Grooper | ACE

Architect · Consultant · Engineer

ARCHITECT TRAINING MANUAL VOL.1

OVERVIEW AND CONCEPTS

13900 N. Harvey Ave ● Edmond, OK ● 73013 ● 405-507-7000

www.grooper.com

Grooper | ACE

Architect · Consultant · Engineer

TABLE OF CONTENTS

Grooper™

In this Overview and Concepts introductory guide to **Grooper**, we will cover making a simple batch process from beginning to end. We'll take a rather linear, and simplified approach to build a foundation of concepts and general understanding. This will help establish the general mindset one should develop when working with their own documents.

Raw data comes in, and well organized and easily accessed information comes out. For this magic to happen, there are Five Phases of a the **Grooper** process that should be understood:

- **Phase 1 - Acquire**
- **Phase 2 - Condition**
- **Phase 3 - Organize**
- **Phase 4 - Collect**
- **Phase 5 – Deliver**

PHASE 1 – ACQUIRE

Before anything can happen, content must be put into **Grooper**. There are many methods by which **Grooper** can acquire content, but for the purposes of this document, we'll be scanning. Before we scan, we need to have a place for our documents to go so they can be processed, and we call that container a batch.

WHAT IS A BATCH AND WHY WOULD I WANT TO PROCESS IT?

A batch in **Grooper** is simply a container for a collection of raw pages and images in one place. This raw input is useless in this state, but that's where the Process comes in. The careful organizing and defining of these raw pages and images into documents (a very specific, but important term in **Grooper**) with appropriate indices is the true purpose of the batch process.

Batch processes can vary greatly in complexity, which all depends on the overall goal and destination of the documents, and how much or how little human interaction is required. Ultimately, the less a person needs to be involved in repeated tasks a computer can perform, the better; and taking the time to fully plan out and properly structure a batch process can save unprecedented amounts of time and money.

BATCH PROCESS DESIGN AND UNDERSTANDING

We could talk about concepts and ideas all day, and honestly, you'd forget it all by tomorrow (or by the next time you ate something, if we're honest.) So, for this all to make sense and stick with you, let's just get into **Grooper** and make a batch, get it processed, and go from there.

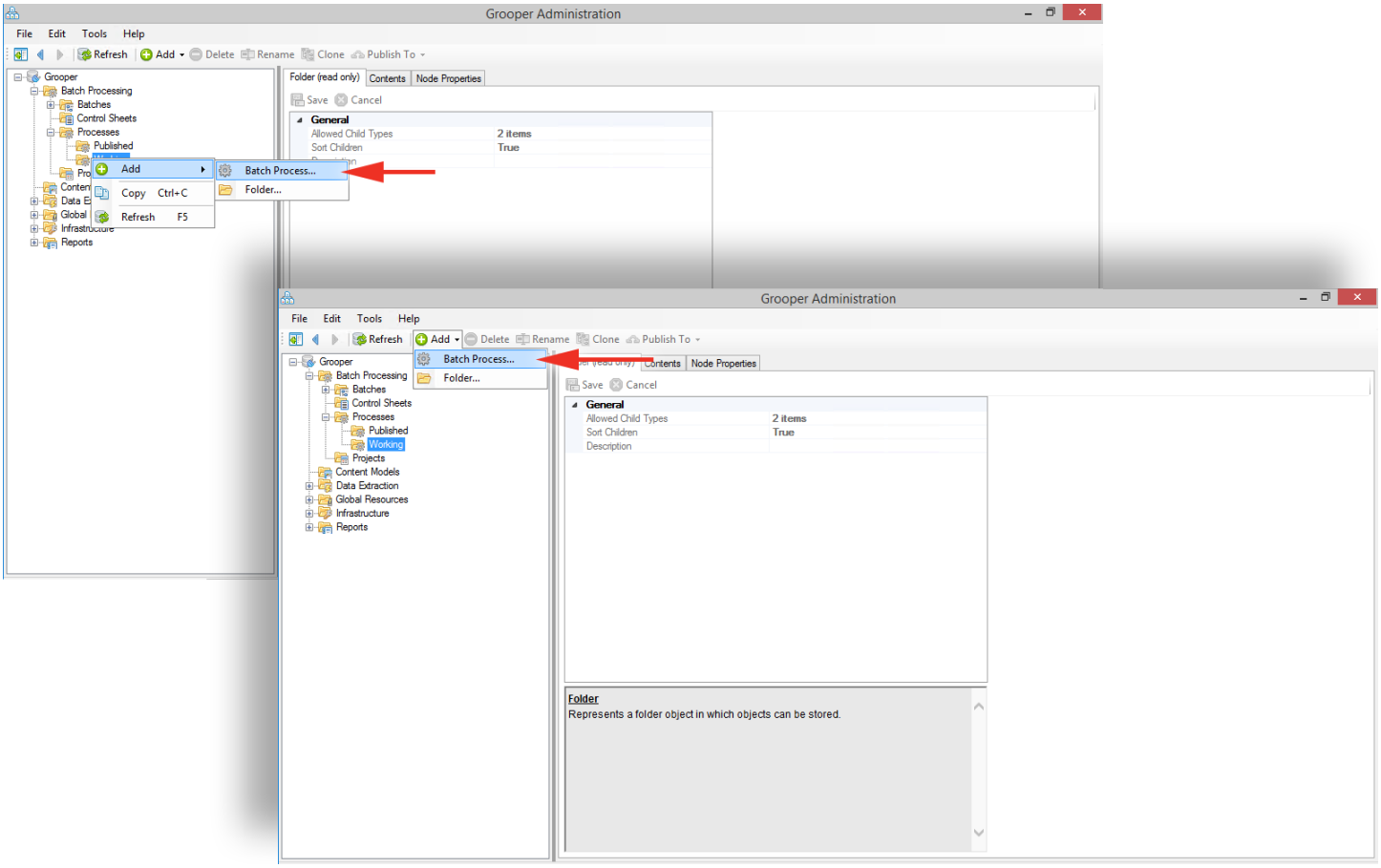
This document is assuming you have already installed and successfully configured and licensed **Grooper**.
Click here for more information.

The work that will be done in this document will rely on set of documents that can be obtained by the following link.
Grooper A.C.E. – Architect Training Vol.1 - Over and Concepts - invoices

BATCH PROCESS CREATION

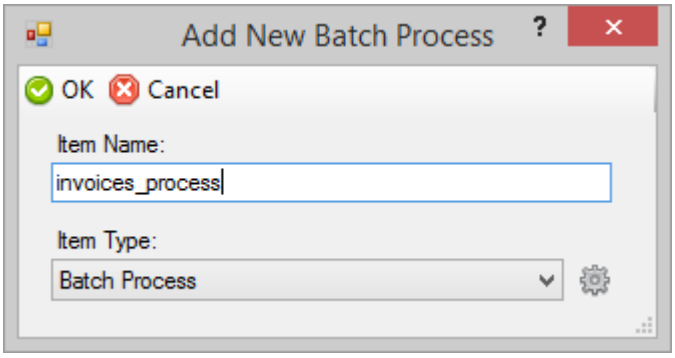
STEP 1 – ADD BATCH PROCESS

In Grooper Administration, expand the node tree to Grooper - Batch Processing - Process and select the Working node. You can either right-click this and highlight Add and select Batch Process, or select Add at the top and then select Batch Process from the dropdown.



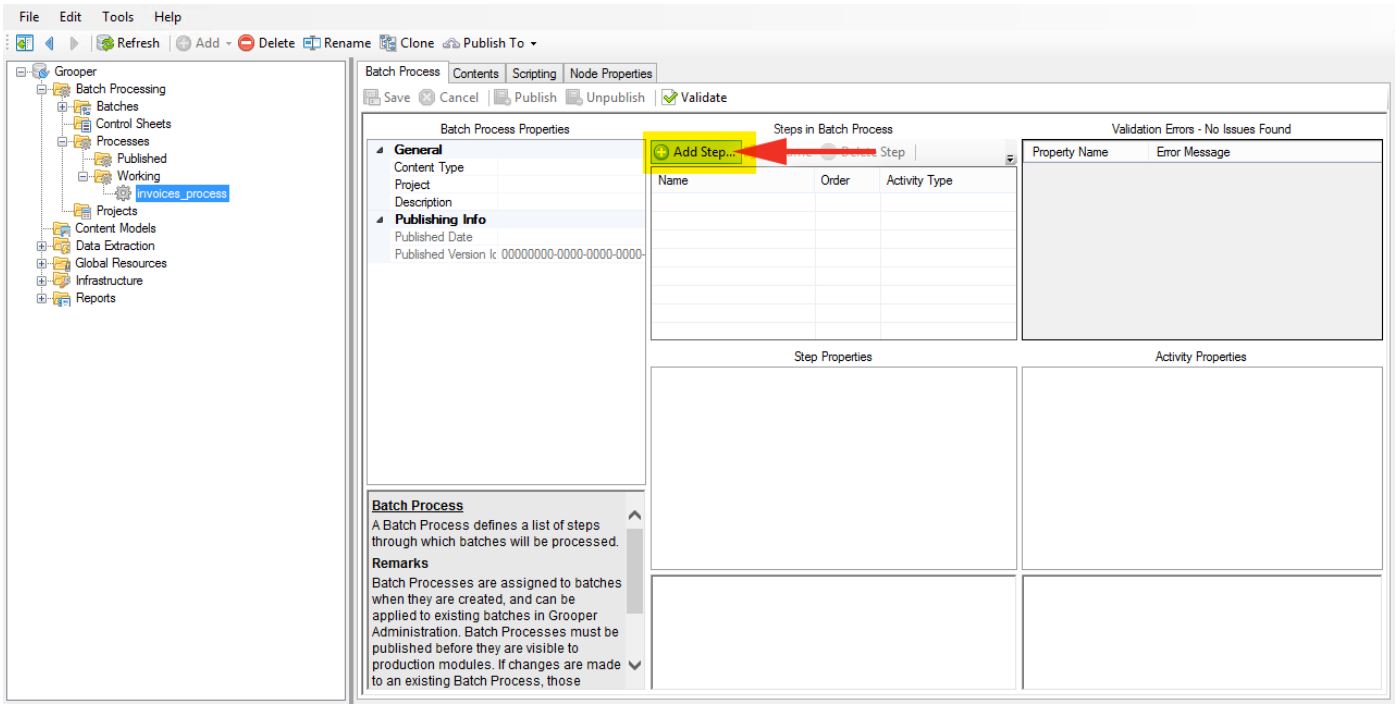
STEP 2 – ADD NEW BATCH PROCESS WINDOW

The Add New Batch Process window will appear. Name the batch process invoices_process, and click OK.



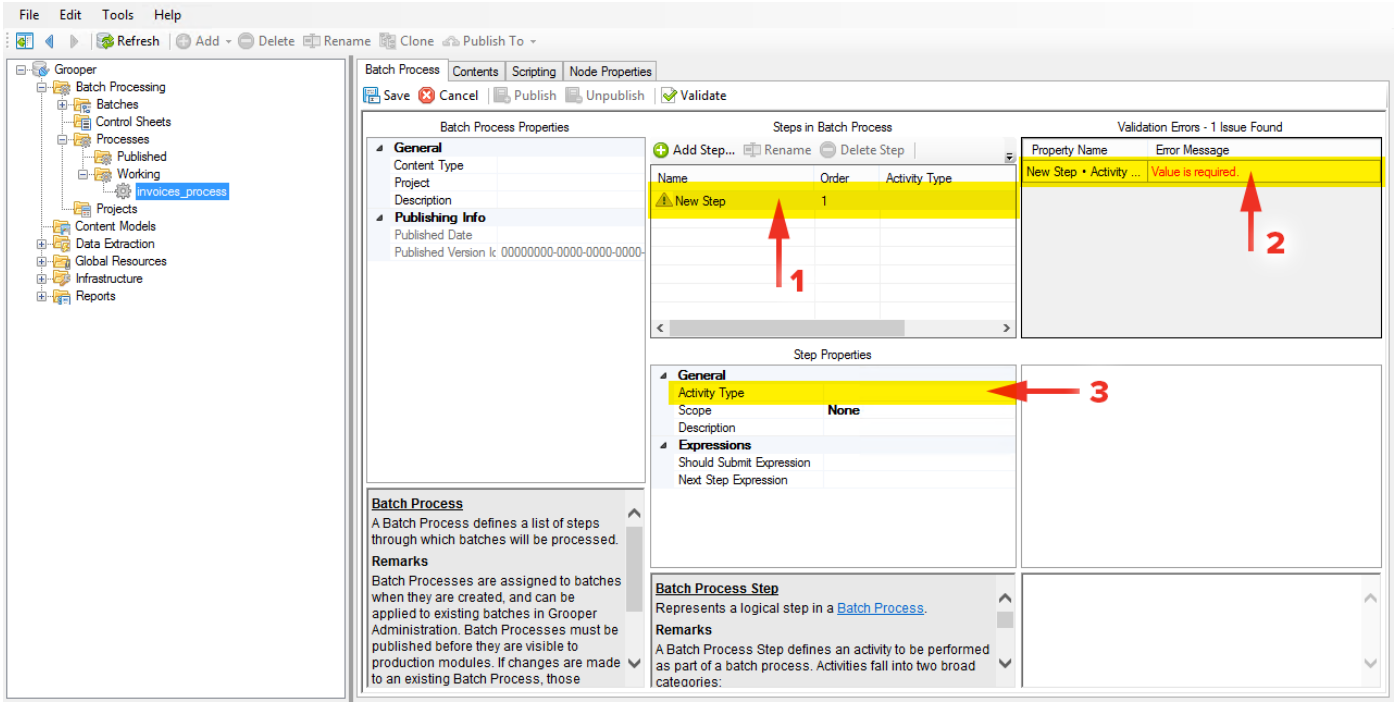
STEP 3 – ADD BATCH PROCESS STEP

This will make an empty batch process, so we'll need to add our first Step. Click the Add Step button.



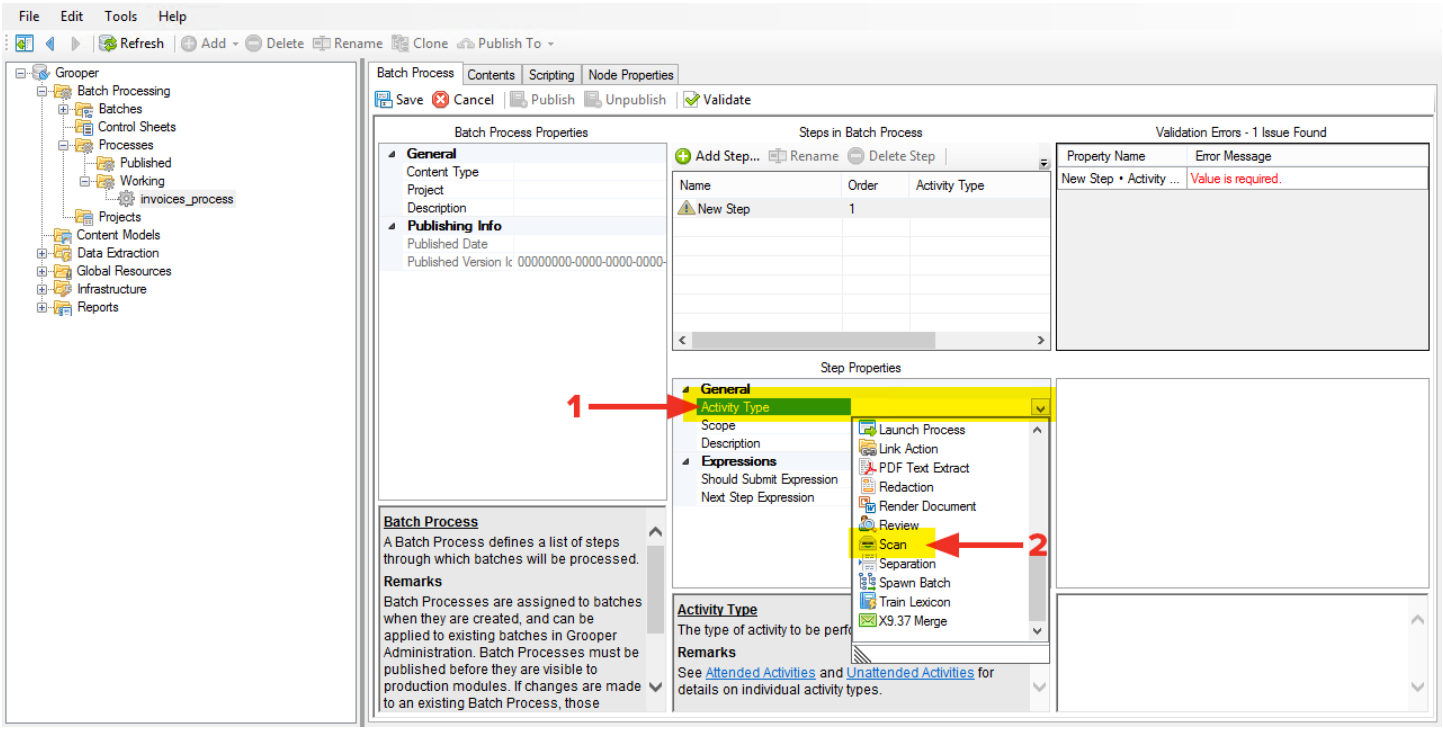
STEP 4 – NEW BLANK STEP

(1) A step labeled New Step will be added to the Steps in Batch Process List View. **(2)** This step will have a warning icon, as well as list an issue in the Validation Errors List View. **(3)** This is due to the Activity Type being blank.



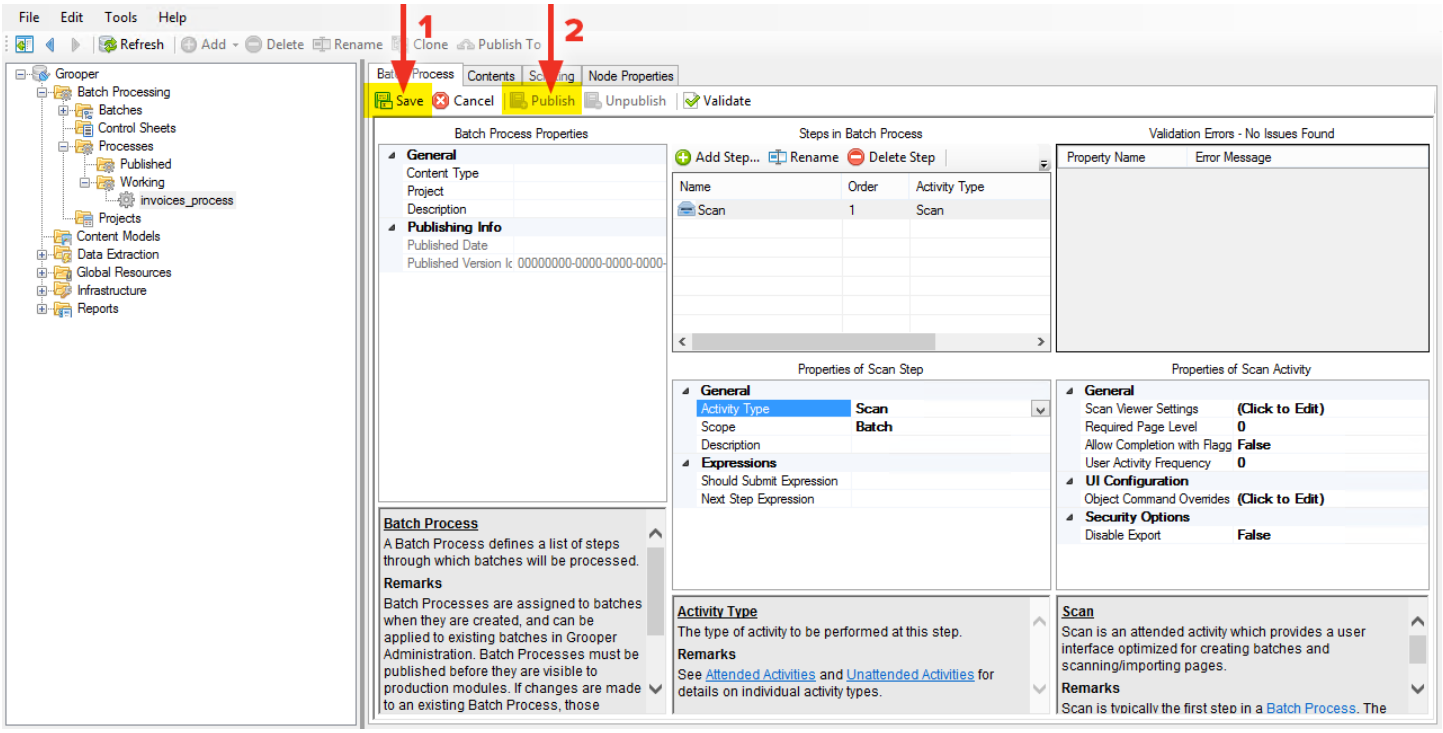
STEP 5 – SET BATCH PROCESS STEP ACTIVITY TYPE

(1) Select Activity Type, and a dropdown arrow will become available. **(2)** Scroll down to Scan and select it.



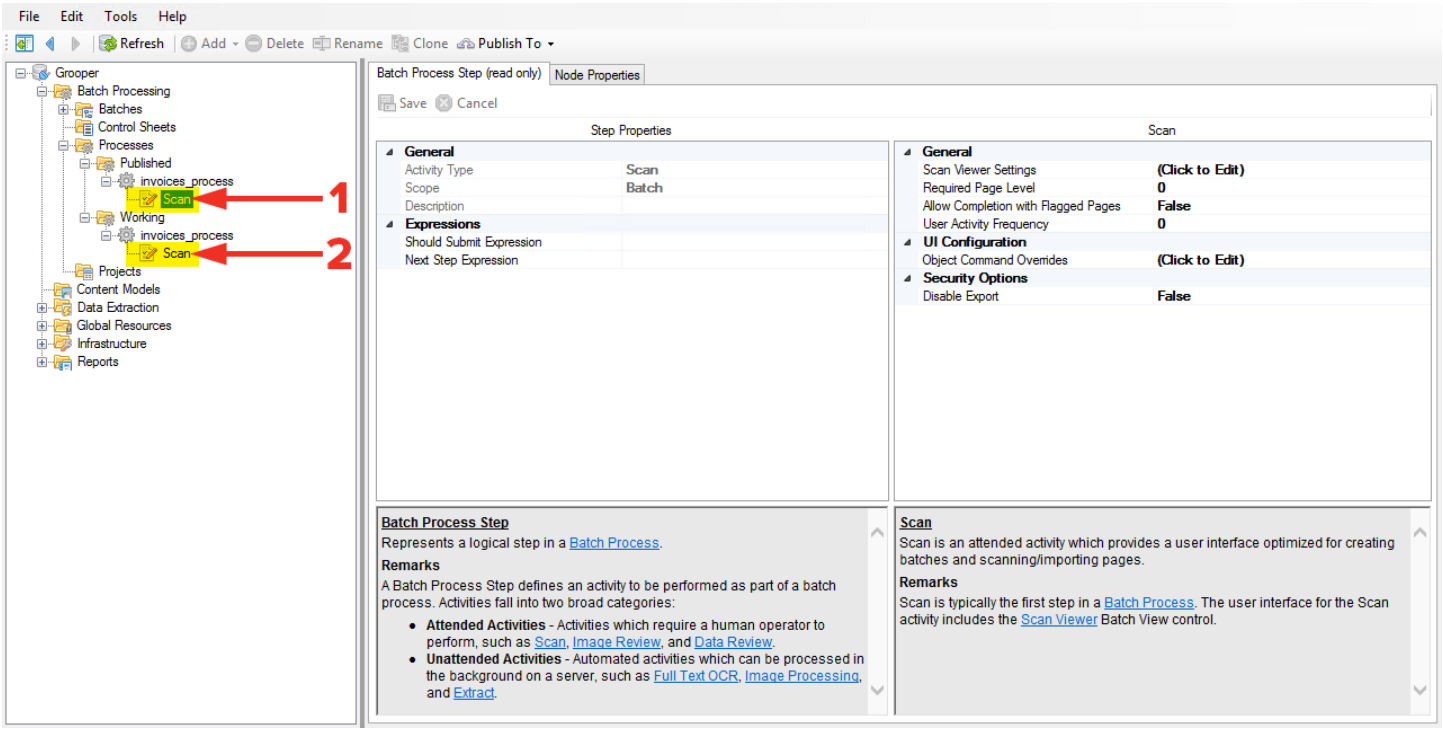
STEP 6 – SAVE AND PUBLISH BATCH PROCESS STEP

We now have a scan step with settings that can be adjusted, but we'll leave them defaulted for now.
(1) Click the Save button. With the batch saved, **(2)** click the Publish button.



STEP 7 – WORKING PROCESS VS PUBLISH PROCESS

Clicking the Publish button creates a copy of the working process. **(1)** Working processes are not exposed to production until published. **(2)** Published processes cannot be edited. If changes to the Published process are desired, you will need to adjust the working process, and republish.



SCAN

WHAT IS SCANNING?

The above question might seem silly, but it's worth covering. Most scanners use technology that you would find in a modern digital camera, so you are literally taking a picture of a page (or pages) and storing them somewhere on a computer. It's in this digital form that **Grooper** can process them.

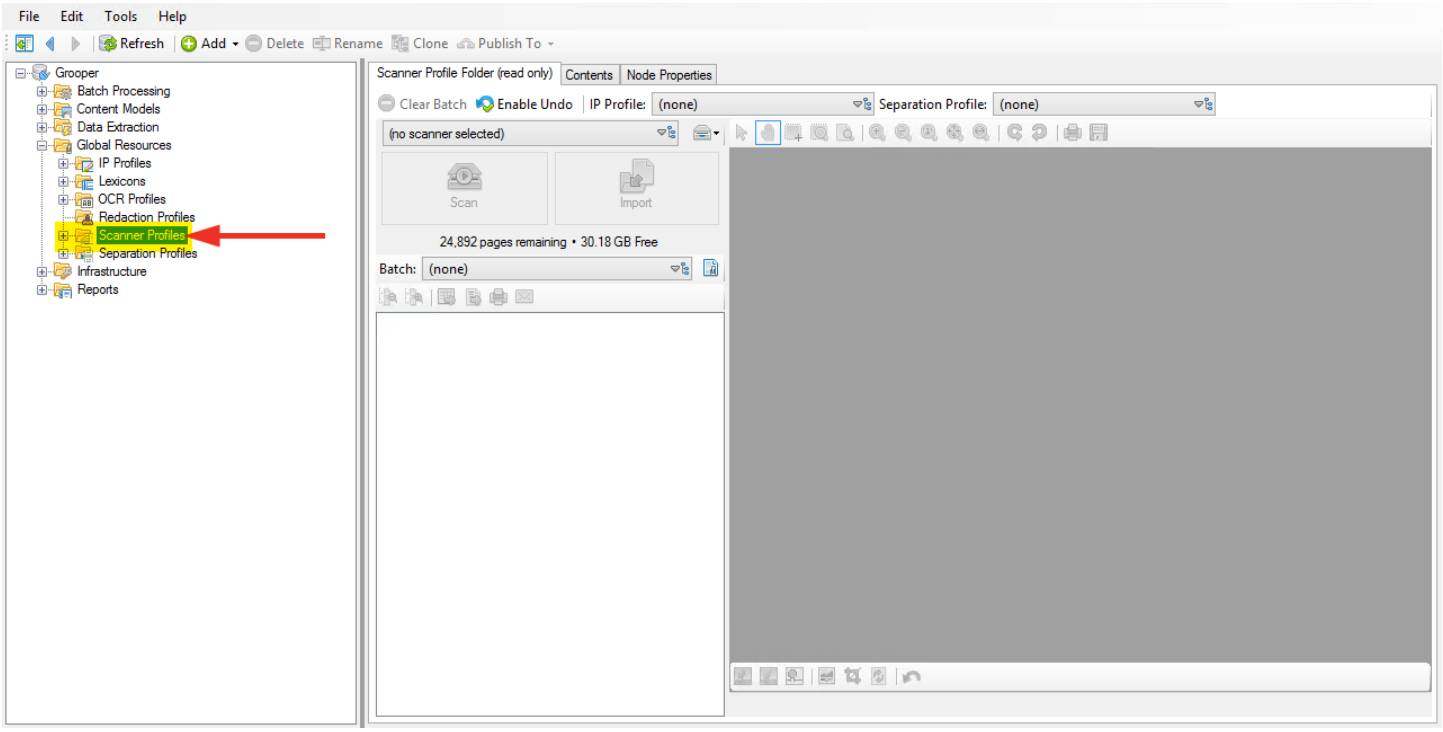
WHAT IS A SCANNER PROFILE?

A Scanner Profile in **Grooper** is an object that contains configuration information for your document scanner and how it will interact with **Grooper**. Scanner model and driver settings for things such as color depth and page rotation are all part of this profile. A Scanner Profile is needed before we can scan with **Grooper**.

HOW TO CONFIGURE AND PERFORM SCANNING

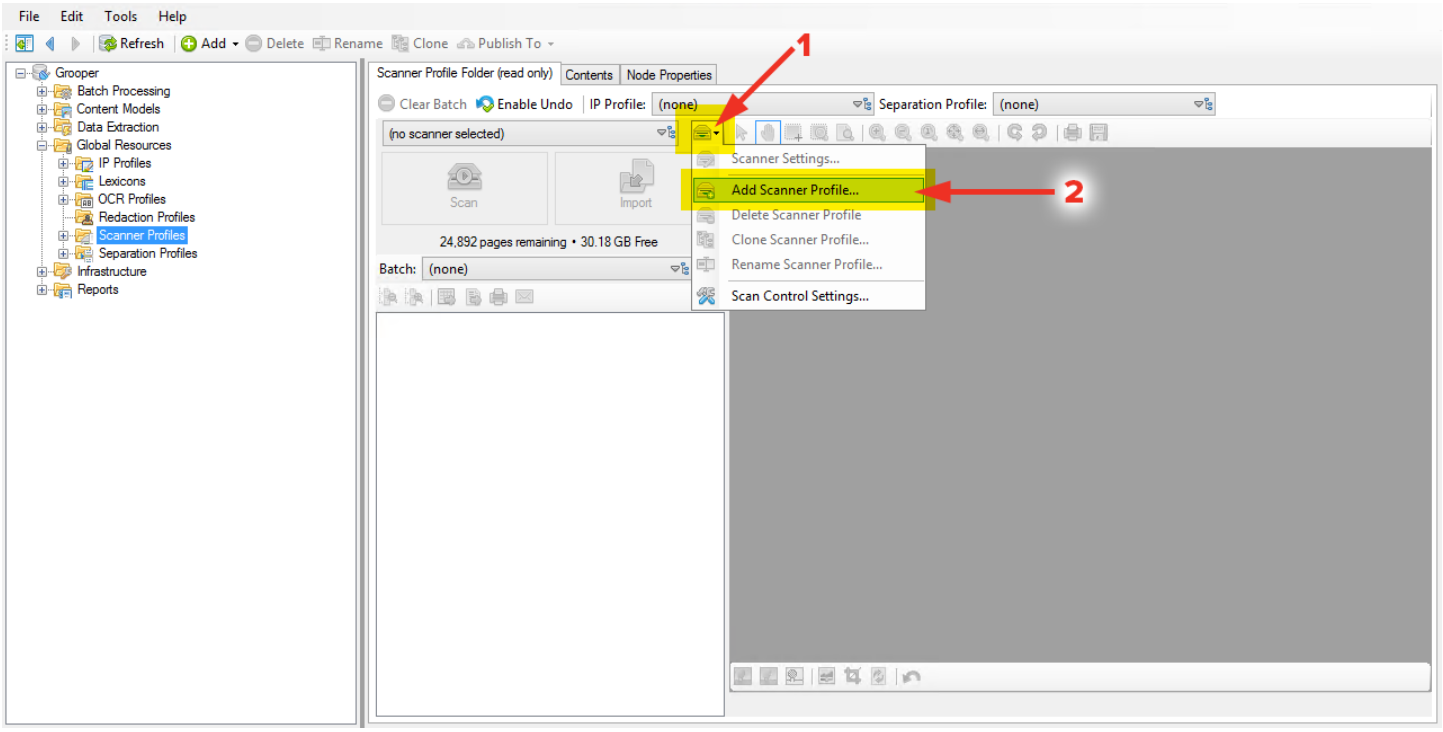
STEP 1 – NAVIGATE TO SCANNER PROFILES

In Grooper Administration expand the node tree to Grooper – Global Resources and select Scanner Profiles.



STEP 2 – ADD SCANNER PROFILE

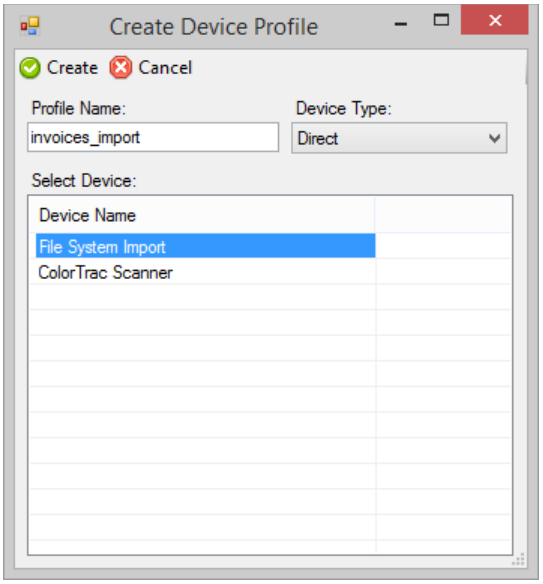
(1) Click the Scanner Profile radial button, and **(2)** from the drop down select Add Scanner Profile.



STEP 3 – CREATE DEVICE PROFILE

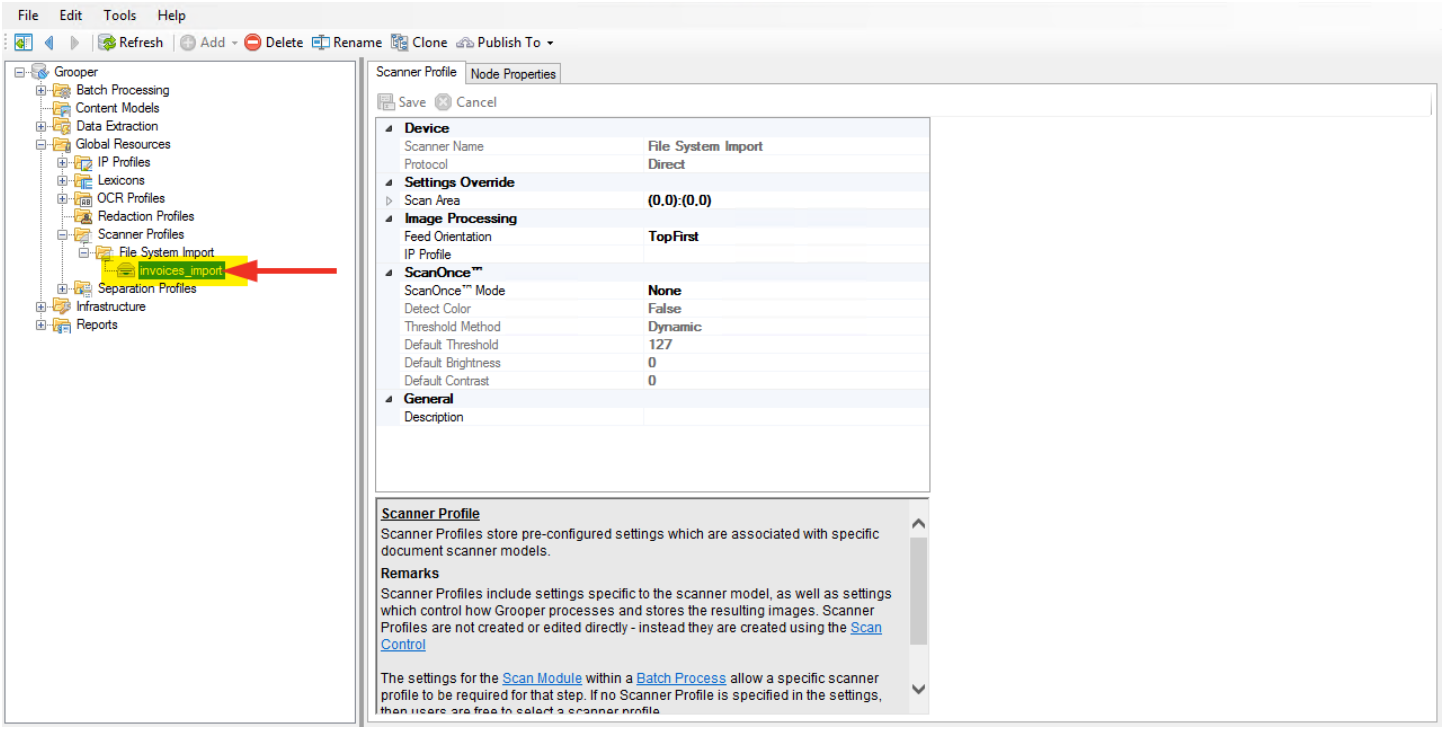
The Create Device Profile window will appear. Fill out invoices_import for the Profile Name, set Device Type to Direct, and choose File System Import for the device name. Click Create.

Were a scanner connected to the computer, the type of driver available (such as TWAIN or ISIS) would be an option from Device Type, and an associated scanner available from Device Name.



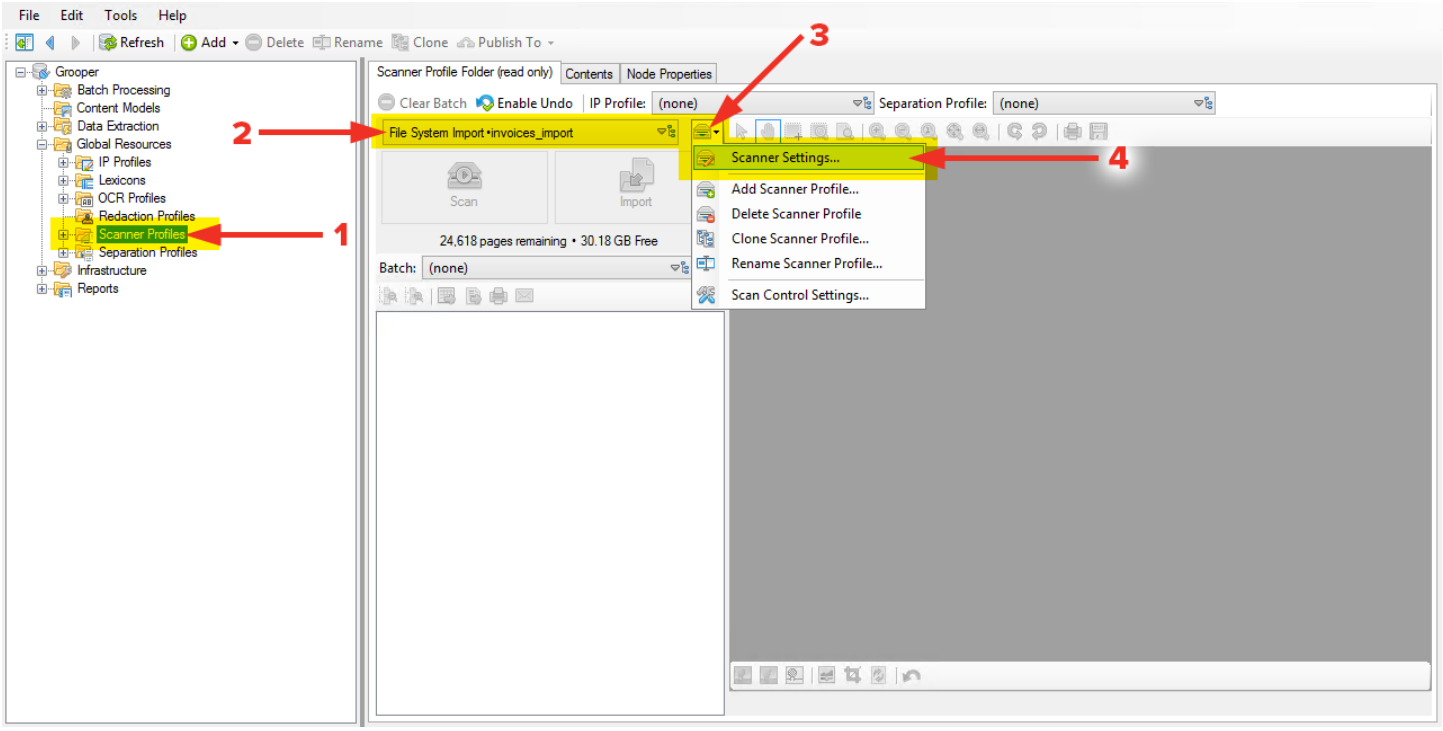
STEP 4 – SCAN DEVICE ADDED

Notice that a new node is added in the node tree.



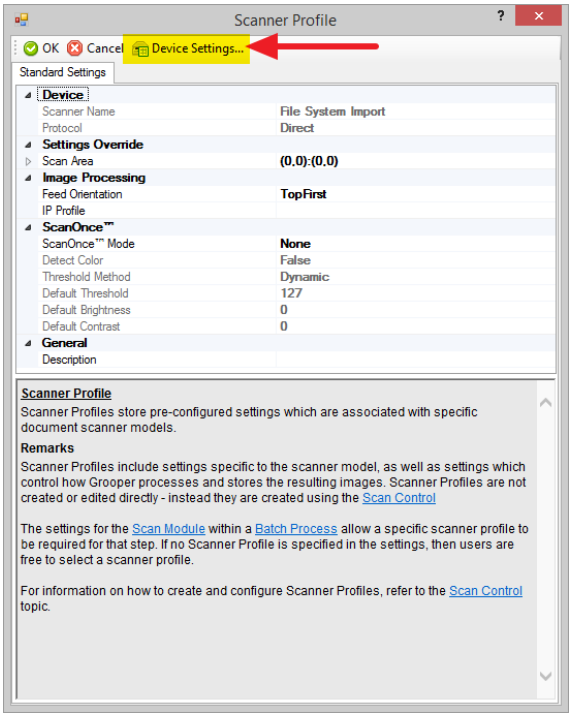
STEP 5 – SCANNER SETTINGS

(1) Click the node for Scanner Profiles in the tree node. **(2)** Make sure the File System Import-invoices_import is populating the Scanner Profile drop down. **(3)** Click the Scanner Profile radial button and **(4)** choose Scanner Settings…



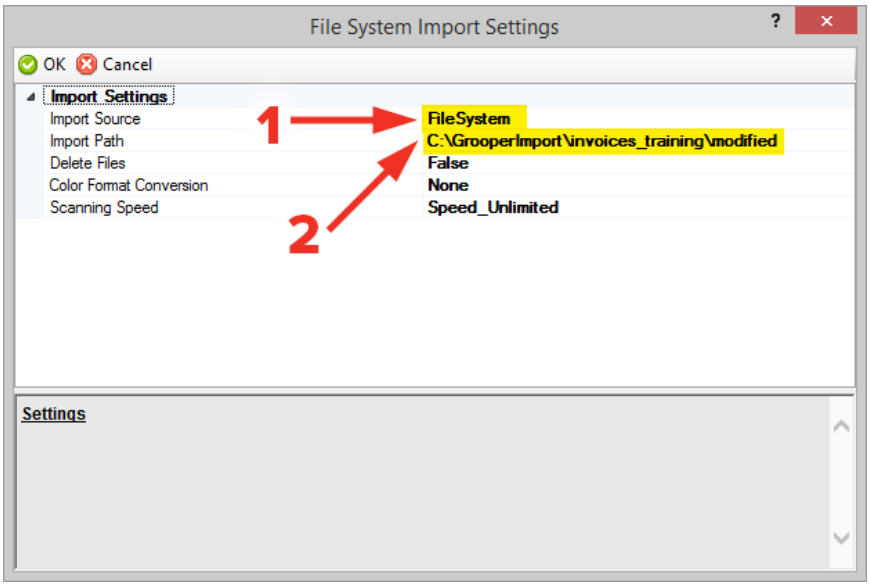
STEP 6 – SCANNER PROFILE WINDOW

The Scanner Profile window will appear. Click the Device Settings… button.



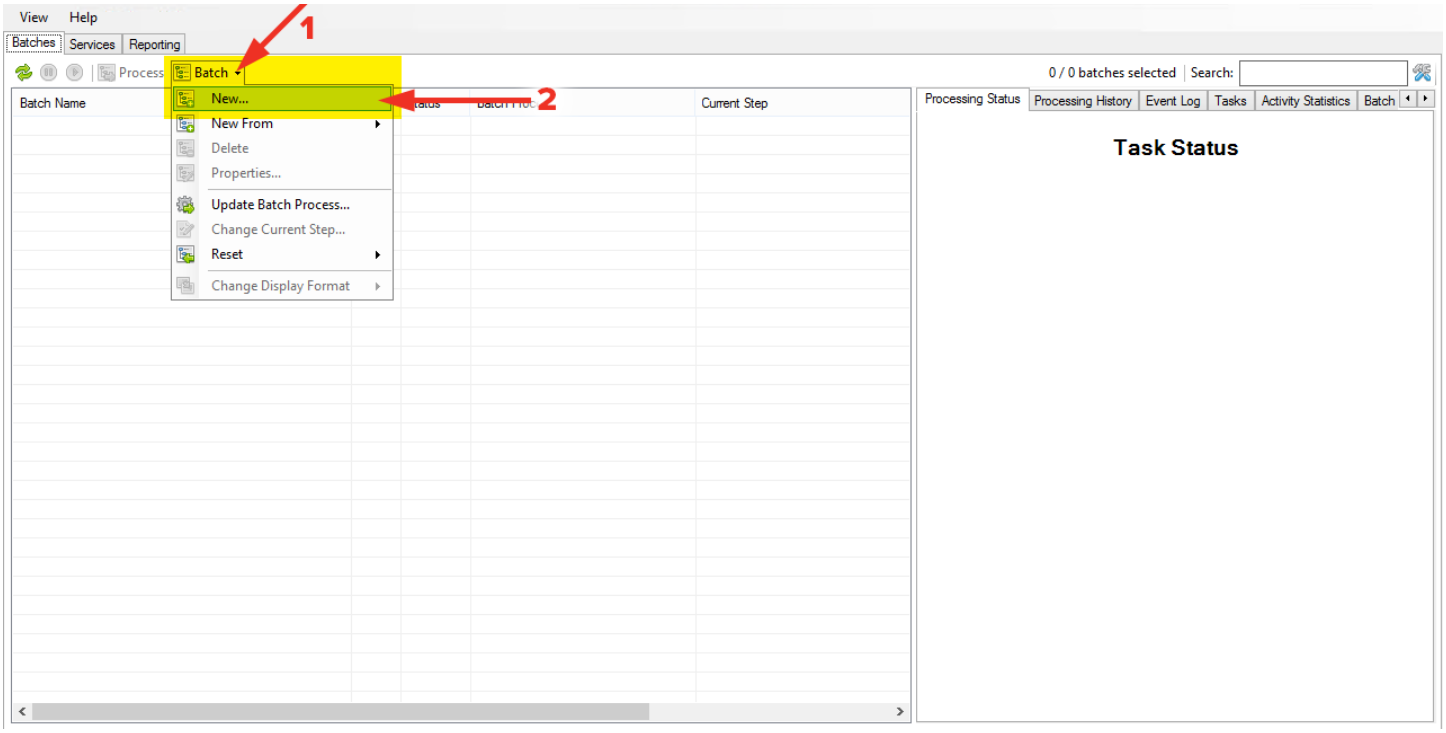
STEP 7 – FILE SYSTEM IMPORT SETTINGS

The File System Import Settings window will appear. **(1)** Set Import Source to FileSystem. Sample documents for this demonstration have been provided. Point Import Path to where those files have been saved. **(2)** The screenshot below is pointing to C:\GrooperImport\invoices_training\modified, but this is specific to the environment. Yours will be different depending on where you save the files. Make sure the remaining fields match the image as well.



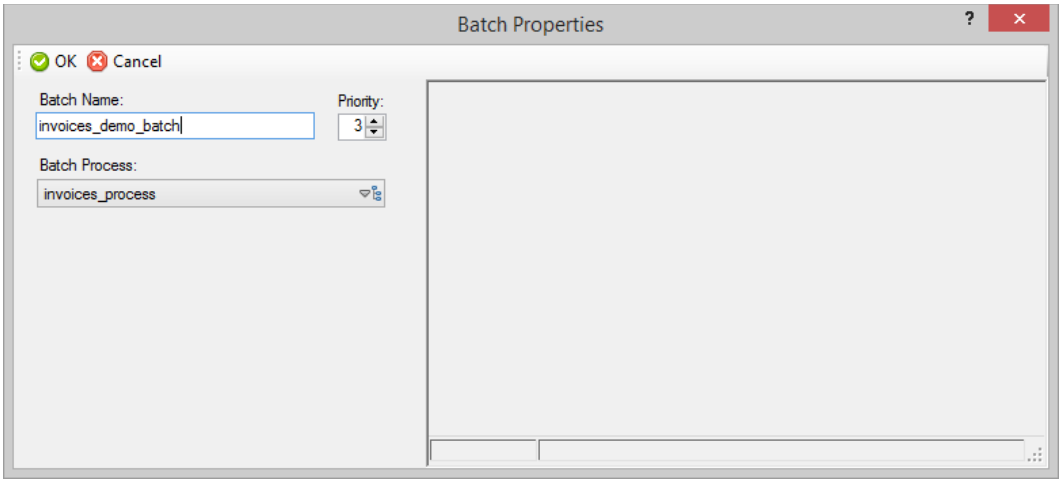
STEP 8 – GROOPER DASHBOARD

Close Grooper Administration and take a look at how a floor supervisor might see managing batches with Grooper. Open Grooper Dashboard. **(1)** Click the Batch radial button and **(2)** select New...



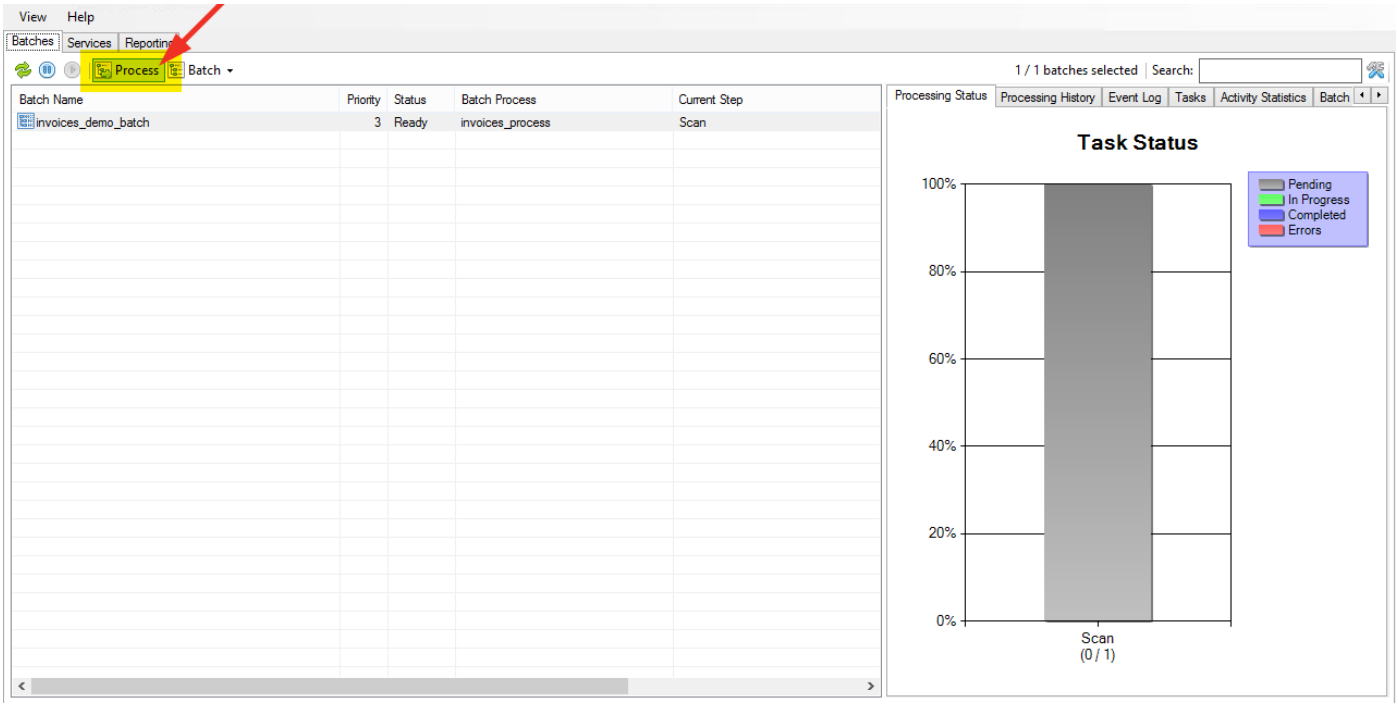
STEP 9 – BATCH PROPERTIES WINDOW

The Batch Properties window will appear. Batches have default values for their name that usually include the time and such, but let's just rename the batch to invoices_demo_batch so it's clear what this batch is. If your environment has had other batch processes made, we'd need to select the process we just made from the Batch Process: drop down, but this demonstration only has the one process, so it's already selected. We'll leave the priority alone. Click OK.



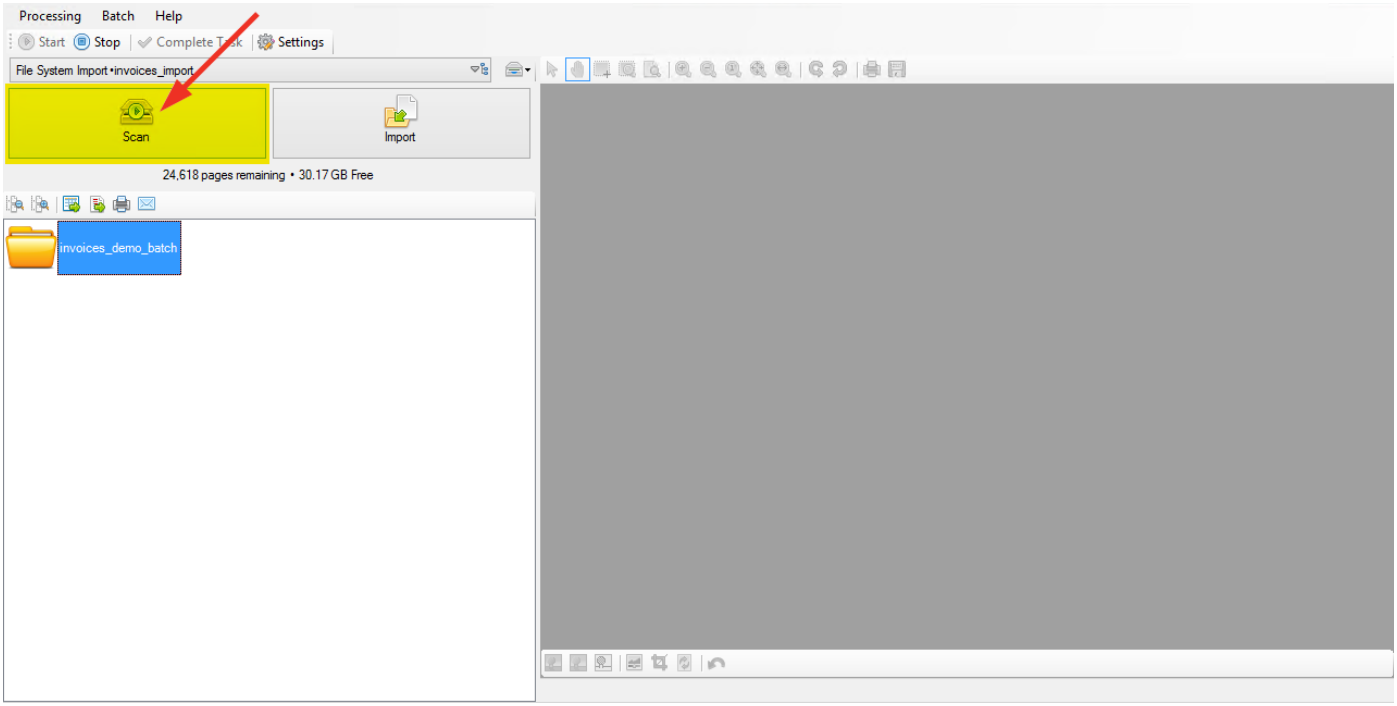
STEP 10 – OVERVIEW OF BATCH VISIBILITY

The Dashboard should now be populated with a batch. The columns on the left side of the screen give some pertinent information regarding this batch such as name, priority, status, etc. Since our batch consists of only one step at the moment, you'll see in the Task Status portion of the screen on the right that there's only one bar representing the progress of that step. Were there more steps in this process, we'd see more bars in the graph. There's also a legend for the color coding of the bars. To move forward, press the 'Process' button.



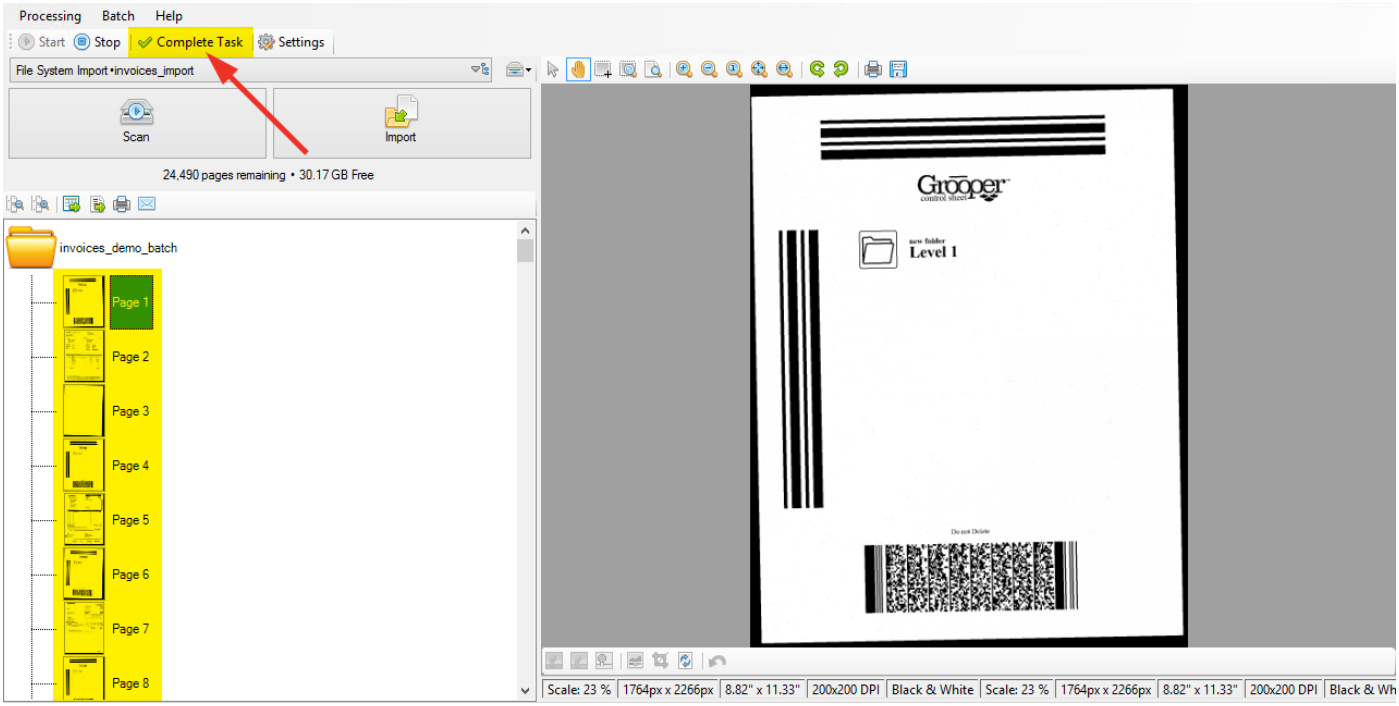
STEP 11 – GROOPER ATTENDED CLIENT - SCAN

This will launch the Grooper Attended Client, and given we are processing a scan step, the Grooper Attended Client will be set in its scan mode. Go ahead and click the big Scan button.



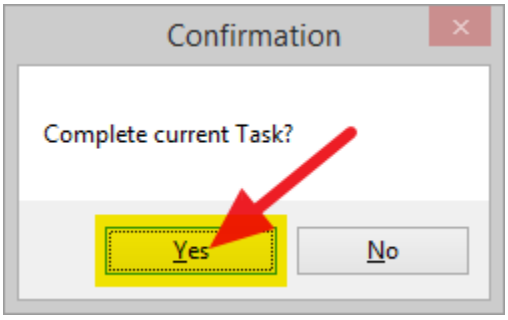
STEP 12 – EXPLAINATION OF IMPORTING VIA SCAN

The scanner profile we established is leveraging the Direct – File System Import, so our scan process will work as an import, instead of a physical scan. With the import done, we can move on to cleaning up the images we just scanned in. Click the Complete Task button.



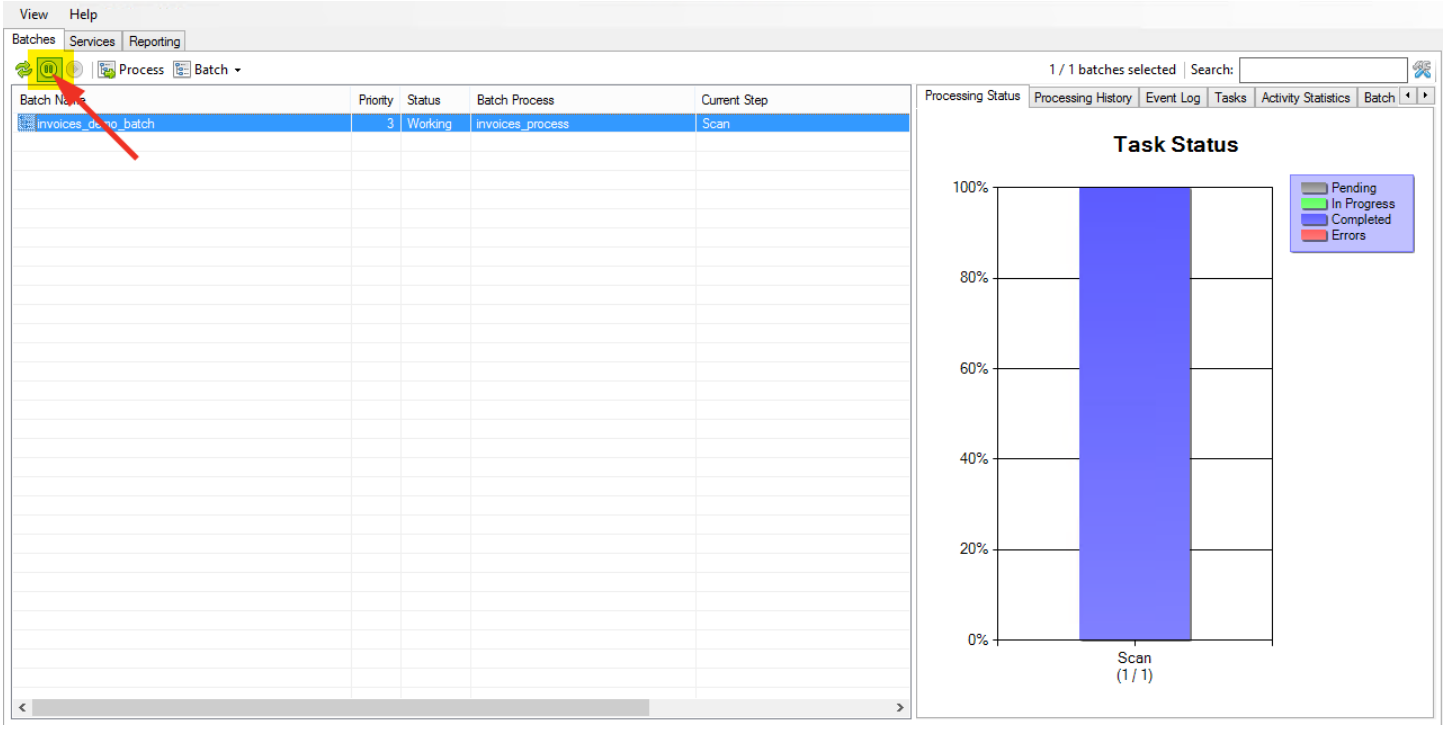
STEP 13 – CONFIRMATION WIDNOW

A Confirmation window will appear. Click Yes to continue.



STEP 14 – SCAN COMPLETE

The Grooper Attended Client module will close and bring us back to the Grooper Dashboard. You'll notice the bar for the scan step is now blue, indicating that the step is Completed. Pause the batch and close Grooper Dashboard.



PHASE 2 - CONDITION

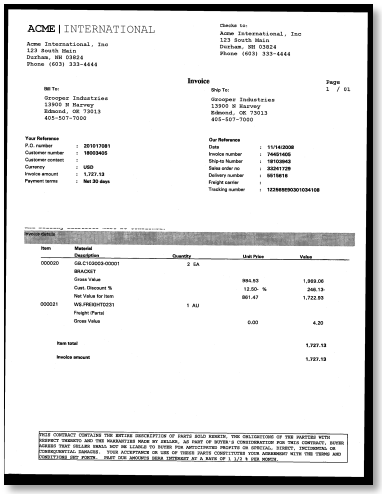
Now that content has been acquired by **Grooper**, it is important that we set that content up for future success. Depending on how the content was captured, defects could have been introduced to it that can get in the way of processing that's soon to come. In order to mitigate any problems, we need to make sure the content is in good shape.

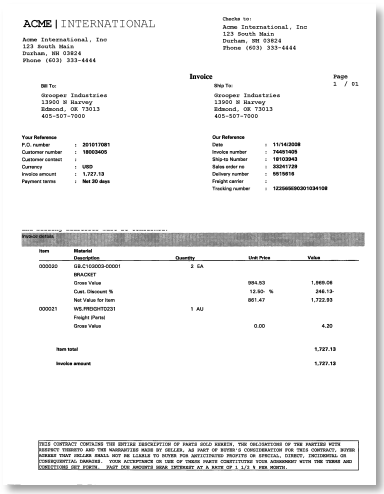
Along with 'cleaning it up', we need to have **Grooper** 'read' the content through a process known as Optical Character Recognition. This step is critical and will embed information into what we captured, allowing for the powerful processing that's soon to come.

IMAGE PROCESSING

WHAT IS IMAGE PROCESSING?

"Photoshopping" has become one of those ubiquitous terms we've come to know to mean - manipulating a digital image for one reason or another. In the case of **Grooper**, it's called Image Processing and it's used to accomplish a similar goal. This cleanup is especially applicable to scanned images, because there can be any number of issues we may want to clean up in a scanned image. From de-skewing and border cleanup to adjusting brightness and contrast and removing artifacts like blobs and speckles, **Grooper**'s Image Processing capabilities are vast. Also, an amazing feature of **Grooper**'s Image Processing is its ability to detect features like stamps, barcodes, and shapes. The recognition of these features is important because it can tag a page with specific information (or metadata) and manipulate it in a specific way later in our batch process.





PERMANENT VS. OCR-ONLY

There are two ways Image Processing can be applied. One method is used to permanently alter the image, such as cleaning up problems with artifacts you might see in a scan process (for example de-skewing and border cropping.) The other method involves leveraging Image Processing during the OCR step. This method alters the images temporarily, so text can be extracted for the OCR process. For the purposes of our demonstration, we will be permanently altering our images. Defects that may happen during scanning were added to these pages to help demonstrate what IP Profiles can do for us.

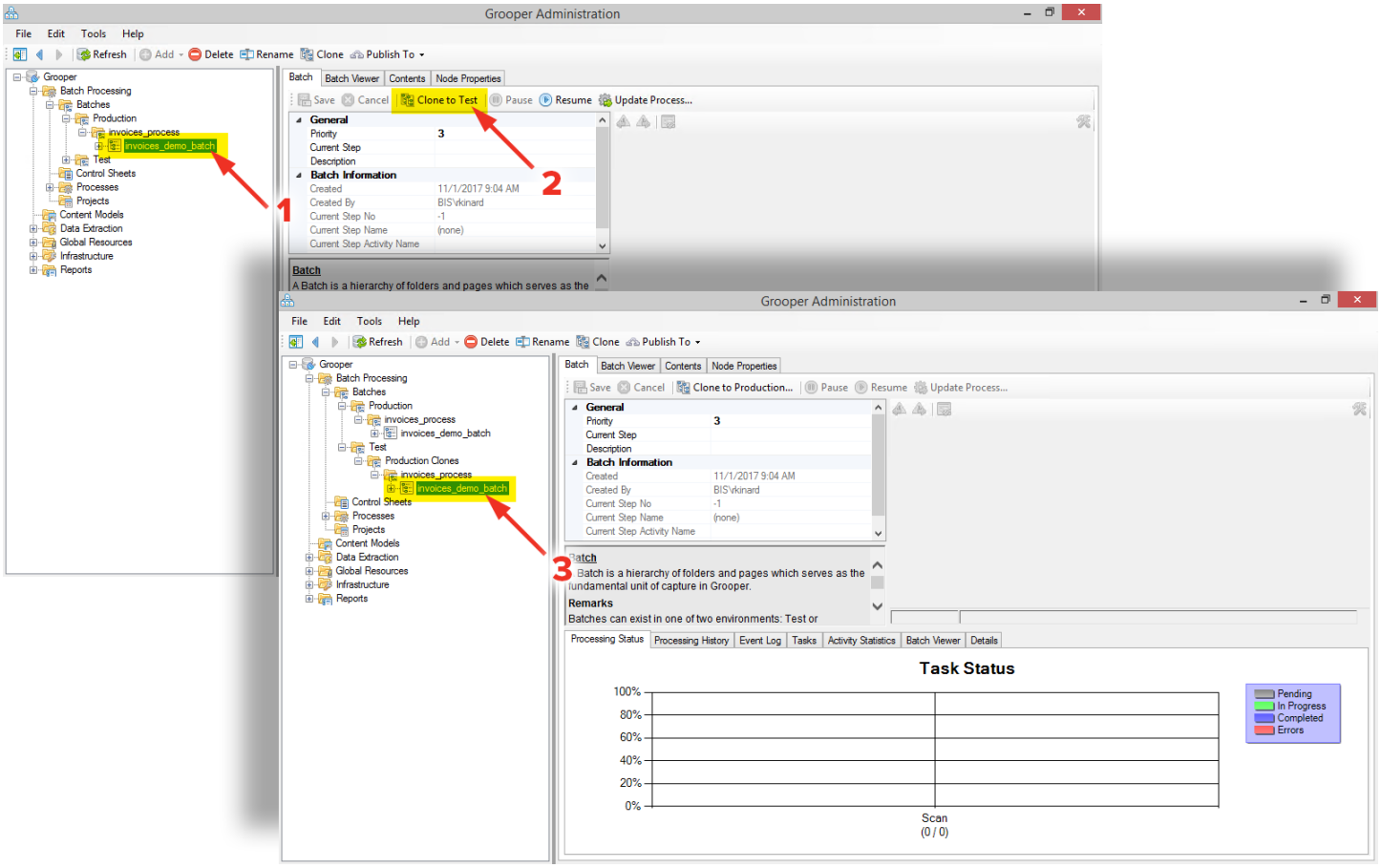
WHAT IS AN IP PROFILE?

Much like the Scanner Profile, an IP Profile is a node that exists in **Grooper** that defines the steps that will be taken to adjust our image and or tag it with metadata (for example, whether the image contains a barcode.)

HOW TO CONFIGURE AND PERFORM IMAGE PROCESSING

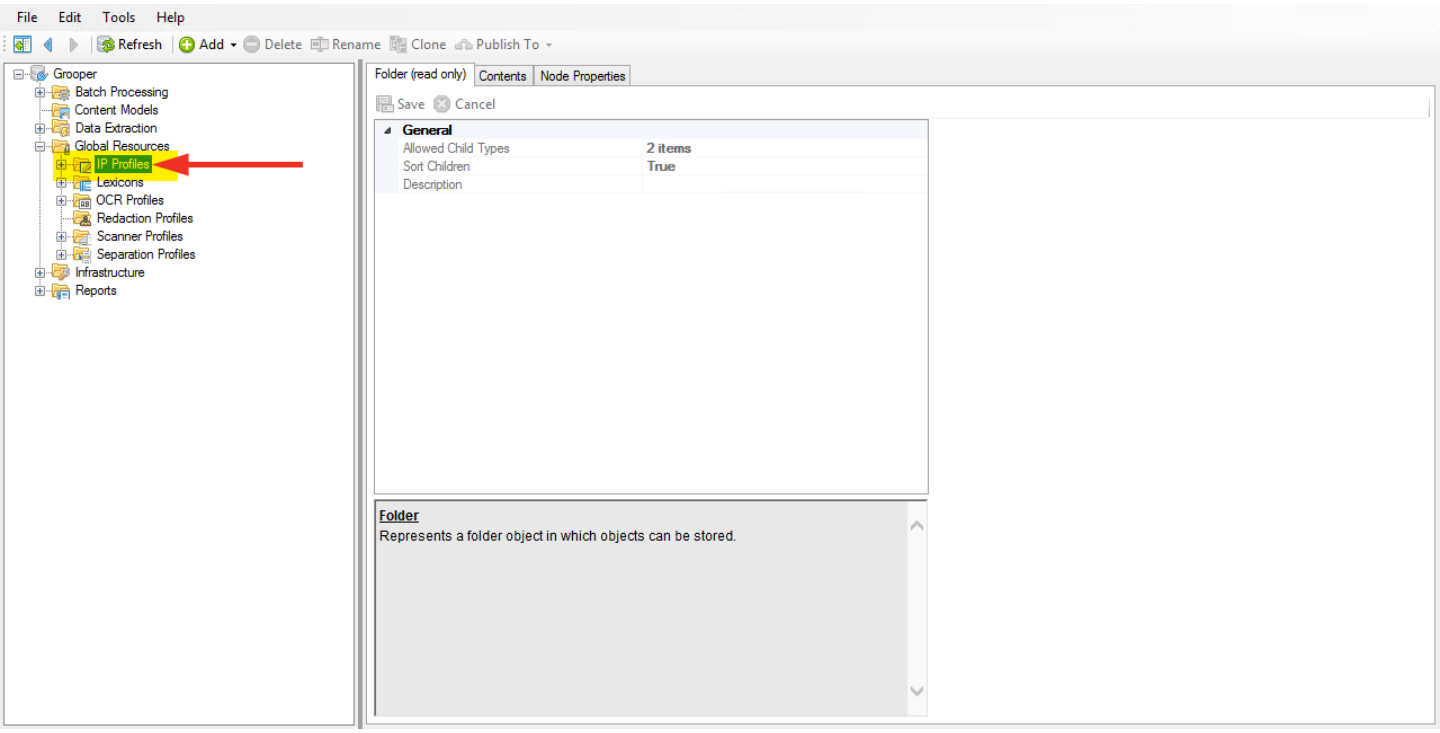
STEP 1 - CLONING A BATCH FOR TESTING

So that we can see how our IP Profile will affect these pages, let's start by cloning our batch to a test version (this is required for testing in the IP Profile, as you cannot point to a Production batch.) Within Grooper Administration, expand the node tree to Grooper – Batch Processing – Batches – Production – invoices_process.
(1) Select the node for invoices_demo_batch. With it highlighted, **(2)** push the Clone to Test button.
(3) A copy of this batch will now exist within Grooper – Batch Processing – Batches – Test – Production Clones – invoices_process.



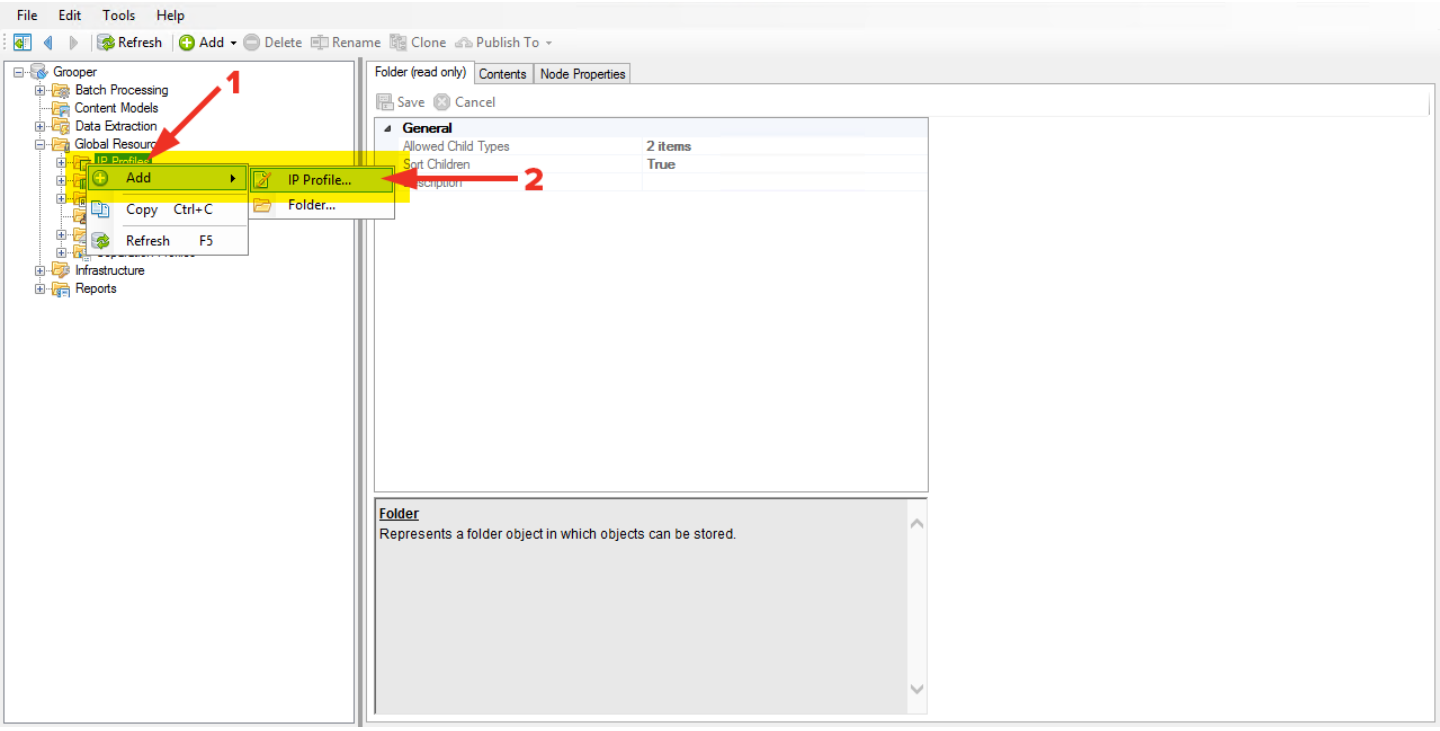
STEP 2 – NAVIGATE TO IP PROFILES

In Grooper Administration expand the node tree to Grooper – Global Resources and select IP Profiles.



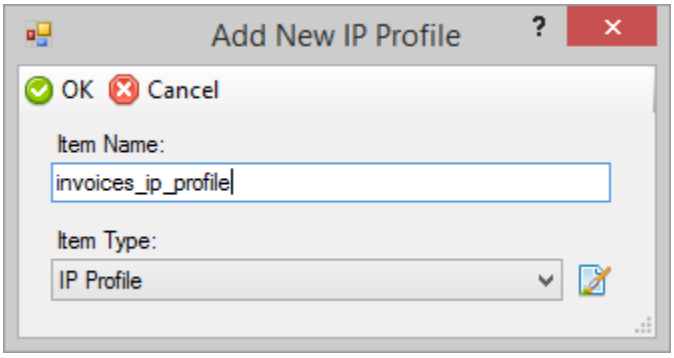
STEP 3 – ADD IP PROFILE

You can either **(1)** right-click this and highlight Add and select IP Profile, or select Add at the top and then **(2)** select IP Profile from the dropdown.



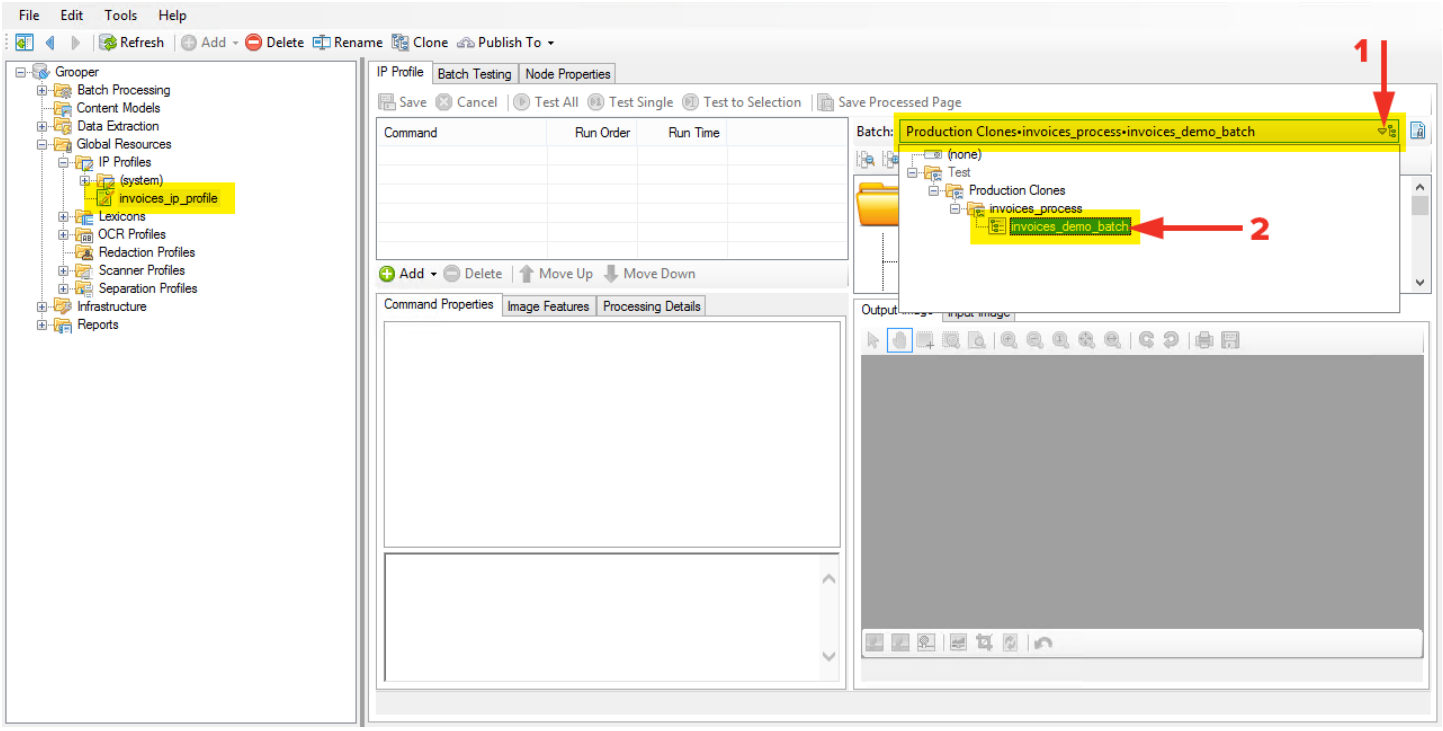
STEP 4 – ADD NEW IP PROFILE WINDOW

The Add New IP Profile window will appear. Name the profile invoices_ip_profile.



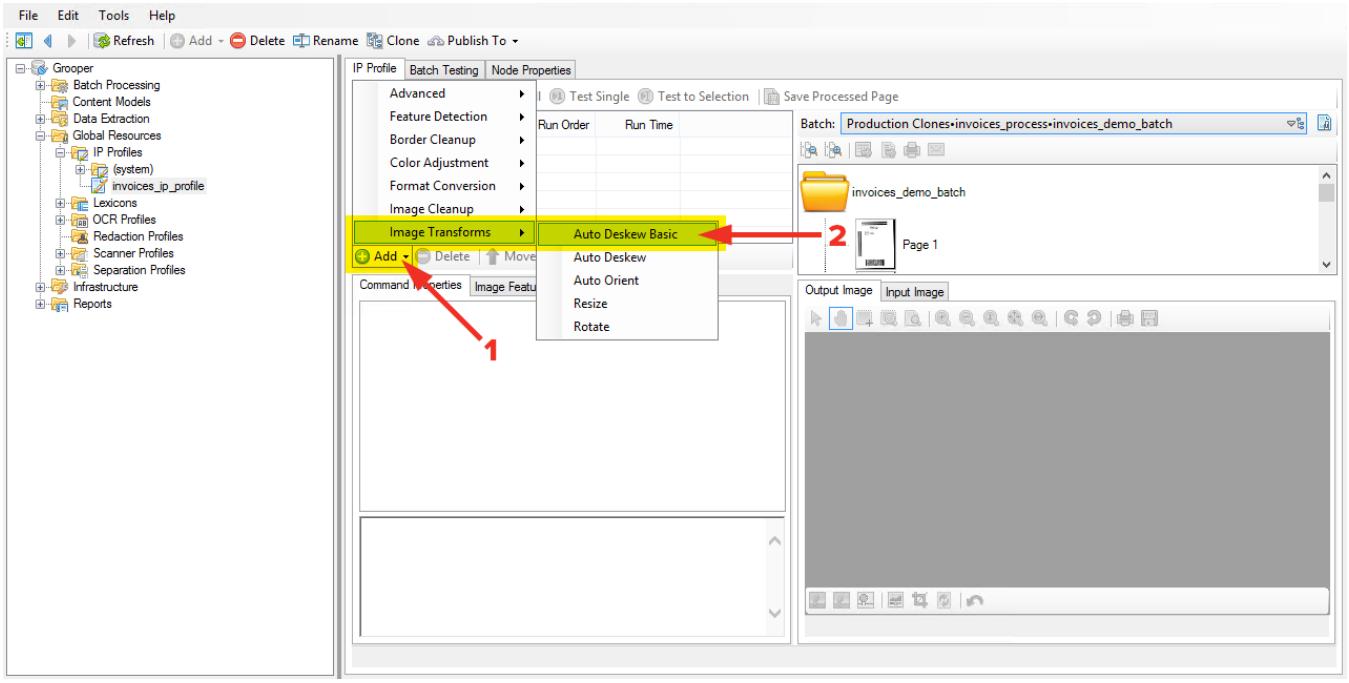
STEP 5 – POINT TO TEST BATCH

We need to point this newly created IP Profile to a test batch to see how the steps we add will affect the pages of our batch. With the new invoices_ip_profile selected from the node tree, **(1)** select the dropdown next to Batch: and **(2)** select our invoices_demo_batch.



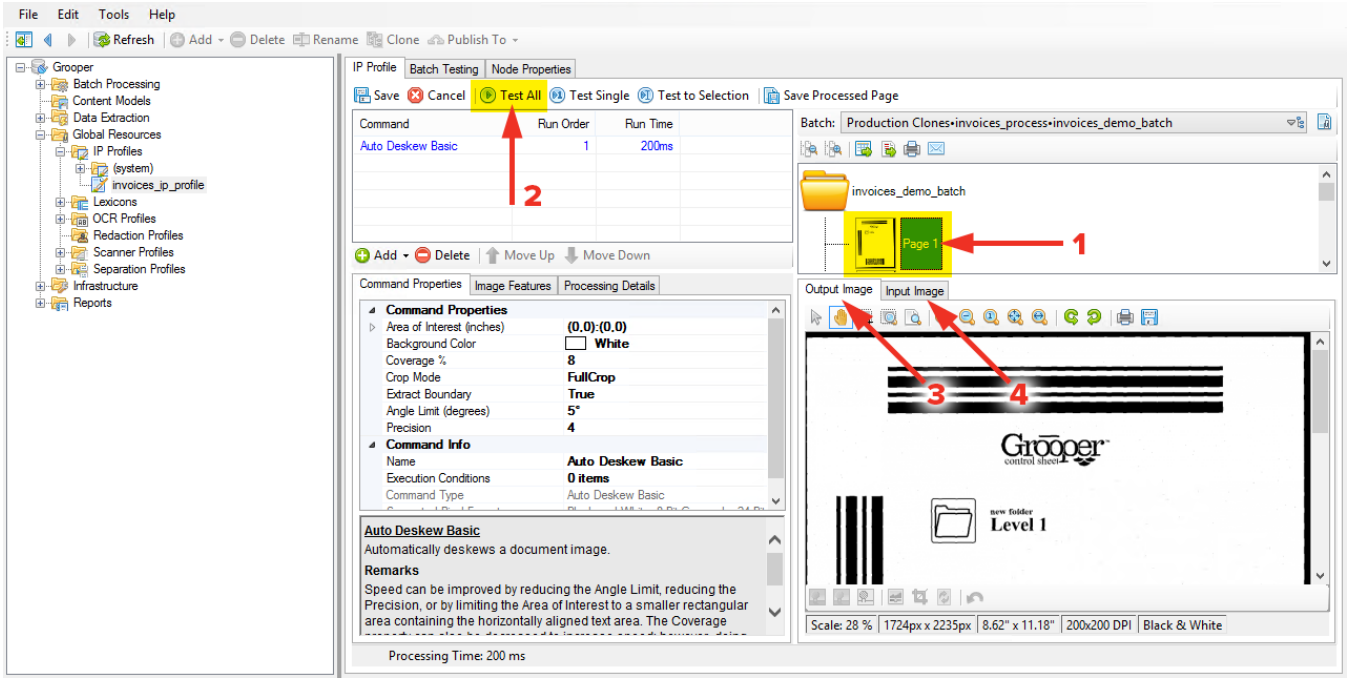
STEP 6 – ADD DESKEW STEP

We need to do a few things to get these images in a good place. We need to de-skew them, crop and remove any black borders, remove the speckling, and perform an analysis of the images that will give us some important information we can leverage later. **(1)** Click Add, mouse over the Image Transforms section, and **(2)** select Auto Deskew Basic. Use default settings.



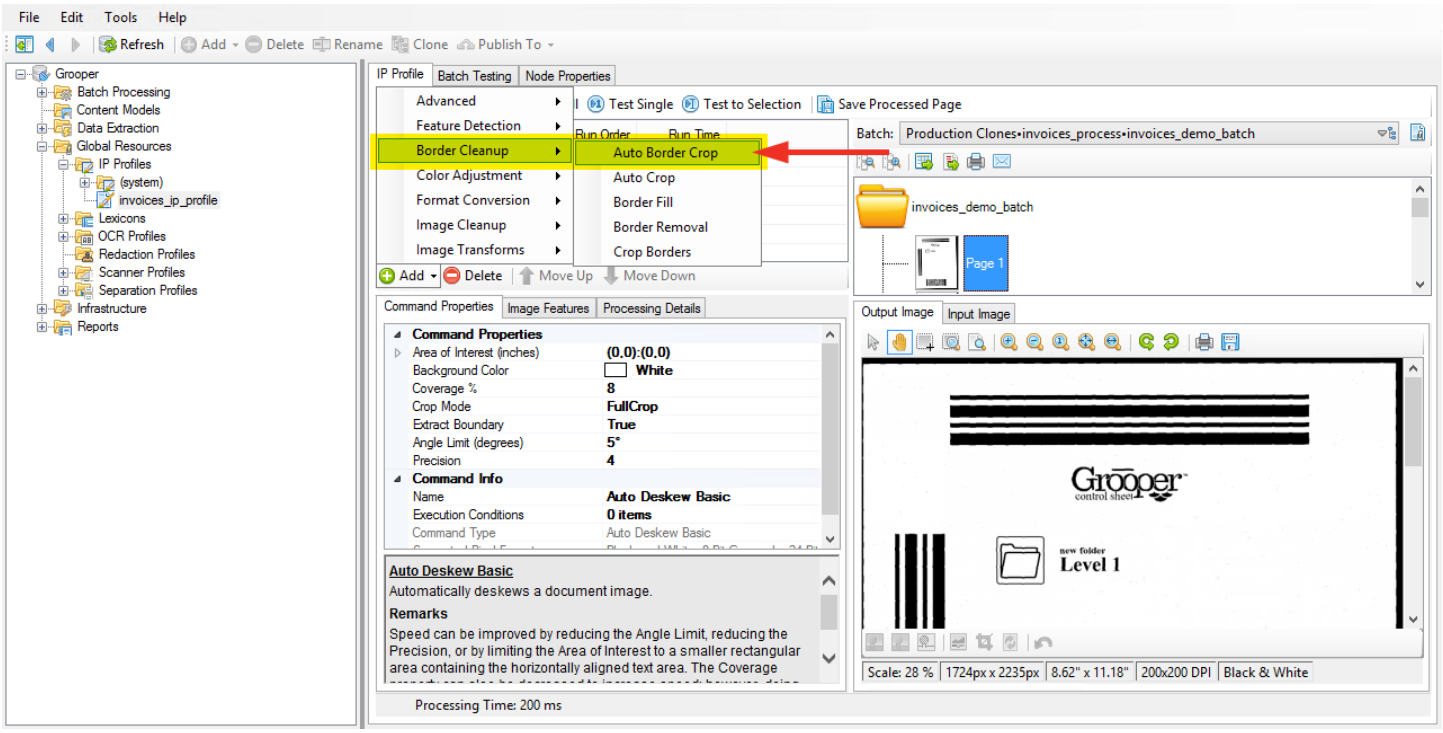
STEP 7 – TEST ALL

(1) Select a page in the Batch Viewer, and **(2)** click the Test All button. In the bottom right portion of the screen, the Page Viewer, you can see the after **(3)** in the Output Image, and a **(4)** before in the Input Image.



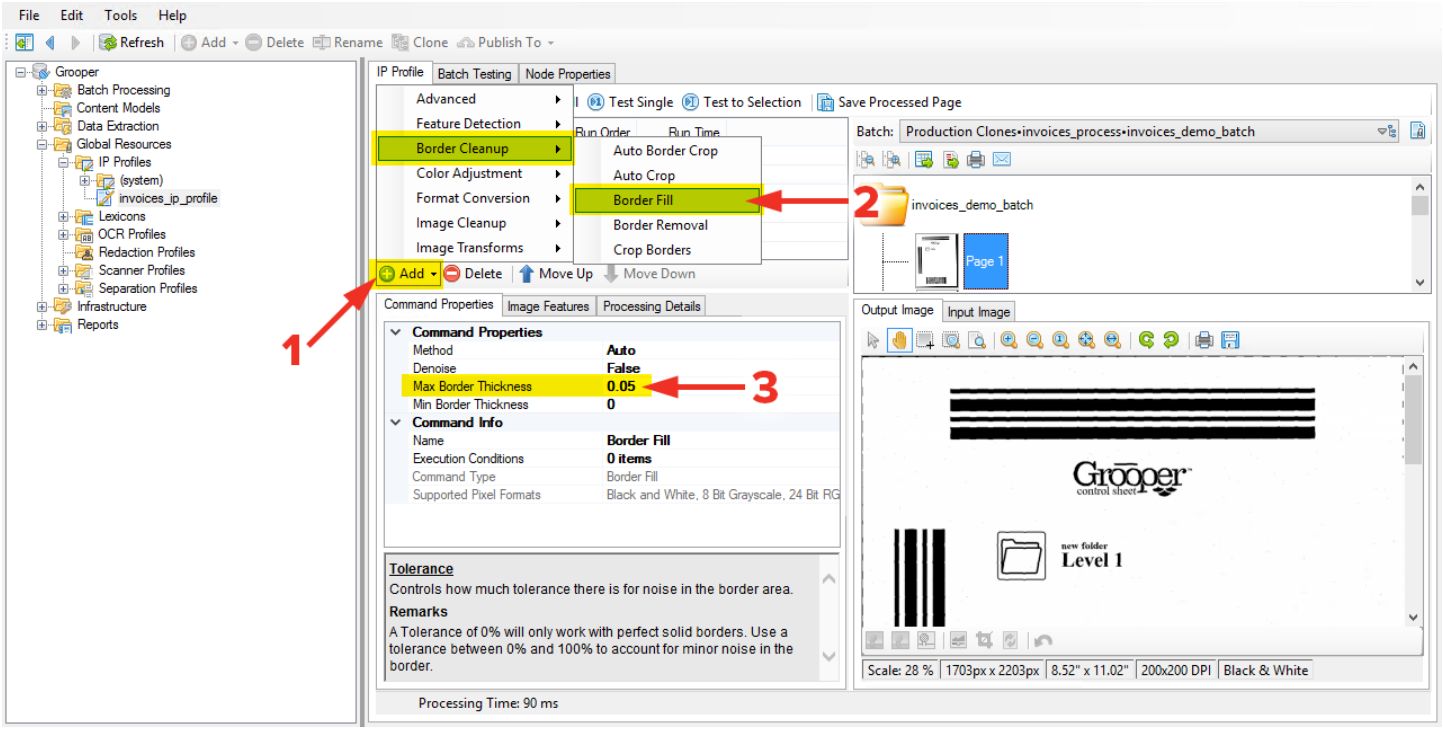
STEP 8 – ADD AUTO BORDER CROP

Click Add, mouse over the Border Cleanup section, and select Auto Border Crop. Use default settings.



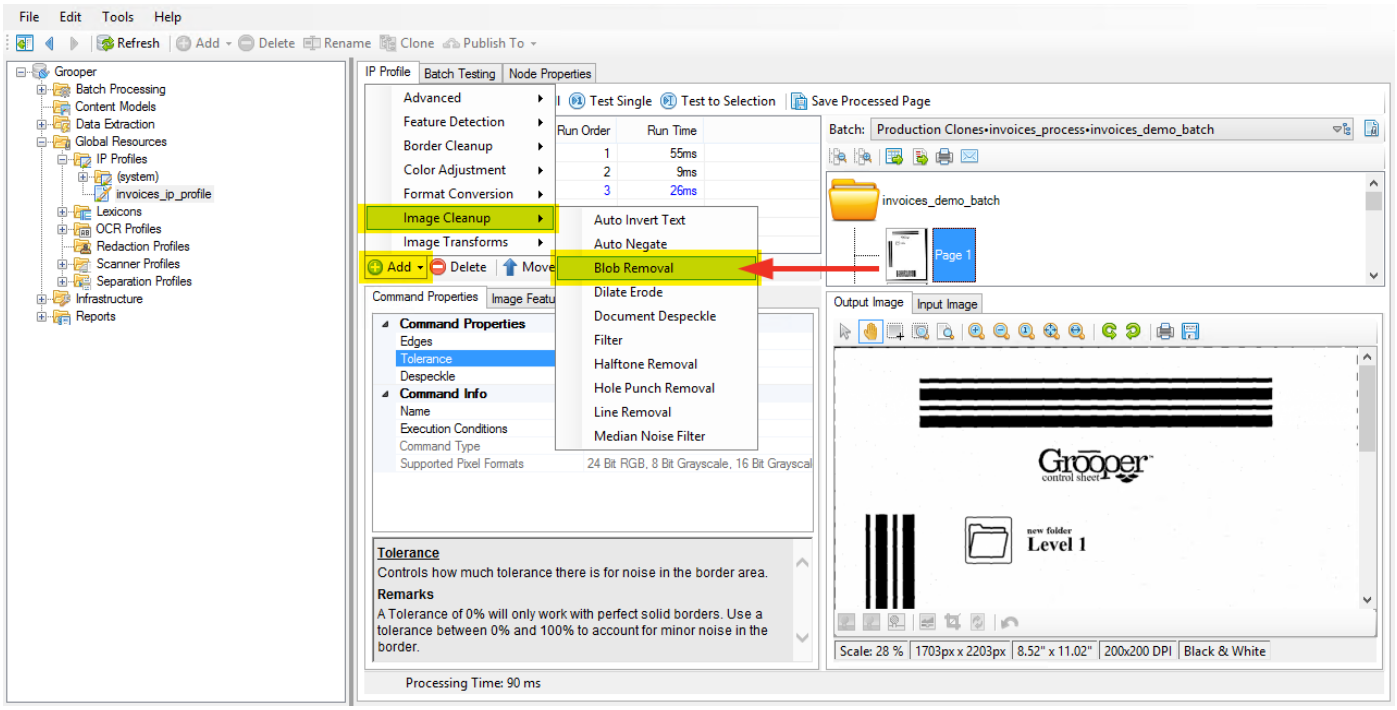
STEP 9 – ADD BORDER REMOVAL

(1) Click Add, mouse over the Border Cleanup section, and **(2)** select Border Fill.
(3) Set Max Border Thickness to 0.05.



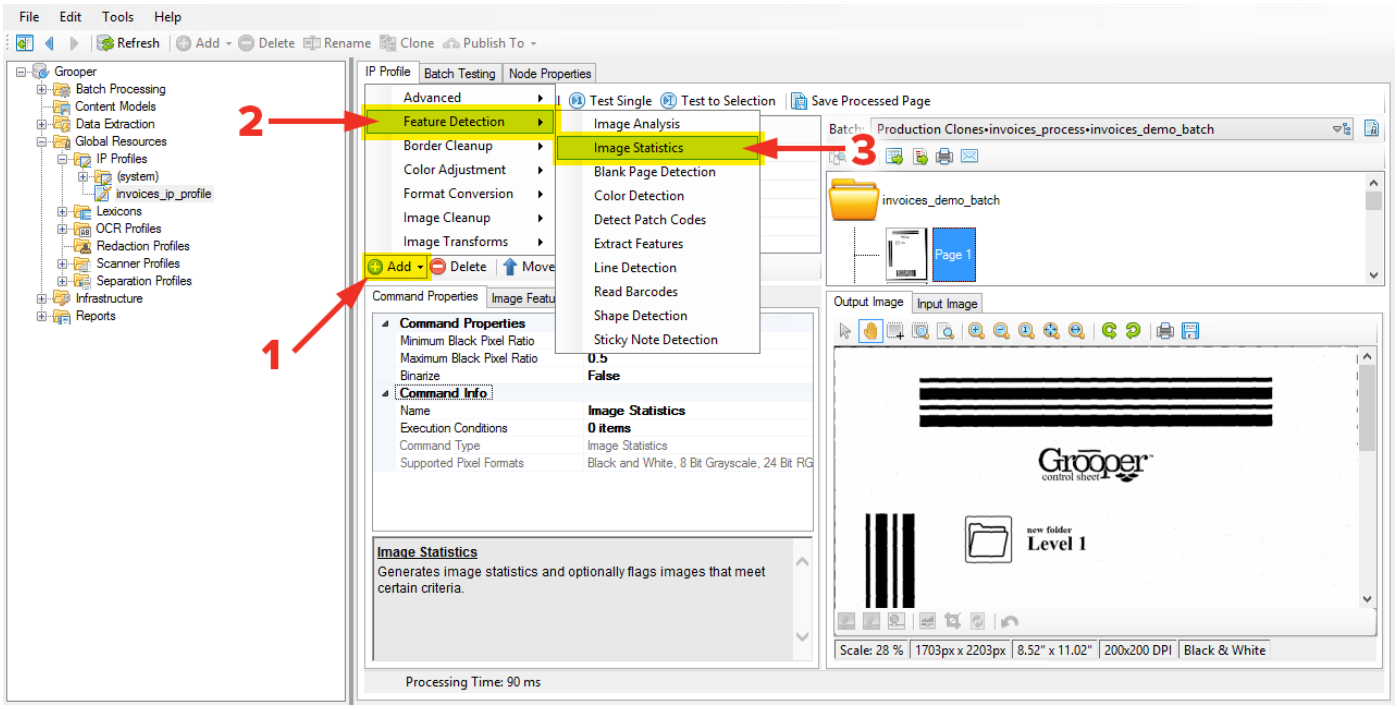
STEP 10 – ADD BLOB REMOVAL

Click Add, mouse over the Image Cleanup section, and select Blob Removal. Use default settings.



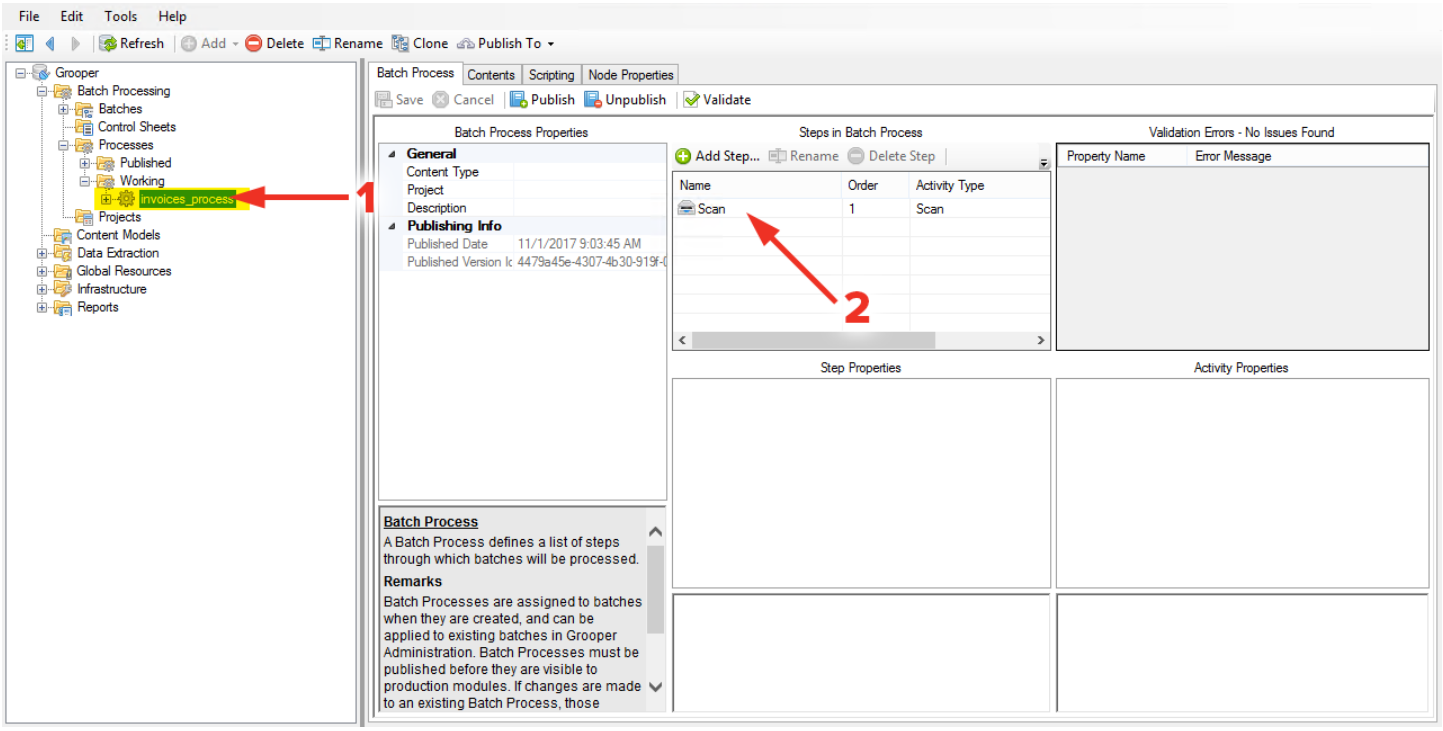
STEP 11 – ADD IMAGE STATISTICS

Finally **(1)** click Add, **(2)** mouse over the Feature Detection section, and **(3)** select Image Statistics. Use default settings. And, with all our steps added, click the Save button. Try the Test All button again to see how our images has been cleaned up. Keep in mind, while we're creating a profile that can be used in a batch process (or a stand-alone Activity Process), the Test All button is not saving the transformations to any of the pages.



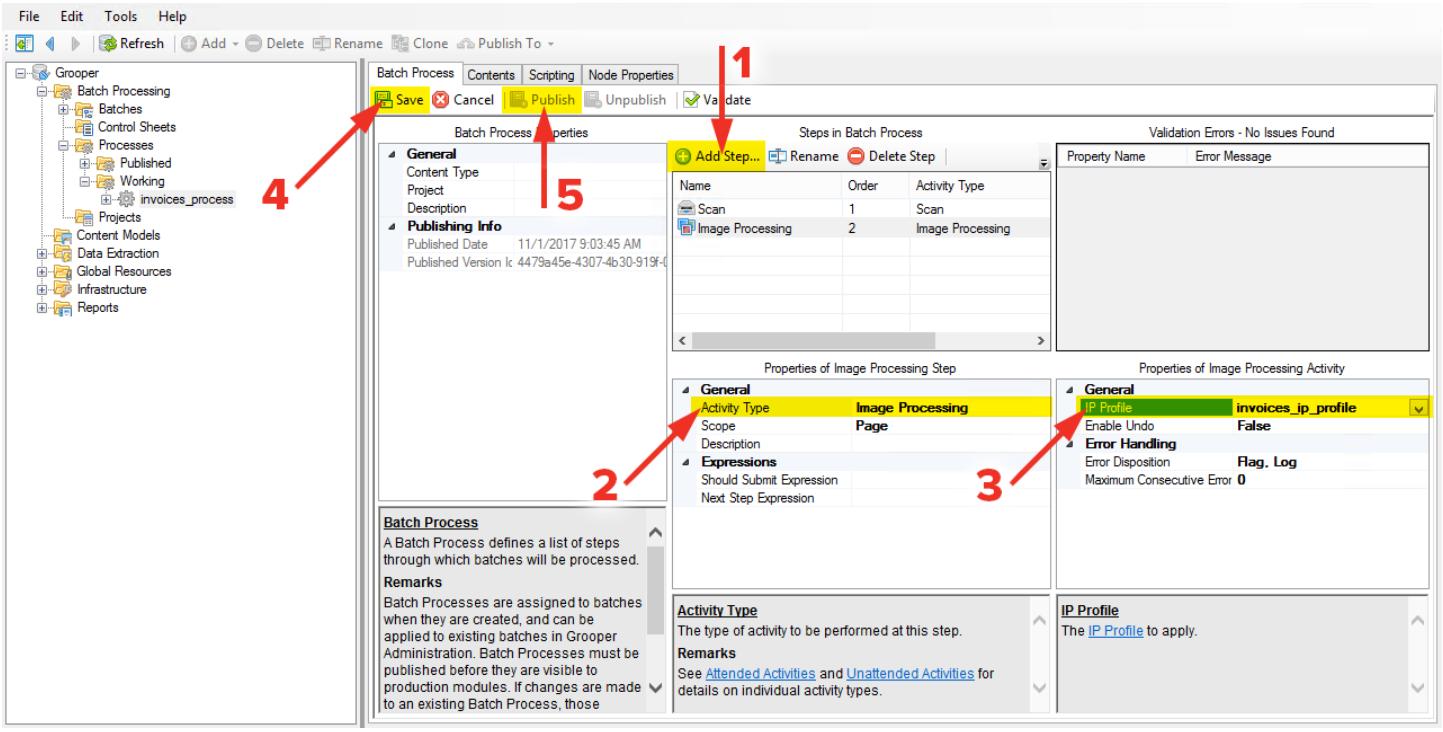
STEP 12 – NEED TO UPDATE BATCH PROCESS

Contents of a batch are affected by their batch process activities, or by individually executed activities. We need to add a step to the batch process to allow for the use of the IP Profile we just created. Open the node tree up to Grooper – Batch Processing – Processes – Working and **(1)** select the invoices_process batch process. As it is right now, **(2)** there's only one step, and it's for Scan.



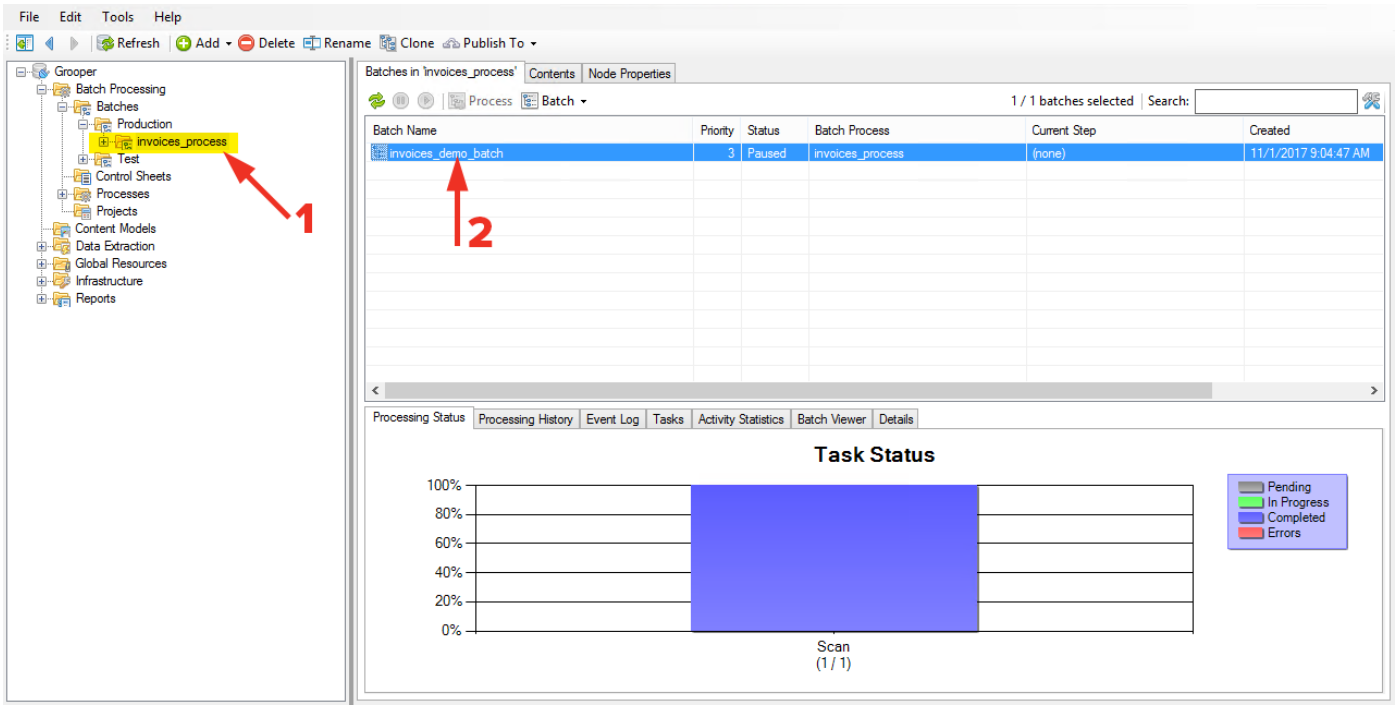
STEP 13 – ADD IP PROFILE STEP

(1) Add a Step, **(2)** set its Activity Type to Image Processing, **(3)** set its IP Profile to the profile we created which is invoices_ip_profile. **(4)** Be sure to save the process and **(5)** publish it (confirm this choice in the window that comes up.) And, to reitterate, Production batches cannot use Working Batch Processes, only published ones. Were we to have added this step, and simply saved without publishing, the published version of this batch process would only contain the Scan activity. Having published this working process, the published version will be updated to reflect the new change we've made, and as a result, expose this change to the Production batches.



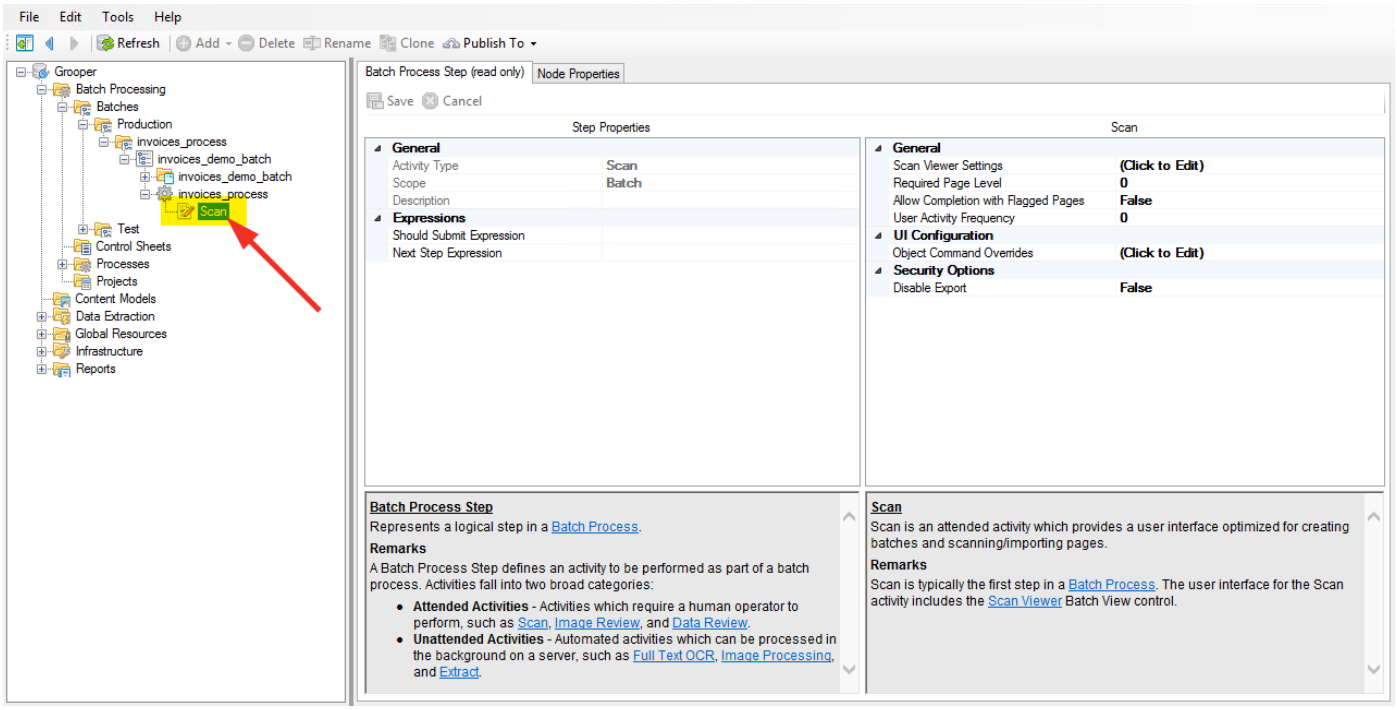
STEP 14 – NAVIGATE TO BATCH

Open the node tree to Grooper – Batch Processing – Batches – Production and **(1)** select the invoices_process folder node. **(2)** Our batch that finished scanning is still here.



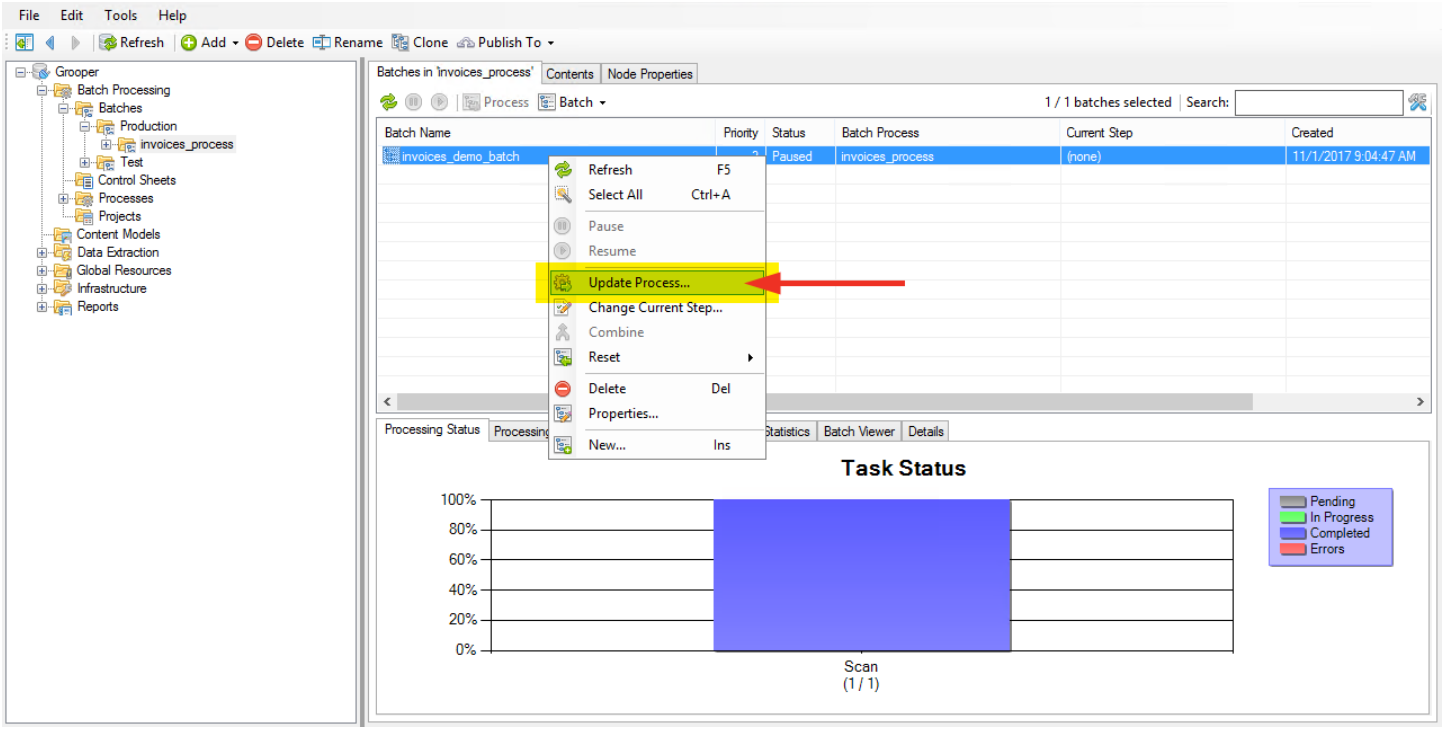
STEP 15 – BATCH STEP UPDATE RECOGNITION

Expand the node tree a bit further to Grooper – Batch Processing – Batches – Production – invoices_process – invoices_demo_batch – invoices_process, and notice there's only still a scan step. The Batch Process needs to be updated to reflect the new IP step that was added to the Working and subsequently Published batch process.



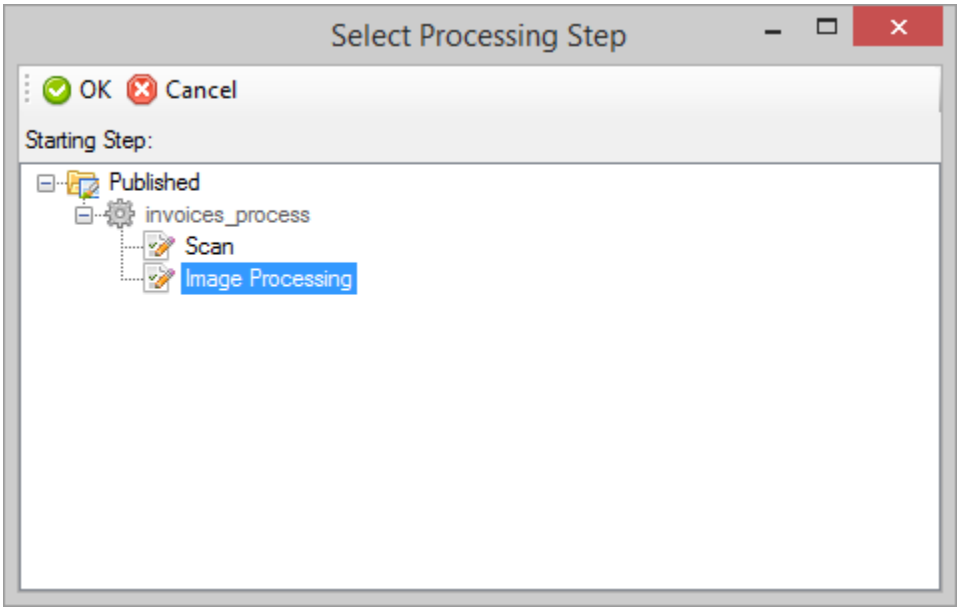
STEP 16 – UPDATE BATCH

Go back to Grooper – Batch Processing – Batches – Production and select the invoices_process folder node. In the list of batches on the right, select the invoices_demo_batch batch. From here either right-click and in the drop down select Update Process... or click the Batch drop down and select Update Batch Process...



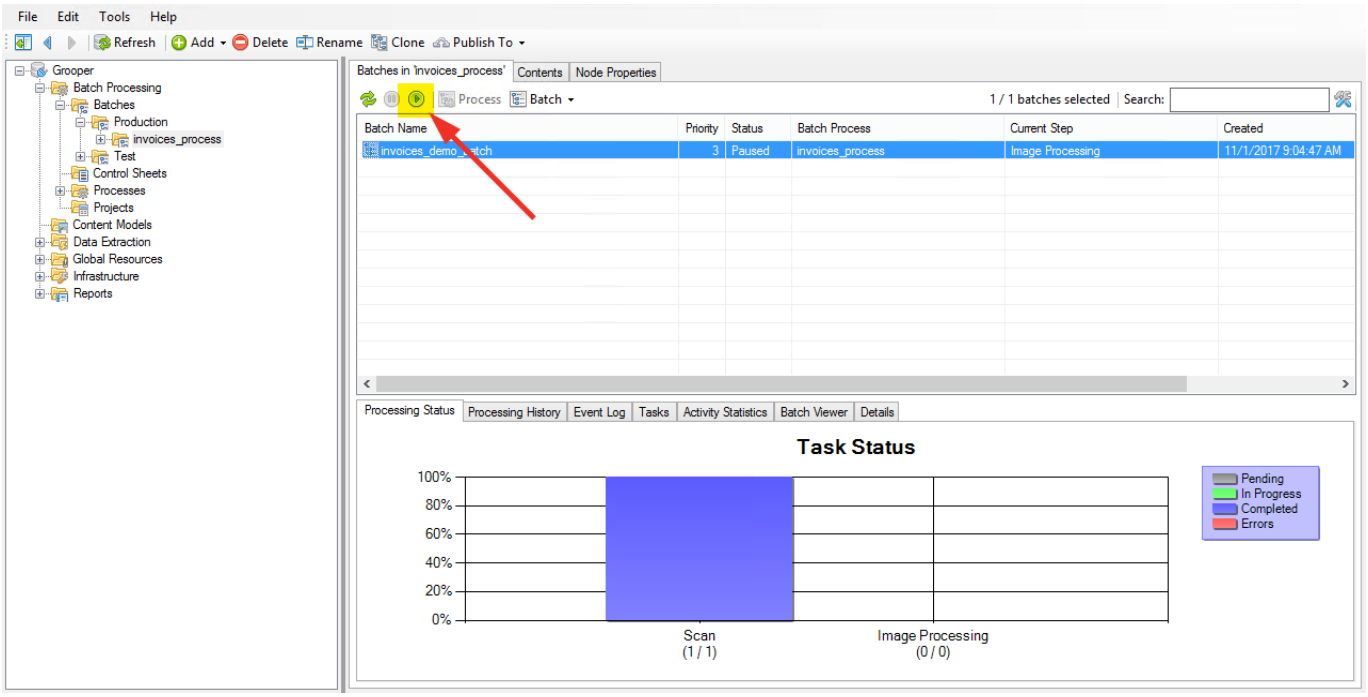
STEP 17 – SELECT IMAGE PROCESSING

In the Select Processing Step window that appears, select the new Image Processing step and click Ok. This will not only add the new step that was added in our batch process publish, but set our batch process to that step.



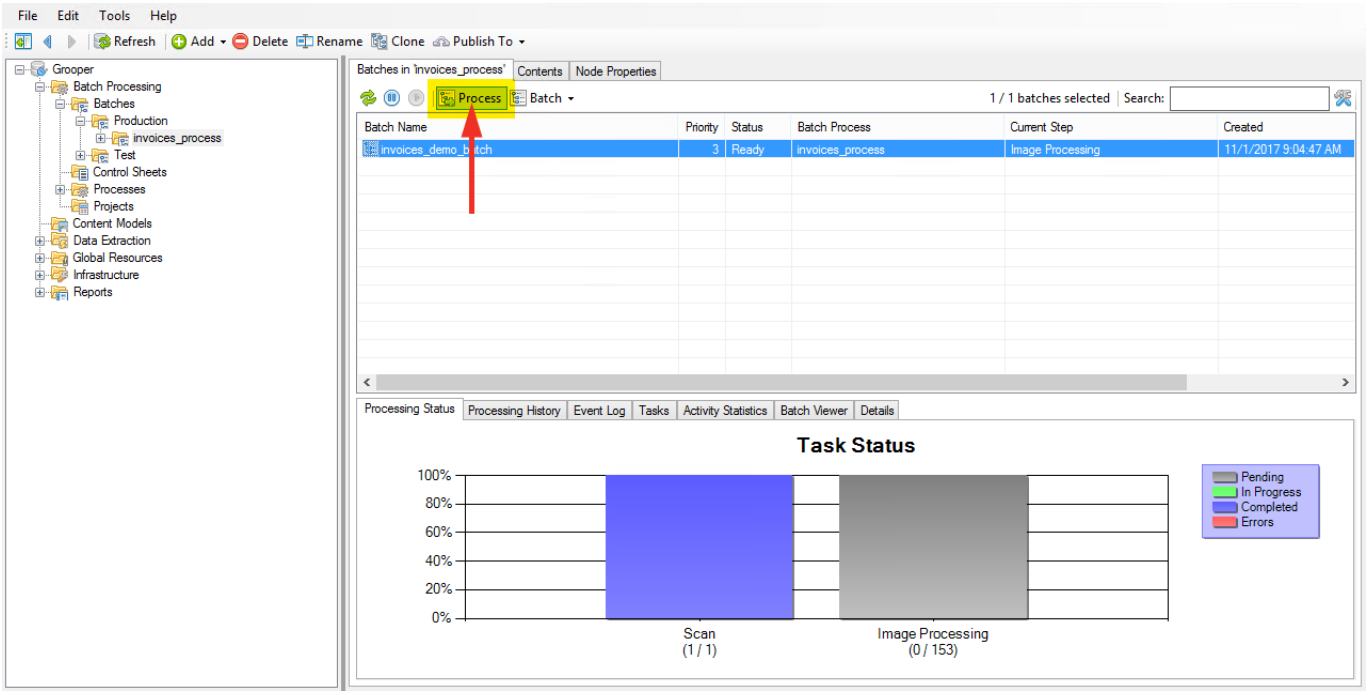
STEP 18 – BATCH PROCESS UPDATED

You'll notice in the Task Status section there's now a step reflecting Image Processing, but no progress bar. We paused our batch before leaving Grooper Dashboard earlier which is good because it's required to update a batch or reset a step, but it needs to be un-paused to move forward. Press the play button to un-pause the batch. Click Yes in the confirmation window that appears.



STEP 19 – PROCESS NEW BATCH STEP

Our Task Status section now has a bar to represent Image Processing but it is not being processed. Press the Process button.



STEP 20 – IMAGE PROCESSING COMPLETE

The Grooper Unattended Client window will appear. This is an application that runs to accomplish tasks a human won't interact with. You can configure Grooper Configuration to have services that run in the background that pick-up processes this application is intended for, but we'll cover that later. As the process runs, you'll see the status bar for the Image Processing step fill up, and some information will update in the Grooper Unattended Client window. When the process is complete, the Grooper Unattended Client window will close.

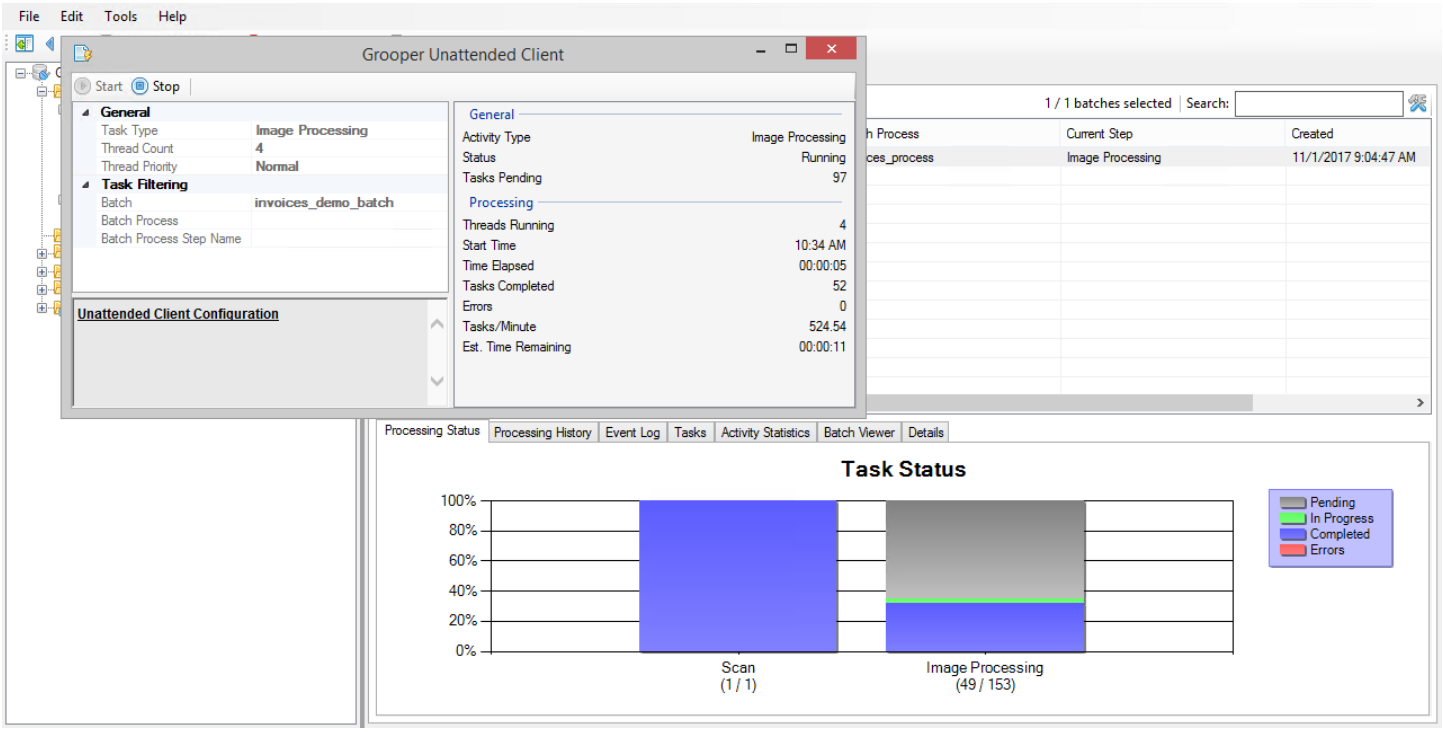


IMAGE REVIEW

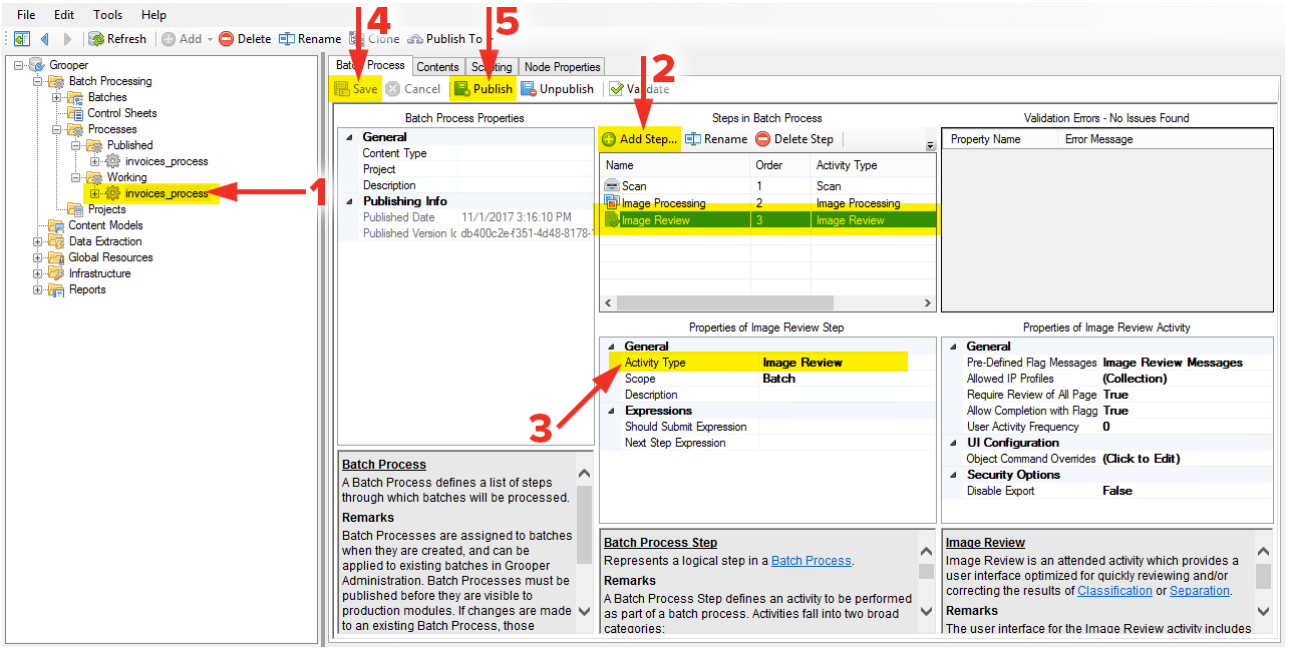
WHAT IS IMAGE REVIEW?

We've now seen both an attended step (in our case from earlier, we performed scanning) and an unattended step. While the main goal of automation and procedural approaches is to limit or eliminate human interaction, there are cases where the discretion of a person is required. With the images cleaned up, we'll want to verify they meet our standards, and with an example batch like this, also make sure there aren't pages present we don't need, like blank ones. We'll now step into another attended step, Image Review.

HOW TO CONFIGURE AND PERFORM IMAGE REVIEW

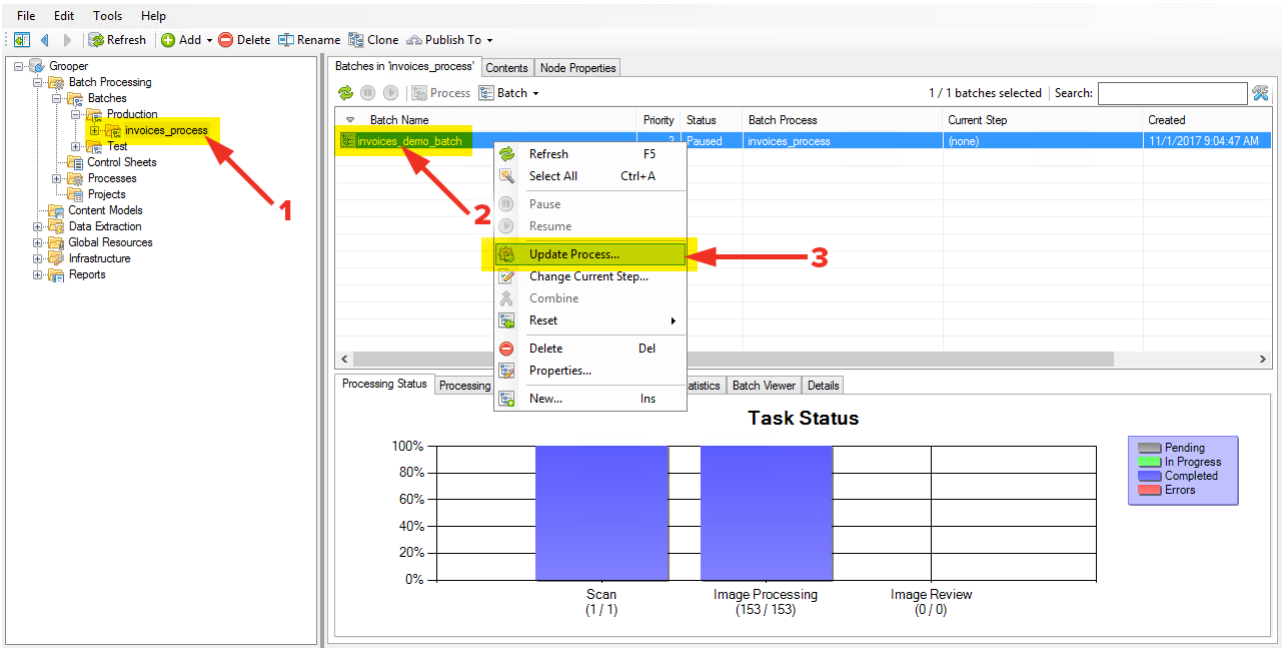
STEP 1 – ADD IMAGE REVIEW TO BATCH PROCESS

Open the node tree up to Grooper – Batch Processing – Processes – Working and **(1)** select the invoices_process batch process. **(2)** Add a step, **(3)** set its Activity Type to Image Review and leave the rest of the fields as default. **(4)** Be sure to save the process **(5)** and publish it.



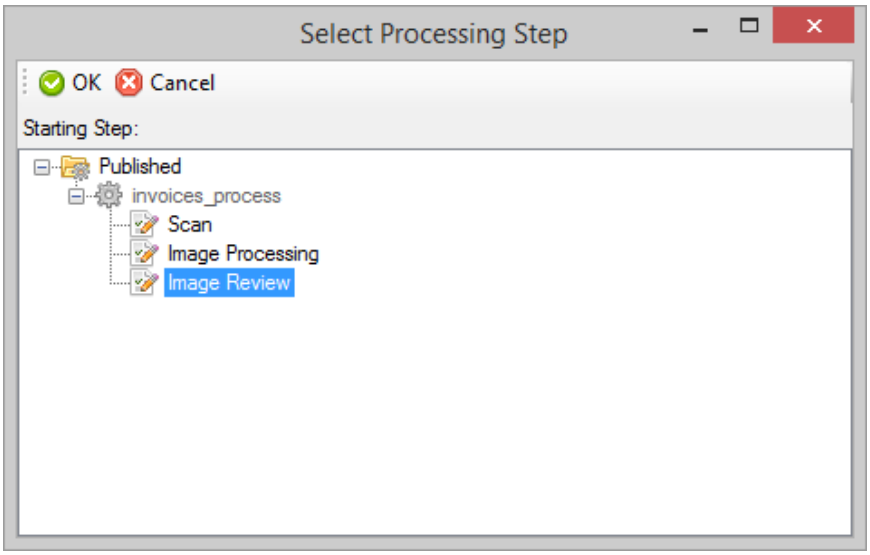
STEP 2 – UPDATE BATCH

Go back to Grooper – Batch Processing – Batches – Production and **(1)** select the invoices_process folder node. In the list of batches on the right, **(2)** select the invoices_demo_batch batch. Make sure it's paused, and using either the Batch drop down, or the right-click context menu method, let's **(3)** update our process again.



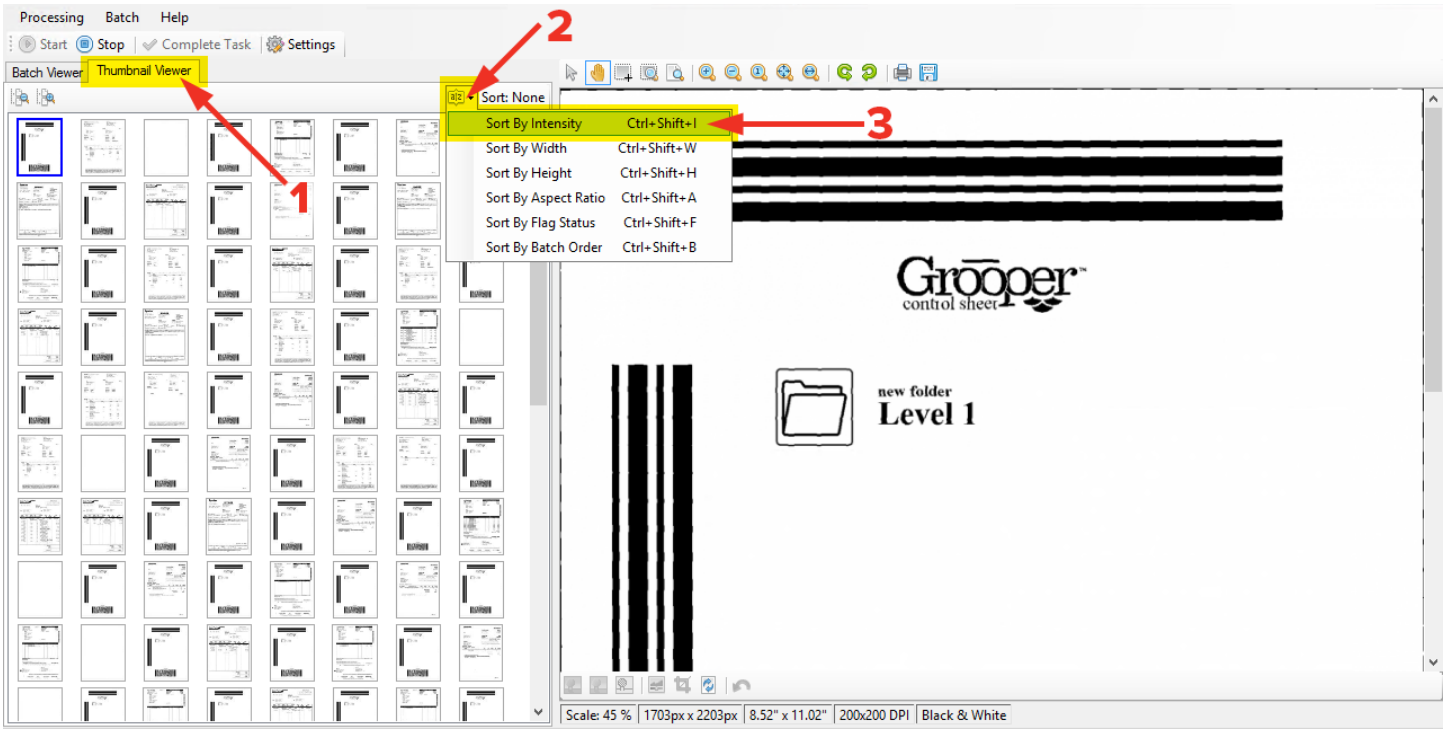
STEP 3 – SELECT PROCESSING STEP

In the Select Processing Step window that appears, select the new Image Review step and click Ok. When the window closes, start the batch back up by pressing the Play button, and follow that up by pressing Process.



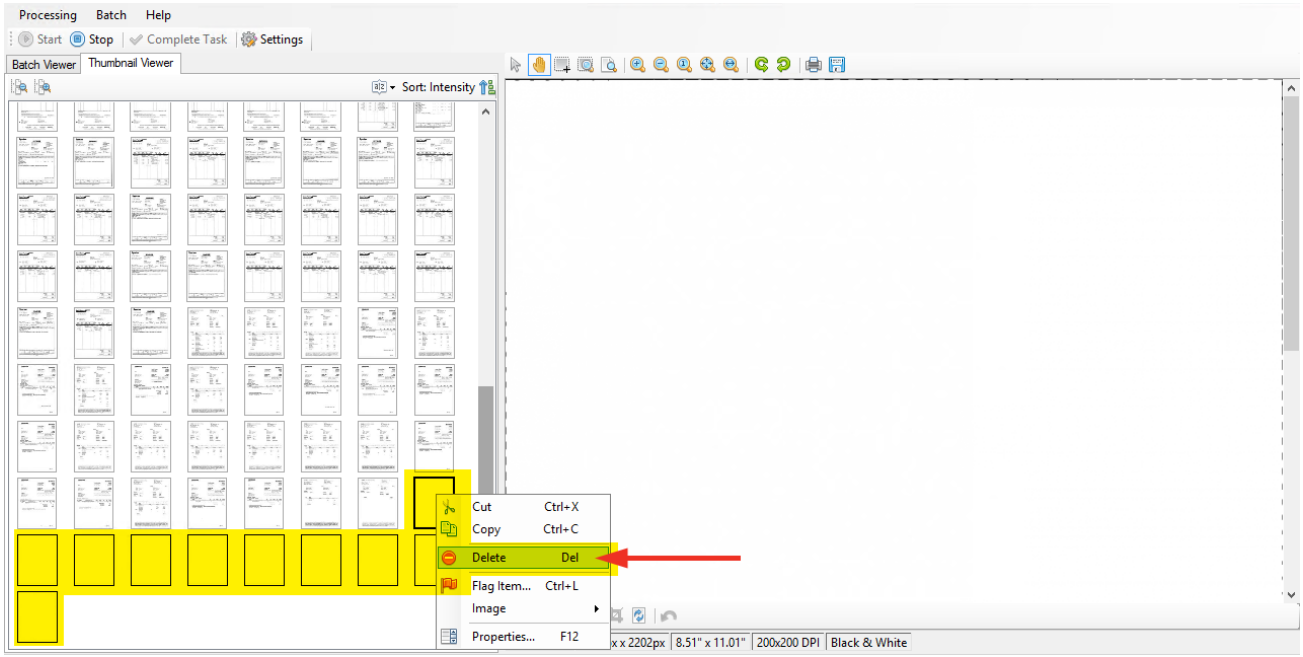
STEP 4 – SORT BY INTENSITY

The Grooper Attended Client application will launch, and this time in its Image Review configuration. **(1)** Click on the Thumbnail Viewer tab, which is used to see the batch contents from a different perspective and allow changes that won't affect the structure of the batch. From here it's easy to delete all the blank pages because of a previous step we ran during Image Processing. All blank pages will have a higher intensity value than pages that have any amount of other color, so if we sort by intensity, we'll have all our blank pages in one area, making it easy to delete them. **(2)** Click the sort dropdown button, and **(3)** select Sort by Intensity.



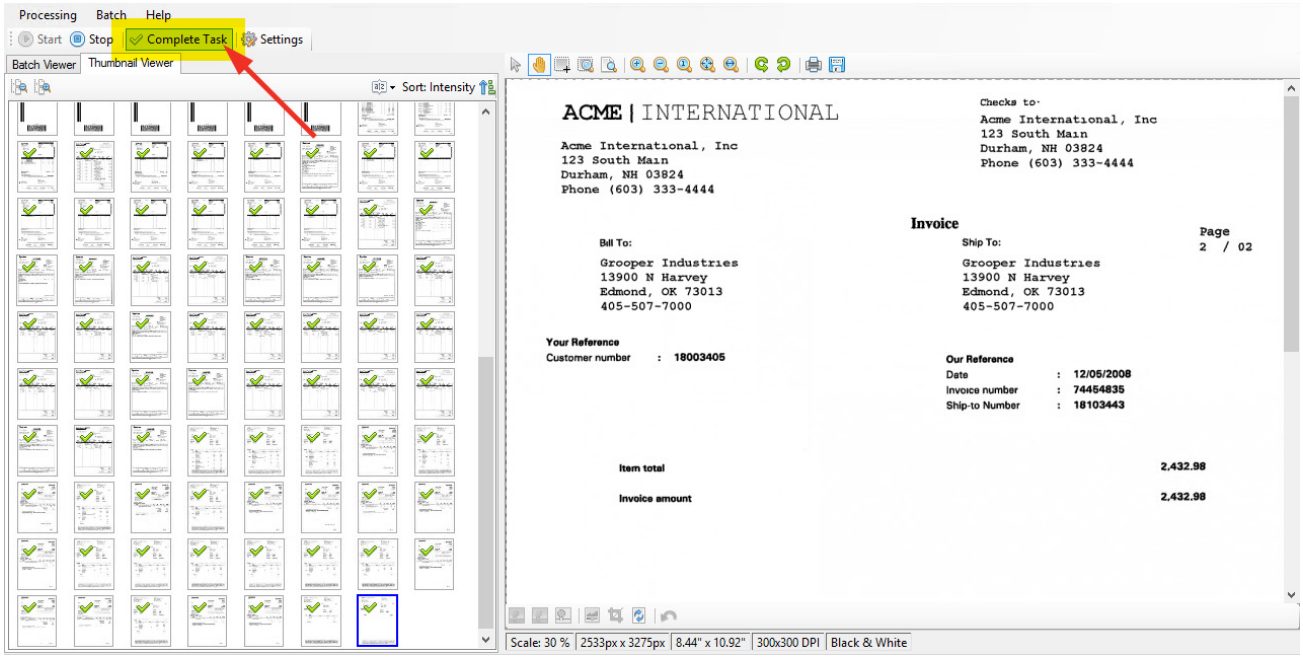
STEP 5 – DELETE BLANK PAGES

All of the blank pages will be at the bottom of our thumbnail list. Select them all, right-click one of them, and in the drop-down menu select Delete.



STEP 6 – COMPLETE TASK

To move forward, all the pages need to be reviewed. This involves simply pressing Enter on a page. In the Thumbnail Viewer tab, you can start at the top and hold Enter until all the images have green checks on the pages representing they've been reviewed. (In the step we added to our process, requiring all pages to be reviewed is an option we could have disabled.) With all the pages reviewed, click Complete Task. Click Yes in the Confirmation window. This will close the Grooper Attended Client, and back in Grooper Administration, go ahead and pause the batch.



FULL-TEXT OCR

WHAT IS OCR?

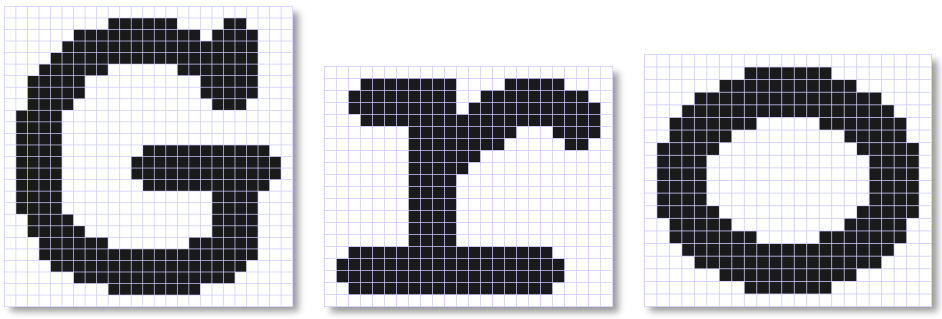
OCR is an acronym for optical character recognition. To put it simply, it's how the computer reads letters from images. I would imagine before someone learns to read, that letters on paper simply look like a bunch of symbols without meaning. For a computer it's even worse. For a scanned document, a computer doesn't even know the symbol is a letter, but instead just a combination of pixels. The OCR process is how the computer takes an image and, line by line, finds combinations of pixels that it ultimately determines are letters, spaces, special characters, etc. – i.e. machine print.

WHEN DO I NEED TO OCR?

To use all the power of **Grooper**, pages must be OCRed. There are perhaps very simple batch processes that could be created that utilized human interaction for every step, and in that case OCR would not be required, but that's like buying a Ferrari to take trips to the grocery store – not practical. All the power of **Grooper** from separation, to classification, to extraction, and so forth, leverages the computer's ability to read the document, and to that end require a page be OCRed.

HOW DOES OCR WORK?

Check out this video on Wikipedia. It's a nice demonstration of a very manual approach to OCR. You can see the person live scanning the document, line by line, and the computer "reads" the letters as the scanning happens. Well, with **Grooper**, the scanning of the whole page happens at once, so, during the OCR process the image is broken into vertical and horizontal pixel lines to find individual letter characters and spacing lines to find individual letter characters and spacing.



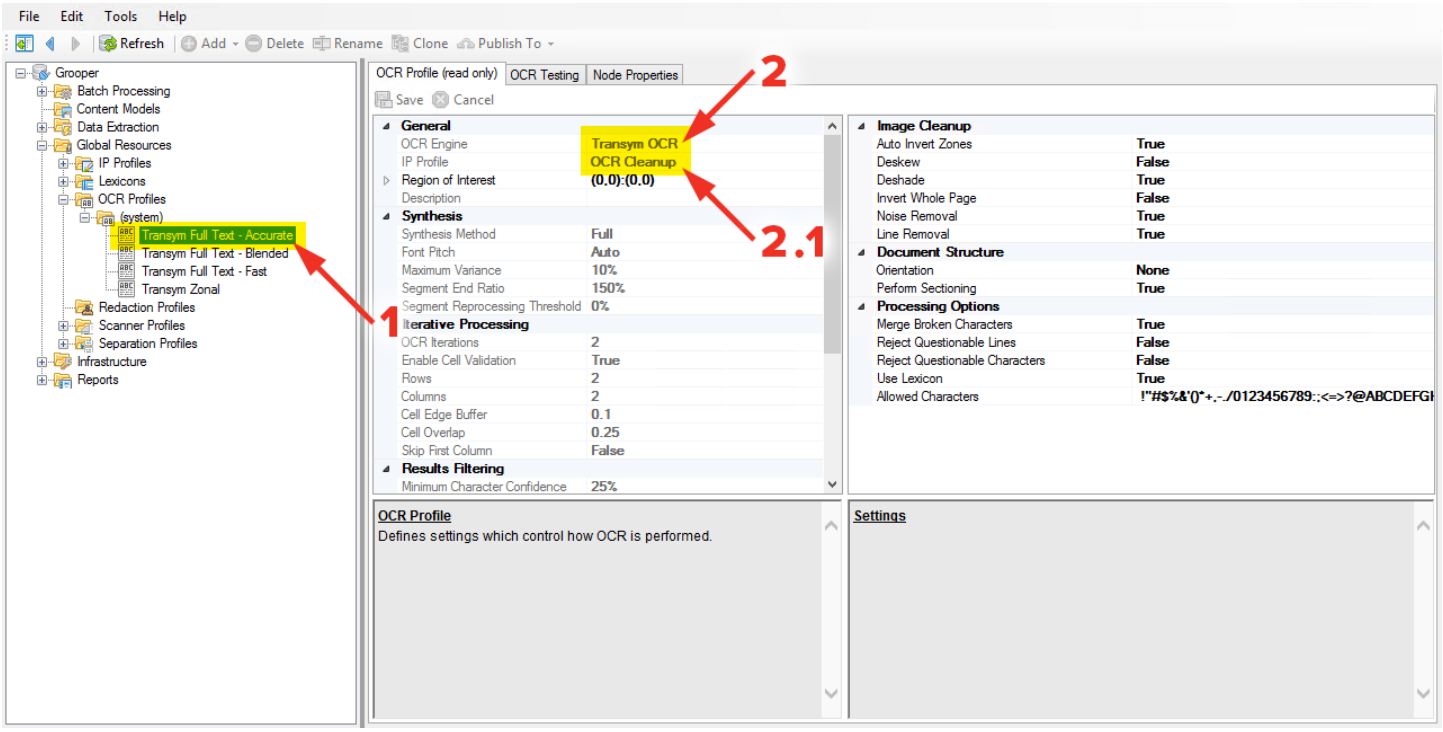
HOW TO CONFIGURE AN OCR PROFILE

Let's continue with the batch we've been working on to get a practical understanding of how to setup a simple OCR Profile.

STEP 1 – OCR SYSTEM PROFILE

Okay, so this one is going to be easy, mostly because it's as good a time as any to introduce system profiles. The **Grooper** development team has built into the software a set of pre-configured profiles, Data Types, lexicons and more to make administration in **Grooper** as streamlined as possible. These nodes always exist in folders labeled (system) and are locked to prevent editing. It's important to note that one should never create profiles and store them in (system) folders.

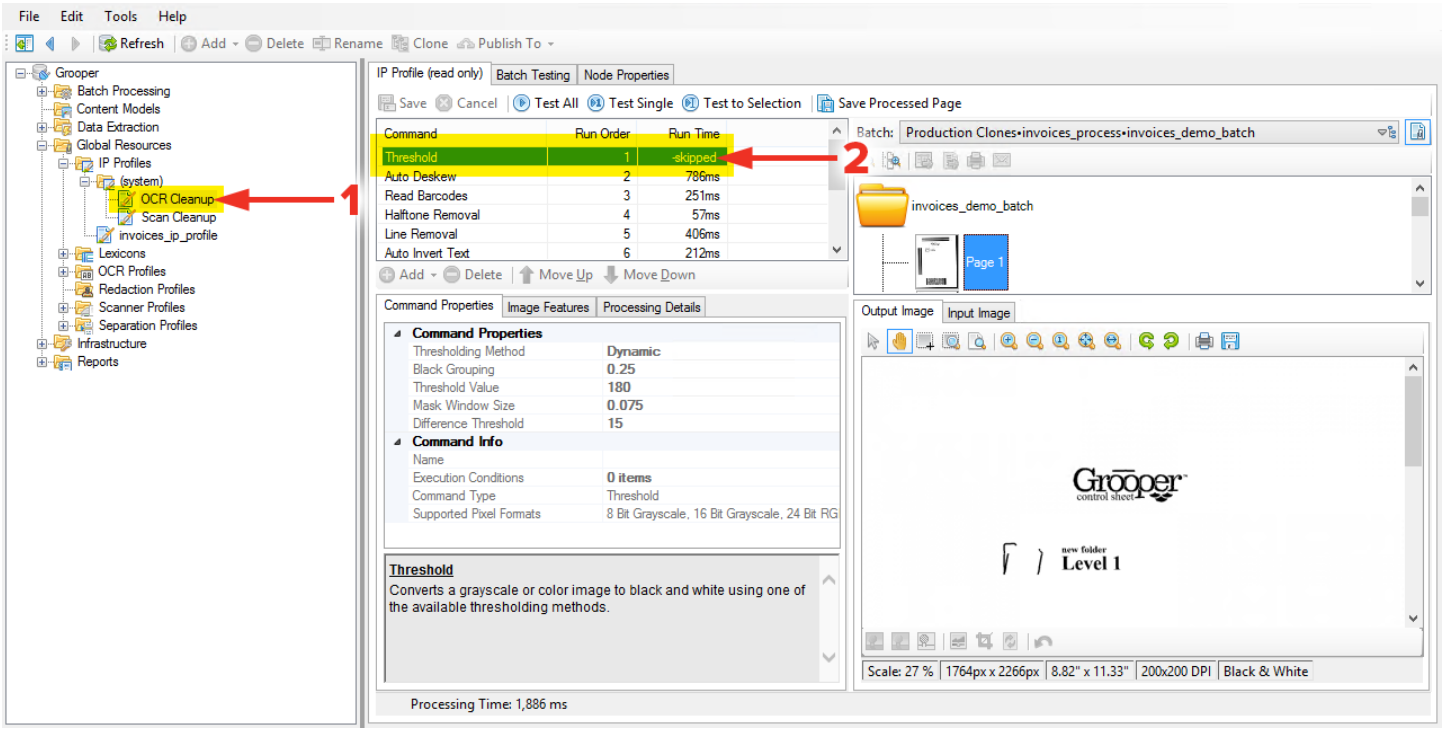
Expand the node tree out to Grooper – Global Resources – OCR Profiles – (system) and **(1)** select the Transym Full Text – Accurate node. Like other things covered so far, there are a lot of settings in here that will be covered in more detail in future lessons, and since it's a system profile it'll all stay as is. **(2)** It's worth pointing out the two most important settings, however: OCR Engine and IP Profile.



STEP 1B – IP PROFILE FOR OCR

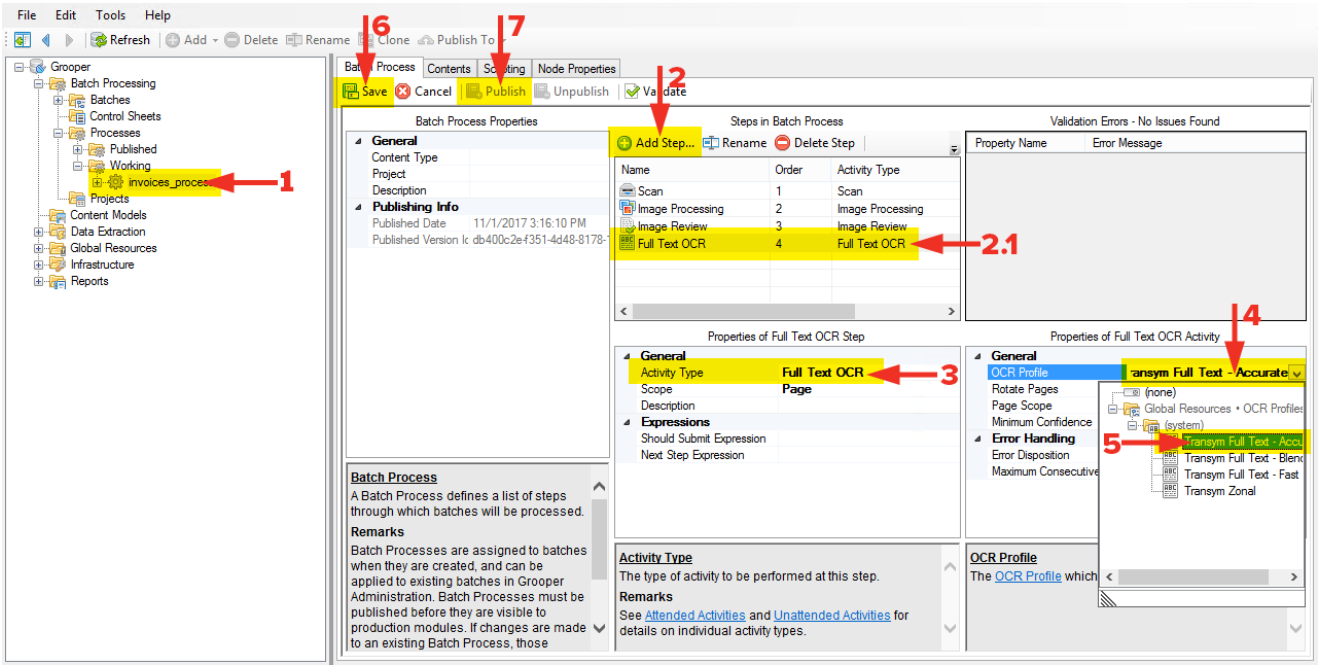
Earlier we discussed permanent vs OCR-only IP Profiles. This is where we set the latter. The system OCR Profile we selected has an IP Profile set (OCR Cleanup), so let's look at it. Navigate to Grooper – Global Resources – IP Profile – (system) and **(1)** select OCR Cleanup. We've setup an IP Profile now, so this should look familiar. We don't need to adjust this IP Profile, but I wanted to call attention to **(2)** the most important step in this profile for the OCR process: Threshold. This converts an image to black and white only, no other values of gray. OCR will only run on black and white images.

So we've looked at the (system) profiles, but haven't set anything yet. Let's move on to setting up an OCR step for our batch process.



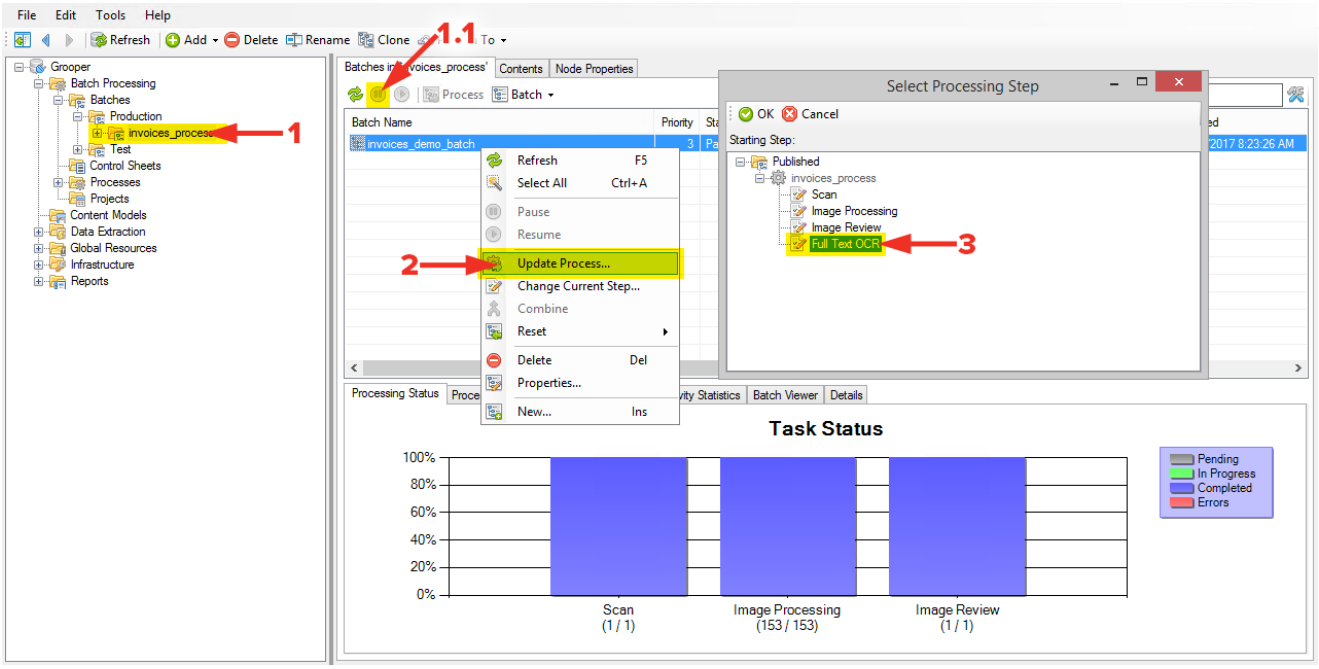
STEP 2 – ADD OCR BATCH PROCESS STEP

(1) Navigate to Grooper – Batch Processing – Processes – Working and select the invoices_process node. **(2)** Add a new step. Change the **(3)** Activity Type to Full Text OCR. **(4)** Click the dropdown for OCR Profile and **(5)** select the (system) profile Transym Full Text – Accurate. Click **(6)** Save and **(7)** Publish the process. Click yes in the Confirmation window.



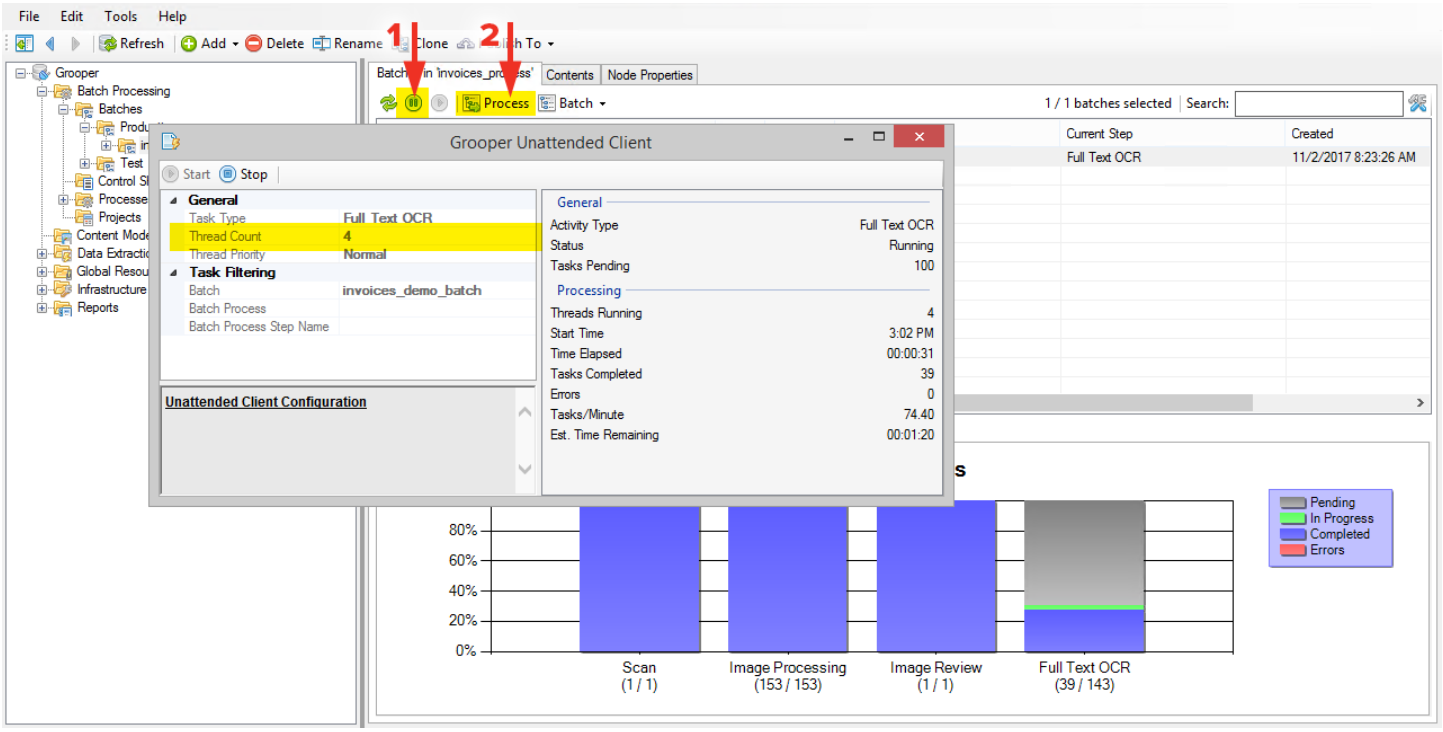
STEP 3 – UPDATE BATCH PROCESS-FULL TEXT OCR

(1) Navigate to Grooper – Batch Processing – Batches – Production, select the invoices_process folder node, and make sure the batch is paused. With invoices_demo_batch selected, **(2)** using either the batch drop down, or the right-click context menu method, let's **(3)** update our process again.



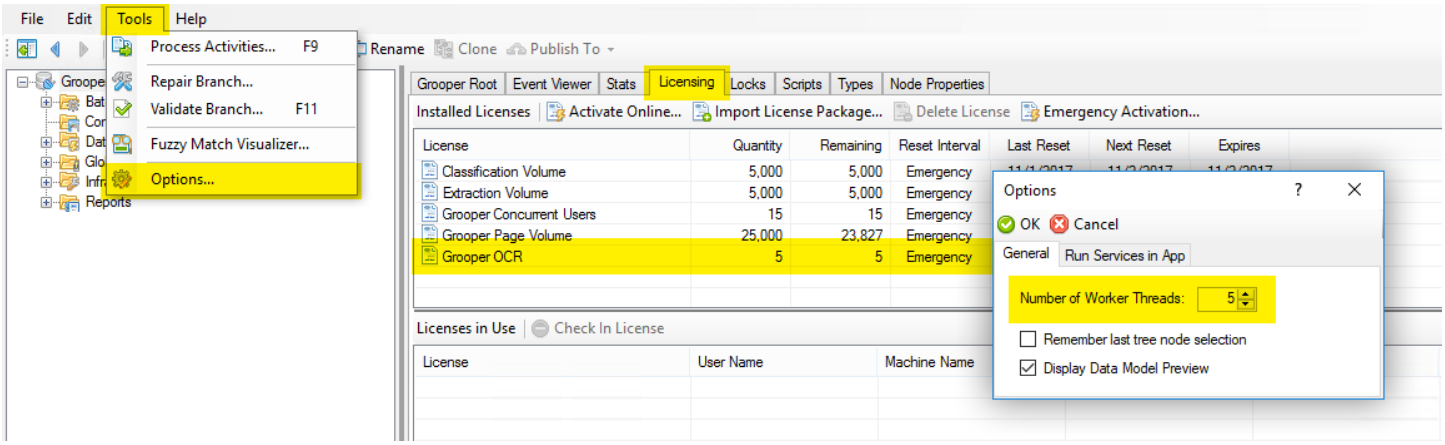
STEP 4 – START AND PROCESS OCR STEP

(1) Start the batch and **(2)** press Process, and we'll again see the Grooper Unattended Client window appear and begin to process our OCR step. When it's done processing and the window closes, we'll now have critical data that we can leverage to move on to the next important steps in our batch process.



Please take note of the highlighted area in the Grooper Unattended Client window (above). It's important that the Thread Count not exceed your maximum Transym OCR Concurrent Processes, otherwise it will error out. From the **Grooper** root node, select the Licensing tab and check your licensing allocation.

From the Grooper Root node you can click the Licensing tab to see your concurrent OCR licenses. Click the Tools>Options menu and make sure the Number of Worker Threads doesn't exceed the total quantity of your OCR licensing.



PHASE 3 - ORGANIZE

With content brought in, and prepared for processing, it's important to get it organized. In order to organize the documents, we'll need to identify them as well. The following section will discuss this process.

DOCUMENT SEPARATION

WHAT IS A DOCUMENT?

If I ask Google what a document is I get: *a piece of written, printed, or electronic matter that provides information or evidence or that serves as an official record*. Well, that's dry and boring, so to put it simply a document is an organization of things (in our case pages) that are related into a single idea. A document can be a single page, and without definition it is just that, a page. Let's take, for example, a single page resume. We've probably all seen a resume before, so we'd probably know if we saw one again (that recognition, that's critical). It's the fact that the page is accomplishing a specific goal, obviously with a resume it's to apply for a job, and that I recognize this goal that the page is no longer just a page, but a defined idea, and to that end a document. This could be true of a multipage resume, but because I know the pages are related, they all come together as ONE document.

WHAT IS SEPARATION?

Document Separation is the process of transforming a batch of raw pages into a hierarchy of folders which represent individual documents. Separation is most commonly needed when documents are acquired through scanning, or when raw pages are imported into **Grooper**. This process is typically a precursor to classifying the folders as a content type.

WHY DO I NEED TO DO SEPARATION?

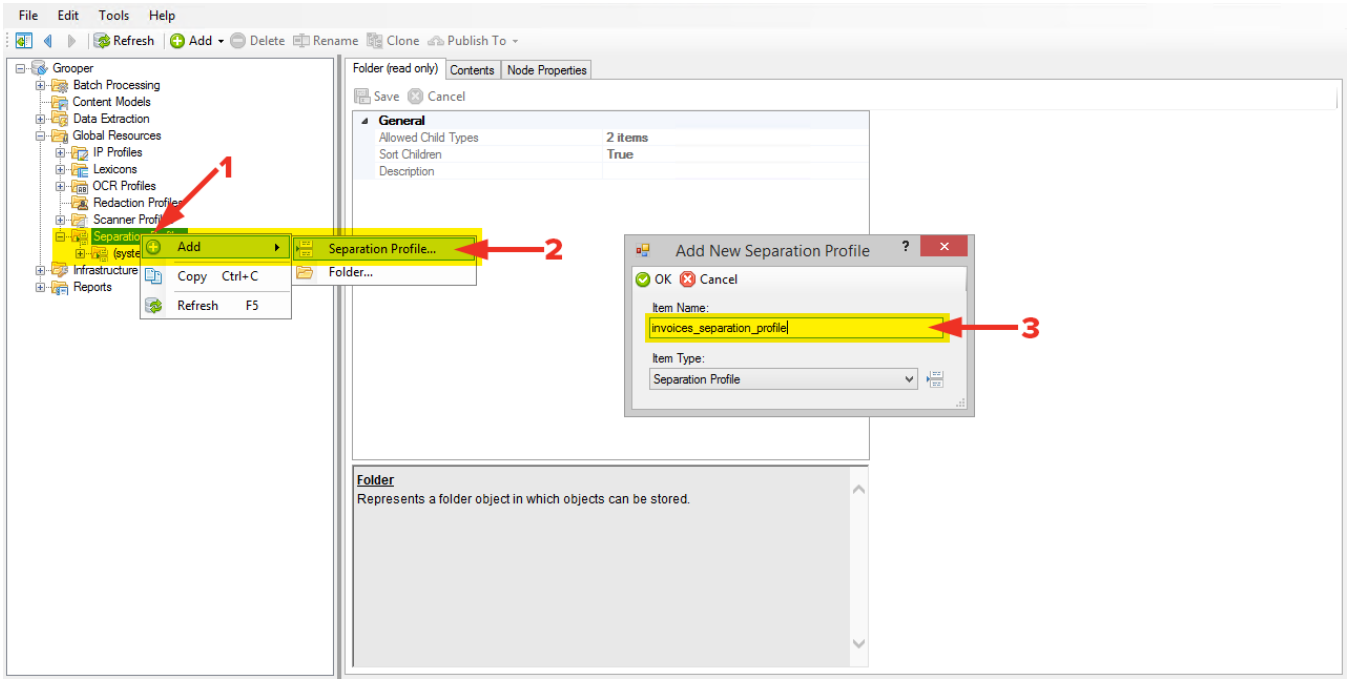
Think about whenever you've had a stack of paper, and decided to group them and either staple them together or put them in folders. This grouping is done as a desire to recognize those individual pages as an ordered group with a shared meaning. This implied meaning we'll discuss in a moment as Classification. There's a technical requirement that something be recognized as a document, and separated as such, before we can classify it and ultimately assign fields to it that we can extract information into and later organize as a result.

You can separate beyond one level as well. For example, I can have a person's set of documents that would belong to the Human Resources department of a specific company. The company would be organization level 1, the department would be level 2, and the individual person would be level 3.

HOW TO CONFIGURE AND PERFORM SEPARATION

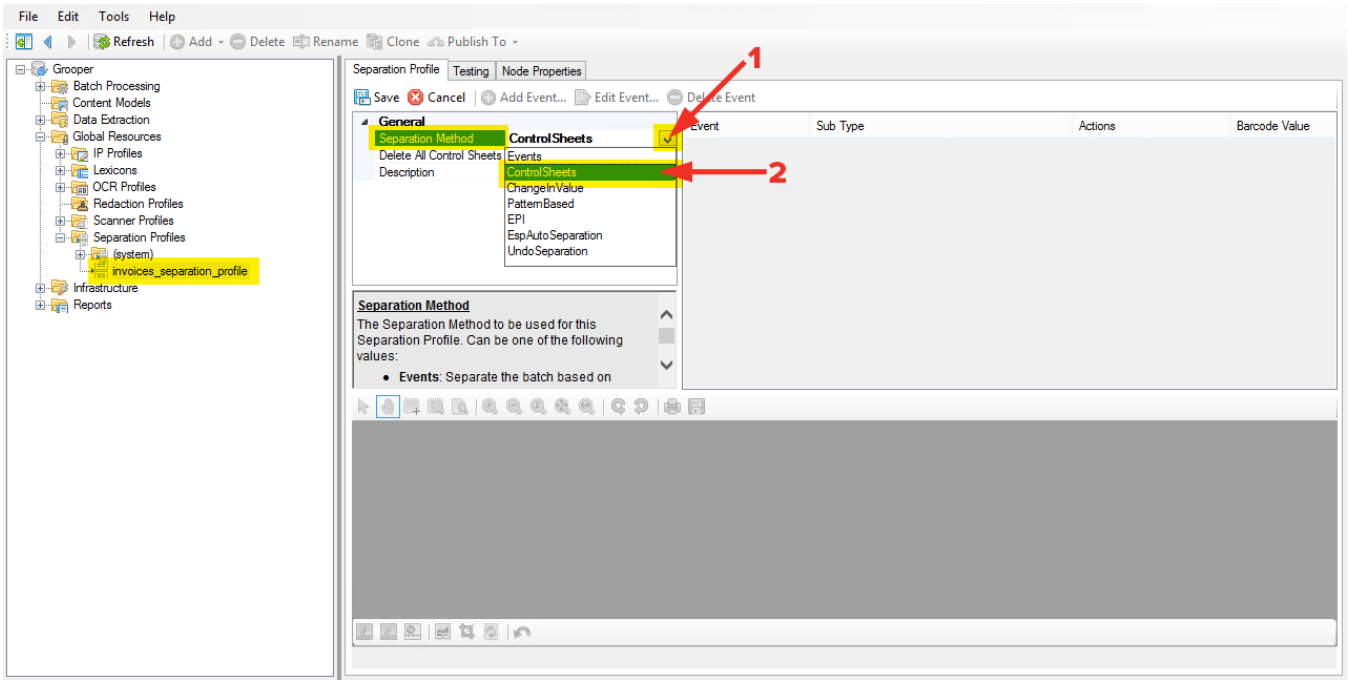
STEP 1 – ADD SEPARATION PROFILE

(1) Navigate to Grooper – Global Resources then select and expand Separation Profiles. **(2)** Use the Add dropdown or right-click and select Add > Separation Profile. **(3)** In the Add New Separation Profile window, name the profile invoices_separation_profile.



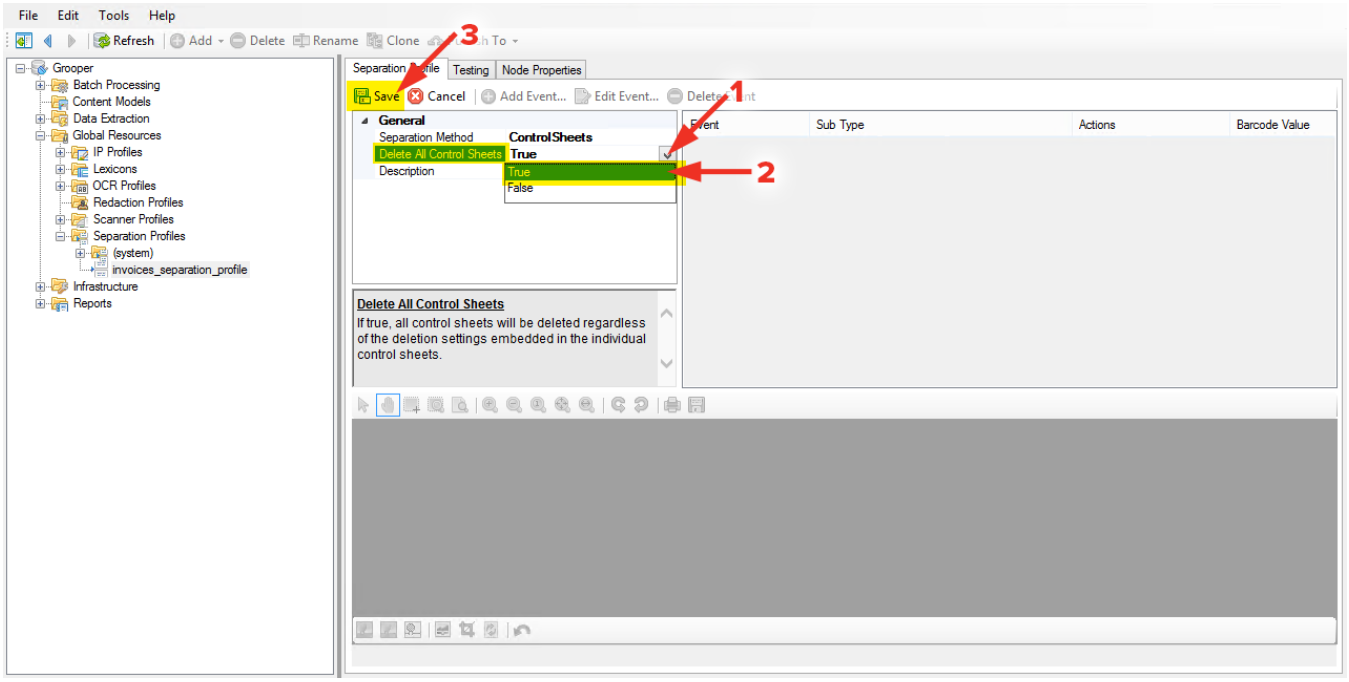
STEP 2 – SETTING SEPARATION METHOD

(1) Click the dropdown for Separation Method and **(2)** choose ControlSheets.



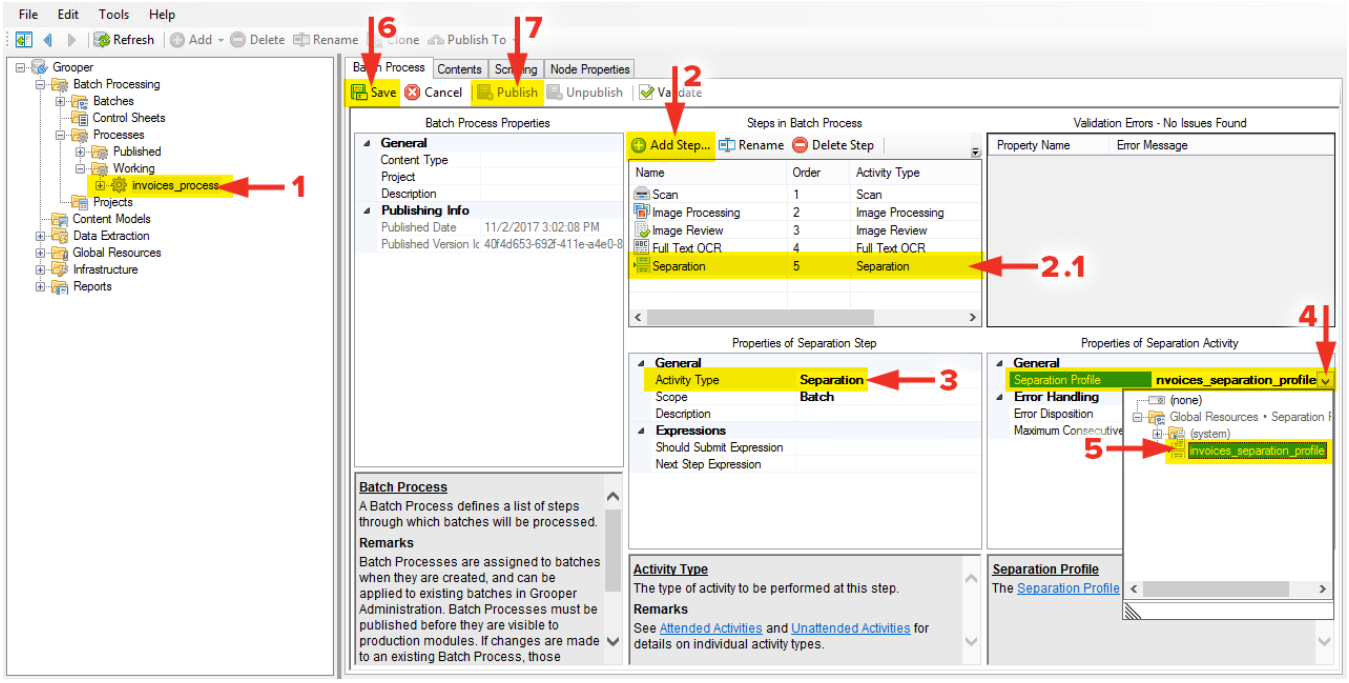
STEP 3 – DELETE CONTROL SHEETS

Going forward, I know we won't need the control sheets to exist in our batch as they were leveraged for this specific purpose, and nothing else. **(1)** On the dropdown for Delete All Control Sheets option, choose **(2)** True. **(3)** Save the profile.



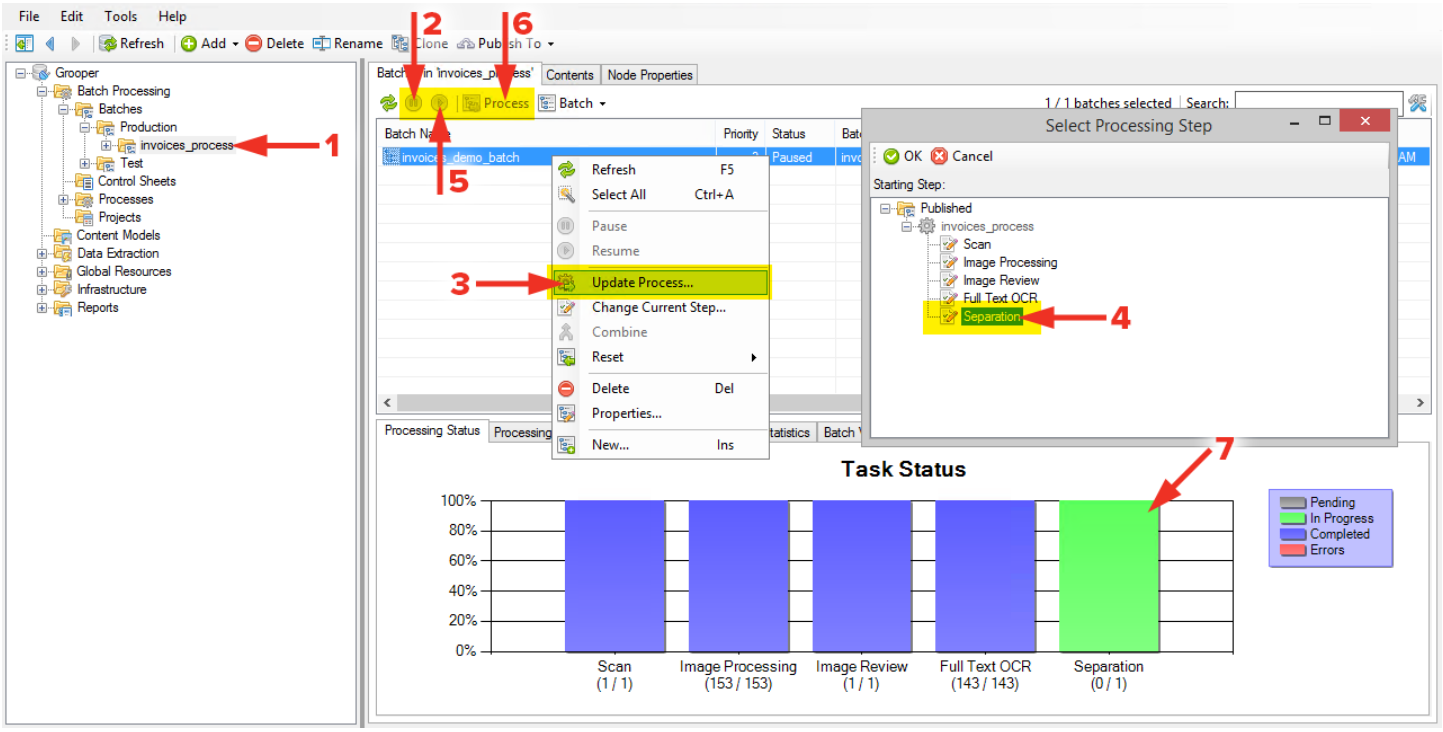
STEP 4 – UPDATE PROCESS

(1) Navigate to Grooper – Batch Processing – Processes – Working and select invoices_process. **(2)** Add Step and **(3)** set the Activity Type to Separation. **(4)** Click the dropdown for the Separation Profile and **(5)** select invoices_separation_profile. **(6)** Save the process and **(7)** Publish it.



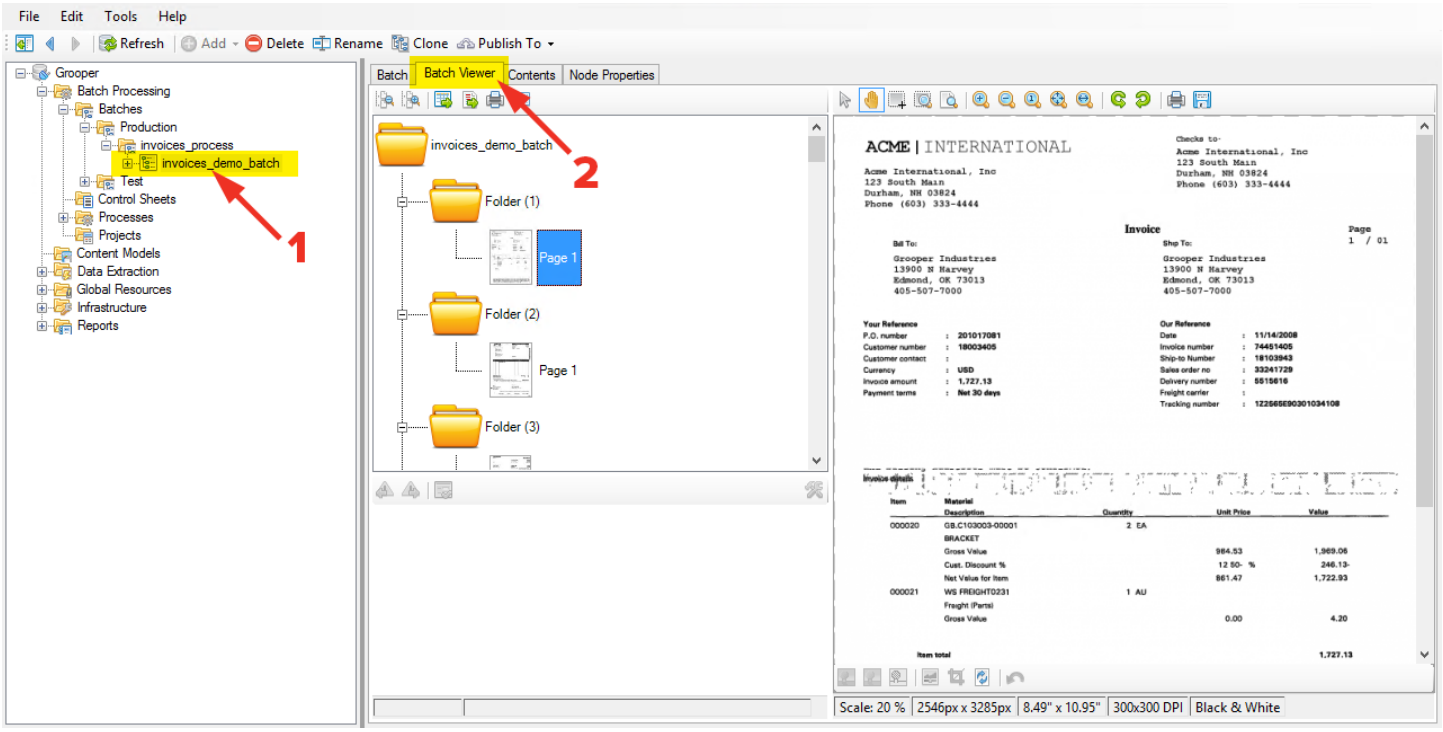
STEP 5 – UPDATE BATCH PROCESS

(1) Navigate to Grooper – Batch Processing – Batches – Production and select invoices_process. **(2)** Select the invoices_demo_batch and make sure it's paused. **(3)** Use the Batch dropdown menu, or right-click and select Update Process. **(4)** Within the Select Processing Step window, select Separation and click Ok. **(5)** Resume the batch and **(6)** press Process. The Grooper Unattended Client window will appear and show the progress of the step running, and you'll also see **(7)** the new progress bar for the separation step give feedback. Pause the batch when the activity completes.



STEP 6 – SEEING RESULTS

Navigate to Grooper – Batch Processing – Batches – Production – invoices_process and **(1)** select the invoices_demo_batch in the node tree, then **(2)** click the Batch Viewer tab. You'll notice now what was loose pages are now organized in folders. Had our documents had more than one page between our Grooper Control Sheets, our folders would have multiple pages in them. Pages being in folders is important because defining a page, or group of pages, within this folder as a document can now occur. Loose pages cannot have Document Types assigned to them. Classification is coming up next, which will help understand why a folder needs to be assigned a Document Type.



CONTENT MODEL

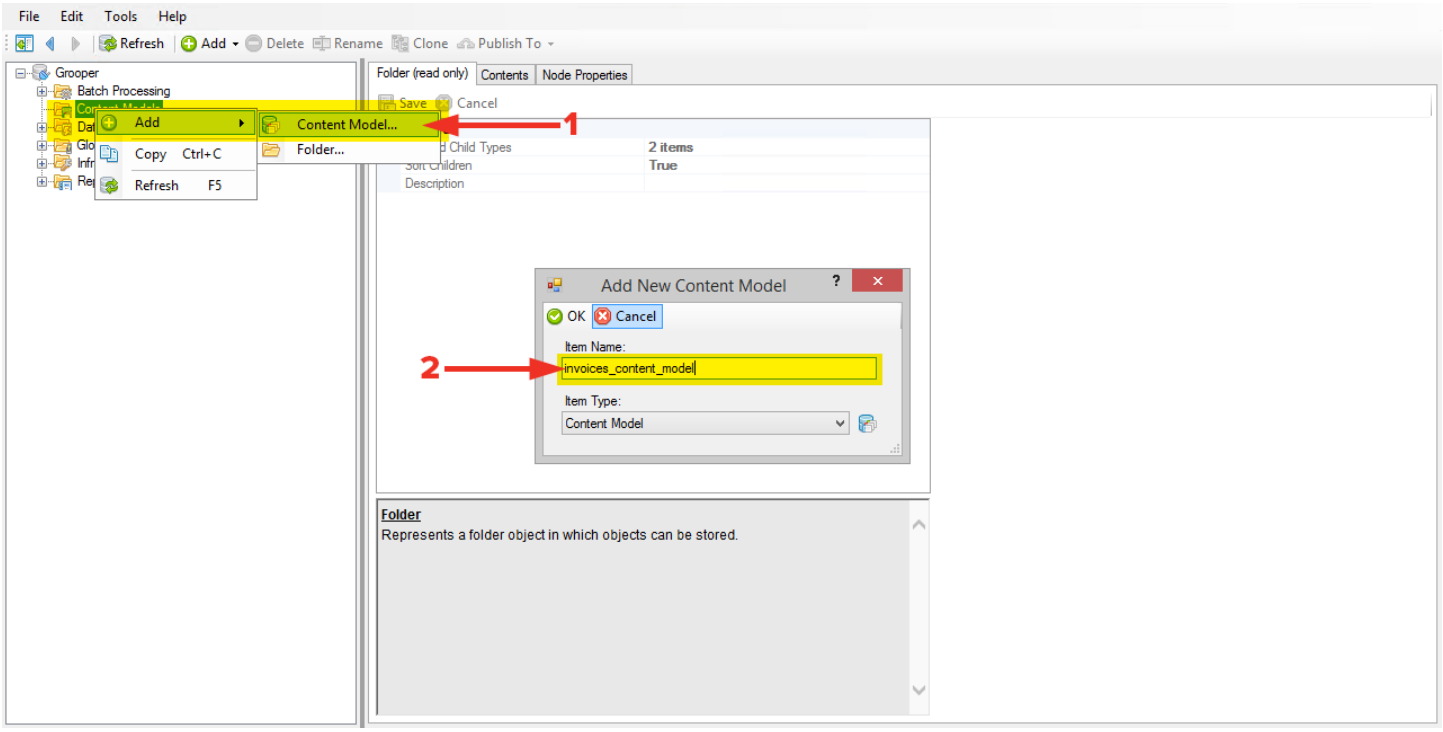
WHAT IS A CONTENT MODEL

The Content Model in Grooper is the main component that houses the information that defines what our documents are, why they are what they are, and fields of information that we want to get from these documents. To elaborate, it describes the taxonomy of a set of documents. It can be a simple as a list of Document Types, or as complex as the records management strategy of a large organization. Content Models are hierarchical in nature, and describe a set of documents in terms of Content Categories and Document Types.

HOW TO CONFIGURE A CONTENT MODEL

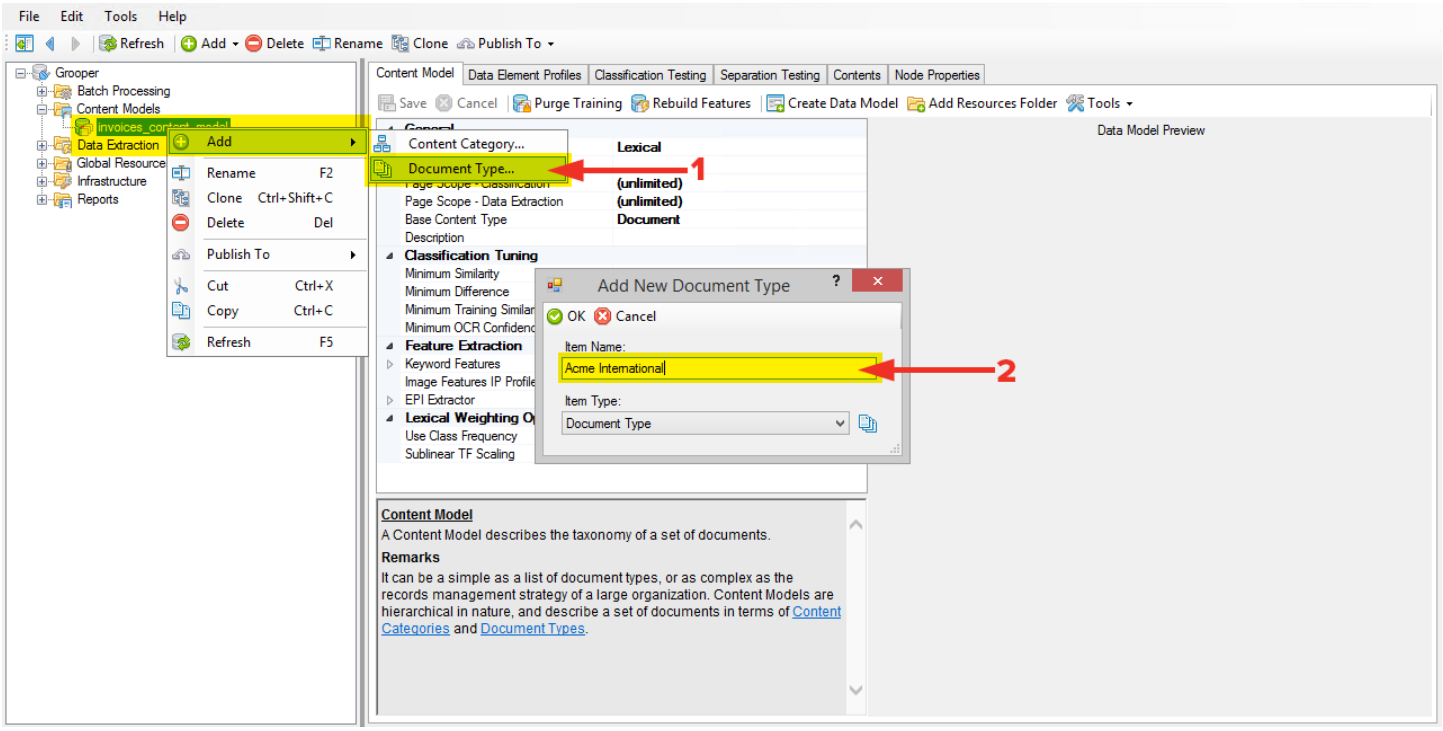
STEP 1 – CREATE NEW CONTENT MODEL

(1) In the **Grooper** root node, select Content Models, and use the Add drop-down, or right-click and select Add > Content Model. **(2)** In the Add New Content Model name it invoices_content_model and click Ok.



STEP 2 – DOCUMENT TYPES

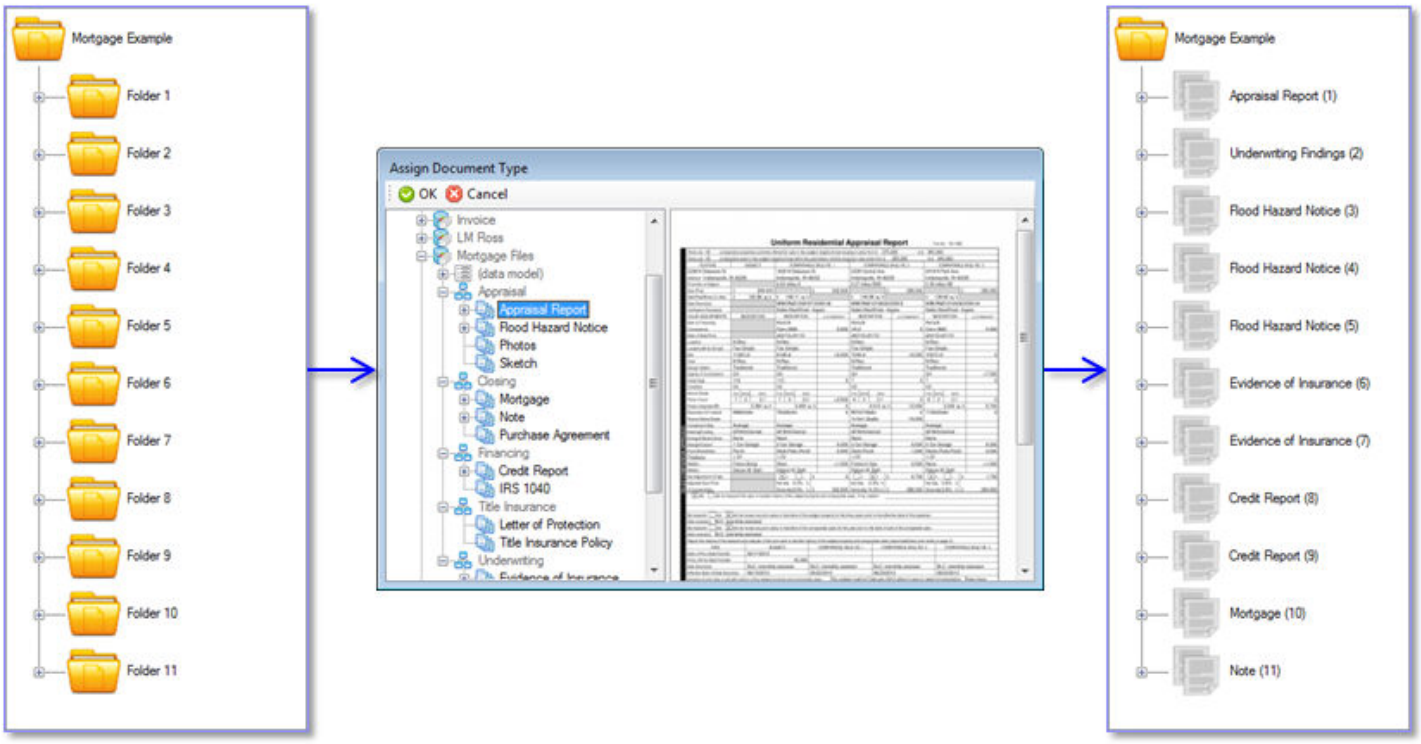
(1) With the new Content Model created and selected, either use the Add drop down menu, or right-click Add > Document Type... . The Add New Document Type window will appear. **(2)** Give the new Document Type a name of Acme International, and click OK. Repeat this process of creating a new Document Type a few more times and add names of the following: Standard, Express, Spartan, and Enid Parts.



CLASSIFICATION

WHAT IS CLASSIFICATION

Classification in **Grooper** is the process of assigning a Content Type to a batch folder. When a folder in a batch is assigned a Content Type, it is thereafter considered a document, and will be displayed in the Batch Viewer using a document icon rather than a folder icon, as shown in the example below.



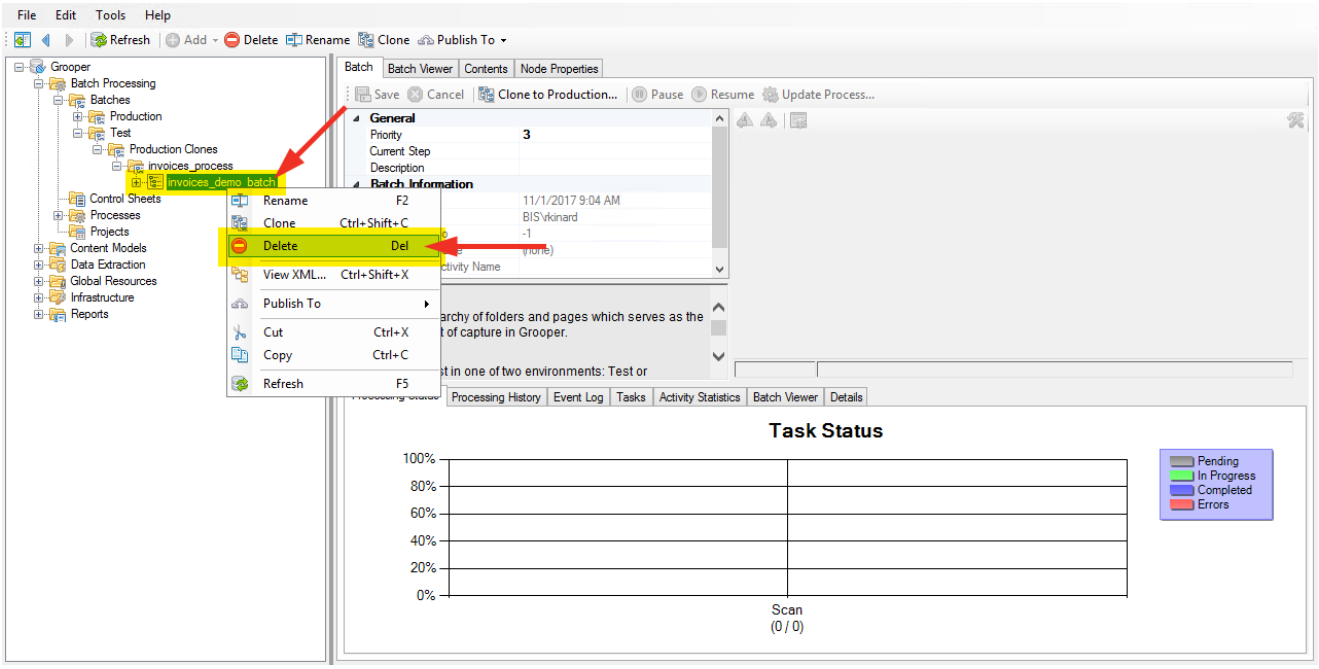
Look back at what was said about Classification during the discussion of Separation. The concepts are mostly covered there, however, it's worth expanding upon here.

So you've got a stack of paper. You've been told to get this stack of paper filed away into three different file cabinets. First, you'd separate the individual pages into folders. With the pages organized, you can now place them into their appropriate file cabinets, but to put them in the correct cabinet, you must understand which one goes where. This idea is essentially what Classification is. Understanding and recognizing a document to be of a specific type so you can further process it.

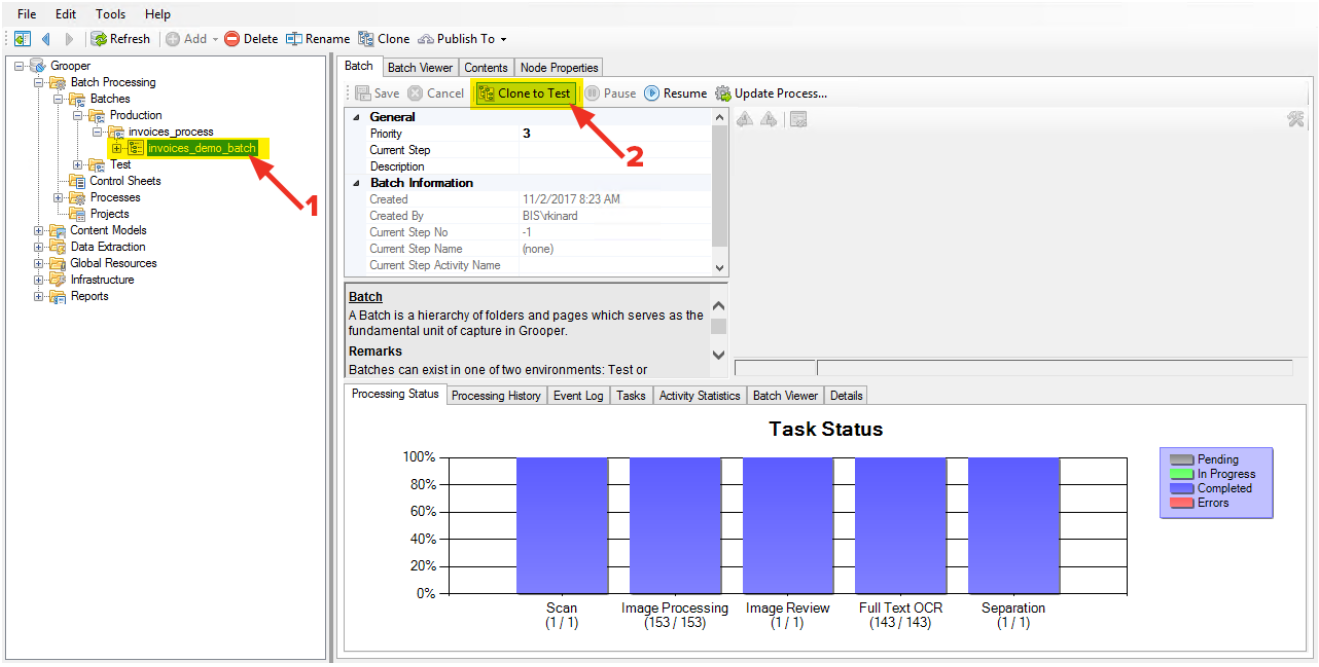
HOW TO CONFIGURE AND PERFORM CLASSIFICATION

STEP 1 – CLONE PRODUCTION BATCH FOR TESTING

We cloned our production batch a while back, but since then have made changes to the production one. To keep things simple, let's delete the clone in our test branch first. Navigate to Grooper – Batch Processing – Batches – Test – Production Clones – invoices_process and select the invoices_demo_batch batch node. Use either the Delete button at the top, or right-click and select Delete.

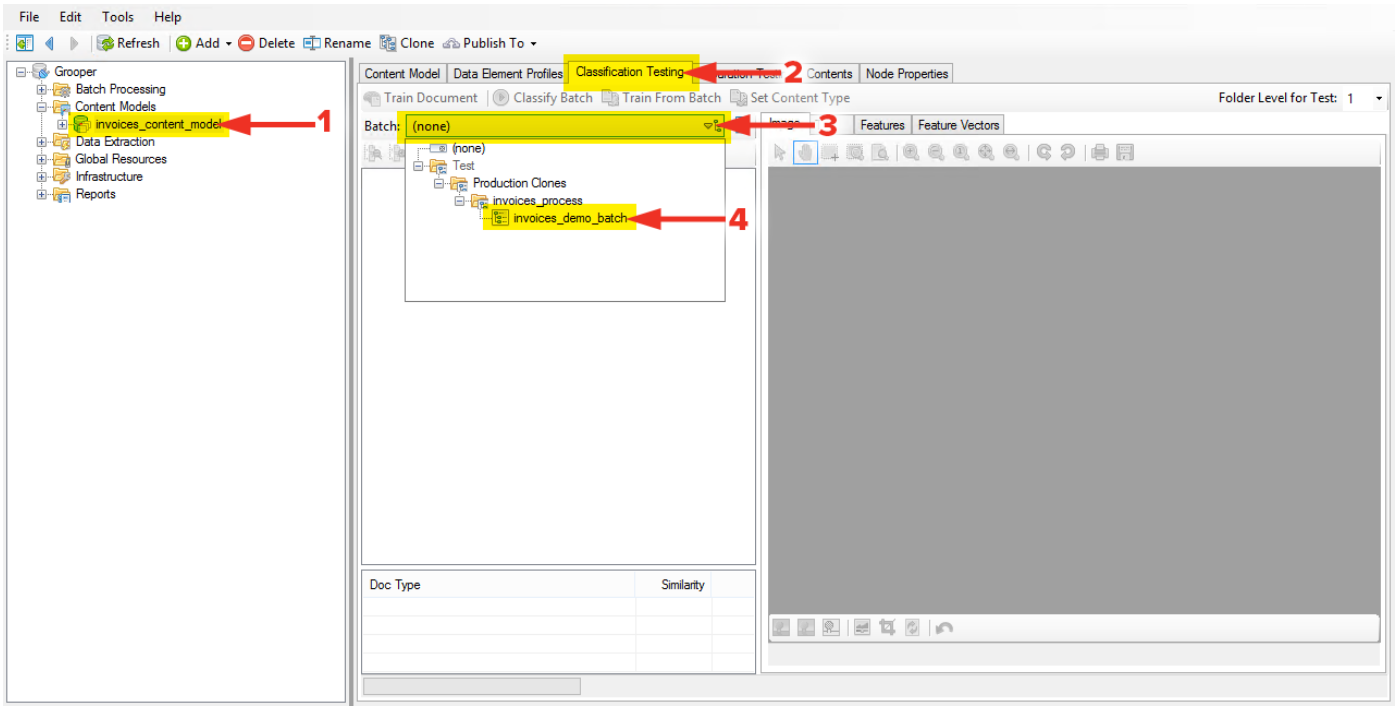


With the old test clone deleted, **(1)** navigate to Grooper – Batch Processing – Batches – Production – invoices_process and select the invoices_demo_batch batch node **(2)** and Clone to Test.



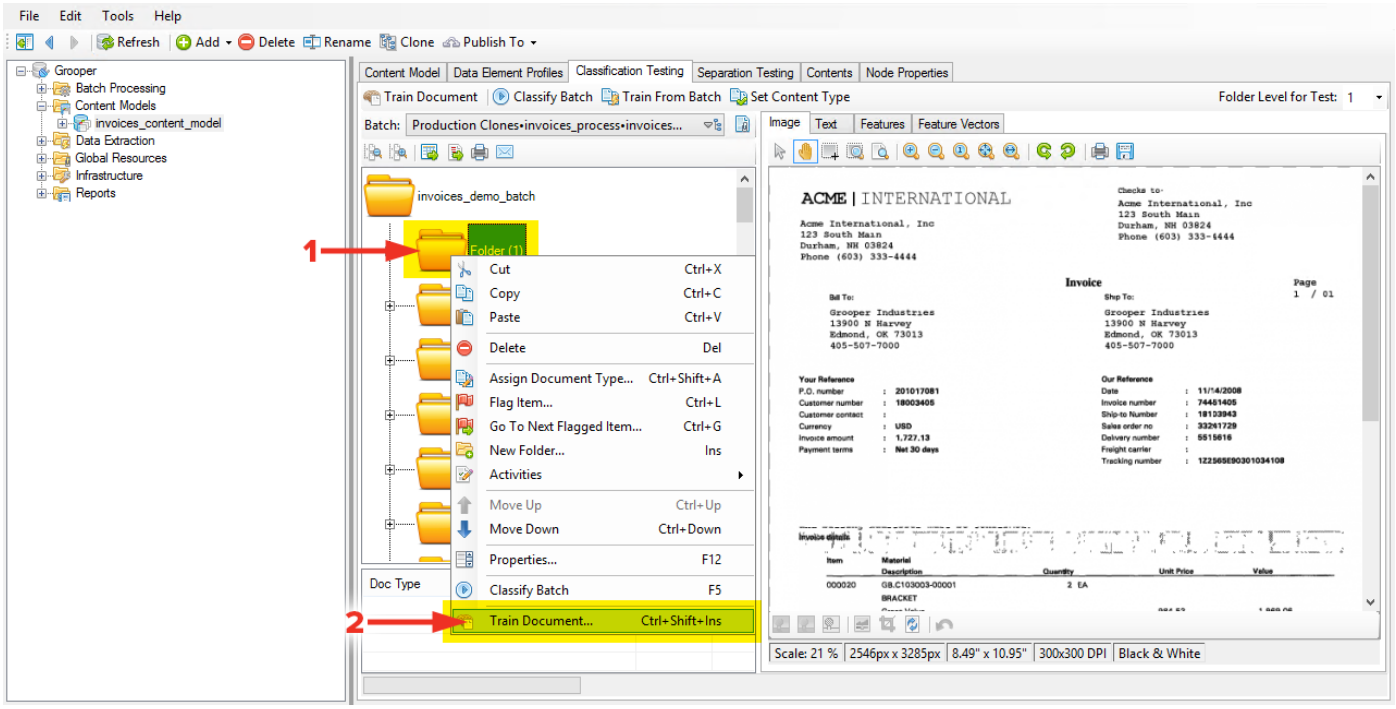
STEP 2 – TRAINING DOCUMENT TYPES

(1) Go back to our Content Model we created and **(2)** select the Classification Testing tab. **(3)** In the Batch: drop down **(4)** select our newly cloned batch.



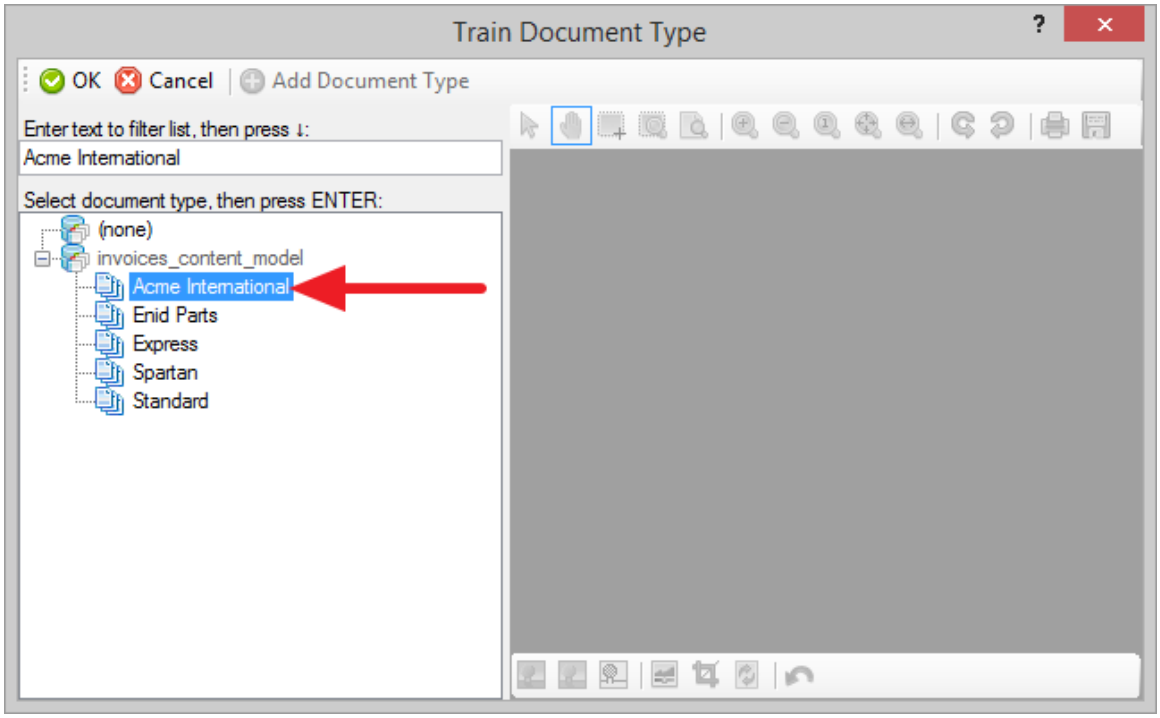
STEP 3 – TRAIN OUR FIRST DOCUMENT

(1) Select the first folder, and we'll see an image of our document on the right. This is clearly a document from Acme International, so either **(2)** use the Train Document button at the top, or right-click and select Train Document...



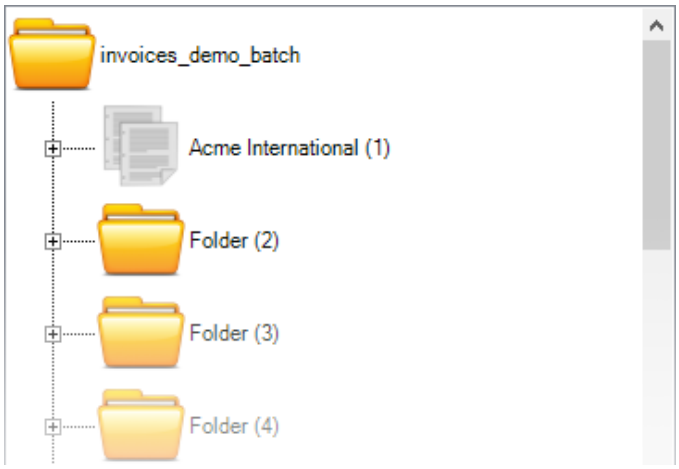
STEP 4 – TRAIN DOCUMENT TYPE WINDOW

The Train Document Type window will appear. Select the Acme International Document Type within our Content Model and click OK. An information window will pop up after confirming the action took effect and that the document is trained.

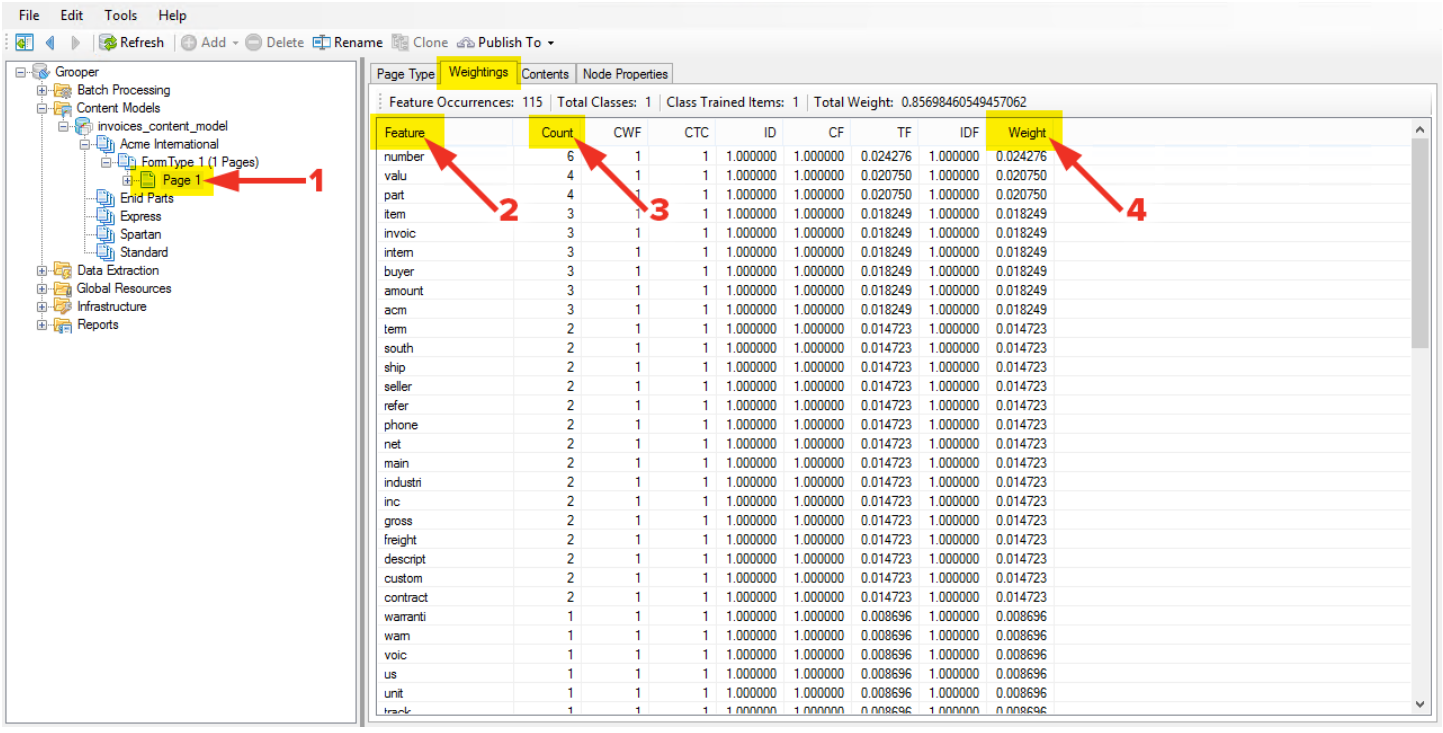


STEP 5 – VIEW WEIGHTINGS

First off, you'll see that the folder icon for what was Folder (1) has now been replaced by a document icon, and that item has been renamed to the Document Type we applied the training to, in this case Acme International.



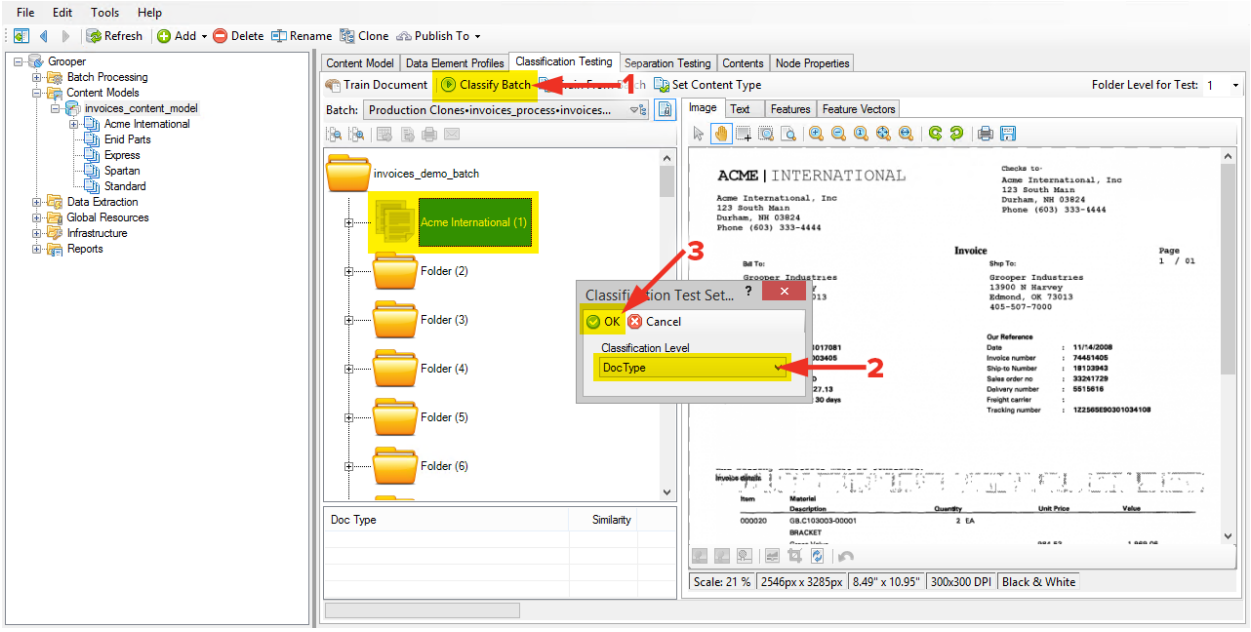
Next, let's take a quick look behind the scenes at what has happened as a result of this classification training. **(1)** Navigate to Grooper – Content Models – invoices_content_model – Acme International – Form Type 1 (1 Pages) and select the node for Page 1. Click on the Weightings tab. **(2)** There are a lot of words in the Feature column, and several columns with information that probably seems like gibberish. We won't concern ourselves with the algorithm that makes up this information, but just focus on the most important ones to consider: Count and Weight. **(3)** Because of the Count of the word (and other factors) **(4)** we get a Weight at the end that determines how and why this document will be considered an Acme International document moving forward.



STEP 6 – CLASSIFICATION TEST ONE

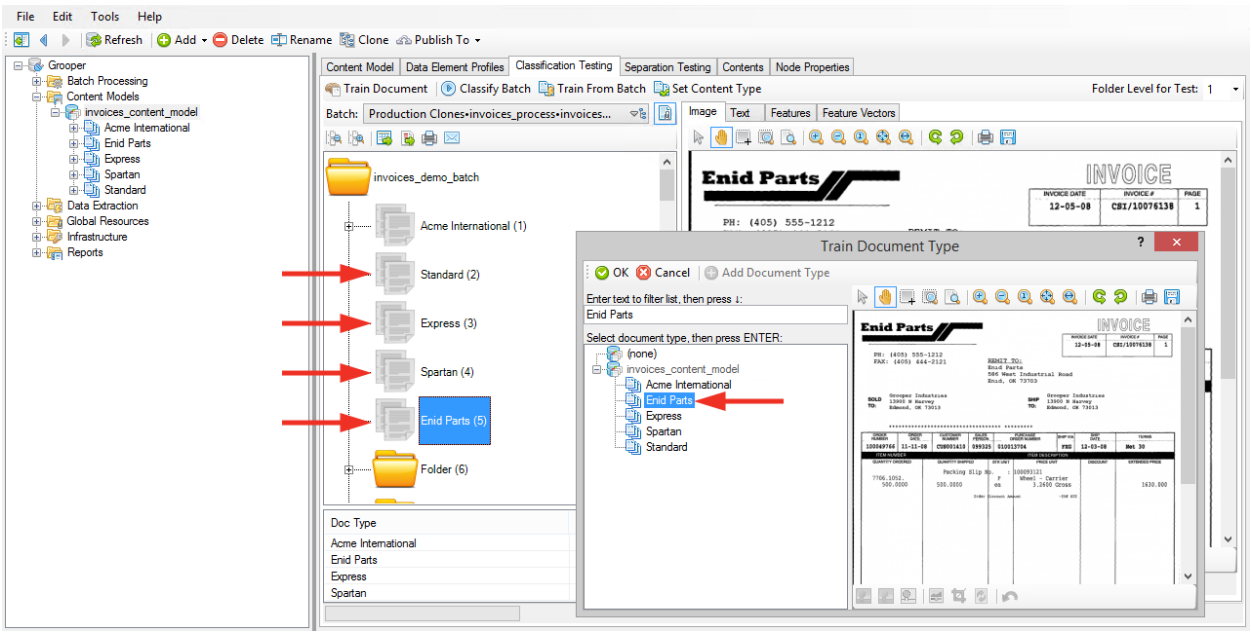
So we've only trained one of five Document Types, and our batch consists of 70 documents. The weightings system works so well, however, that out of all 70 of those documents, having only trained the Acme International document once, **Grooper** will accurately classify all Acme International documents if tested.

Go back to the Classification Testing tab within our invoices_content_model. Try the Classify Batch button. The Classification Test Set... window will appear. Leave it set to DocType and press OK. You will see in the list now what Acme International documents were classified. Click through them and verify.



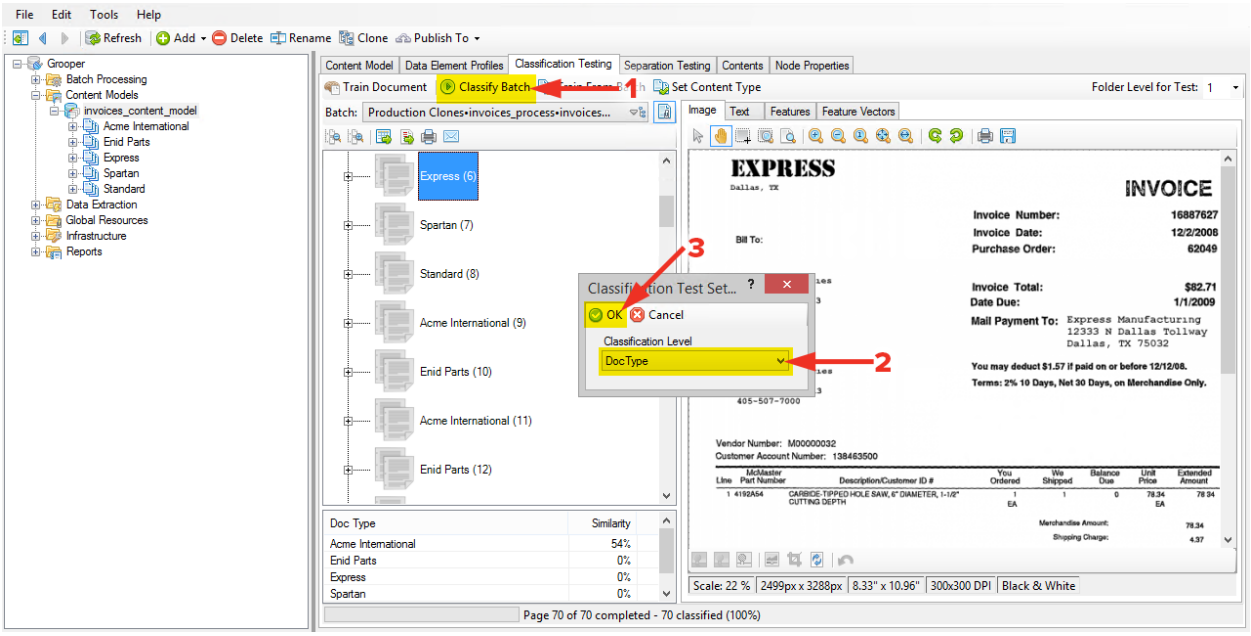
STEP 7 – TRAIN THE REMAINING FOUR DOCUMENTS

As before, click the Train Document button, or right-click and Train Document... for Folder (2), Folder (3), Folder (4), and Folder (5) and train them to Standard, Express, Spartan, and Enid Parts, respectively. Once again, you'll notice their icons, as well as their titles change. Feel free to go look at their weightings too, if you're curious.



STEP 8 – CLASSIFICATION TEST TWO

Classify the batch again and observe it accurately classify the remaining documents. Click or arrow key through them and verify for yourself.



We could leave it here, but it would be prudent to update our batch process and allow it to handle the classification. I also want to introduce the Classification Review section coming up. We'll wrap the two up together.

CLASSIFICATION REVIEW

WHAT IS CLASSIFICATION REVIEW?

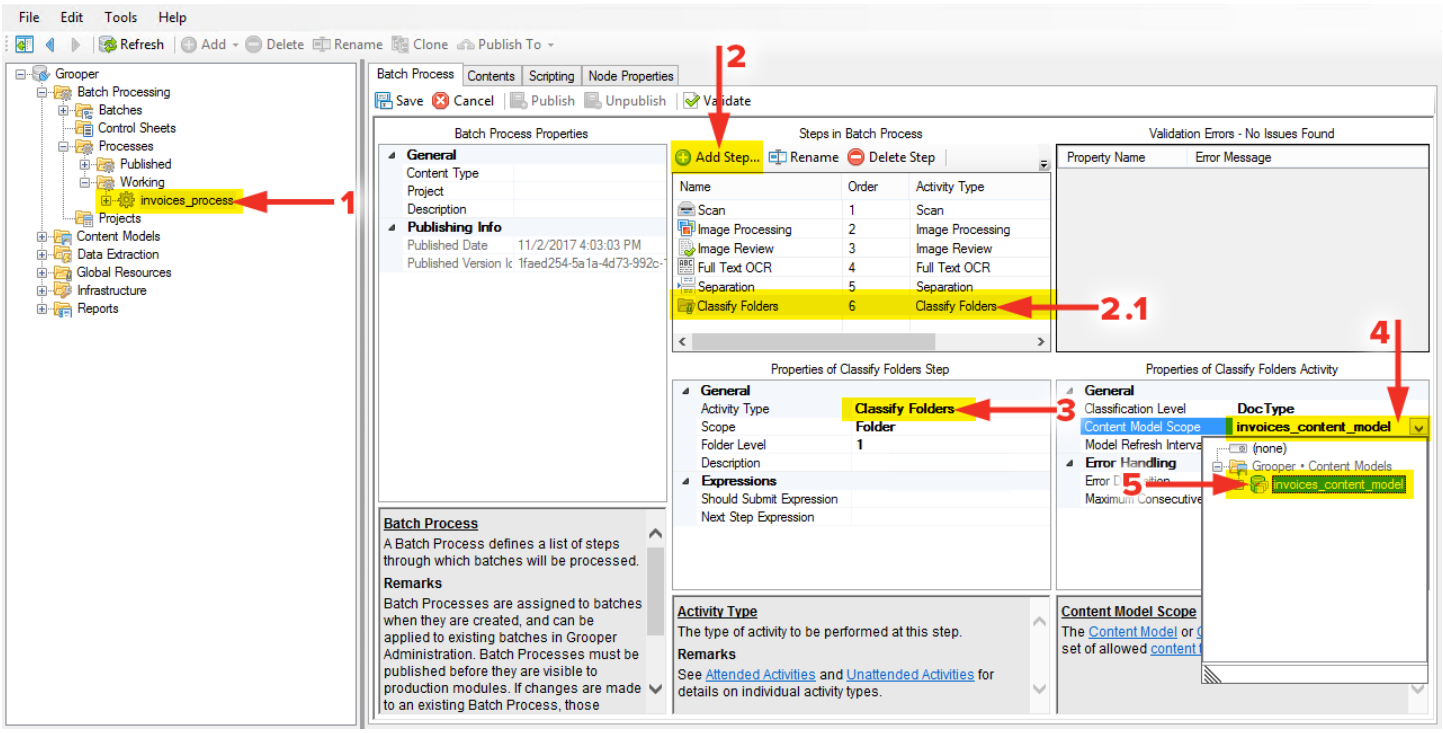
Like Image Review before it, Classification Review is an attended step that allows someone to review the batch structure and make sure its foldering and subsequent classification of that folder is accurate. Adjustments can be made on the fly, if required.

HOW TO CONFIGURE AND PERFORM CLASSIFICATION REVIEW

STEP 1 – UPDATING BATCH PROCESS FOR CLASSIFICATION

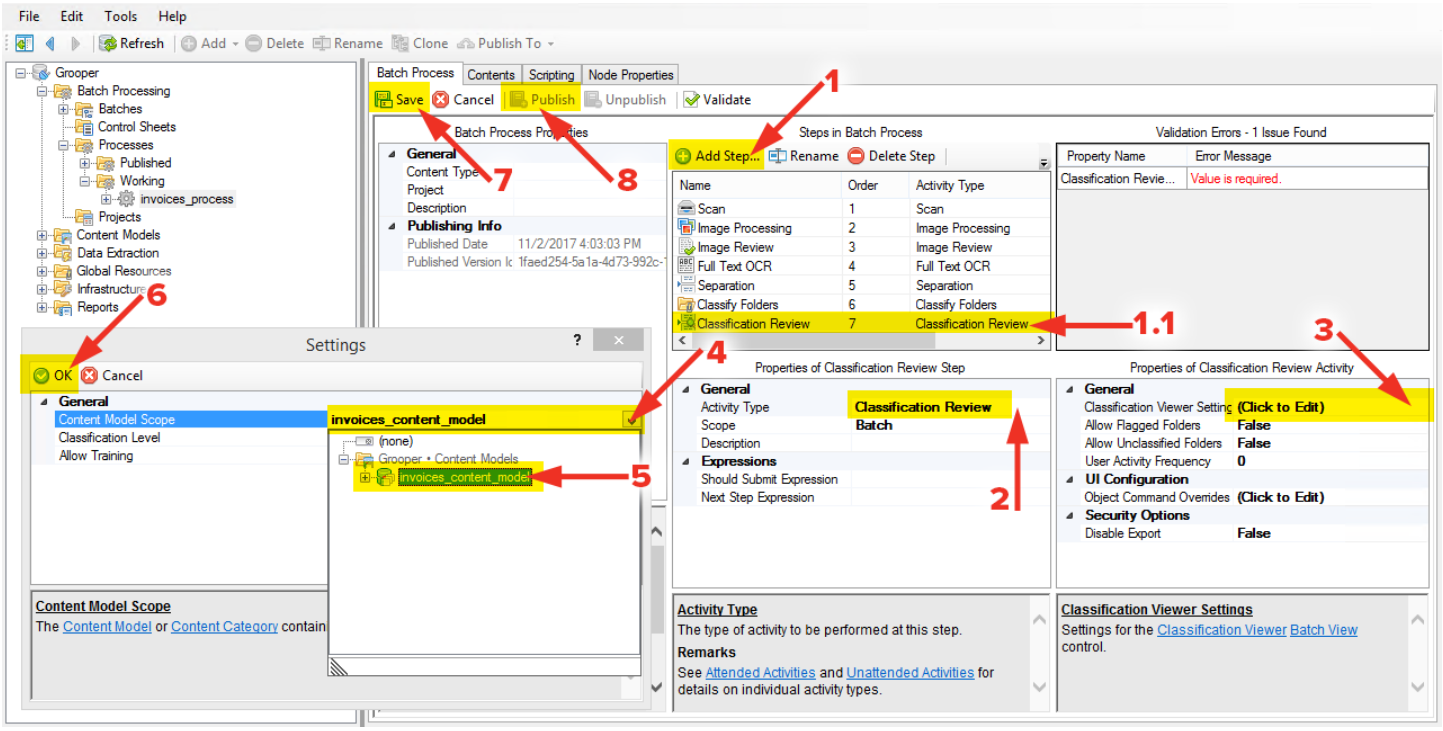
As was mentioned earlier, adding Classification to our batch process and implementing Classification Review are going to be handled all at once here.

(1) Navigate to Grooper – Batch Processing – Processes – Working and select the invoices_process node. **(2)** Add a step and **(3)** set its Activity Type to Classify Folders. **(4)** Click the dropdown for Content Model Scope and **(5)** select our Content Model invoices_content_model.



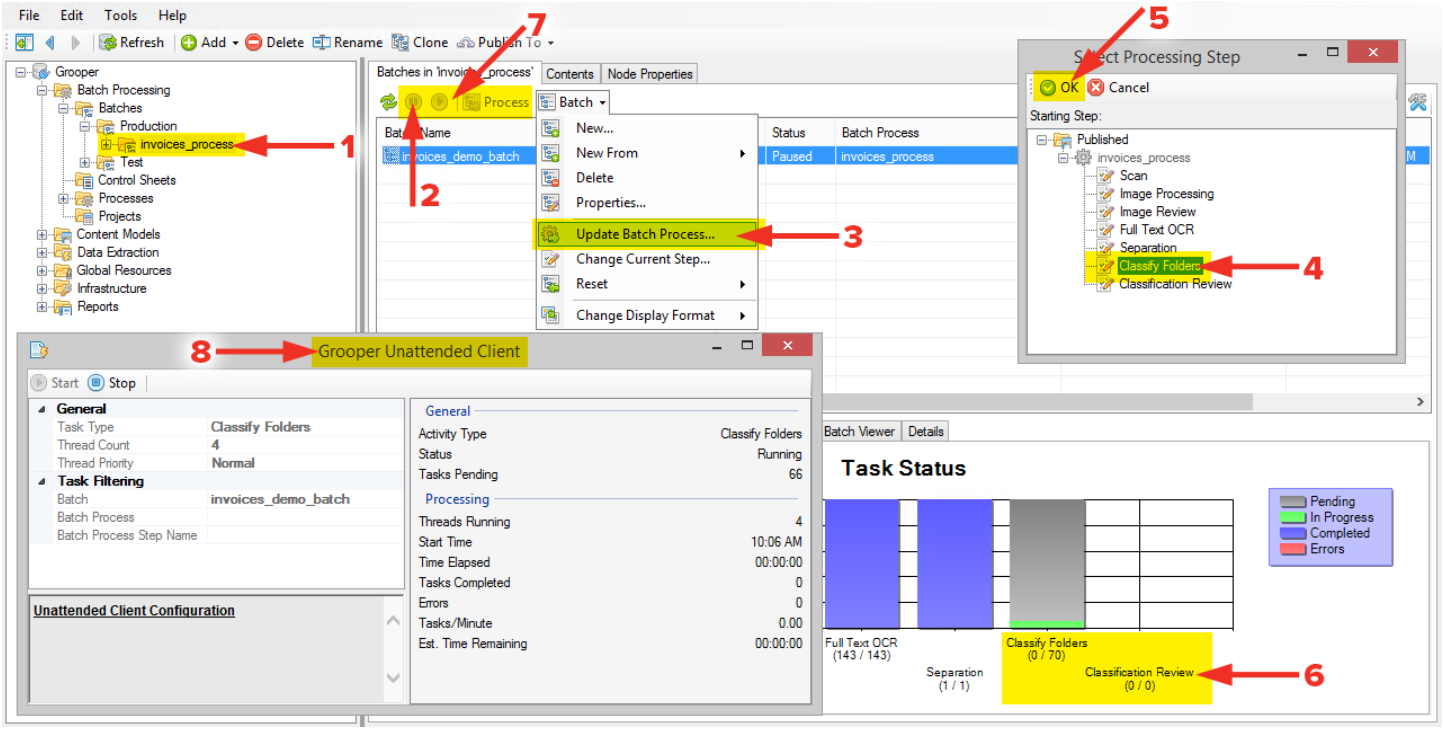
STEP 2 – UPDATE BATCH PROCESS FOR CLASSIFICATION REVIEW

(1) Add a new step and **(2)** set its Activity Type to Classification Review. **(3)** In the Classification Viewer Settings field, click the ellipsis box to open the Settings window. In this window click the drop down for the Content Model Scope field and select our invoices_content_model, then click OK. **(7)** Save and **(8)** Publish the batch process.



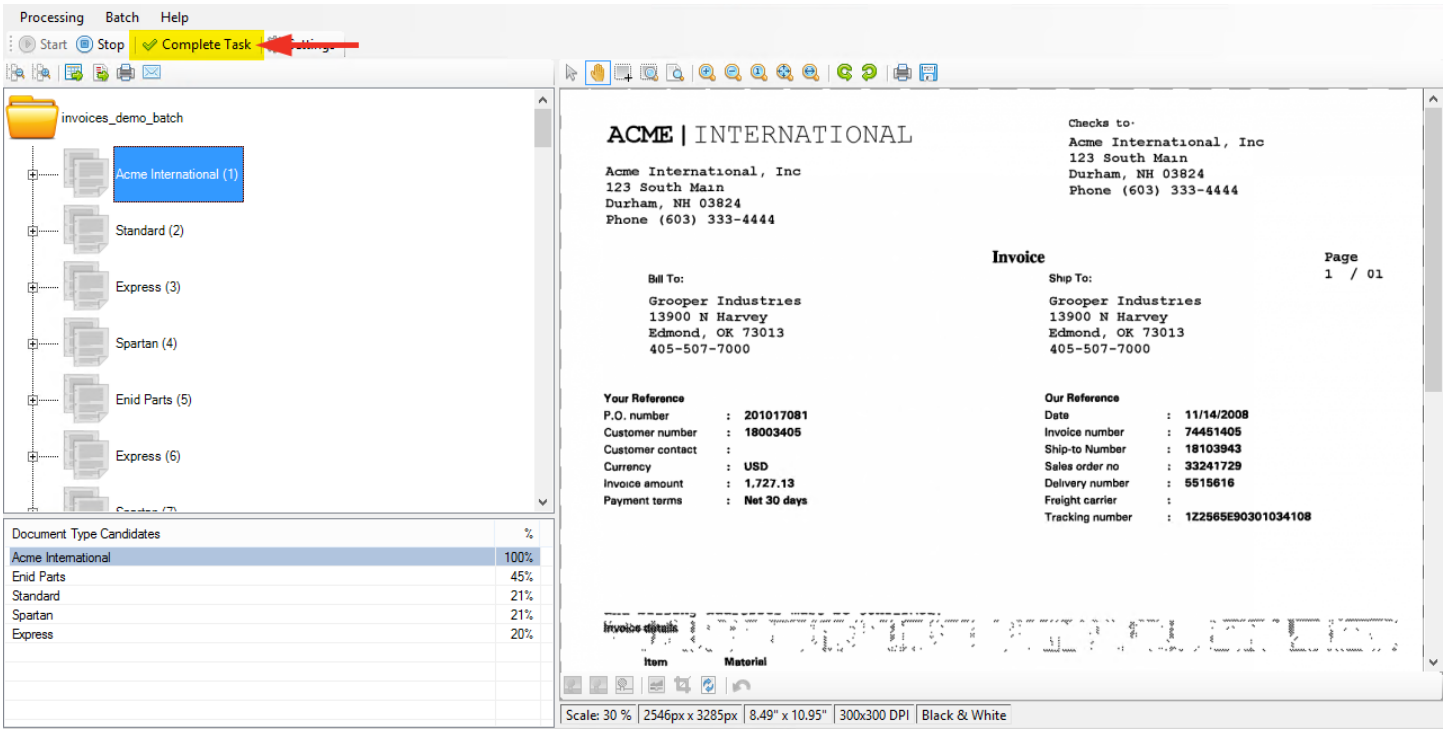
STEP 3 – UPDATE BATCH PROCESS

(1) Navigate to Grooper – Batch Processing – Batches – Production and select invoices_process. **(2)** Select our batch and make sure its paused. **(3)** Use either the Batch dropdown, or right-click and Update Process... In the Select Processing Step window that comes up select **(4)** Classify Folders and **(5)** click OK. **(6)** You'll see the new steps added to the process. **(7)** Press Play (and confirm) and Process the batch. **(8)** The Grooper Unattended Client window will appear and begin to process Classify Folders. When the process finishes, press Process again to launch the Classification Review module (as it's the next step in our batch process.)



STEP 4 – PLAY AND PROCESS BATCH

While Classification Review can allow you to make modifications to your batch to correct for errors, we won't need to in this case. Feel free to click through the documents again and verify they're classified properly. Click Complete Task when you're satisfied and confirm in the following window. Be sure to pause our batch back in Grooper Administration.



PHASE 4 - COLLECT

Other than being put into neatly sorted groups, our documents also contain information that makes them unique. In this section we'll discuss configuring **Grooper** to be able to recognize this data and pull it from the documents.

EXTRACTION

WHAT IS DATA EXTRACTION?

Data Extraction is the process of extracting data elements from documents in a batch. Previously, we separated our pages into folders and classified them to be specific documents from the Content Model. When a folder is assigned a Document Type, it becomes a logical document and inherits all the data elements defined for that Document Type, allowing it to be indexed manually by a user, or automatically using data extraction techniques.

DATA MODELS – DEFINING DATA ELEMENTS

A Data Model describes the structure of data inside a specific document using a hierarchy of data elements such as Sections, Tables, and Fields. A simple Data Model can be nothing more than list of fields, but **Grooper**'s Data Models are hierarchical in nature, allowing complex documents to be broken down into any number of sections, sub-sections, tables, and fields. A Data Model and its child data elements collectively describe all the data elements on a Document Type with which **Grooper** will interact, and also define properties which control the behavior and appearance of the data elements in the Index Panel control.

Data Models are defined as children of individual Content Types, and can be created using the Create Data Model command on the property panel for any Content Model, Content Category, or Document Type object. Once a Data Model has been created, Sections, Tables, and Fields may be created as children of the Data Model. The data elements defined in a Data Model apply to the content type on which the Data Model is defined, as well as any child content types which inherit from that content type.

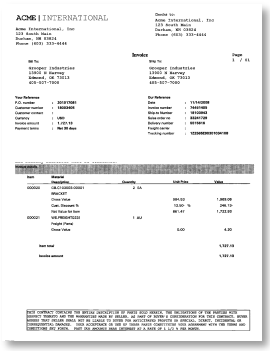
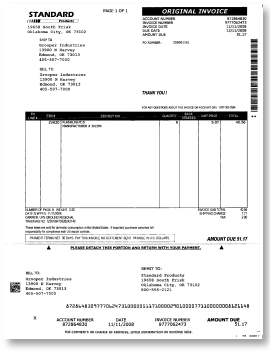
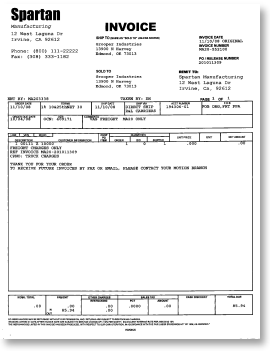
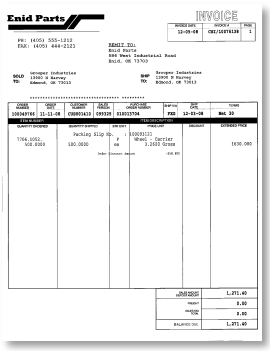
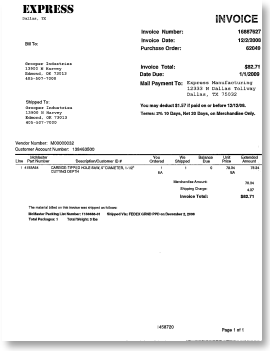
HOW TO CONFIGURE AND PERFORM EXTRACTION

It's prudent to start this section by taking a moment to mention, briefly, a very important topic: Regular Expression. To put it simply, it is a sequence of characters with a syntax that define a search pattern. Moving forward, finding repeatable patterns, and being able to leverage them to find the information we want to extract is incredibly powerful.

The scope of this document is not enough to even tell you all the places you can go to learn about regular expression, much less teach it to you regular expression in its entirety. I could mention which methods have worked for me, but they may not work for you. The good news is that the amount of information available on the internet or in printed form is so vast, you should have little to no trouble finding educational material on it. Another bit of good news is that **Grooper** can be an excellent place to learn regular expression, as its built-in tools for developing patterns are excellent.

STEP 1 – UNDERSTANDING THE DOCUMENTS

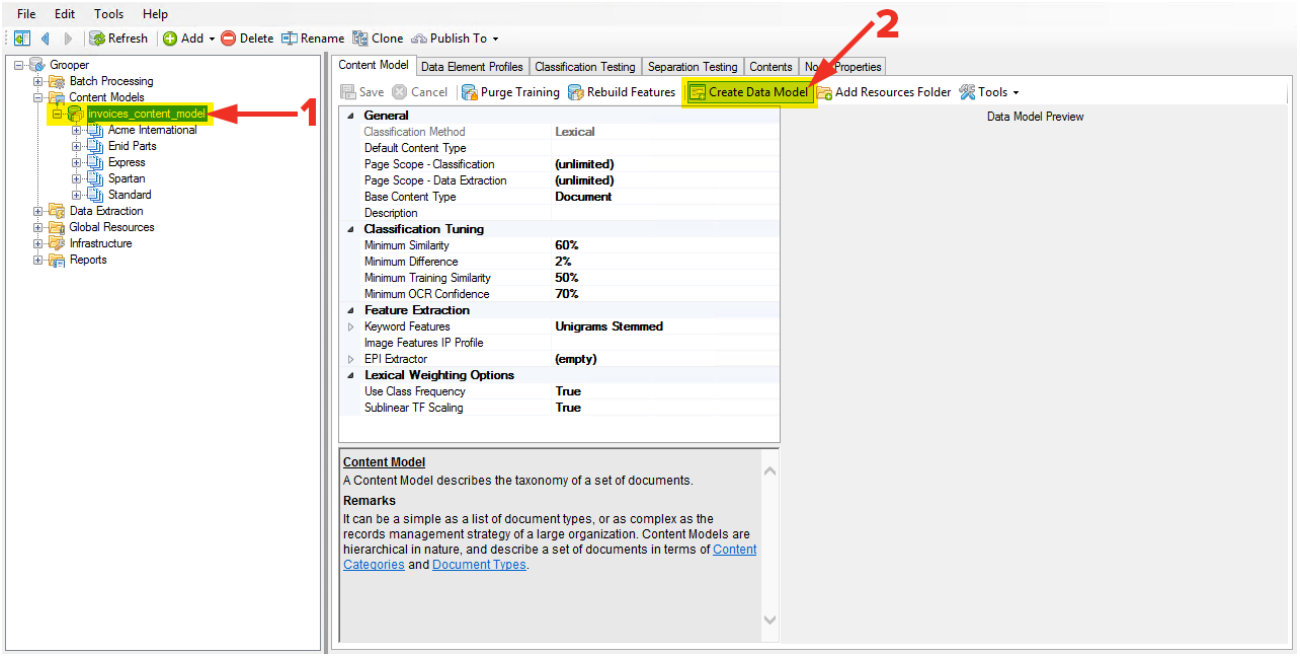
Within our batch so far, we've had 5 different Document Types, represented by the companies to which they belong: Acme International, Enid Parts, Spartan, Express, and Standard.

If it hasn't been clear based on the naming of our objects thus far, or by reviewing the documents, I'll state now these are representations of invoices. There are particular bits of information that each document houses that we'll want to extract. Knowing what to extract can depend on certain things. Starting from scratch, as an individual, you'd simply have to ask yourself what information on these documents is important to you. You may simply want to get the invoice number from each document. Do you have an established database with information and you simply want to add documents to it? In that case you'd need to marry the Data Model to match that database. Has a department in your company given you these documents and told you specifically which fields of information are important to them, and they don't want to manually input that information anymore? For our purposes, we'll assume the final option, and say this department simply wants the Invoice Number from these documents.

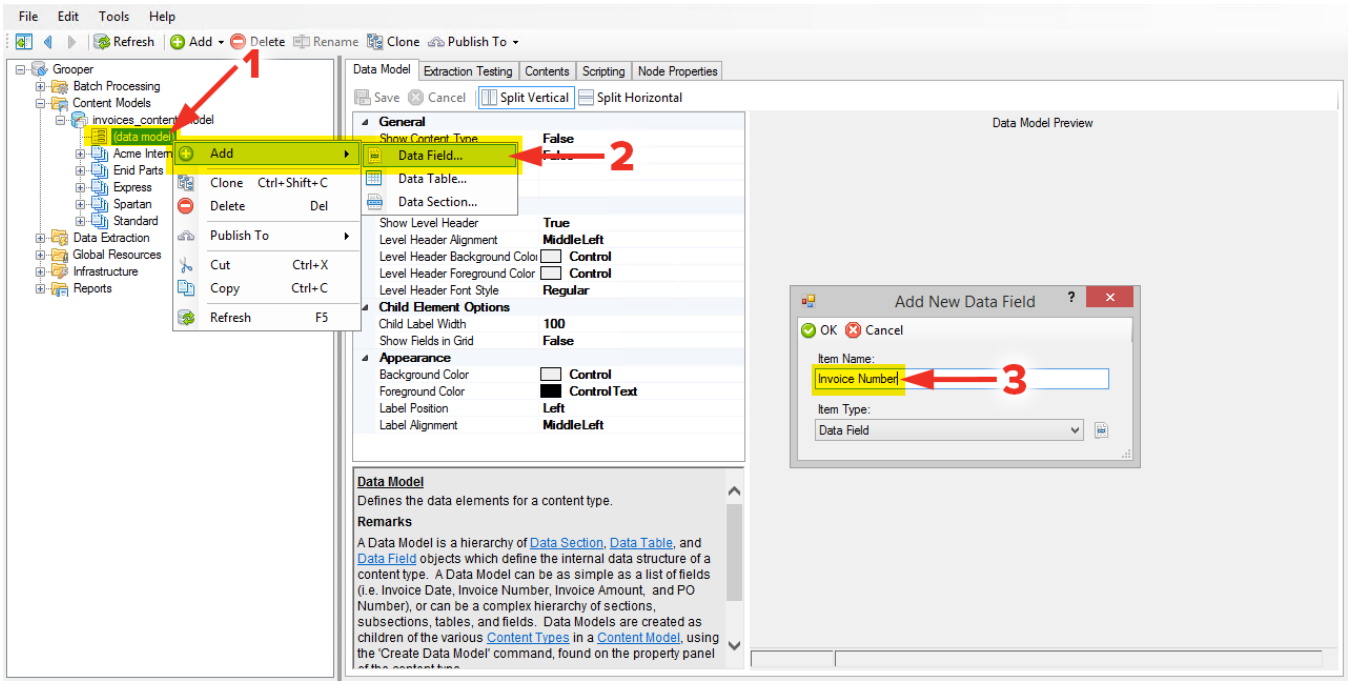
STEP 2 – CREATING A DATA MODEL

Navigate to Grooper – Content Models and **(1)** select invoices_content_model. **(2)** Click the Create Data Model button. This will create a parent Data Model (leave the names of (data model)s alone) within our Content Model that will house indices for all of our Document Types. As expressed earlier, we could create further Data Models within each document to have indices specific to them, but for now we won't.



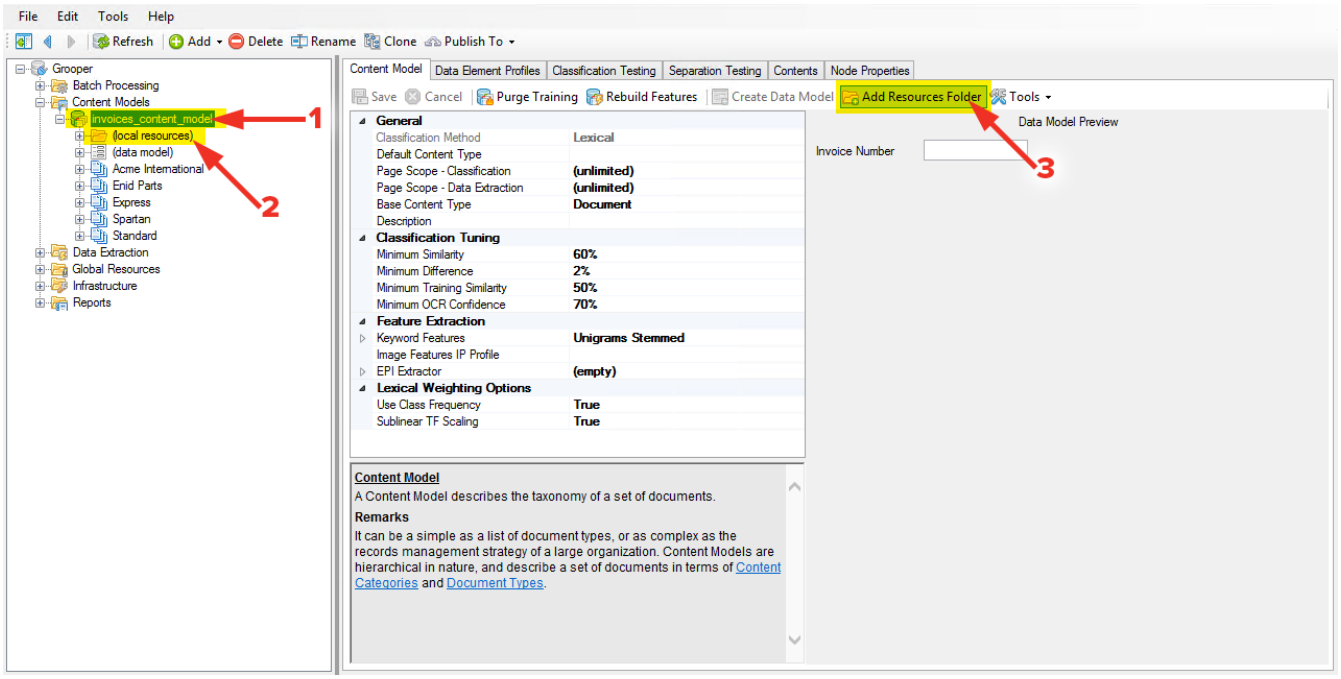
STEP 3 – CREATING DATA FIELDS

(1) Select the (data model) and use the Add drop down, or **(2)** right-click and Add > Data Field... In the Add New Data Field window that appears, **(3)** name the field Invoice Number.



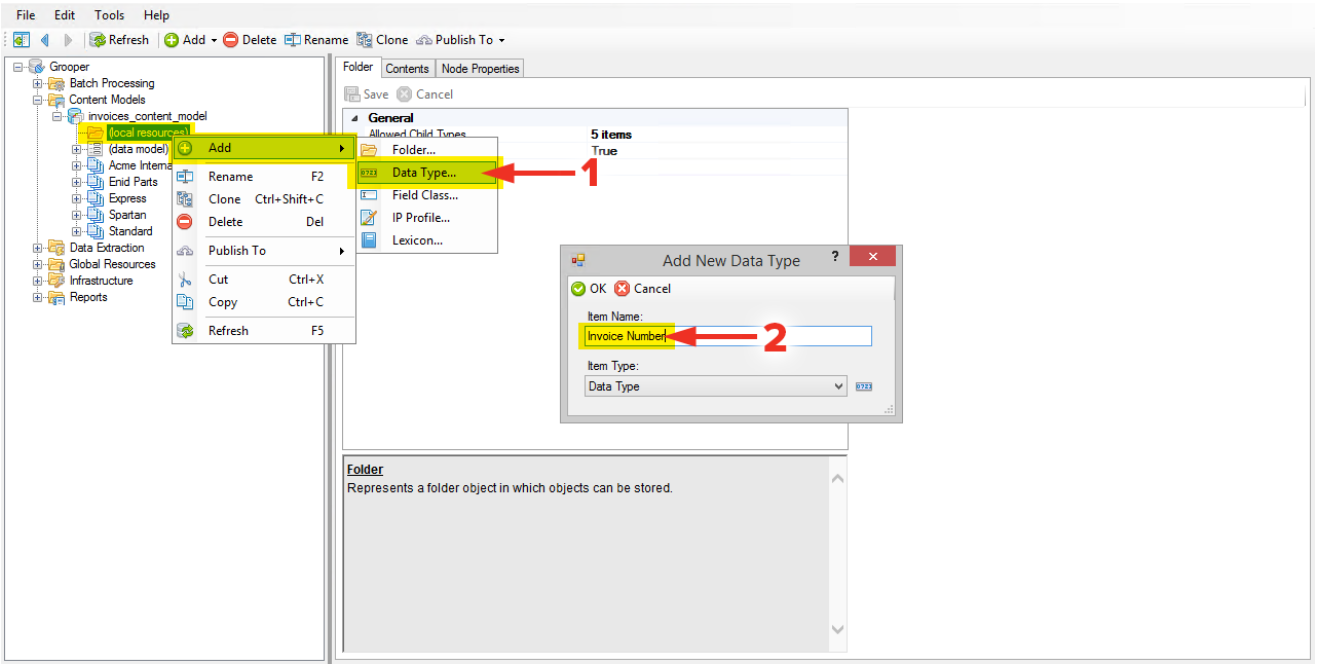
STEP 4 – CREATE LOCAL RESOURCE FOLDER

(1) Select invoices_content_model and **(2)** click the Add Resources Folder button. **(3)** You will see the (local resources) folder node appear in the Content Model. This folder acts to allow for organization of objects that can be used by all child objects of the Content Model.



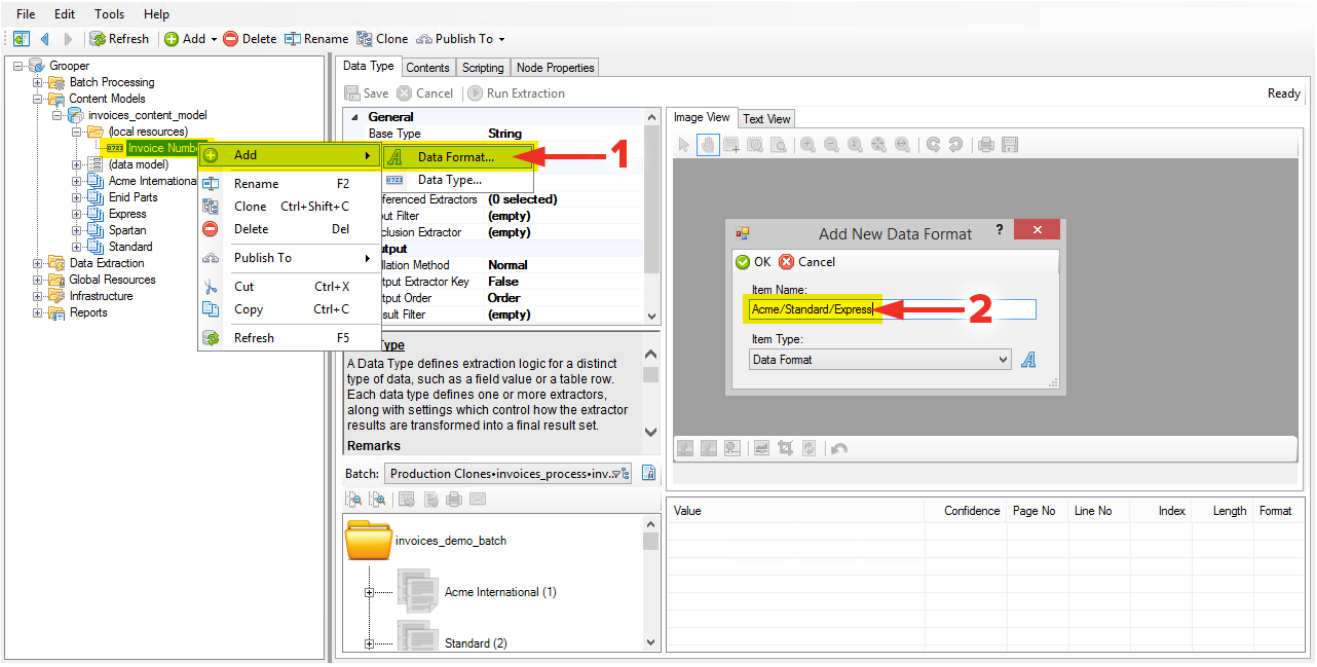
STEP 5 – CREATE DATA TYPE

Select the newly created (local resources) folder and use the Add dropdown or **(1)** right-click and Add > Data Type... In the Add New Data Type window that appears, **(2)** name it Invoice Number. Data Types contain the logic necessary to extract information from our documents, and can do so directly, but we will be adding child objects which it can leverage to extract multiple results.



STEP 6 – CREATING DATA FORMAT OBJECTS

Select the newly created Data Type and use the Add dropdown or **(1)** right-click and Add > Data Format... In the Add New Data Format window that appears, **(2)** name it Acme/Standard/Express. Data Format objects are child objects of Data Types that define the pattern that will be leveraged for extraction

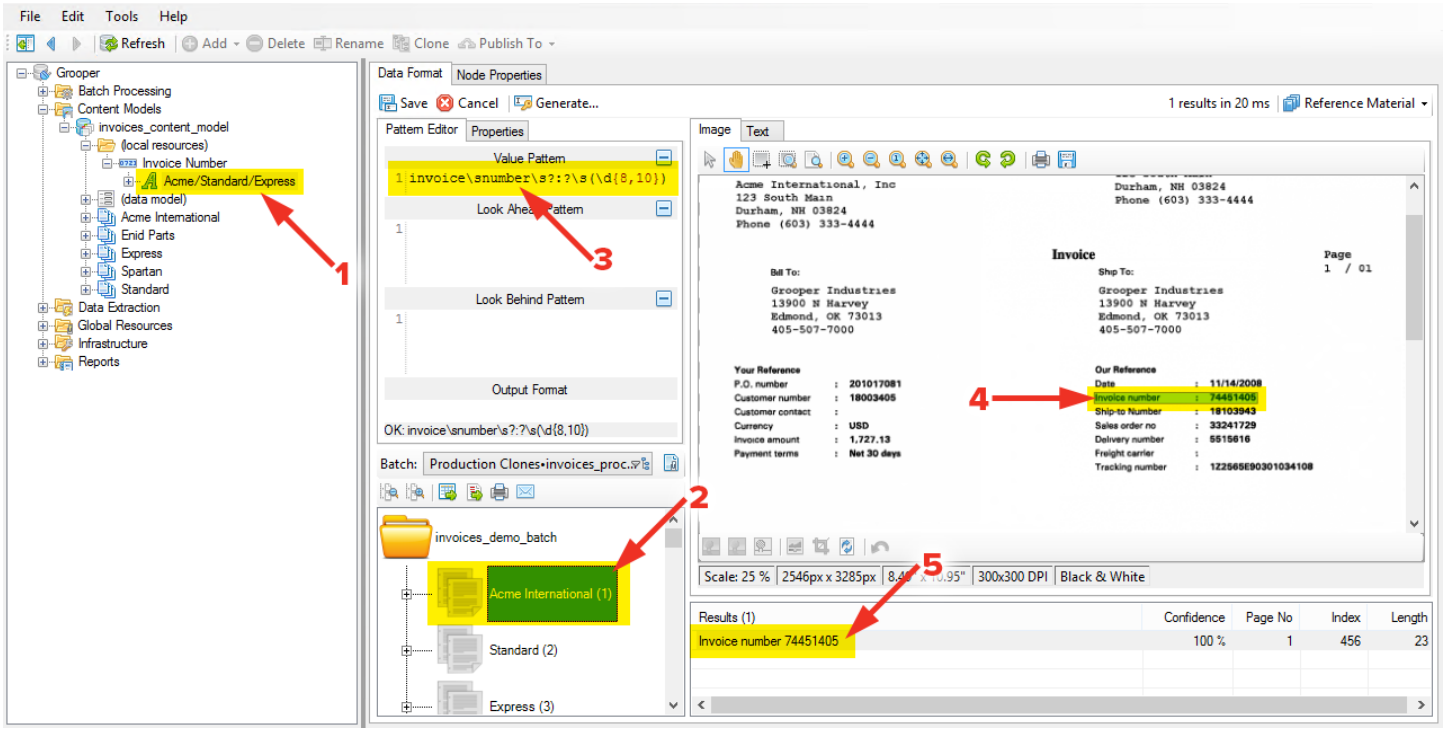


STEP 7 – WRITING OUR FIRST REGULAR EXPRESSION PATTERN

(1) Select the newly created Data Format object, **(2)** then select the Acme International (1) document. **(3)** In the Value Pattern area, type the following exactly:

```
invoice\snumber\s?:?\s(\d{8,10})
```

(4) You should notice a green box appear on the right side, highlighting the results of our pattern, **(5)** as well as a result in the bottom. We're looking for a literal string of **invoice** followed by a space character **\s** and another literal string of **number**. After that we're looking for another (optional) space character **\s?**, a colon character **:** and saying it may or may not be there **?** followed by another space character **\s**. Following this we're opening a group with an open parenthesis **(** and looking for a number character **\d** with a specified length of a minimum of 8 characters up to a maximum of 10 **{8,10}**, then closing the group with a close parenthesis **)**.

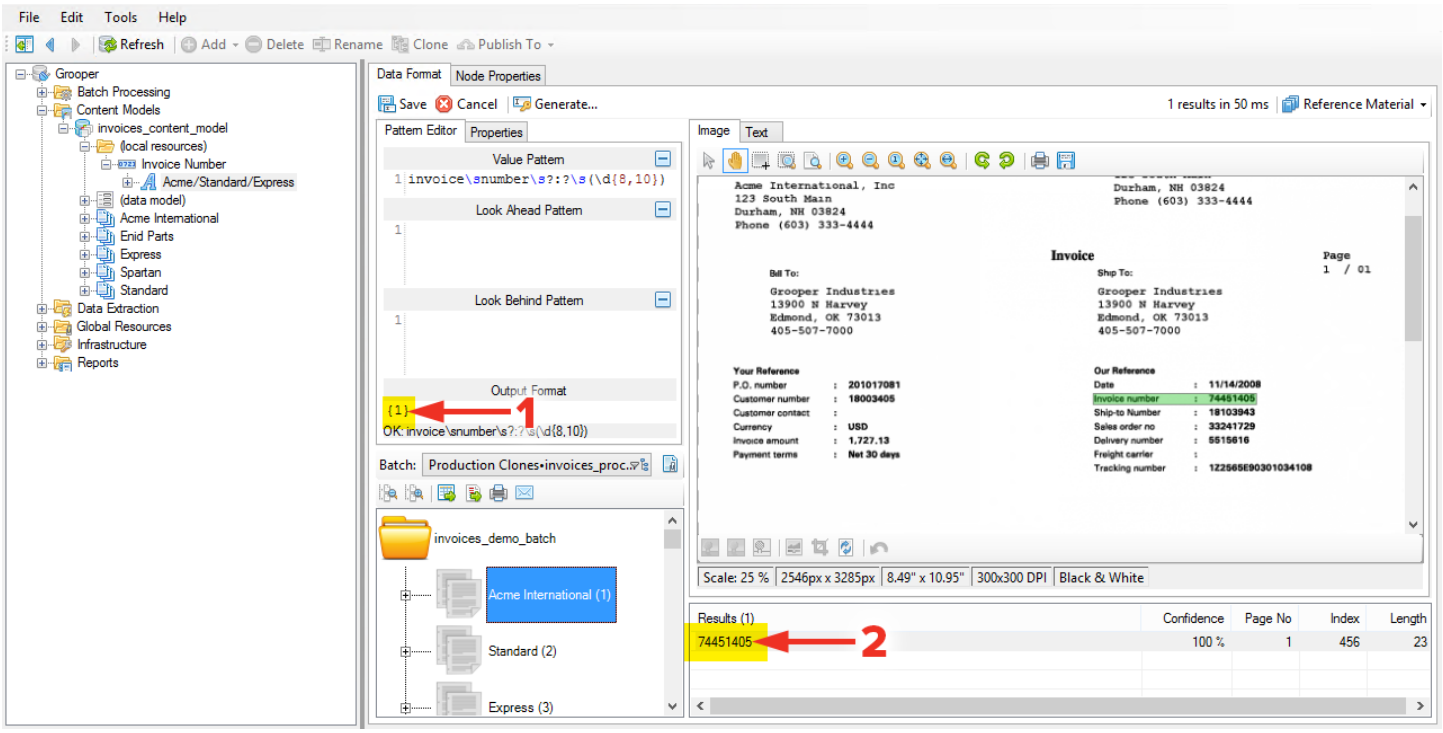


STEP 8 – USING GROUPS FOR CLEANER RESULTS

The only problem with this pattern is that while it gets what we want, it returns more than we need. If you notice in the previous image **(5)** the result contains not only the number we want, but also the words Invoice number. We can resolve this by leveraging regular expression capture groups, which we've already done by using the open and close parenthesis. All we need now do is **(1)** insert the following into the Output Format:

```
{1}
```

This will restrict the resulting output to only what is in the parenthesis, and **(2)** you should see that reflected in the Results area. Go ahead and Save our changes. Feel free to click on the other documents to see that this pattern gets results on them, except for …



STEP 9 – NEW DATA FORMATS

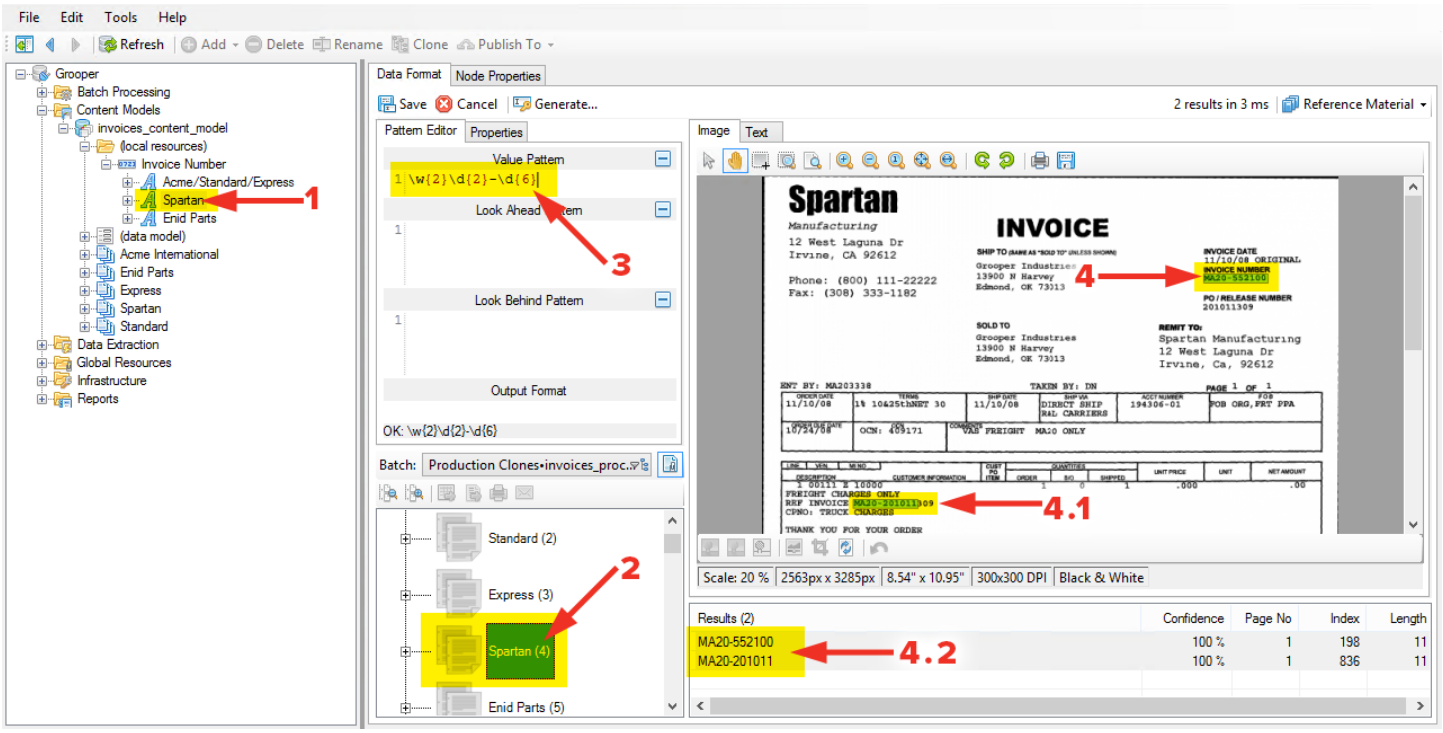
As the name of the Data Format may have suggested, its pattern will work for Acme International, Standard, and Express documents, and them only. We have two other documents we need to capture the invoice number for, but we need different patterns to capture them.

Refer back to step 6, and make two new Data Format objects. Name them Spartan and Enid Parts respectively.

(1) Select the newly created Spartan Data Format object, **(2)** then select the Spartan (4) document. **(3)** In the Value Pattern area, type the following exactly:

```
\w{2}\d{2}-\d{6}
```

So this pattern is looking for two of any word character **\w{2}** , followed immediately by two of any number character **\d{2}** , followed immediately by a hyphen character **-** , and finally six of any number character **\d{6}**. **(4)** You should see the number we want highlight, but also one we don't want (this will be reflected in the results area in the bottom too.)



STEP 10 – LOOK BEHIND PATTERN

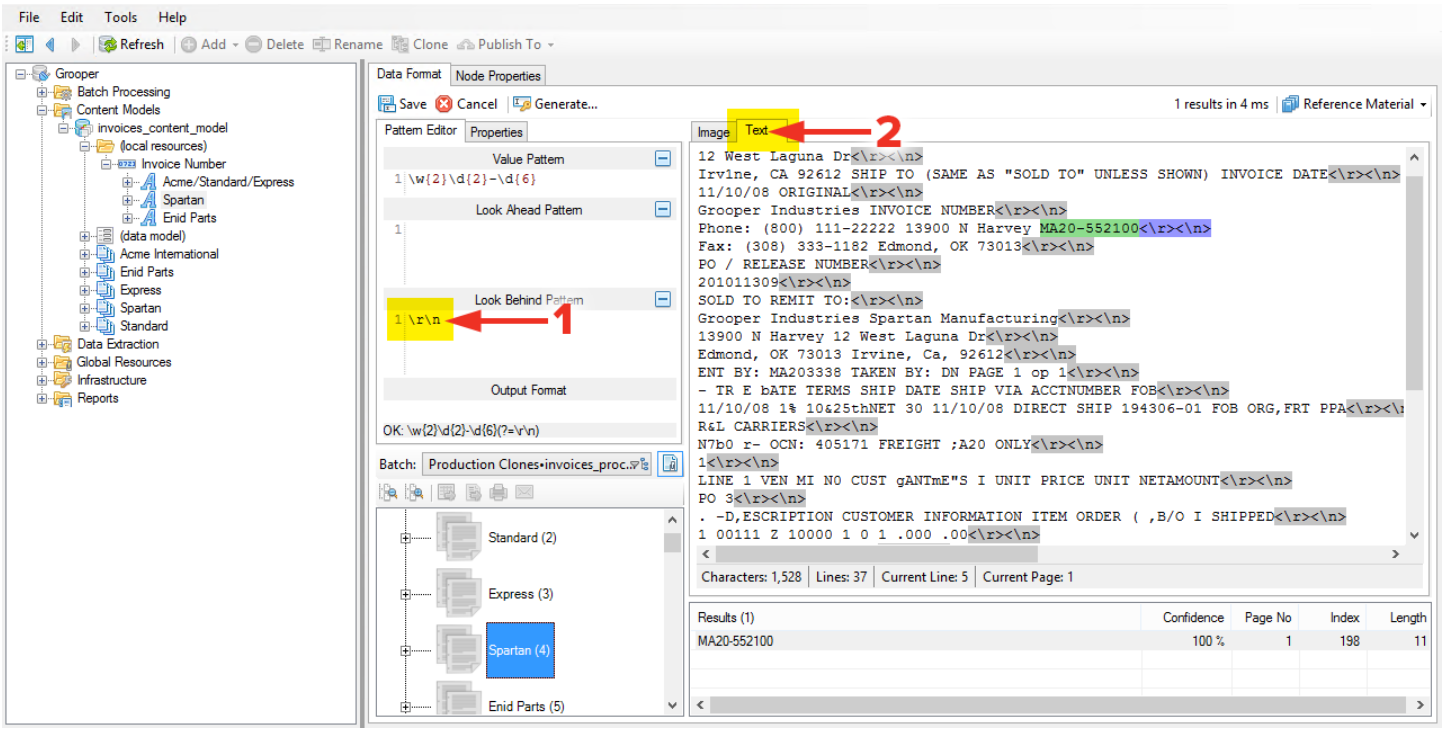
Click on the Text tab and you'll see the OCR text data. Notice how the first pattern is followed by a return line and a new line feed, while the second pattern is not.

Let's make the pattern more specific by giving it something to look to the right of the pattern for that the second result doesn't have. To do this, we'll use a Look Behind Pattern.

(1) In the Look Behind Pattern area type the following:

```
\r\n
```

(2) This will highlight the return and newline features in blue, and tell our pattern work only when followed by these characters.



STEP 11 – ENID PARTS AND LOOK AHEAD PATTERN

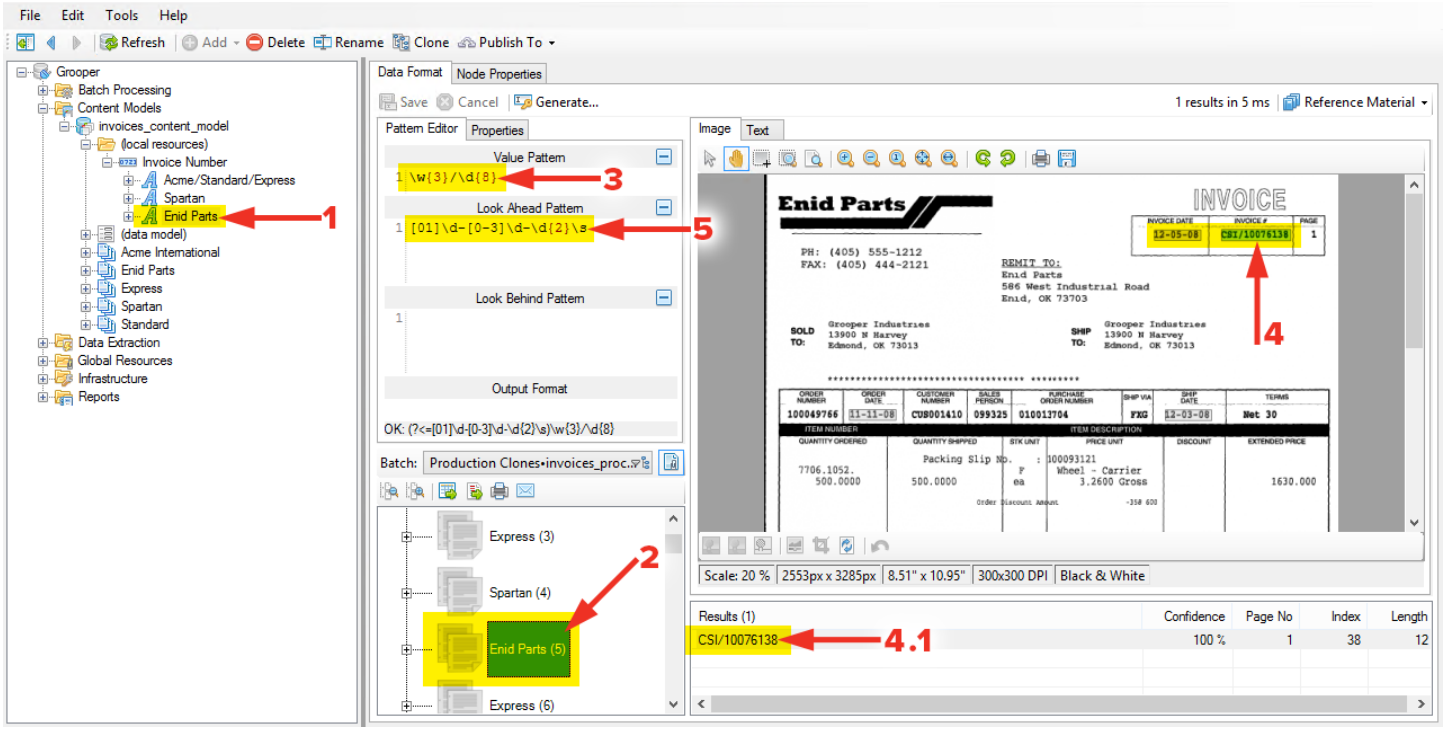
(1) Select the Enid Parts Data Format then **(2)** select the Enid Parts **(5)** document. **(3)** In the Value Pattern enter the following:

```
\w{3}/\d{8}
```

(4) This will give us the invoice number we want, but let's anchor it to that date that happens to exist near it by using something very similar to the Look Behind Pattern. As its name suggests (and opposite of the Look Behind Pattern), the Look Ahead Pattern will find something ahead of, or to the left of our pattern. **(5)** So, for the Look Ahead Pattern, type the following:

```
[01]\d-[0-3]\d-\d{2}\s
```

Based on the regular expressions you've seen so far, see if you can't discern what exactly these patterns are doing.

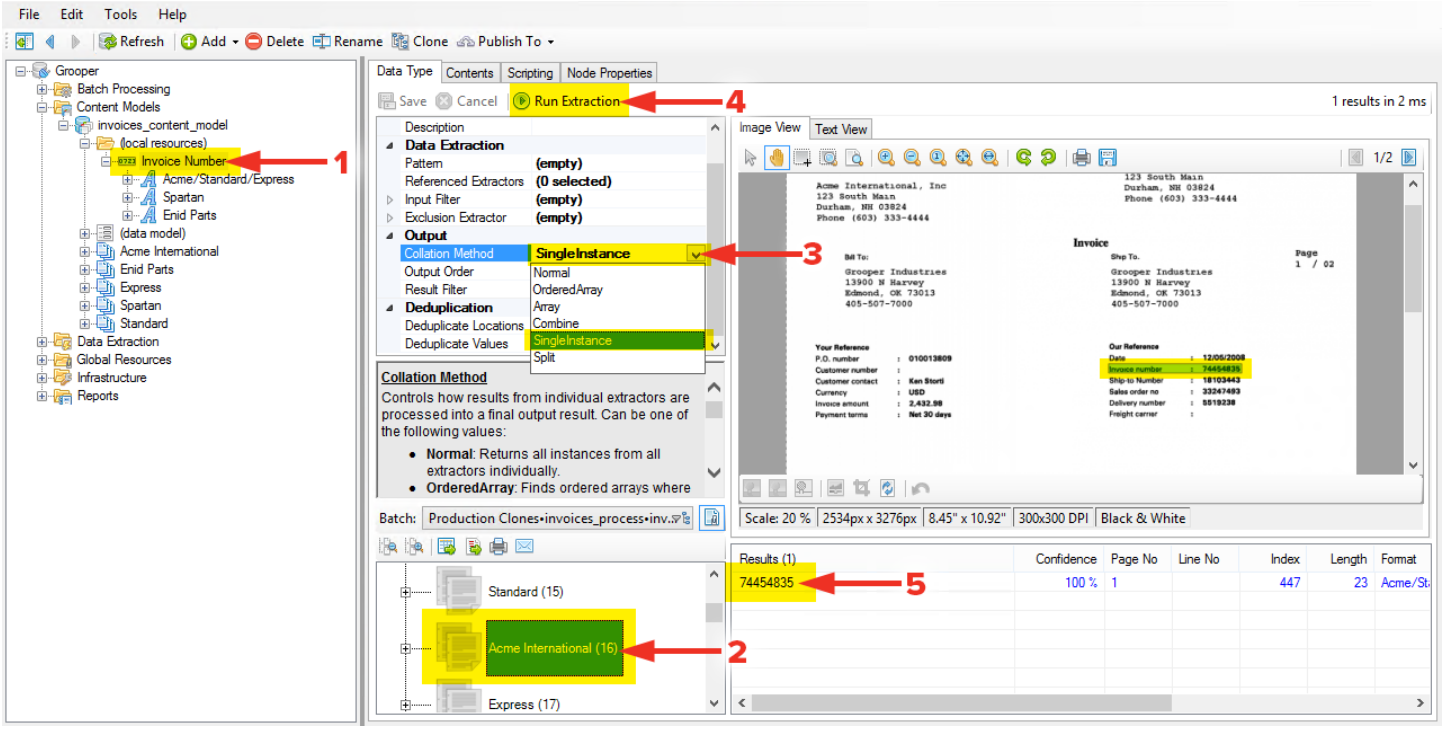


STEP 12 – CHECKING OUR WORK

At this point, we have three Data Formats that extract the information we need. They are child objects of the parent Data Type, Invoice Number. **(1)** The parent object can perform the extraction of its three child objects, so select the Invoice Number parent Data Type. **(2)** Click the first document in the Batch Viewer and use your down arrow key to move through the documents sequentially, and you should see the desired information being returned.

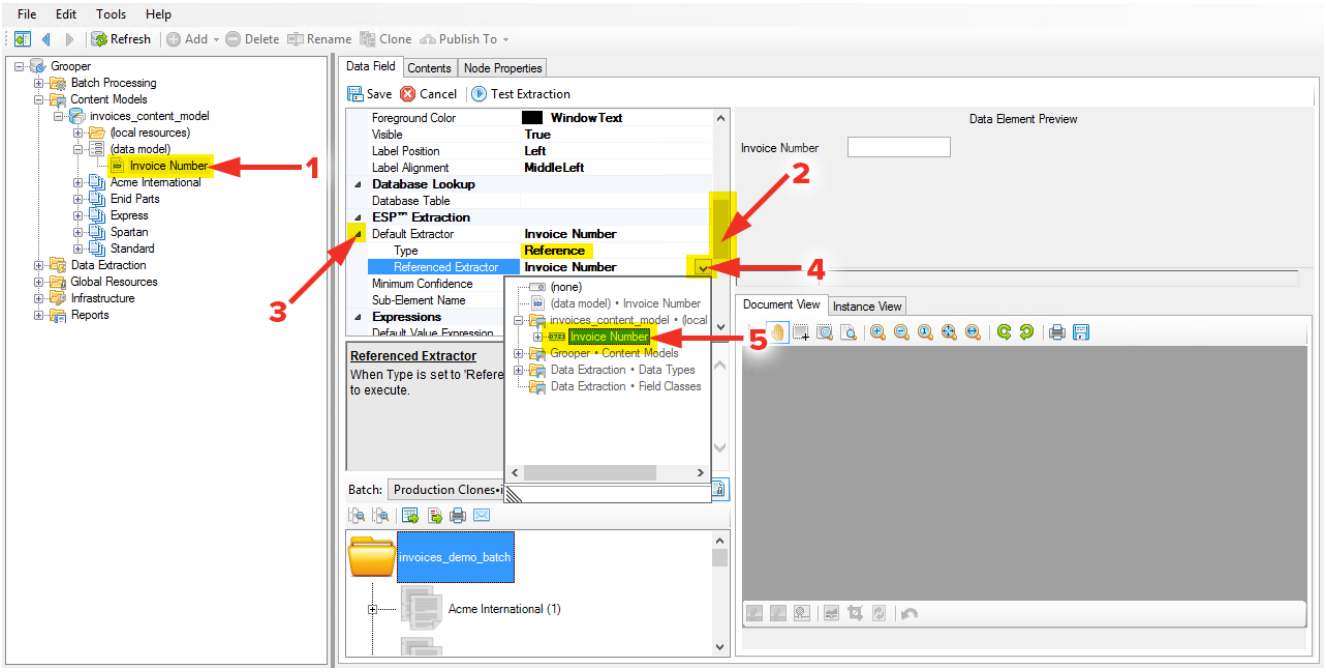
There are two Spartan documents that won't return values, and we'll cover those in the next section.

You'll eventually come to an Acme International document that has two returned values, one of them being on page two. We only want one of the values, and there's an easy method to remedy this. In our Data Type, in the Output section, **(3)** click the dropdown for Collation Method and select SingleInstance. Leave the Output Order to Order. This will tell the Data Type to only accept one instance of returned information, and consider the first one it comes across above all others. **(4)** Save the changes and Run Extraction. You should notice now only one value being returned.



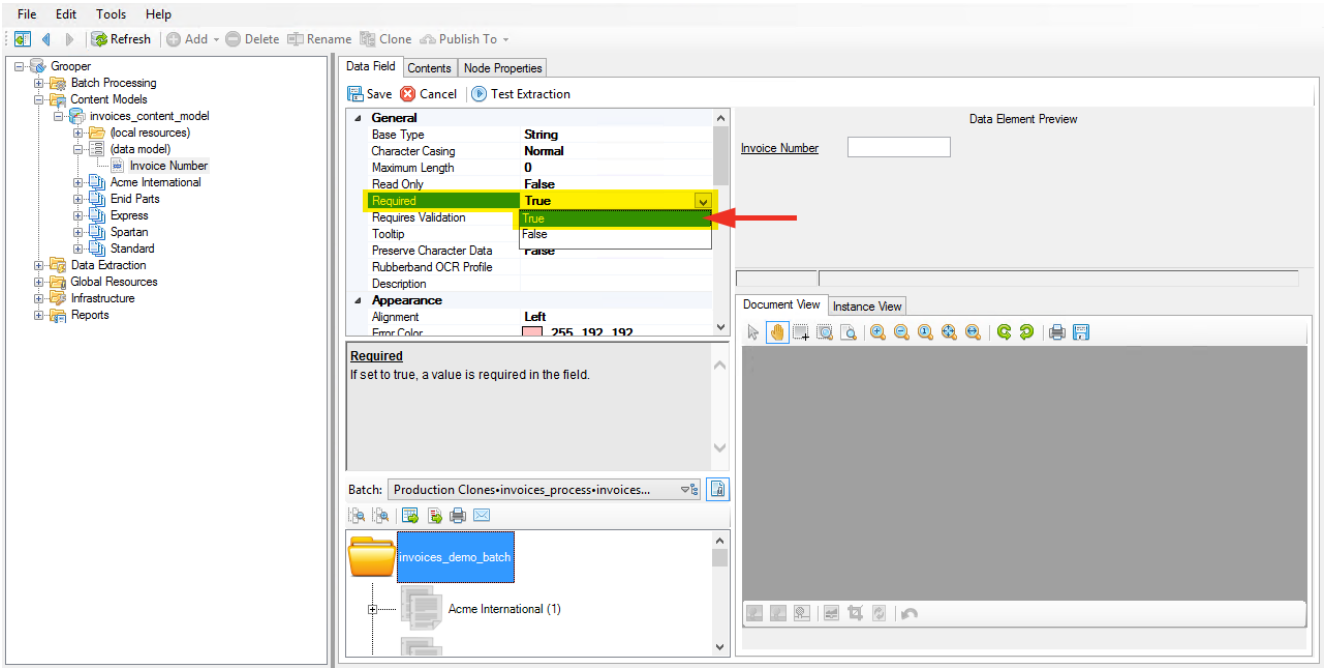
STEP 13 – DATA FIELD AND DEFAULT EXTRACTOR

We need to marry our extractor to our Data Model. **(1)** Expand the (data model) node and select the Invoice Number Data Field. **(2)** Scroll the Data Field tab down a bit and **(3)** expand the Default Extractor area. **(4)** Set the Type to Reference (via the dropdown) and this will expose the Referenced Extractor field. Click its dropdown and **(5)** set it to the Invoice Number Data Type within the invoices_content_model • (local resources) area.



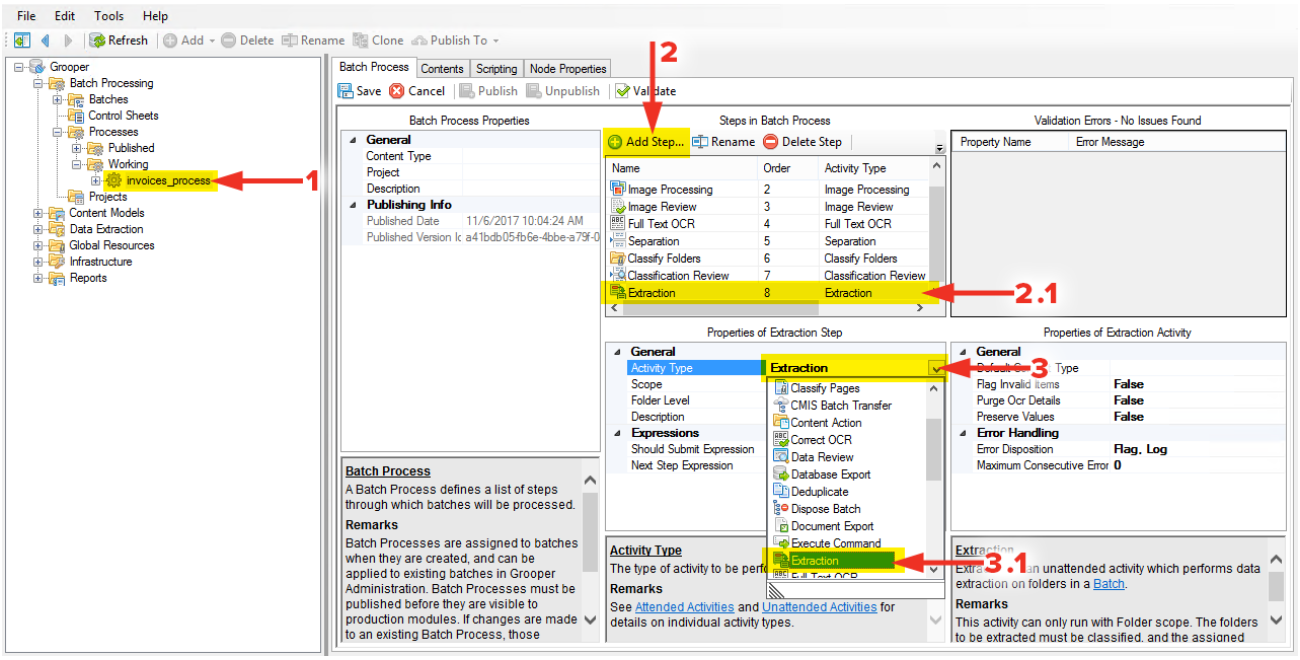
STEP 14 – SETTING FIELD TO REQUIRED

While still on our Data Field, scroll back up and set the Required field to True. This is a validation check that flags documents that do not have information in this field. This will be more apparent in a moment when we perform Data Review. Save the changes to the Data Field.



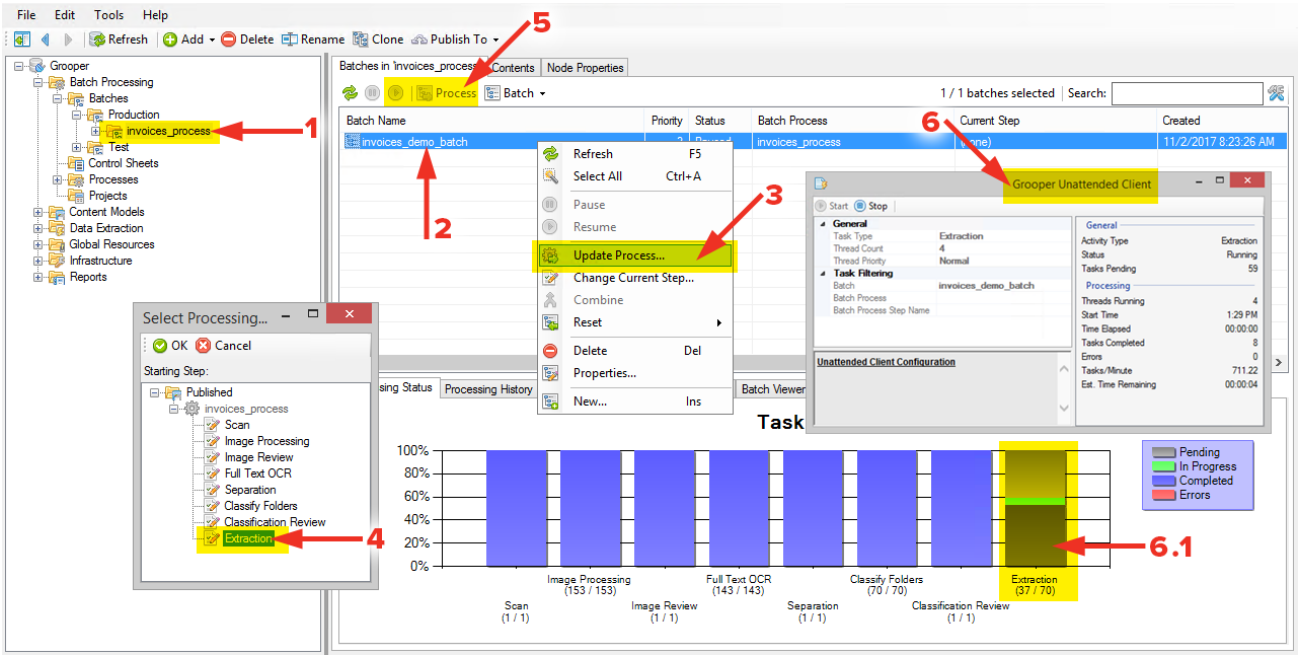
STEP 15 – UPDATE BATCH PROCESS FOR EXTRACTION

(1) Navigate to the Working process node and select the invoices_process batch process. **(2)** Add a step and **(3)** set its Activity Type to Extraction. Save and Publish the batch process.



STEP 16 – UPDATE BATCH PROCESS

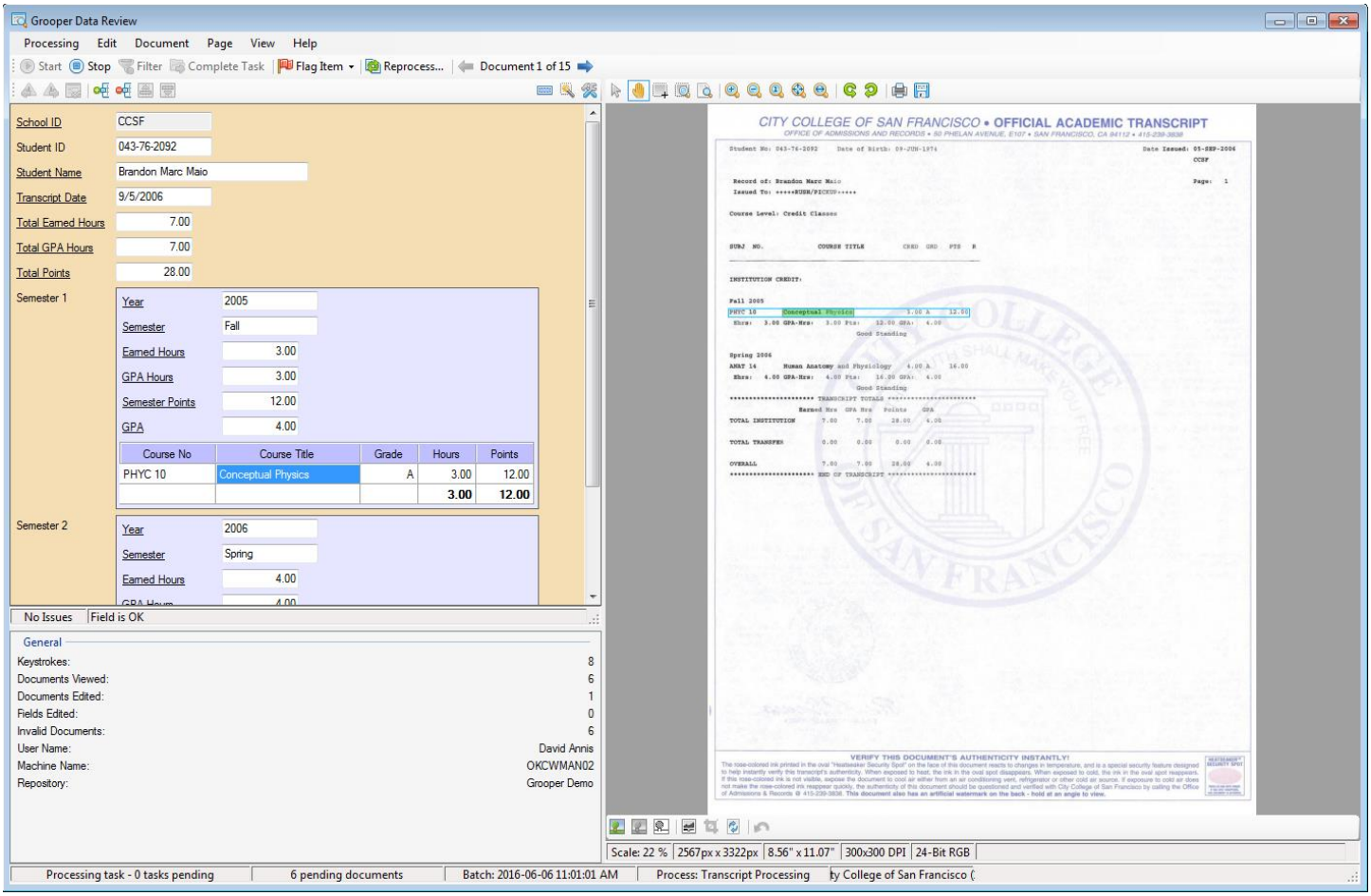
With our batch process updated and published, **(1)** go to the invoices_process node in the Production Batches area. **(2)** Select the invoices_demo_batch and make sure it's paused. **(3)** Using the Batch dropdown, or right-click and select Update Process... **(4)** In the Select Processing Step window select the Extraction step and click OK. **(5)** Start the batch and Process it. **(6)** The Unattended Client window will appear and begin processing the Extraction step. When the processs completes go ahead and pause the batch.



DATA REVIEW

WHAT IS DATA REVIEW

Data Review is an attended step that allows a user to review the extraction data from documents performed by the Extraction Activity. How a user would accomplish this is via an Attended Client module called Grooper Data Review.

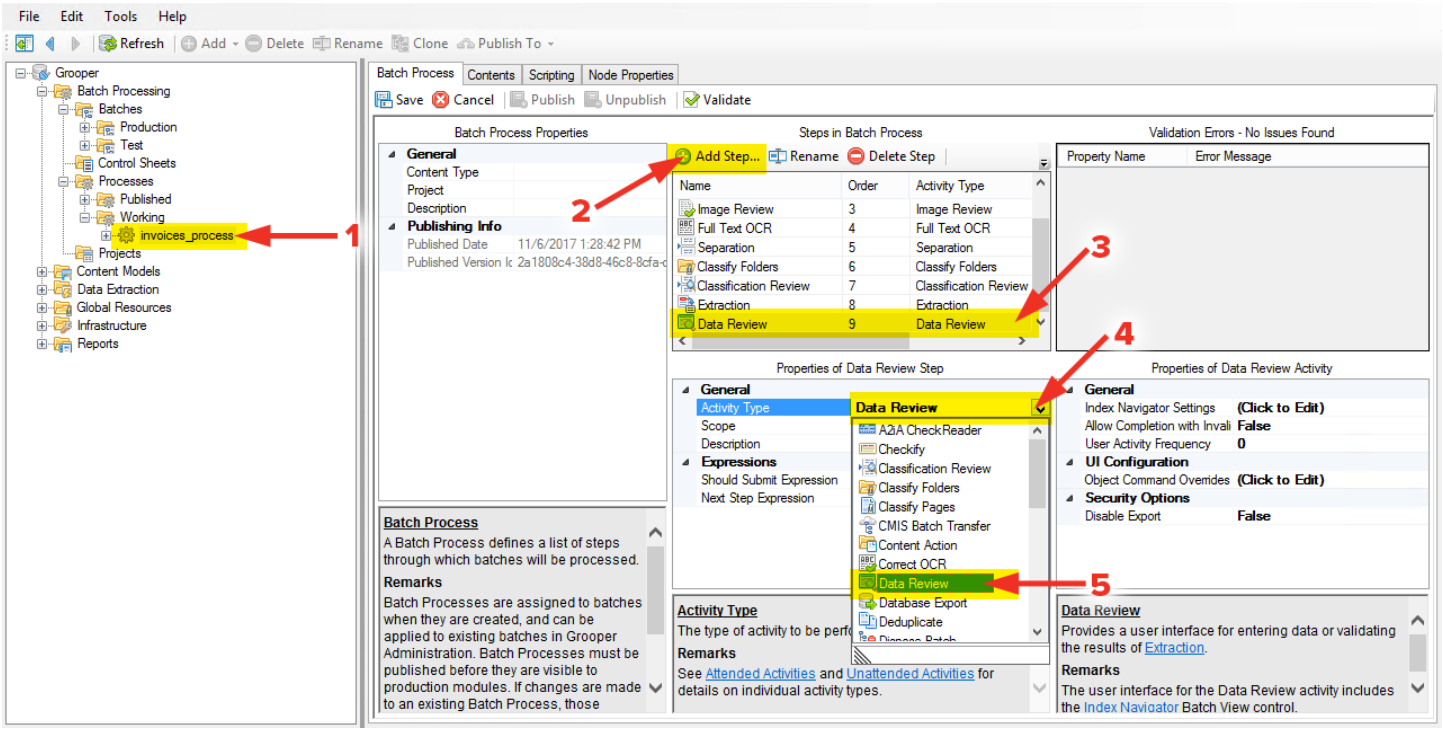


HOW TO CONFIGURE AND PEROFORM DATA REVIEW

STEP 1 – ADD DATA REVIEW STEP

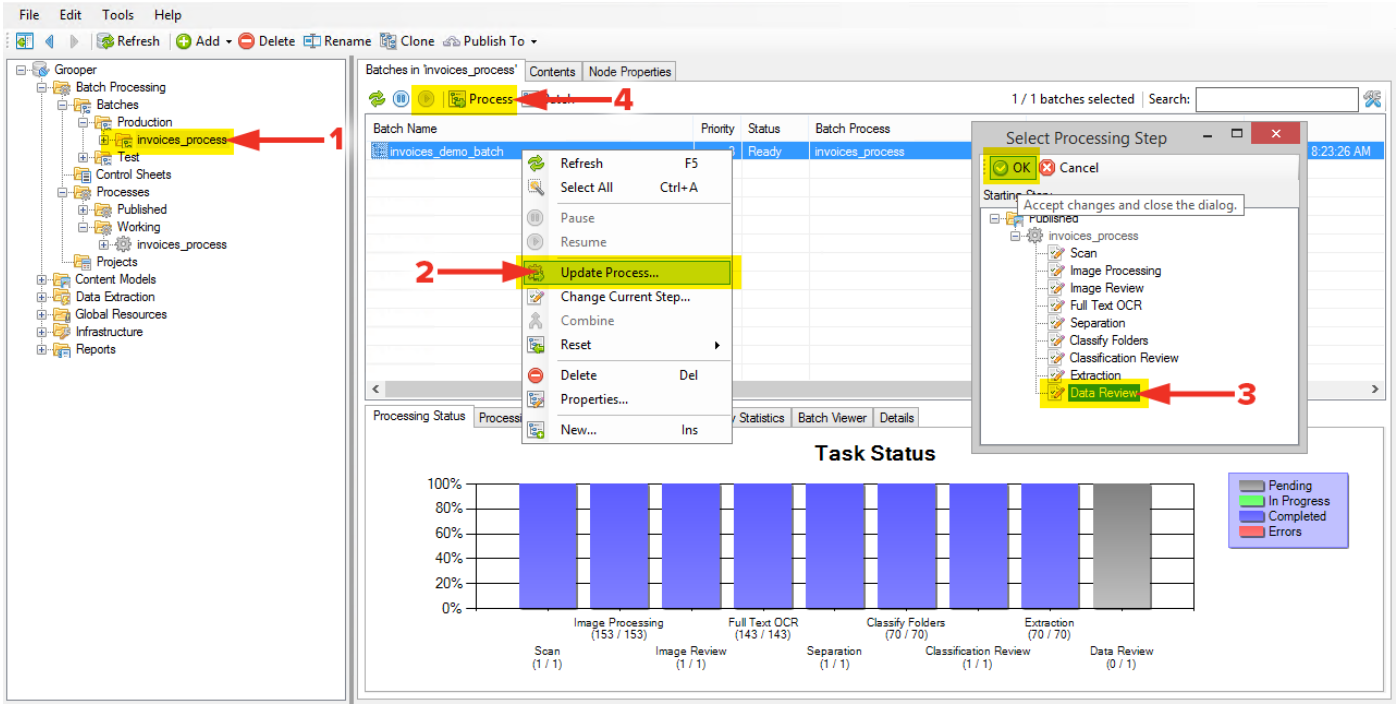
Like Classification Review (see Classification Review) before it, not much is required to setup Data Review. We simply need to add the step to our batch process.

(1) Go to and expand the Working processes node and select the invoices_process node. **(2)** Add a step and **(3)** set its Activity Type to Data Review. It's not seen in the image below due to the dropdown covering it, but be sure the Scope is set to Batch. Save and Publish the process.



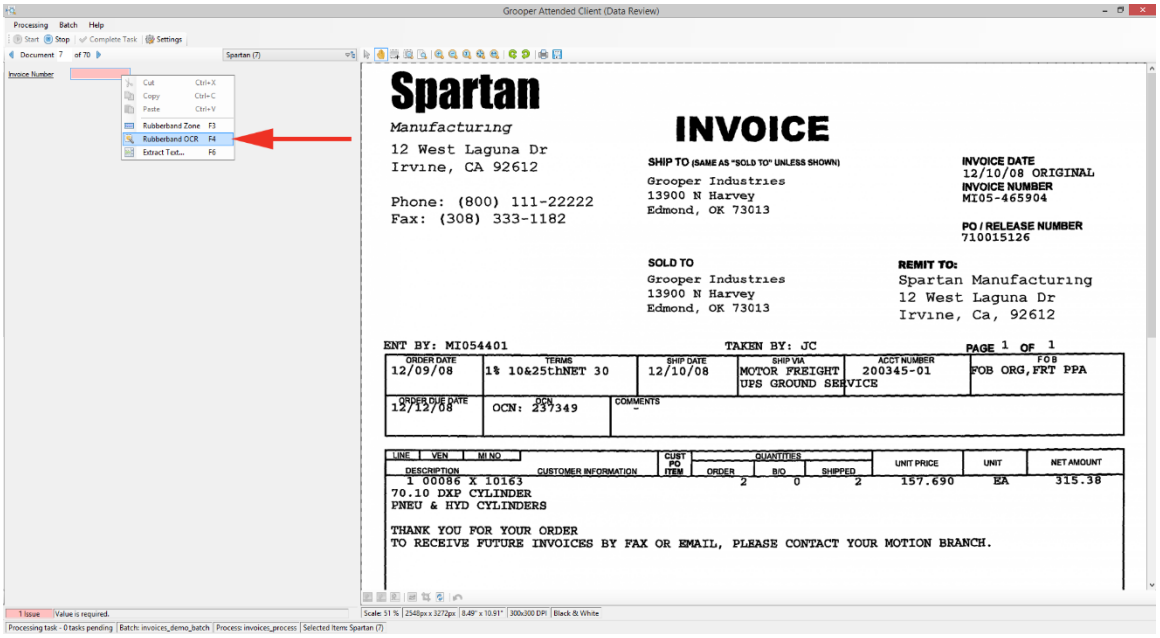
STEP 2 – UPDATE BATCH PROCESS

(1) Select the invoice_process node and in the Batch Manager view select the invoices_demo_batch. Make sure it's Paused and using either the batch dropdown, **(2)** or right-click and select Update Process... **(3)** In the Select Processing Step window, select Data Review and click OK. **(4)** Start the batch and Process it.



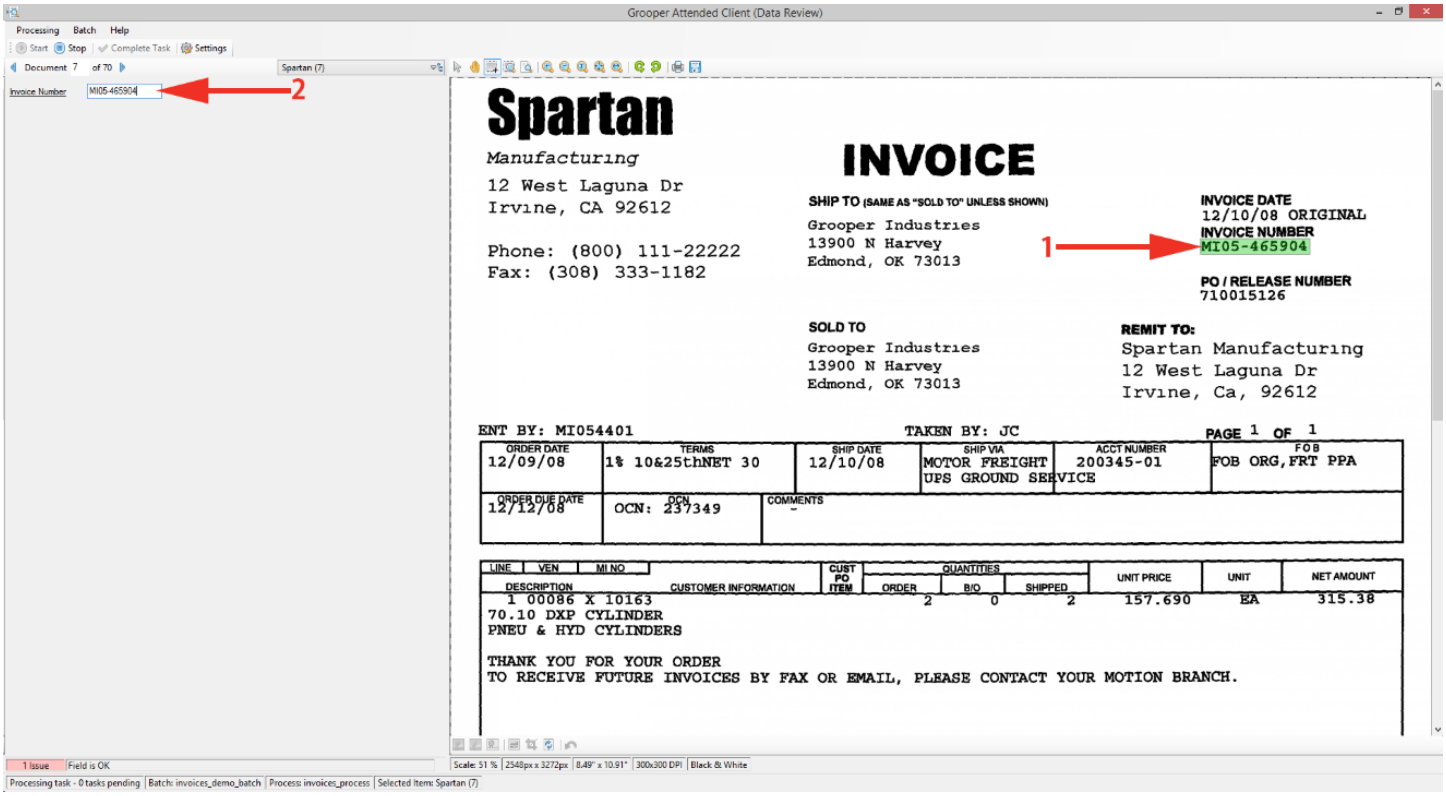
STEP 3 – PROCESSING DATA REVIEW – VALIDATION ERROR

Processing from the previous step will bring up the Grooper Attended Client in its Data Review configuration. Because we set our Data Field to required (Extraction step 14), and nothing was successfully extracted, this document gets flagged and is brought immediately to our attention in Data Review. Right click the field (it's made red as a result of there being a validation problem) and select the Rubberband OCR option...



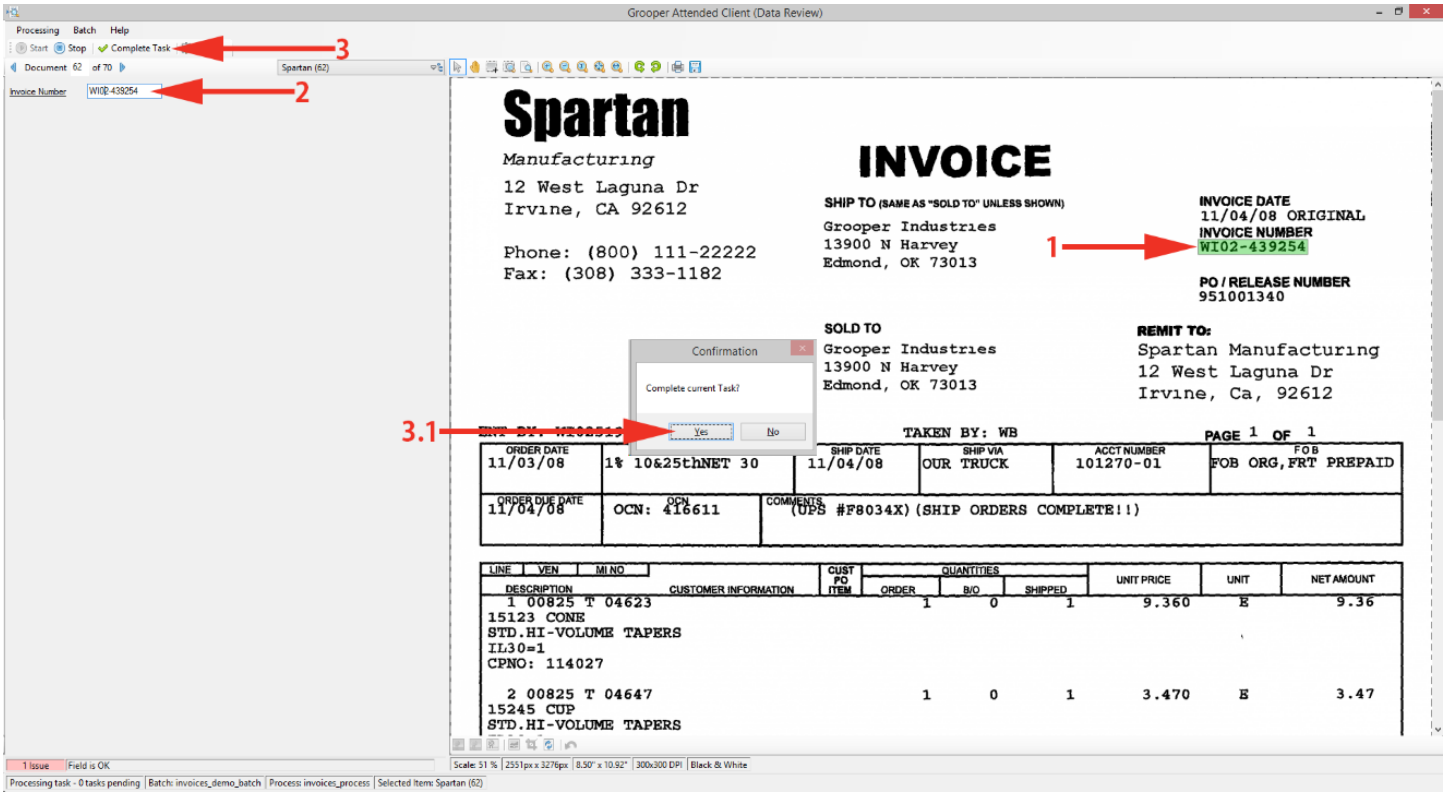
STEP 4 – CONTINUING WITH RUBBERBAND OCR

Having selected Rubberband OCR nothing will seem immediately apparent, but the tool is active. **(1)** In the Image Viewer, drag a box around our desired value. Because there is OCR data there, **(2)** the data that OCR read will populate the Invoice Number field on the left. You may notice that the information inserted is incorrect as OCR picked up a capital O instead of a 0, and a capital S instead of a 5. Go ahead and change these to their correct values. Once you're satisfied, click TAB to move out of the field, and this will take us to the next document with an error.



STEP 5 – FINAL VALIDATION AND CONFIRMATION

This document has a similar problem, so **(1)** Rubberband OCR it as well, and **(2)** replace the capital O with a 0. With this field validated, and there being no more validation issues, you'll see the Complete Task button become available. **(3)** Click the Complete Task button and confirm. This will close the Grooper Attended Client module and bring us back to Grooper Administration. Go ahead and Pause the batch.



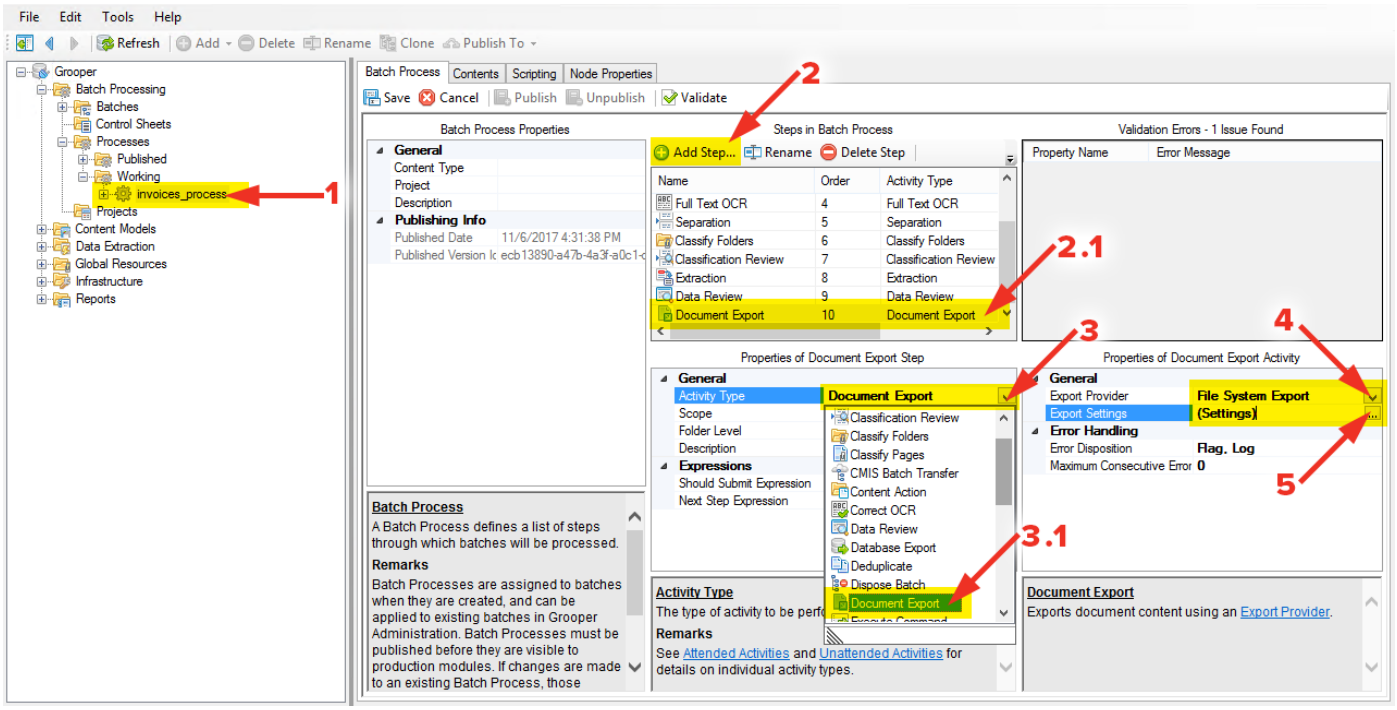
PHASE 5 - DELIVER

The final phase of the document processing in **Grooper** is sending the files to their final distention, or put simply: export. There are numerous end results for a dizzying array of different types of documents from databases to simple file systems. The use case and system infrastructure of the end user will determine how the documents are finally stored.

HOW TO CONFIGURE AND PROCESS FILE SYSTEM EXPORT

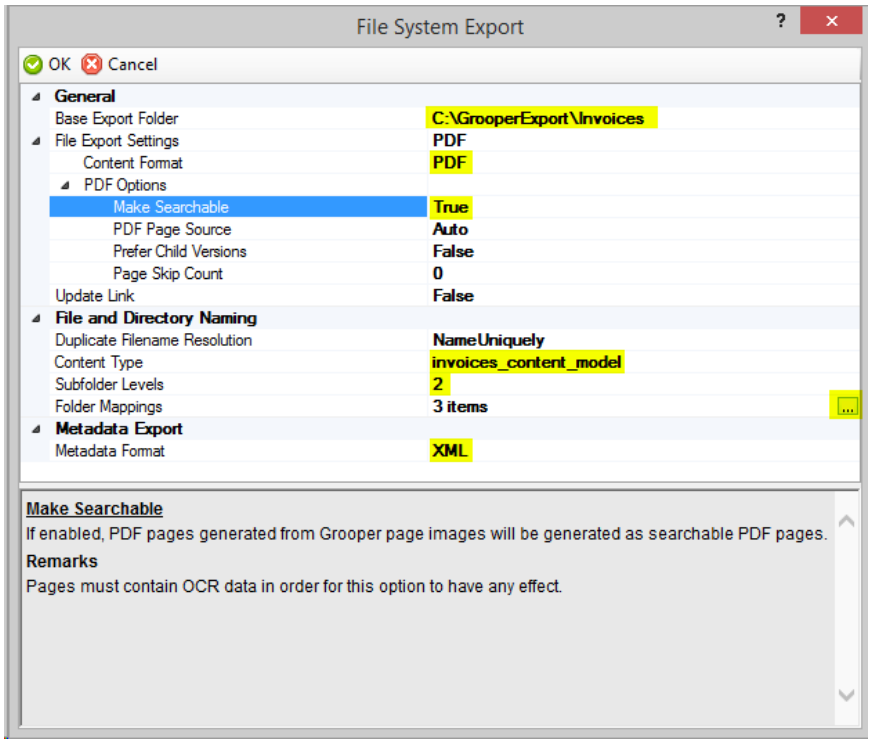
STEP 1 – CONFIGURE BATCH PROCESS STEP

(1) Navigate to the invoices_process batch process node within the Working folder. **(2)** Add a step and **(3)** set its Activity Type to Document Export. **(4)** Set Export Provider to File System Export and click the ellipsis button to open the Export Settings window.



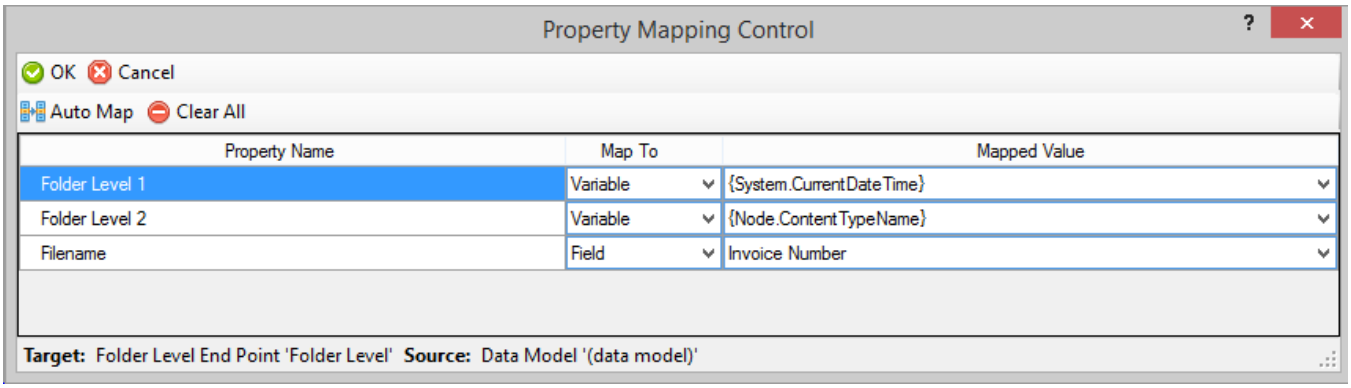
STEP 2 – EXPORT SETTINGS

In the File System Export window that opens, there are several settings that need to be set that will dictate how and where our files will get exported. The first, Base Export Folder, is an arbitrary path for the base export. For my system I created the path C:\GrooperExport\Invoices, and this could be different for your environment. Expand File Export Settings and change the Content Format to PDF. Expand the PDF Options and Make Searchable True. Click the dropdown for Content Type and choose the invoices_content_model. Set Subfolders to 2. Set Metadata Format to XML. Finally, click the ellipsis to open settings for the Folder Mappings.



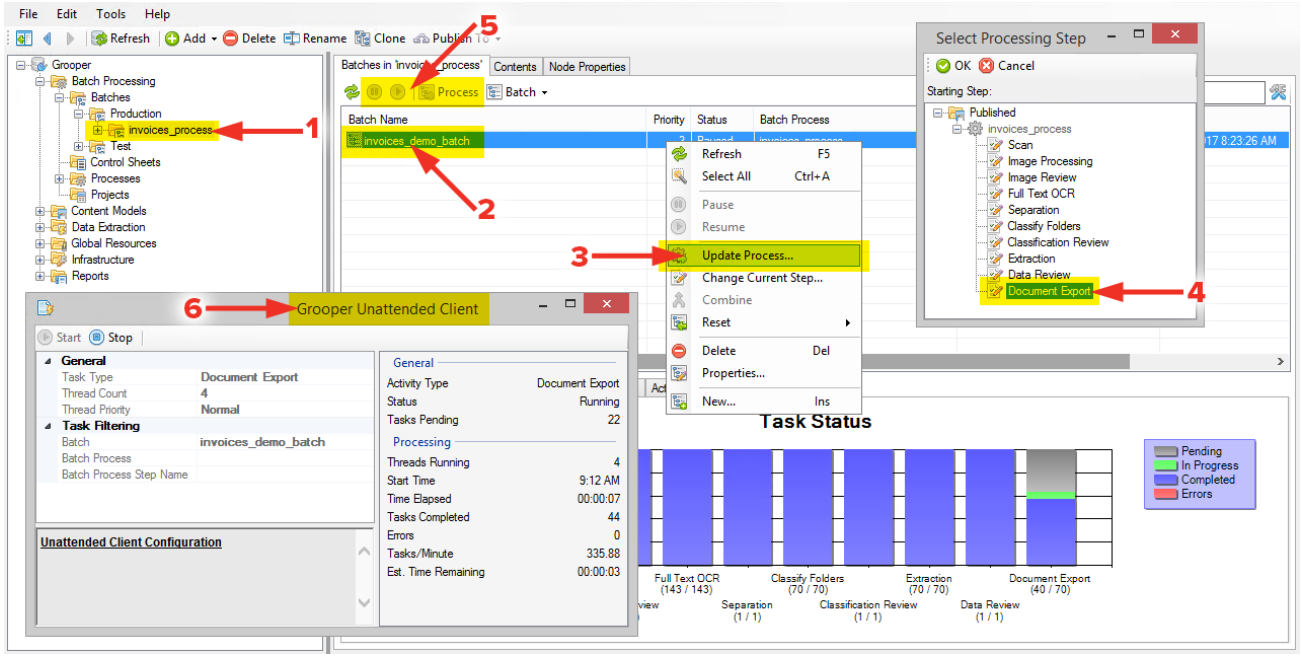
STEP 3 – PROPERTY MAPPING CONTROL

In the Property Mapping Control window there are two settings that procedurally determine the subfolder naming that will be created from the Base Export Folder. For Folder Level 1 set Map To to Variable and Mapped Value to {System.CurrentDateTime}. For Folder Level 2 set Map To to Variable and Mapped Value to {Node.ContentTypeName}. Finally, for Filename set Map To to Field and Mapped Value to Invoice Number. Click OK to close both open windows. Save and Publish the batch process.



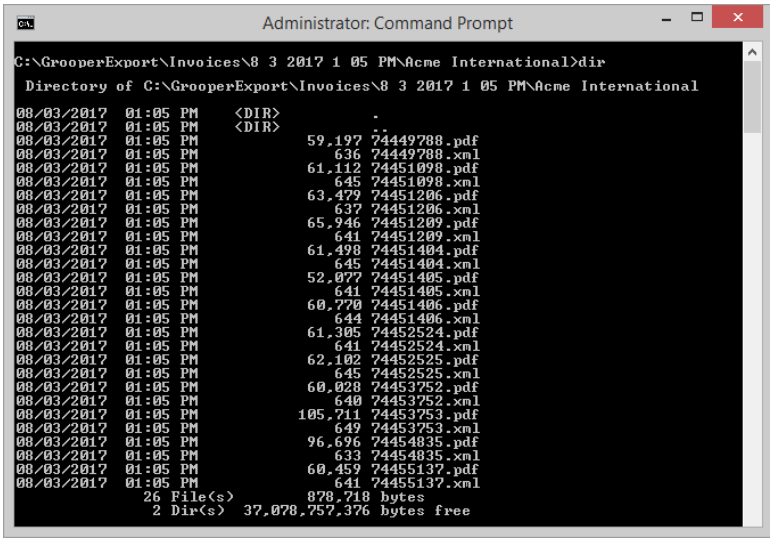
STEP 4 – UPDATE BATCH PROCESS

With our batch process updated and published, **(1)** go to the invoices_process node in the Production Batches area. **(2)** Select the invoices_demo_batch and make sure it's paused. **(3)** Using the Batch dropdown, or right-click and select Update Process... **(4)** In the Select Processing Step window select the Document Export step and click OK. **(5)** Start the batch and Process it. **(6)** The Unattended Client window will appear and begin processing the Document Export step. When the processs completes, the window will close and our batch process will be complete.



STEP 5 – VERIFY SUCCESSFUL EXPORT

With the Document Export step successfully completed, let's check the directory that was setup. In the image below, I've navigated to C:\GrooperExport\Invoices and within that folder there is a folder with the current system Date and Time. Within that folder is another with the Document Type that was assigned during the Classification step (there should be folders for each Document Type), and within this are files for the PDFs that successfully exported (named after their invoice number), and XML files with the index data of those documents.



A FINAL NOTE

For the scope of this project, this process is complete, and our work could be considered done. The batch that was created will stay in the Production list until deleted. One could add a Dispose Batch step to automatically delete batches upon successful completion, if so desired. The batch process that was created during this exercise will remain as a template to process documents of this variety moving forward. That is ultimately the goal of what we just walked through: creating and understanding the batch process.

While getting the small number of documents processed and successfully exported was a result, it's having this template, this batch process, and most importantly the Content Model available to process all future documents of this type moving forward that is most important. It's also worth mentioning that the linear approach that was taken in this document to create the batch process is not the typical means by which one would approach making a batch processes. Honestly, the creation of the steps of the batch process are usually the last thing you'll do, as ingesting documents and creating the Content Model are usually the first things you'll do. Setting up a solid Content Model with accurate Separation, Classification, and Extraction constitute the bulk of the work you'll do with **Grooper**. And, as a result, future training material will be heavily comprised of information to master these, among other, aspects of **Grooper**.

Finally, thank you for taking the time to read through this document and begin your journey to mastery of this incredibly powerful application. Welcome to the world of...