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

PSI**capture**

ACE (Accelerated Classification Engine)

Quick Start Guide

Getting Started

Welcome to PSIcapture!

This Quick Start Guide is designed to help you through the process of using the Accelerated Classification Engine (ACE) to configure a Document Type to recognize and process new varieties of documents within a PSIcapture batch. This guide includes sample images that you will process using the Document Type that you will design. 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 PSIcapture software as well.

Make sure you have completed the **00 - Installation Quick Start Guide** and used it to get PSIcapture 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.
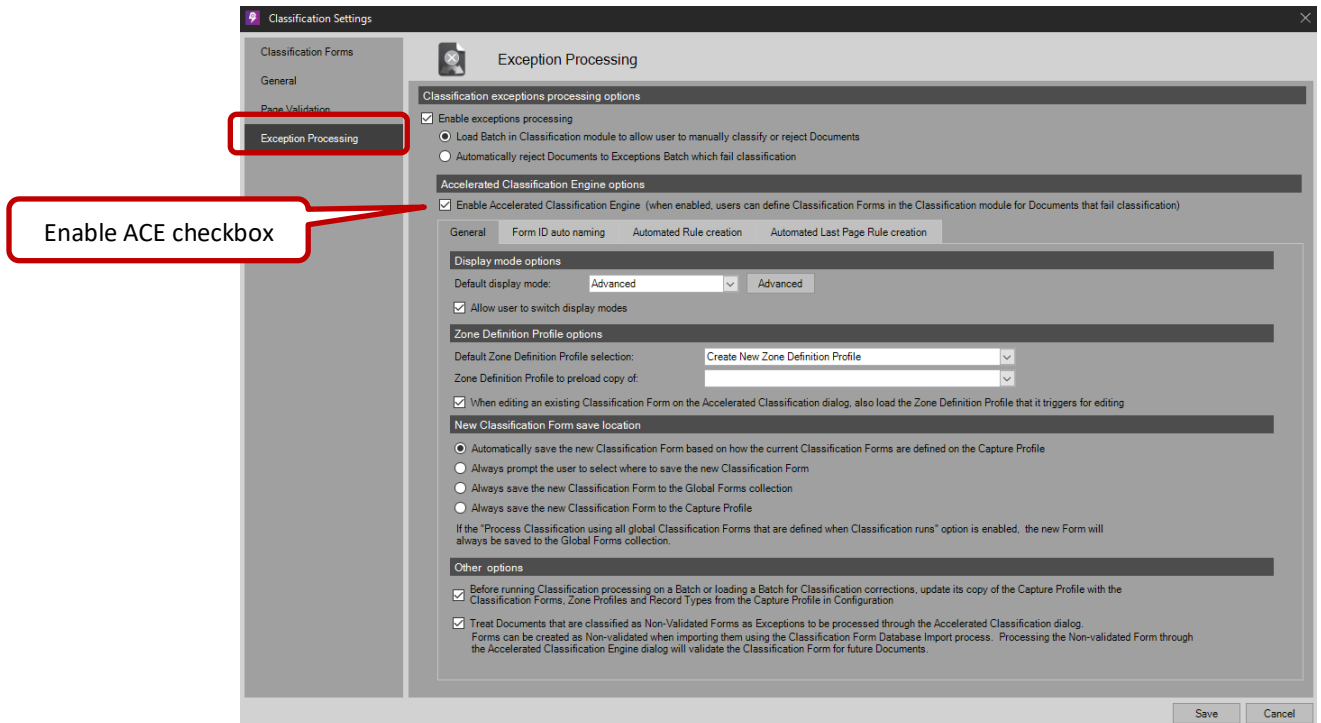
Requirements:

- Completed **00 - Installation Quick Start Guide**
- Ensure your computer meets the min. requirements
- Running version 7.6.x of PSIcapture or newer
- PSIcapture license with Classification module

What is the Accelerated Classification Engine (ACE)?

ACE is a feature of the Classification module that allows users to classify, as well as configure zone profiles for unclassified documents on-the-fly within a batch. This means that users can permanently modify the Capture Profile to which the batch belongs without opening the PSIcapture Configuration module. Once ACE has been used to classify a document, the Capture Profile will be able to successfully recognize documents of the same variety when they are encountered in future batches.
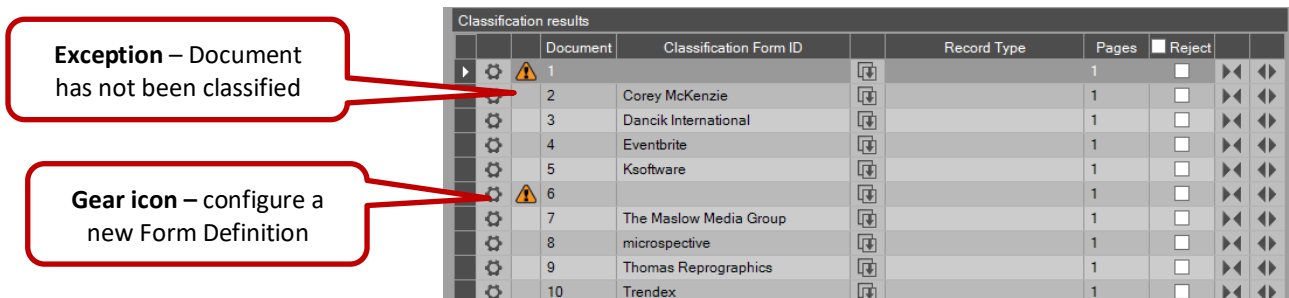
In order to enable ACE, open the Workflow tab and in Classification select the Exception Processing tab, and check the box to **Enable Accelerated Classification Engine**.

Enable ACE checkbox

This tab also contains other options that pertain to ACE. For instance, you can specify which features end users will be able to access using ACE. You can specify which Zone Profile will be used as the default for new forms. You can also specify whether newly created Classification Forms will be stored globally or only within the current Capture Profile.

When the Classification module fails to recognize certain documents, PSIcapture can be configured to display the list of all documents in the Classification Validation window. From this window, the user can manually classify unrecognized documents as belonging to any of the existing Form IDs within the Capture Profile. With ACE enabled, the user also has the option to configure brand new Classification Forms. This is done by clicking the gear icon within the Classification Results window.

Exception – Document has not been classified

Gear icon – configure a new Form Definition

After opening the ACE configuration, the user has access to two major functions. The first of those is the ability to add new Form Definitions, including Regular Expression rules that will be used to classify the documents. The process of configuring new forms from this tab of ACE is very similar to the process that you would use within the Classification workflow module configuration.

Name of the new form

Point-and-click text on the image to create the Regular Expression rule

The second major function of ACE is the capability to create a Zone Definition Profile at the same time, which will be associated with the newly created Form Definition. This means you can specify exactly where to find the data to populate each index field for that form. The process of configuring new Zone Profiles from this tab of ACE is very similar to the process that you would use within the Zone Configuration from the Indexing tab in the Configuration Wizard.

Options to create new Zone Profiles or Copy Existing Profiles

Another PSIcapture feature that can be leveraged within ACE is **Auto Zone Creation.** When this feature is activated for an index field, the software will attempt to *automatically* create a Smart Zone that covers the proper anchor label for that field, and attempt to create a Child Zone that encloses the piece of index data that you would most likely intend to extract.

Once the zones have been automatically created, all the user must do is confirm them, validating their placement and size. If the automatic zone placement turns out to be erroneous, the user can make manual adjustments to correct for this problem.

Auto Zone Creation is activated on a per-field basis from Indexing tab of the Configuration Wizard, on the **Zone** tab.

The **Auto Zone Creation Configuration** window allows the user to specify the type of data that PSIcapture will look for when determining where to place the zones.

Many of these options will default to the ideal settings based on the **Form Field Type** that is selected.

Type of field to look for (uses built-in Regular Expressions)

Enter any additional Custom Regular Expressions

The format of data to look for near the anchor (when placing the child zone)
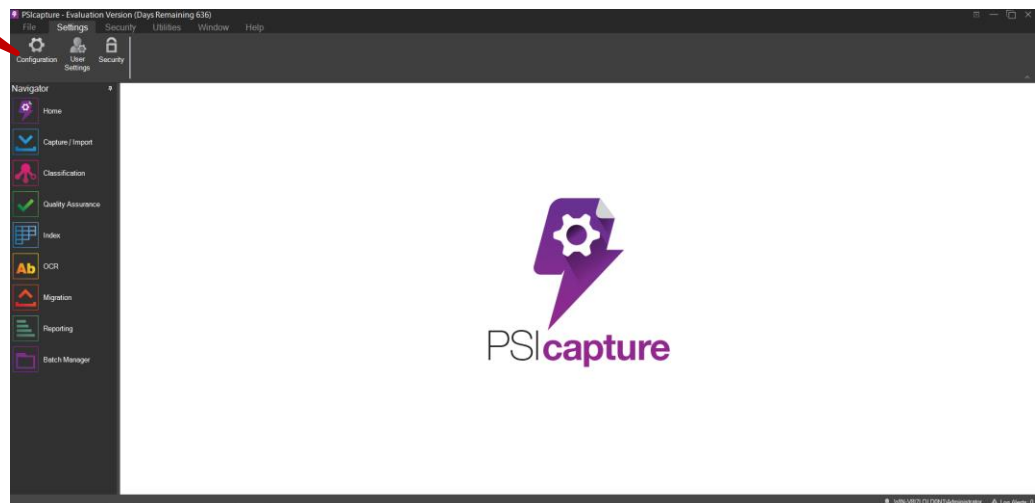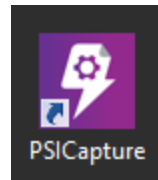
How will the Capture Profile function?

You will build a Capture Profile to process invoices, which will include the 4 standard header Index Data Fields. You will add a Classification module to the workflow (leaving all default settings), as well one vendor's invoice. Initially, this Capture Profile will serve only to recognize one single vendor. You will enable ACE and turn on Auto Zone Creation for the index fields. After you finish building the Capture Profile, you will import sample documents from several vendors, then use ACE to classify them and build zone profiles on-the-fly. This will effectively modify the Capture Profile so that it will be able to automatically process documents from those same vendors when they are encountered in future batches.

Creating your Capture Profile

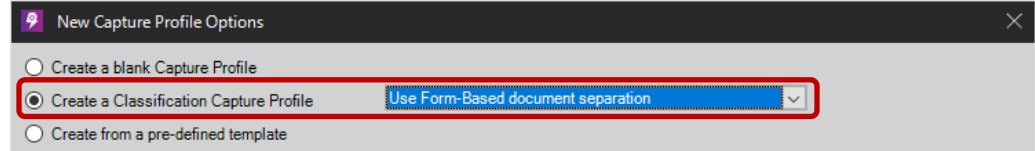First, you'll need to get a basic Capture Profile Configuration started. Open PSIcapture.

After PSIcapture opens you will start in the Configuration module to create the Capture Profile.
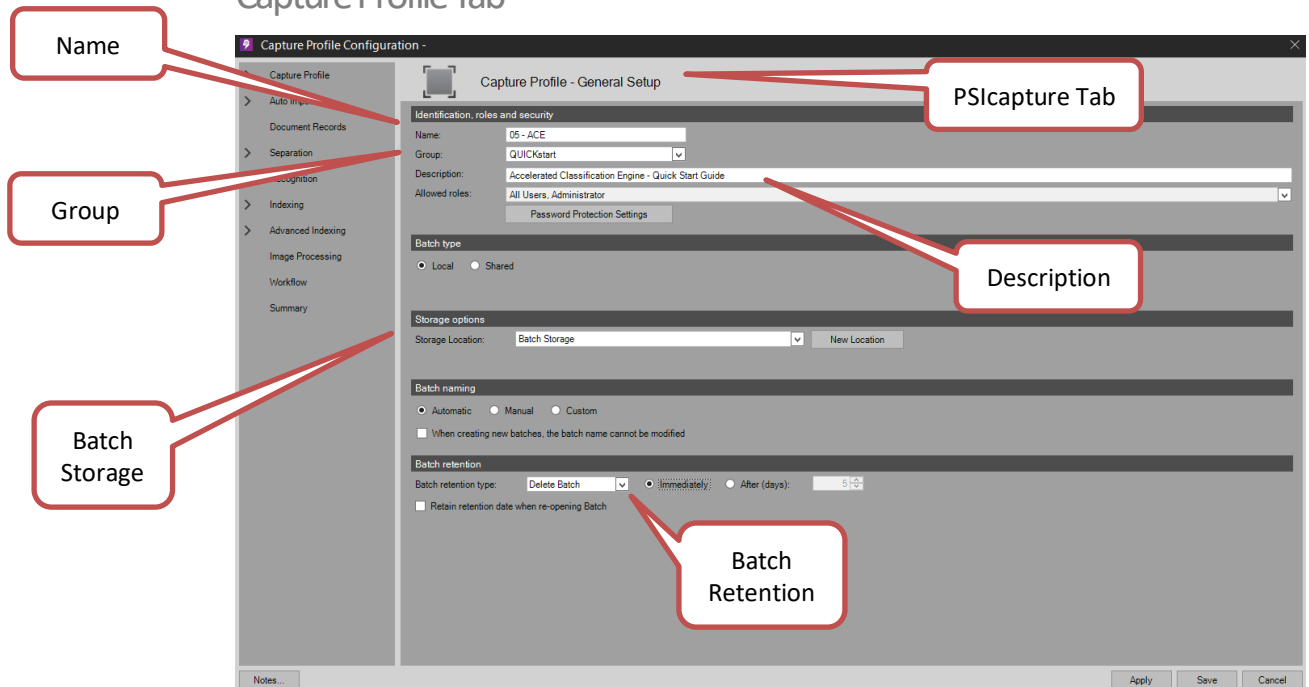
Once the Configuration module is open you can begin work on your first Capture Profile. Click the **Add** button on the right side of the window to proceed.

Change selection to **Create a Classification Capture Profile** and click **Create.**
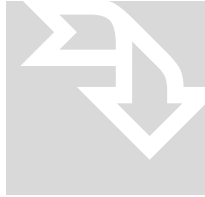

Capture Profile Tab

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 Profile Configurations can be left at their default settings. For additional information about features available on this tab please refer to the PSIGEN Help.
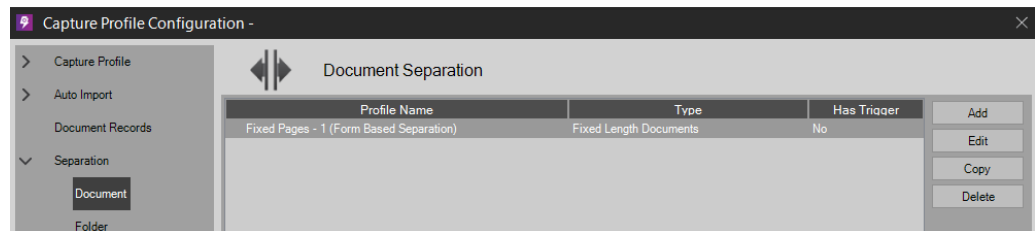
Document Records Tab

Again, leave all the default settings for Document Records, as they are not needed for this Capture Profile Configuration.

Separation Tab

When you first began creating this Capture Profile (see Page 7), you selected the option to **Create a Classification Capture Profile.** This change automatically configured the needed separation profile as Fixed Page – 1 (Form Based Separation).

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

Every invoice in the samples consists of only one page, so this is the only type of separation that will be required in this doctype. Typically, when using Classification, you will setup a Fixed-1 separation profile and let Classification handle the separation and combining of documents. No other changes are needed in this tab.

Recognition Tab

No barcode or other recognition engines are being used in this guide, so you can skip this tab.
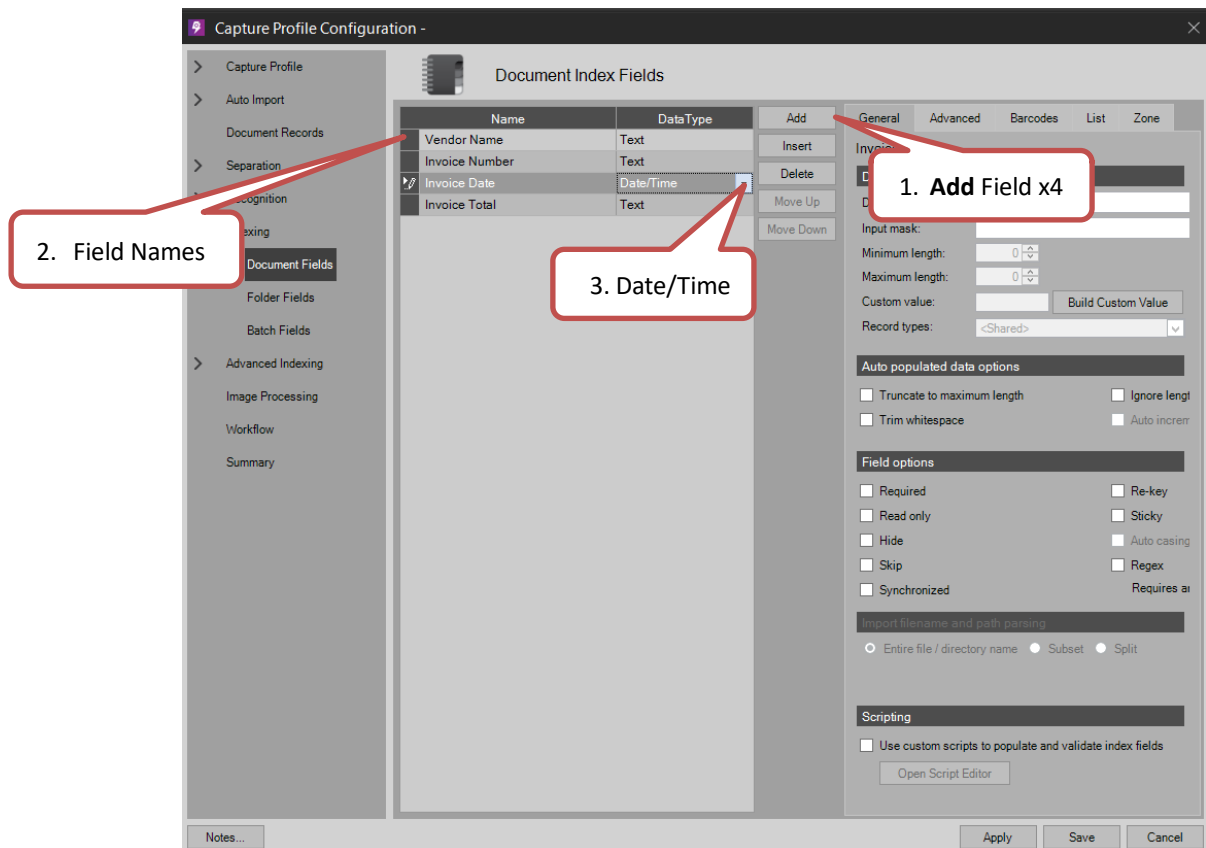
{This area is intentionally left blank, please continue to the next page}

Indexing Tab

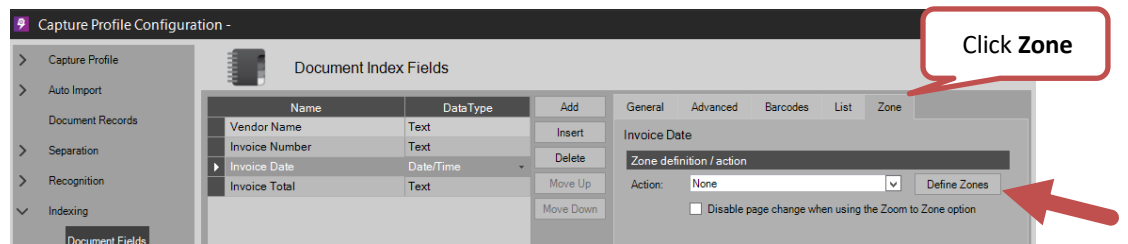Here you will configure the Document Index Fields.

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

- Vendor Name
- Invoice Number
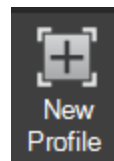- Invoice Date
- Invoice Total

Change the **DataType** dropdown to the right of the Invoice Date field to **Date/Time**.
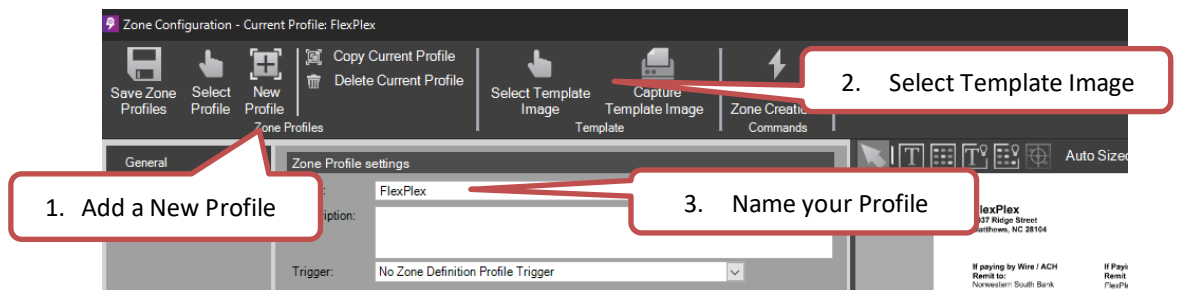
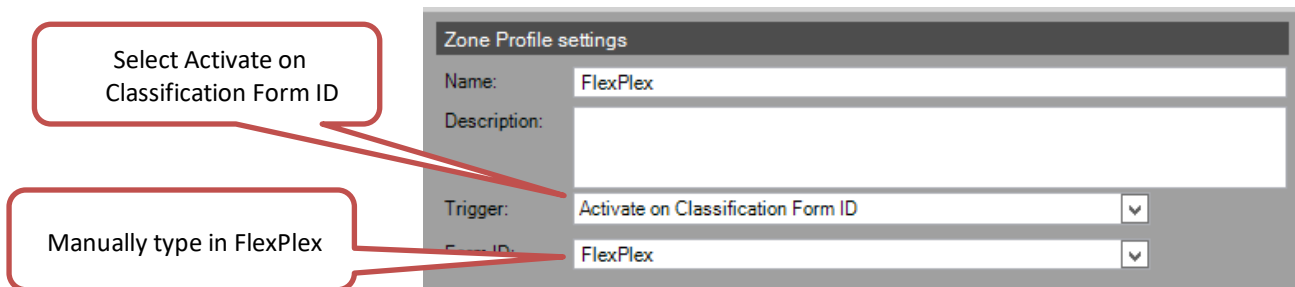Click on the **Zone** tab and click the **Define Zones** button.

At this point, we are going to setup one profile for our first vendor. This will give you a place to create your Smart Zones, to be used with Auto Zones. To add a Blank Zone Profile, click the New Profile icon at the top of the Zone Configuration window.

Then click the toolbar icon to Select Template Image. Navigate to the folder where you saved your files and Select **05 – ACE Guide Samples.tif** as your template Image. Click Yes and select Page 1 (FlexPlex). Name the Profile **FlexPlex**.

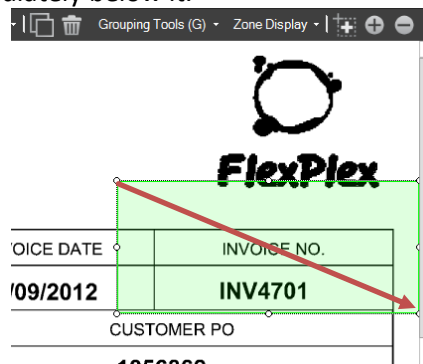Next, we will associate this Zone Profile with a Classification Form ID, which we will be building later.

Next you will create 3 **OCR Smart Zones** for the required Index Fields. Click the Smart Zone tool to draw the zone.

The first Smart Zone that you create will be drawn around the **Invoice Number**. This label will serve as a good **anchor** for the Smart Zone, because the Invoice Number itself is positioned immediately below it.
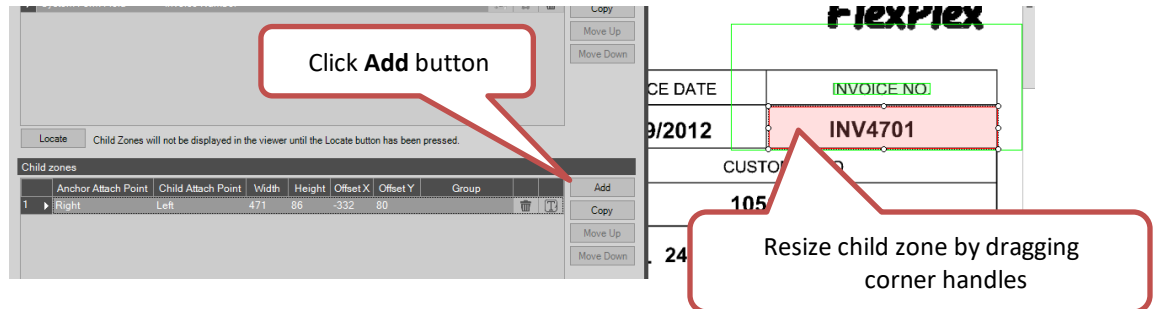

Upon releasing the mouse button, the Smart Zone Configuration window will appear. Select **Invoice Number** from the **Form Field Types** dropdown, then click **Add**, then **Locate.** An alert notifies you that the anchor was located.

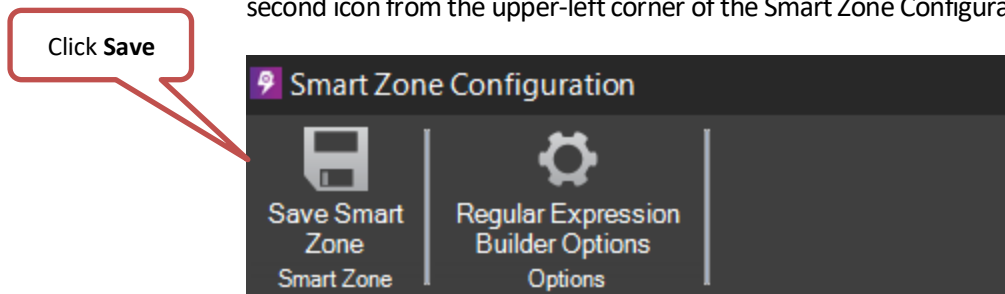Next, you will add a child zone using the **Add** button in the middle of the Smart Zone Configuration window. Reposition and resize the child zone so that it overlaps the entire box around the Invoice Number.
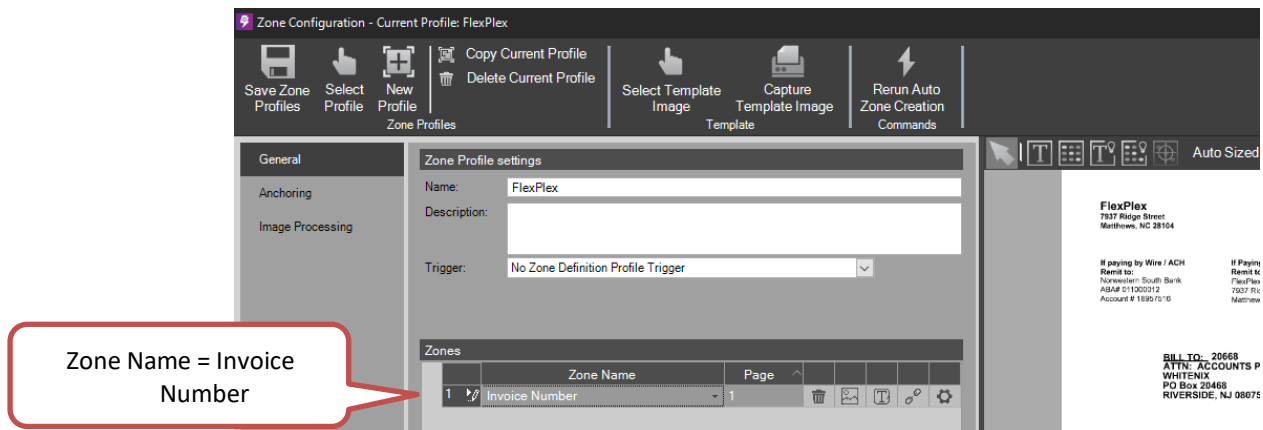

Now that you have finished building your first Smart Zone, click the Save button, the second icon from the upper-left corner of the Smart Zone Configuration window.

This will take you back to the Zone Configuration window.

Now you will assign the name **Invoice Number** to your newly created Smart Zone.
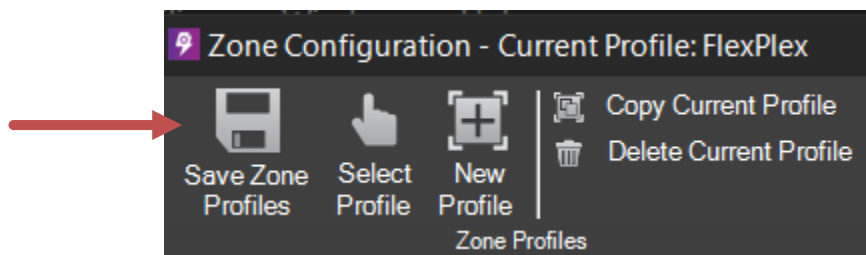
Now repeat the process for the other 2 Smart Zones (Invoice Date and Invoice Total).

1. Draw the Smart Zone.
2. Select the proper Form Field Type and click Add.
3. Click the Locate button to find the anchor on the template image.
4. Click the lower-left Add button to create a child zone.
5. Re-position and re-size the child zone appropriately.
6. Save out of the Smart Zone Configuration window.
7. Name the zone the same as its anchor label.

Once all your Smart Zones have been created, the list of Zone names will look like this:

Now **Save** out of the Zone Configuration.

For each index field _except_ Vendor Name, go to the **Zone** tab and select the **Action** as **OCR** from the dropdown. Choose this option because that is the type of recognition engine that we will be using to read the zones.

Then ensure that each Index Field is mapped to the proper zone.

1. Select each of the bottom 3 index fields

☆ Note:
Index Fields will be automatically mapped to the proper zones if their names match exactly.

2. Choose **OCR** Action

3. Map to proper zone

☆ Note:
Don't forget to complete these steps 3 times (once for each index field that is mapped to a zone).

For each of the **3** bottom zones, you will also select the option to **Auto Create Zones for new Zone Definition Profiles**. Then you will click the **Setup** button to specify the type of data that the automatic zones will adhere to.

From the **Auto Zone Creation Configuration** window, select the proper **Form Field Type** for the field you are currently configuring, then click **Save**. Make sure to do this for all three Invoice fields.

Check box, then click **Setup**

Select proper Form Field Type

Advanced Indexing Tab

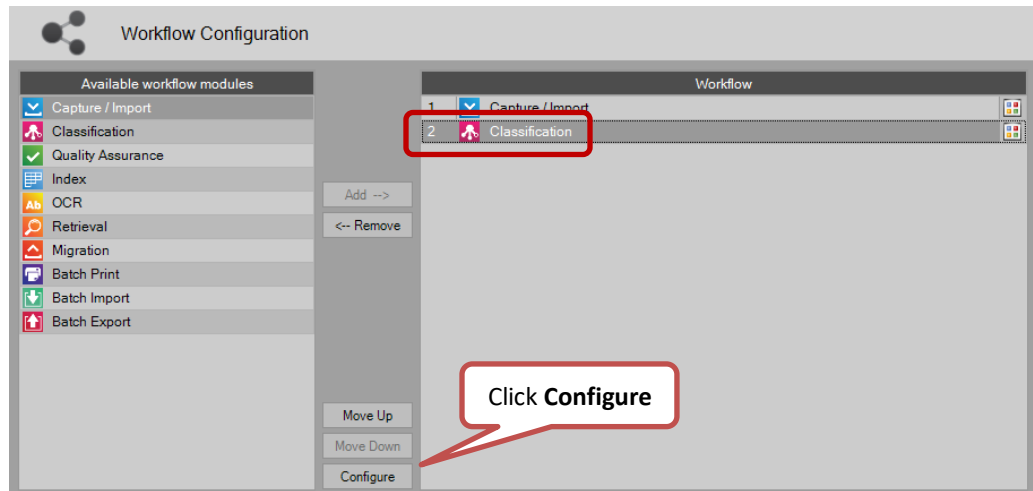Since no changes are required here, you can skip this tab.

Image Processing Tab

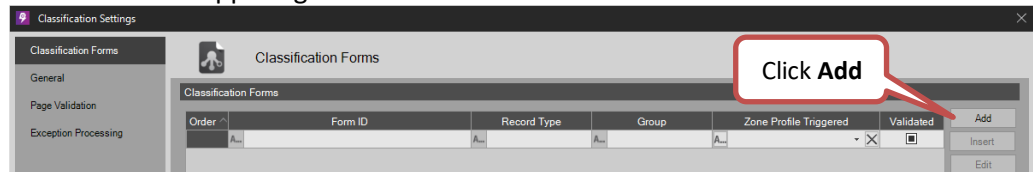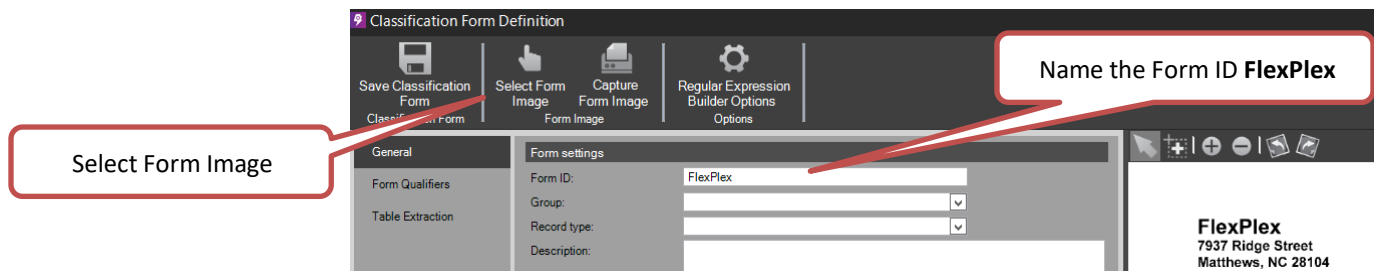Since no changes are required here, you can skip this tab as well.

Workflow Tab

In this tab you will build the PSIcapture workflow. Select the **Classification** workflow module from the list on the *right* and click the **Configure** button to edit this module.
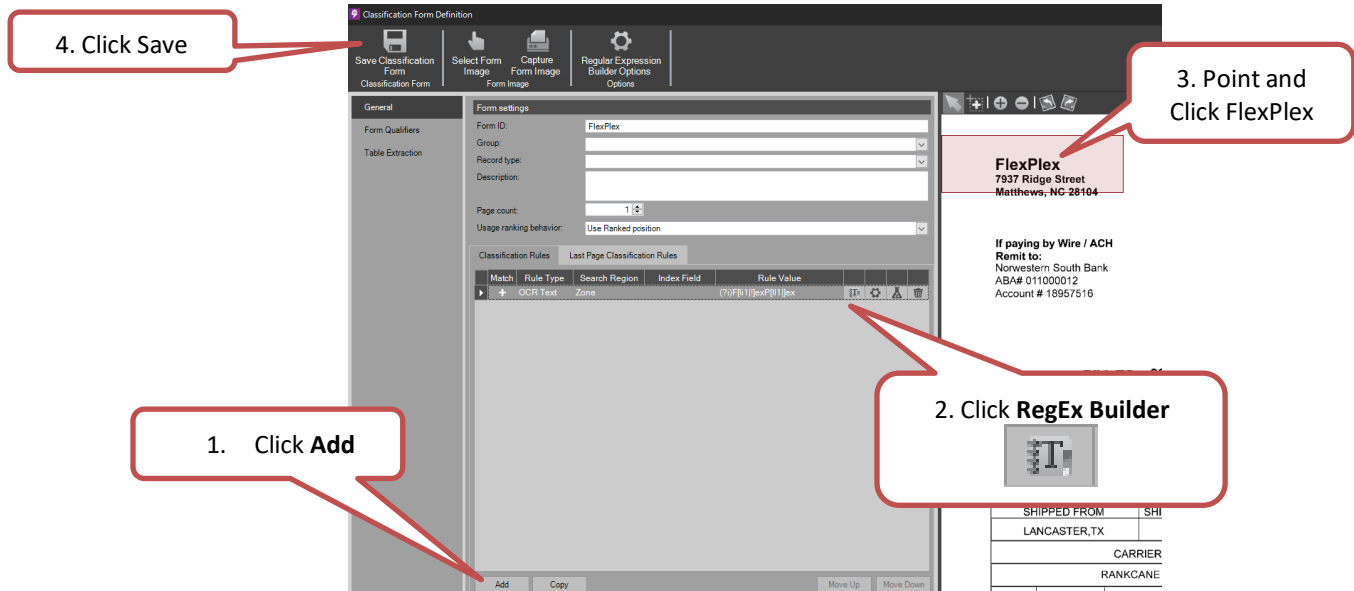

We first need to setup a Classification Form for our vendor, FlexPlex. You will want to click **Add** in the upper right-hand corner.

Name your Form ID **FlexPlex**. Click the button to load in a sample of your first image type so that you can write a rule to identify it. Navigate to the folder where you saved your files and load page 1 of **05 – ACE Guide Samples.tif** as your template image. When prompted, select **Yes** to load only one page of the multi-page document as your template.
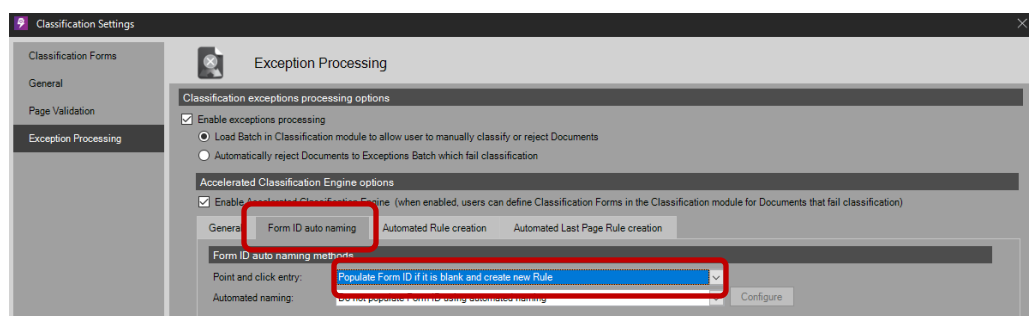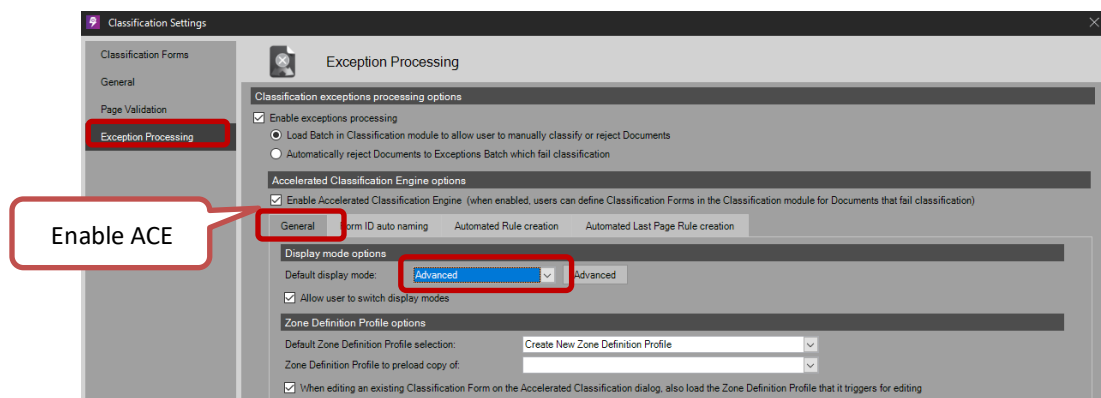
With the template image loaded, you will now add a rule to recognize it. Click **Add**, and then click the **Regular Expression Builder** button. Once rule is created, click **Save**.

Select the **Classification Settings – General** tab. You will make one change near the middle of the window. In the **Index Field to populate with Form ID** dropdown, choose **Document.Vendor Name**. This will take the value of the Classification Form ID that is recognized for a given document and use it to populate the Vendor Name index field for that document.
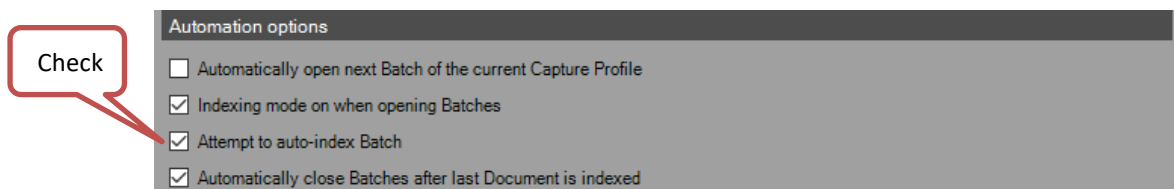

On the **Classification Settings - Exceptions Processing** tab, you will check the box to **Enable Accelerated Classification Engine** and set the **Defualt Display mode to Advanced**. This will give you access to ACE functionality once you start actually processing documents. You will also want to set the Form ID Point and click entry drop down to "Populate Form ID if it is blank an create new Rule". Now click **Save** to close the Classification Settings window.
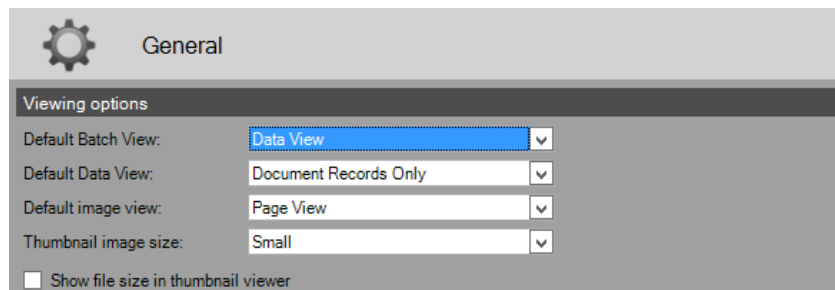
Enable ACE

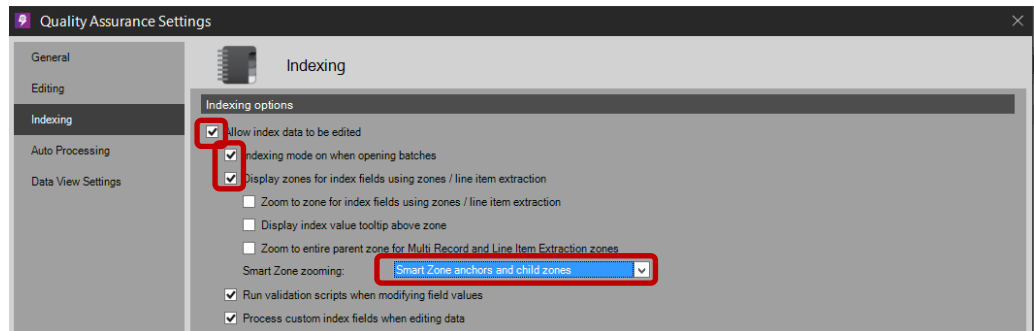Next you will add 2 additional modules to complete the workflow.

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 Options section. This will auto-index each document without requiring user interaction from the Index Module. Click **Save**.
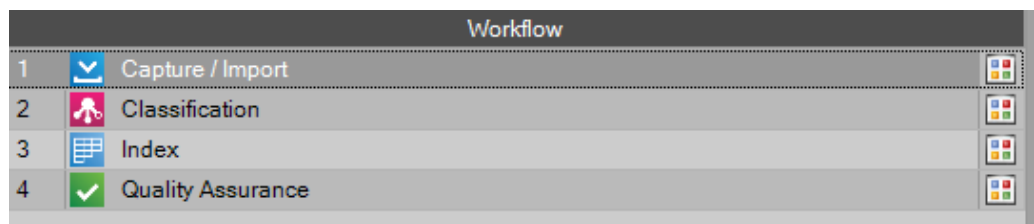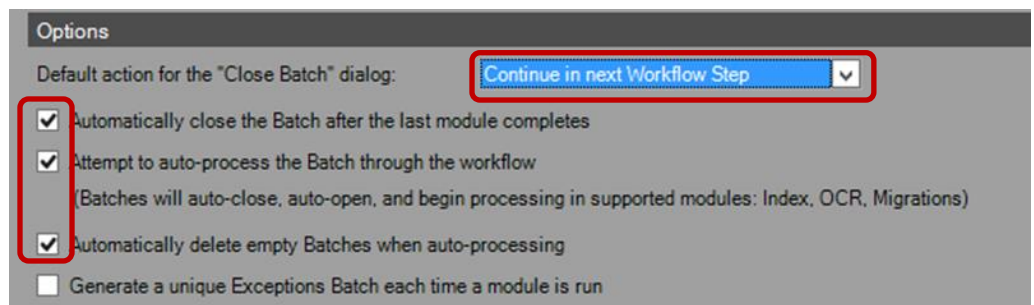
Check

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**.
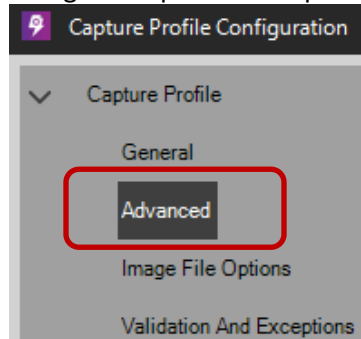

Next, in the Workflow options (at the bottom of the tab), check the top 3 checkboxes and change the dropdown to match below. This will automate the workflow, so that the Capture Profile will process the documents without requiring user interaction after the Capture/Import module.
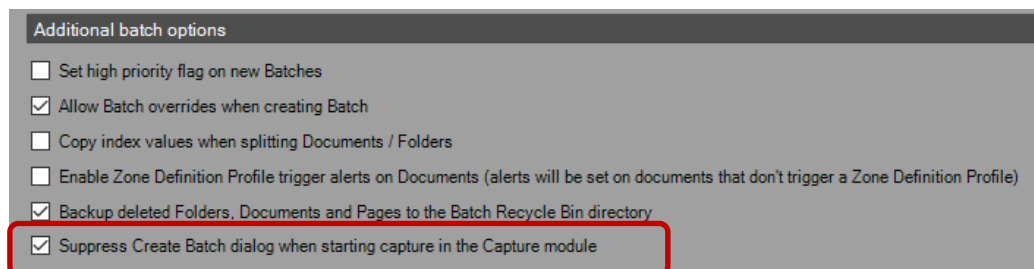

Your workflow on the right will now consist of 4 modules.

Suppress Create Batch Screen

When running a batch through PSIcapture 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.
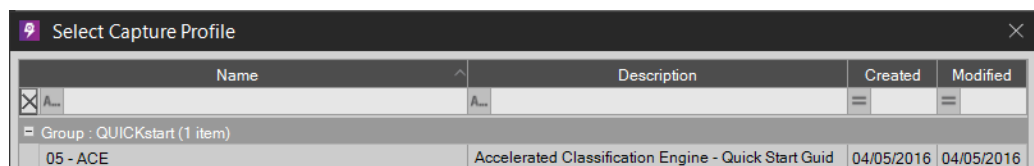

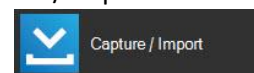In the Additional batch options section, we want to check the 6th checkbox titled "Suppress Create Batch dialog when starting capture in the Capture module"
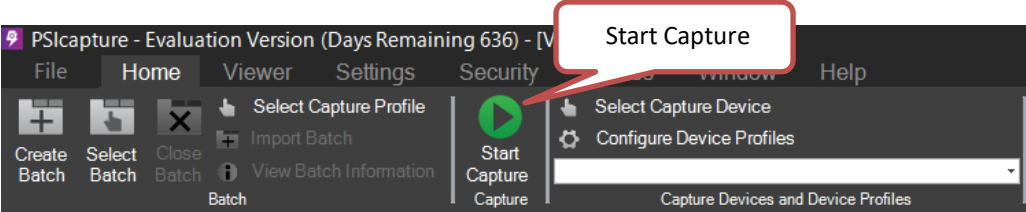

Run the Capture Profile

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.

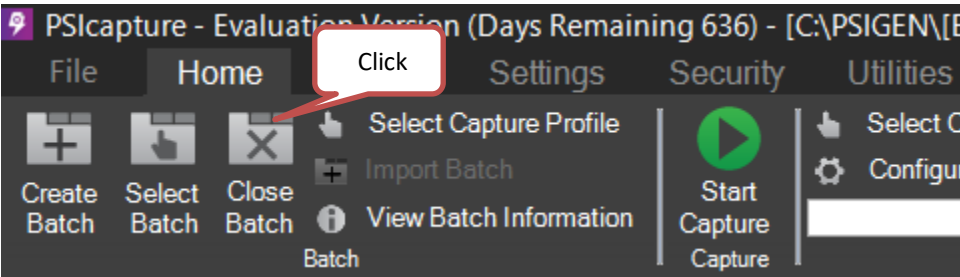Select the **05 – ACE** Capture Profile, Click **OK.**

Using the File Browser window, navigate to where you saved the files for this guide. You will select the folder on top and then select **05 - ACE Guide Samples.tif** from the lower pane. Highlight the file and select the **Start Capture** button or Control+S to begin importing the documents into PSIcapture.
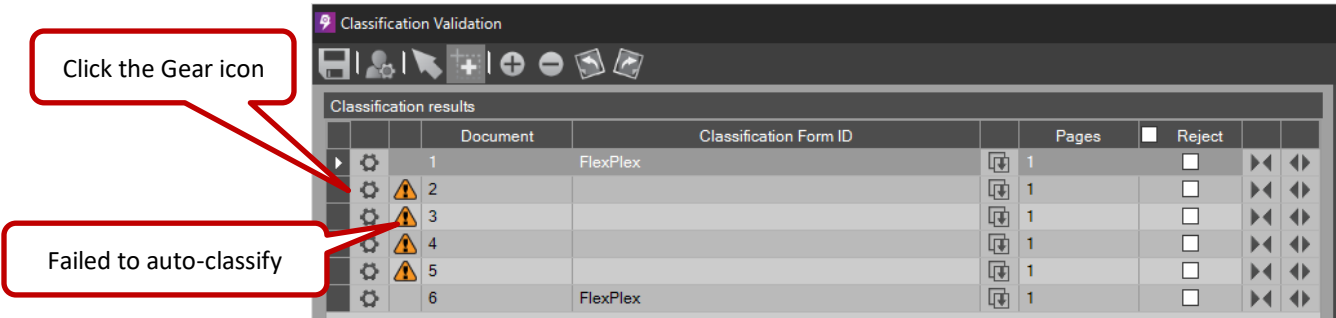
When the file has finished importing, you will click the **Close Batch** button.

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

The Classification module will not be able to classify all the documents, because you have not created any Form Definitions. The **Classification Validation** window will appear, showing four of the documents as Exceptions. Click the Gear icon to the left of the first failed document to build a Form Definition and a Zone Profile.

★ Note:
The value that you enter for **Form ID** will later be used to populate the Vendor Name index field.

The **Accelerated Classification Engine** window now appears. First, you will assign a **Form ID.** Because this document belongs to the vendor **GoldenAce**, that is the name you will enter.

The Regular Expression builder tool is already activated by default, and allows you to point-and-click any string of text on the page that can be used to classify this document. Choose something that will appear uniquely and consistently on

documents that belong to this vendor. This does not necessarily need to be the vendor's name. For this example, we can use the vendor's address instead.

The text that you point-and-click will automatically be converted to a Classification Rule that is flexible enough to match even if the OCR engine happens to misread certain characters.

Next you will click the **Zone Profile** tab so that you can specify the zones from which to extract index information on FlexPlex documents.

Because you are using Auto Zone creation, you will notice that all the Smart Zones are already built and positioned automatically, and you are not required to draw them from scratch.

Because the zones you are using are associated with a specific Form Field Type, the Auto Zone Creation knows what to search for on the page that is most likely to represent the correct anchor label. For instance, in order to position the **Invoice Number** Smart Zone, the software searches the entire page for any permutation of the words **Invoice Number**. When that string is located, an oversized Smart Zone will be automatically drawn around it. Next, the software will automatically look around the anchor for a string of text that matches the expected format of the index value

itself. In the case of Invoice Number, that would be any string of text that contains at least 1 digit. When that string of text is located, a child zone will be automatically drawn around it.
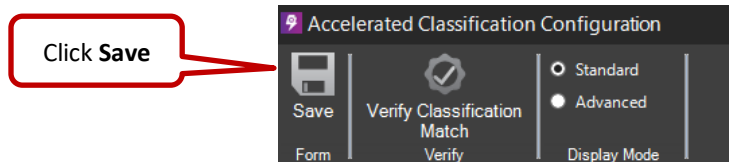
Double-click each green Smart Zone on the page to validate that both the anchor and its child zone are positioned correctly.

Once you have validated all 3 zones for the GoldenAce document, you are finished using ACE to configure this vendor. Click the **Save** button to close out of the Accelerated Classification Engine window.
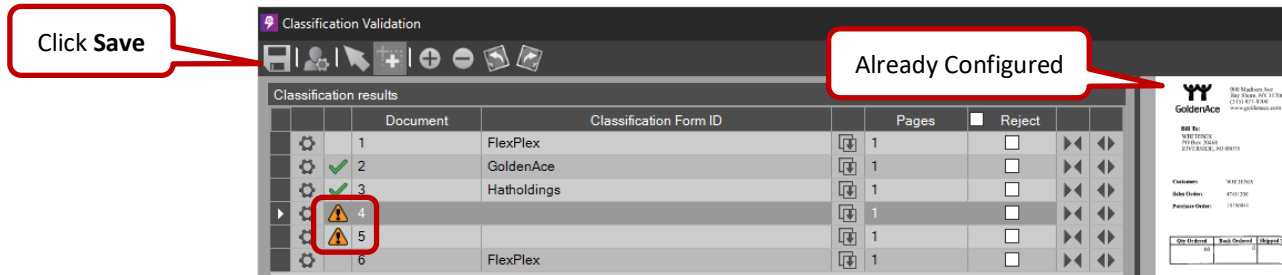

Now repeat the ACE process for Document #3, which represents invoices from the vendor Hatholdings.
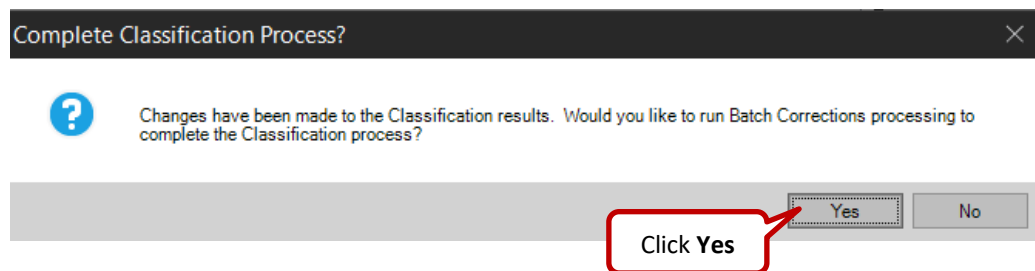
1. From the Classification Validation window, click the Gear icon to the left of the unclassified document.
2. Type the vendor's name as the Form ID.
3. Point-and-click some string of text on the page (e.g. the vendor's address or phone number) that can be used to classify the document.
4. Click the Zone Definition Profile tab.
5. Validate each auto-created Smart Zone and its child zone by double-clicking the zone on the page. Resize child zones if needed
6. Save out of the ACE window once all zones have been validated.

From the Classification Validation window, notice that Documents #4 and #5 are simply different invoices from the same vendors that you have already configured.
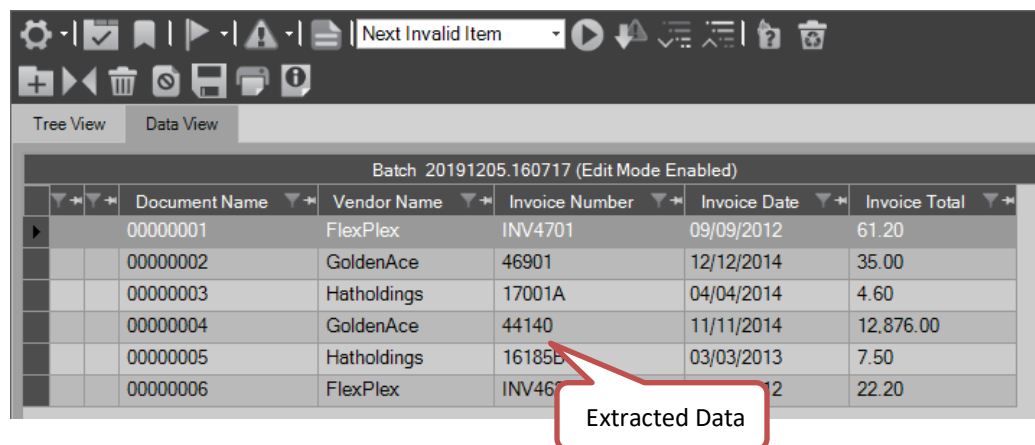
You do not need to use ACE for these documents. Simply click the **Save** button instead.

Click **Save**

Already Configured

Click **Yes** to confirm your changes and Complete the Classification Process.

Click **Yes**

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

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. PSIcapture knows which vendor each invoice belongs to, and where to find the data it's looking for on each vendor's document. Notice that data has been extracted for all six of the documents, not just the first three that we used ACE to configure.

When you are finished reviewing the data, close the batch by clicking the **Home** button.

Next, select the Capture / Import module again to create a brand-new batch. Import the exact same set of documents again. This time, when you pass the batch to the Classification module, it will not stop at the Classification Validation window. This is because the Capture Profile is now configured to classify and extract all the data from these vendors' documents.

CONGRATULATIONS!!
You've just configured and run a Capture Profile using the Accelerated Classification Engine.