Child Zones**.



Advanced Indexing Tab

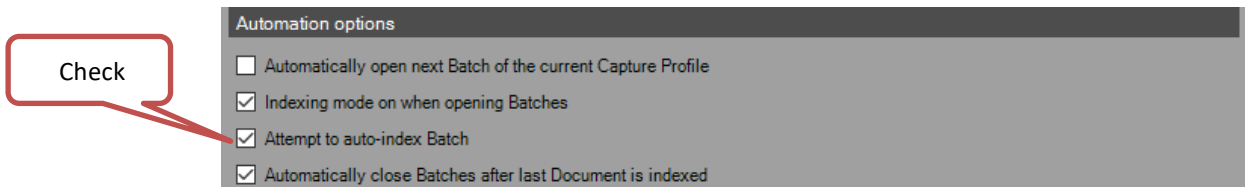
Since no changes are required here, you can skip this tab.

Image Processing Tab

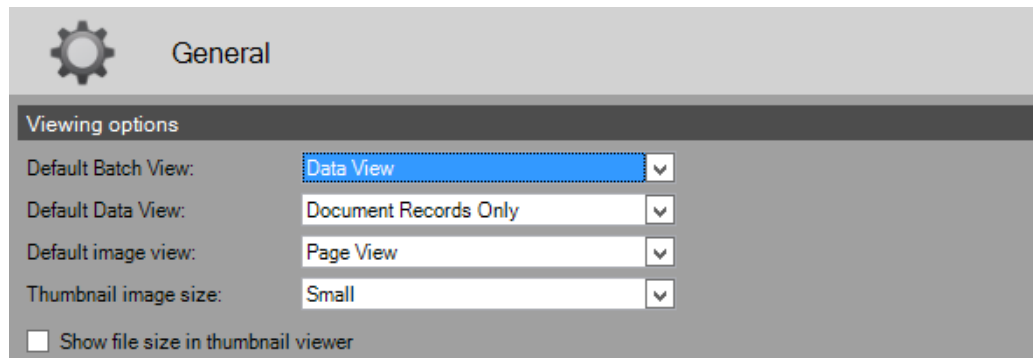
Since no changes are required here, you can skip this tab as well.

Workflow Tab

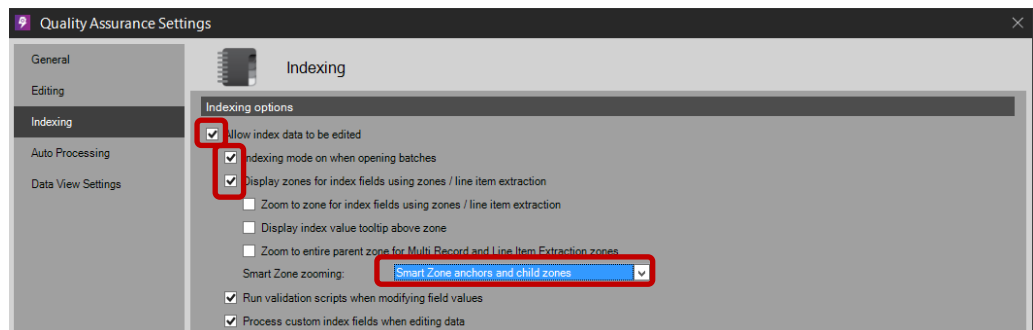
In this tab you will build the PSIs capture workflow. Select **Index** from the left and click **Add**. You will want to make only one change – check the **Attempt to Auto-Index Batch** checkbox from the top in the Automation section. This will auto-index each document without requiring user interaction from the Index Module. Click **Save**.



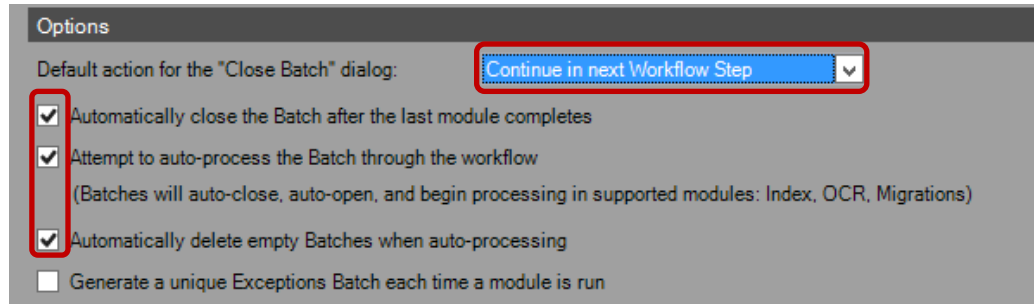
Next you will select **Quality Assurance** from the left and click **Add**. Change the Default Batch View from Tree View to **Data View**. This will cause the values you will have extracted during Indexing to be displayed when you enter the Quality Assurance workflow module.



On the **Quality Assurance – Indexing** sub tab, check the 3 checkboxes outlined in the below screenshot and choose **Smart Zone anchors and child zones** from the dropdown. This will display the exact location from which each index value was extracted as you verify the data from the Quality Assurance screen. Click **Save**.

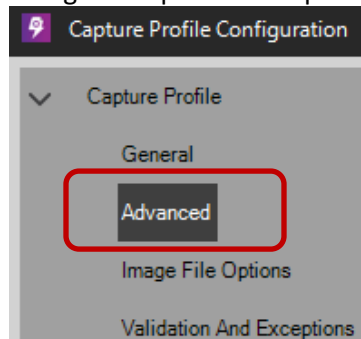


Next, in the Workflow options (at the bottom of the tab), check the top 3 checkboxes and change the dropdown to match below. This will automate the workflow, so that the Capture Profile will process the documents without requiring user interaction after the Capture/Import module.

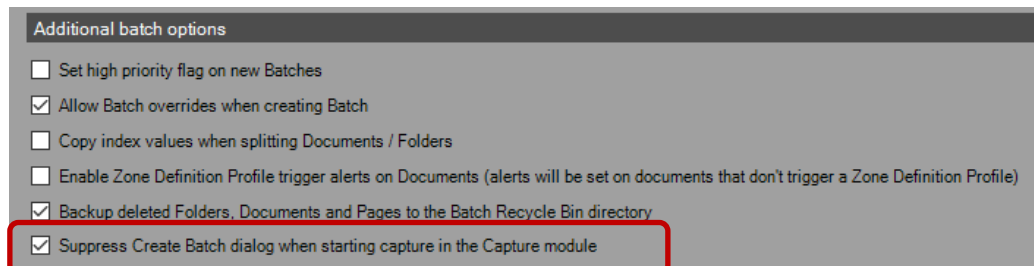


Suppress Create Batch Screen

When running a batch through PSISoftware the Create Batch screen is displayed by default. We will not have a need to change any settings on that screen in this profile, so we are going to suppress that screen from being displayed. On the left-hand Navigator expand out Capture Profile and select the Advanced tab.

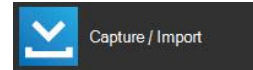


In the Additional batch options section, we want to check the 6th checkbox titled "Suppress Create Batch dialog when starting capture in the Capture module"

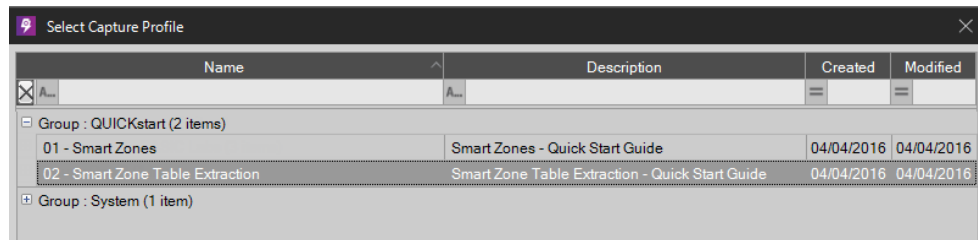


Run the Capture Profile

You can now click **Save** in the lower right-hand corner. Click **Save** again from the Configuration screen and run the Capture Profile by clicking Capture / Import tab.



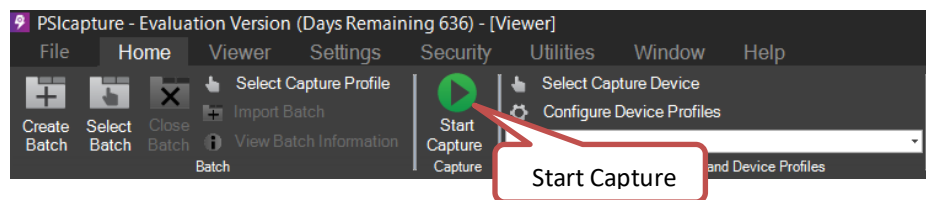
Select the **02 – Smart Zone Table Extraction** capture profile.



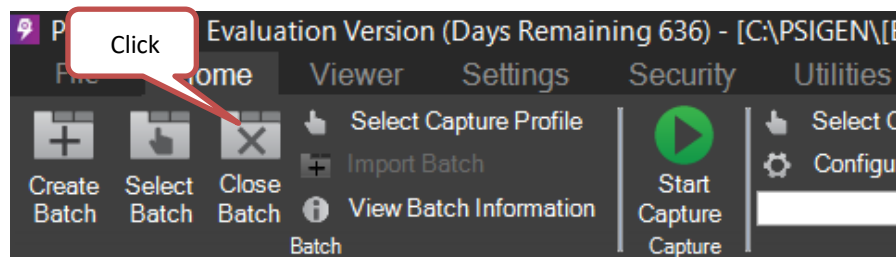
Using the File Browser pane, navigate to where you saved the files for this guide. You will select the folder on top and then select **02 – Smart Zone Table Extraction Sample.tif** from the lower pane. Highlight the file and select the green **Start Capture** or (Control+S) button to begin importing the document into PSICapture.

☆ Note:

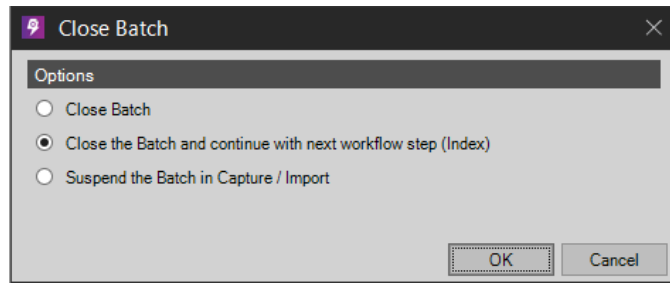
If you do not see the File Browser toward the left of the PSICapture screen, click **Select Capture Device...** icon to switch to the **Import** mode.



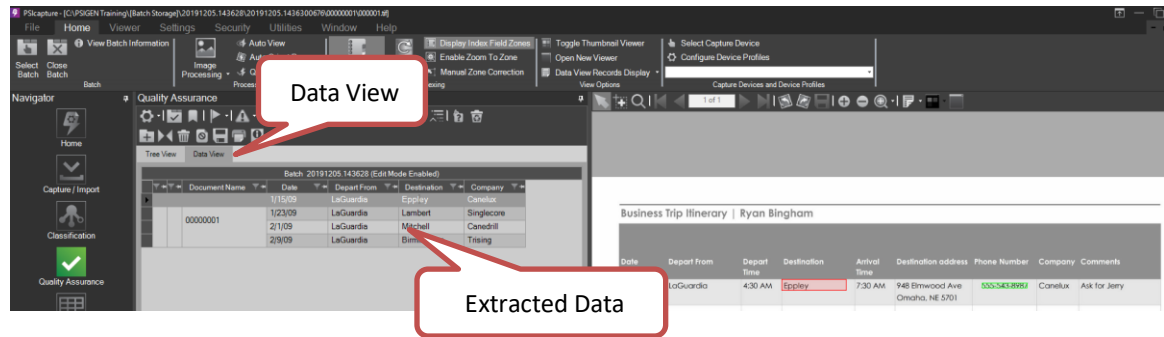
When the file has finished importing, you will click **Close Batch** button.



Ensure the selection for **Close Batch and Continue in next Workflow step (Index)** is selected. Click **OK**.



The Index step will attempt to auto-index the table and will then pass the batch directly to the Quality Assurance step, where you will be able to review the extracted data.



Click inside each individual cell of the Data View to view the exact zone on the page from which the data was extracted.

CONGRATULATIONS!!
 You've just configured and tested a
 Capture Profile using Smart Zone Table Extraction.