



Advanced Capture

Reference Guide

Includes:

Installation Guide

Administration Guide

User Guide

Copyright

Information in this document is subject to change without notice. The software described in this document is furnished only under a separate license agreement and may be used or copied only according to the terms of such agreement. It is against the law to copy the software except as specifically allowed in the license agreement. This document or accompanying materials contains certain information which is confidential information of Hyland Software, Inc. and its affiliates, and which is subject to the confidentiality provisions agreed to by you.

All data, names, and formats used in this document's examples are fictitious unless noted otherwise. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright law, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Hyland Software, Inc. or one of its affiliates.

Hyland®, Hyland Software®, Hyland Healthcare, and Hyland product names are registered and/or unregistered trademarks of Hyland Software, Inc. and its affiliates in the United States and other countries. All other trademarks, service marks, trade names and products of other companies are the property of their respective owners.

© 2021 Hyland Software, Inc. and its affiliates. All rights reserved.

Document NameAdvanced Capture
Department/Group Documentation
Revision NumberFoundation EP5

OVERVIEW

General Overview	1
Advanced Capture	1
Point and Shoot Indexing	2
Applications	3
System Prerequisites	3
Licensing	4
Simplified Licensing	4
Additional Licensing	4

INSTALLATION GUIDE

INSTALLATION

Requirements	6
General Requirements	6
Licensing	6
Pre-Installation	6
Hyland OCR Engine	6
Additional Information	7
Bar Code Processing	7
Installation	7
Command Line Switches	8
Auto-Processing Switches	8
-SCANAUTOINDEX	8
-SCANAUTOQUEUE	8
AutoProcessDelay INI Setting	9
INI Options	9
System	9
OCRLocation	10
OCRRestartLimit	10
Scan	10
AutoProcessDelay	10
AutomatedIndex	10
Diagnostic	10
AnalysisMin	11
AnalysisMax	11
MinProcessHeight	11
INI File	11
Location	12
INI Considerations in a Citrix and Microsoft Windows Remote Desktop Environment	12
Editing the INI File	13
Backup/Recovery	13

Backup	13
Configuration	13
Registry Settings	13
External Files	13
Module-Related INI Options	13
Additional Steps	14
Recovery	14
Configuration	14
Registry Settings	14
External Files	14
Module-Related INI Options	14
Registration	14
Additional Steps	14
Upgrade Considerations	15
Advanced Capture Upgrade Considerations	15
Troubleshooting	15
Common Issues	15
Process Locks	17
Advanced Capture & AutoFill Keyword Sets	18
Contacting Support	19

ADMINISTRATION GUIDE

CONFIGURATION

Configuration Overview	21
OCR and Document Type Configuration	22
Working With AutoFill Keyword Sets in Advanced Capture	22
Ad Hoc Advanced Capture	22
Batch Advanced Capture	23
Point and Shoot Indexing	23
Creating AutoFill Keyword Sets via Advanced Capture	23
Configuring Point and Shoot Indexing	23
A Note About Single and Double-Byte Characters	23
Configuring Advanced Capture	24
Configuring Advanced Capture Forms for Institutional Databases	24
Using Advanced Capture Forms with Institutional Databases	26
Testing Advanced Capture Forms	26
Performing Advanced Capture Processes	26
Granting User Groups Rights for Configuration and Processing	26
Configuring Advanced Capture Forms	27
Accessing the Advanced Capture Configuration Window	27
Opening the Advanced Capture Configuration Window from the Client	29
Opening the Advanced Capture Configuration Window from the Document Imaging Window	30

The Advanced Capture Configuration Window	31
The Form Configuration Pane.....	34
The Result Verification Pane.....	35
The Document Viewer	36
The Forms Toolbar	38
The Batch Documents Pane	40
The Tools Pane	41
Defining Advanced Capture Forms	44
Creating a New Advanced Capture Form	45
Assigning Existing Advanced Capture Forms to Documents in the Batch	47
Modifying or Deleting a Form	50
Modifying a Form.....	52
Form Definition Options	53
Common Options.....	53
Create/Assign Form Options	59
Modify Form Options.....	60
Additional Information on Form Definition Options	60
Configuring Document Breaks.....	61
Specifying Advanced Capture Form Availability by Document Type	62
Configuring Scan Queue or Keyword Value Rules for a Form	65
Configuring Keyword Lookup/Replace Dictionaries.....	73
Configuring Lookup Text for a Form	83
Configuring Image Processing Operations.....	87
Restricting Form Modification Rights By User Group	94
Showing or Hiding Image Processing Operations	97
Updating the Archetype Document	98
Modifying Form Configuration Options for a Document Type	99
Configuring a Form Identification Zone	100
Configuring a Text Value Form Identifier	104
Configuring an Image Match Form Identifier	113
Modifying or Analyzing a Form Identification Zone	119
Modifying or Analyzing Overlapping Zones	121
Configuring A Data Field Zone	121
Configuring a Text Data Field Zone	125
Creating an Entire Page Zone for Find by Tag/Expression	147
Configuring a Line Item Extraction Data Field Zone	150
Configuring Line Item Extraction in Automatic Table Decomposition Mode	152
Configuring Line Item Extraction in Manual Table Decomposition Mode	153
Additional Line Item Extraction Configuration Options.....	162
Configuring a Grouped Line Item Extraction Data Field Zone	169
Identifying Keyword Values from Line Item Data	173
Identifying Keyword Values Using Tags or Regular Expressions.....	181
Configuring Additional Grouped Line Item Extraction Options.....	190
Configuring an Optical Mark Data Field Zone	197
Configuring a Signature Detection Data Field Zone	207

Configuring a Snippet Data Field Zone	214
Modifying or Analyzing a Data Field Zone	219
Modifying or Analyzing Overlapping Zones	221
Changing the Data Field Zone Processing Order	222
Configuring a Page Registration Zone	224
Configuring a Page Registration Zone to Use a Text Value	228
Configuring a Page Registration Zone to Use an Image Match	238
Modifying or Analyzing a Page Registration Zone	244
Modifying or Analyzing Overlapping Zones	246
Configuring Multiple Form Identification or Page Registration Zones for Scaling	246
Limitations and Requirements for Scaling	247
The Regular Expression Library	247
Editing Default Regular Expressions	249
Creating Custom Regular Expressions	250
Configuring Regular Expressions From a Text Field	251
Importing/Exporting Advanced Capture Forms	253
Importing Advanced Capture Forms	253
Resolving Form Template Collision	255
Reconciling Configuration References	256
Exporting Advanced Capture Forms	258
Configuring Form Definition Groups	259
Advanced Capture Form Order Analysis	263
Advanced Capture Form History Analysis	265
Granting Configuration Rights to Additional User Groups	267
Updating Multiple Form Definitions	269
Generating Configuration Reports	272
Configuring XML Renditions	274
Testing Advanced Capture Forms	274
Viewing Document History	277
Batch History	277
Generating a Document History Report	277
Configuring Advanced Capture to Use Intelligent Character Recognition	278
Limitations of Intelligent Character Recognition	280
Improving the Accuracy of Intelligent Character Recognition	280
Configuring a Scan Queue for Advanced Capture	282
Disabling Images of Keyword Values from Displaying During Indexing	284
Disabling Automatic Keyword Value Zoom	285
Navigating Only Between Suspect Keyword Values	285
Configuring Automatic Advanced Capture Processes	286
Automating Advanced Capture at the Scan Queue Level	287
Automating Advanced Capture via a Scheduled Process	289
Automating Advanced Capture via a Command Line Switch	289
Exiting the Advanced Capture Form Configuration Interface	289
System Interaction	289
Automated Redaction	289

Document Imaging	290
Orientation Detection & Automatic Page Rotation	290
Appending Documents to Existing Documents	290
EDM Services	290
Intelligent Character Recognition (ICR)	291
Limitations of Intelligent Character Recognition	291
Improving the Accuracy of Intelligent Character Recognition	292
Unity Client	293
Web Client	293
Workflow	293
Virtual Print Driver	293

SCHEDULING

Scheduling Overview	294
Configuring & Using the Scheduler	294
Requirements for Configuring/Running a Scheduled Process	294
Using the -SCHED and -SCHEDINST Switches	295
-SCHED	295
-SCHEDINST	295
Verifying the Scheduler is Running	296
Running Multiple Scheduled Processes	296
Scheduled Process Configuration Reports	297
Working With Process Formats	297
Creating a Scheduled Process Format	297
Schedule Configuration	300
Calendar	301
Default Daily Schedule	302
Selected Day	303
Processing Options	304
Viewing Scheduled Processes	309
Modifying a Scheduled Process Format	309
Deleting a Scheduled Process Format	310
Running/Suspending a Scheduled Process Format	310
Working With Process Jobs	311
Creating a Job	311
Configuring a Job	313
Scheduling a Job	315
Schedule Configuration	316
Calendar	317
Default Daily Schedule	319
Selected Day	320
Processing Options	322
Viewing a Job	325
Modifying a Job	326

Renaming a Job	326
Deleting a Job	326
Running/Suspending a Job	327
Viewing the Activity Log.....	328
Creating Schedule Templates	330
Creating Schedule Templates	330
Calendar	331
Default Daily Schedule.....	331
Selected Day	332
Configuring Schedule Logging	334
Creating a Scheduler Workstation Group	335
Editing a Scheduler Workstation Group	338
Deleting a Scheduler Workstation Group	341

LEGACY LICENSING

Overview	342
Additional Licensing.....	343

ADVANCED CAPTURE BEST PRACTICES

General Information	344
Recommended Standards for OCR Processing	344
Recommended Standards for ICR Processing.....	345
Date Strings	346
When Documents Can Be Modified	346
When Documents Cannot Be Modified	347
OCR vs. AutoFill Keyword Sets.....	347
Usage	347
Advanced Capture.....	347
Ad Hoc Advanced Capture	347
Batch Advanced Capture	348
Performing Point and Shoot Indexing.....	348
Configuration	348
Advanced Capture Form Configuration	348
Keyword Type Masking.....	349

REGULAR EXPRESSIONS

Commonly Used Characters.....	350
Commonly Used Regular Expressions	351
Personal Information	351
Higher Education.....	352
General Examples	352

QUICK REFERENCE GUIDES

USER GUIDE

USAGE

Advanced Capture.....	360
Performing Ad Hoc Advanced Capture.....	360
From the Document Search Results List	362
From an Open Document	363
From the Document Imaging Window	364
From the Import Document Dialog Box	367
From a Folder	368
From a Workflow Queue	371
From the Virtual Print Driver	371
Performing Batch Advanced Capture	372
On a Newly Imported Batch	373
Resolving In-Progress Batches	375
Index Verification	375
On Documents Queued for Ad Hoc Advanced Capture	379
Keyword Considerations.....	382
Keyword Type Masking	382
Default Keyword Values	382
Point and Shoot Indexing.....	382
Performing Point and Shoot Indexing.....	383
A Note About Single and Double-Byte Characters	388

General Overview

The Advanced Capture module allows documents to be automatically indexed by extracting indexing values (that is, a Document Type identifier, the Document Date, and Keyword Values) directly from documents via optical character recognition (OCR) technology.

The Advanced Capture module consists of two distinct, but equally powerful, functionalities:

- **Advanced Capture.** Advanced Capture automatically compares documents to pre-defined Advanced Capture forms, or templates, that define the areas of the document that are to be processed. Once a document is matched to a form, it is processed and the extracted indexing values are assigned to it.
Advanced Capture can be performed on the following types of documents:
 - Images
 - Uneditable PDFs (image-based or text-based).
- **Point and Shoot Indexing.** Point and Shoot Indexing allows you to select an area of a document to be processed so a Document Date or Keyword Values can be extracted from that area.

Point and Shoot Indexing can be performed on image documents only.

Both Advanced Capture and Point and Shoot Indexing support Keyword data type recognition and Keyword Type masking.

Your Advanced Capture solution is able to identify and process machine-printed data via optical character recognition (OCR). If your solution is also licensed for **ICR Support for Advanced Capture**, your solution is also able to identify and process handwritten text via intelligent character recognition (ICR).

Note: ICR is not supported for Point and Shoot Indexing.

Unlike other forms recognition products, Advanced Capture has no per-page or per-document restrictions or costs, and all Advanced Capture is performed within, and is seamlessly integrated into, OnBase's Document Imaging interface – no external databases or additional hardware or software is required.

Advanced Capture

Like other document processing options, such as Bar Code Processing or Full-Page OCR, Advanced Capture can be performed on batches of documents within a scan queue. It can also be performed on a single document from several places in the OnBase Client module, including the Document Viewer, the **Import** dialog box, or, with additional licensing, from within a Workflow queue.

Advanced Capture can help to streamline most indexing tasks. Advanced Capture can do much more than identify and extract basic indexing data; depending on your configuration, Advanced Capture can also perform these advanced indexing tasks:

- Automatically score and evaluate indexing data read by the Advanced Capture engine to determine if a document needs to be manually reviewed by a user for accuracy. Batches of documents that are determined to have been accurately indexed can be marked as fully indexed and committed without any user intervention.
- Group and store line-item values using Multi-Instance Keyword Type Groups.
- Perform Optical Mark Recognition (OMR) and signature detection, and assign appropriate Keyword Values based on these results.
- Read MICR line data from check documents.

Advanced Capture is easily configured from using a visual, point-and-click interface. Users can create, modify, and delete Advanced Capture forms -- templates that allow a user to define the areas of the document that are processed and the type of indexing data (i.e., a Document Type identifier, the Document Date, or Keyword Value) that is extracted from each area.

A user with the proper rights can create and edit Advanced Capture forms from within the OnBase Client or the Document Imaging interface. Access to the OnBase Configuration module is not needed.

During processing, a document is classified and indexed based on the form matched to it by the Advanced Capture engine. Multiple forms can be created and associated with a single OnBase Document Type, allowing documents with different physical layouts to be processed and assigned to the same Document Type.

For example, an organization might scan and archive all of its shipping receipts, from any number of vendors, in one Document Type. Even though all of these documents have a different physical layout, as long as Advanced Capture is able to match an Advanced Capture form to each one, they can all be properly processed, indexed, and archived in the same Document Type.

If Advanced Capture cannot match a form to a document, or if its built-in quality controls determine that a value returned by the Advanced Capture engine is suspect, the batch containing the document is routed for user review. Documents are reviewed and corrected in the familiar environment of OnBase Document Imaging; no external programs or additional software is required.

Point and Shoot Indexing

Point and Shoot Indexing offers organizations an automated and accurate alternative to manual indexing without requiring a major change to existing scanning and indexing processes. Users can perform Point and Shoot Indexing when indexing individual image documents from within Document Imaging without needing pre-configured Advanced Capture forms.

Point and Shoot Indexing can be performed in any of Document Imaging's indexing batch status queues (for example, **Awaiting Indexing**, **Index in Progress**, **Awaiting Reindex**, or **Awaiting QA Reindex**). Users can capture a value from the scanned image and automatically assign it to the selected Keyword Type simply by selecting the area of the image containing the desired value with the pointer. This area is automatically processed and the selected Keyword field is populated with the extracted data.

Applications

Advanced Capture and Point and Shoot Indexing can be used by any organization that is looking for ways to reduce the cost and chance for error that exist when manually indexing documents.

- An Accounts Payable department can use Advanced Capture to automatically index all incoming vendor invoices. Even though no two of the invoices are physically alike, an Advanced Capture form can be created for each different layout and all of the documents can still be archived under the same Document Type. By automating the indexing process, the clerk or clerks that used to manually index these invoices now have additional time to work on more critical tasks, and the capacity of the department to process invoices can be greatly increased without any additional staff.
- A Human Resources department can take advantage of Point and Shoot Indexing to improve its hiring process. As cover letters and resumes are received from applicants, the HR specialists can scan the paper documents, or sweep batches of electronically-submitted images, and use Point and Shoot Indexing to extract Keyword Values from the images more accurately and efficiently than if they had to manually index them.
- A bank with several regional offices can use Disconnected Scanning to scan documents into a central OnBase system. One or more centralized Advanced Capture processing workstations can be designated to perform Advanced Capture on all batches uploaded to the system from each of the regional offices.
Additionally, to conserve network capacity and bandwidth during business hours, Disconnected Scanning's scheduling functionality can be used to schedule batch uploads and the OnBase Scheduler can be used to schedule Advanced Capture processes to take place at night.

System Prerequisites

In order to take advantage of batch Advanced Capture processing, your OnBase solution must be properly configured to import batches of images or PDF documents into a scan queue.

A number of OnBase modules, such as Document Imaging, Disconnected Scanning, or the Document Import Processor (DIP), can be used to import documents into batches via scan queues.

Additionally, the Hyland OCR Engine software must be installed on each workstation that is to perform Advanced Capture. This software is available from your solution provider, and does not need separate licensing or registration.

Licensing

Beginning in OnBase Foundation EP5, new customers must use simplified licensing to access Advanced Capture functionality. Existing customers upgrading from a version of OnBase prior to OnBase Foundation EP5 can continue to use legacy licensing to access this functionality.

If you are a new customer as of OnBase Foundation EP5 or greater, see [Simplified Licensing on page 4](#).

If you are upgrading from a version of OnBase prior to OnBase Foundation EP5, see [Legacy Licensing on page 342](#).

Simplified Licensing

In addition to a base package license for standard OnBase functionality, one or more add-on package licenses are required to access the following functionalities:

Add-On License	Functionality
Ad-Hoc Advanced Capture	Point and Shoot Indexing and Advanced Capture processing on individual documents
Advanced Capture	Point and Shoot Indexing and Advanced Capture processing on individual and batch documents
ICR Support for Advanced Capture	Identification of handwritten text in an Advanced Capture Data Field Zone using Intelligent Capture Recognition (ICR)

For more information on the packages available to you, contact your account manager.

Additional Licensing

Additional licensing may be required to use the following products in conjunction with Advanced Capture:

- Workflow
- Automated Redaction

For the licensing requirements of a specific product, see that product's documentation.



Advanced Capture

Installation Guide

Requirements

The following sections outline requirement information specific to Advanced Capture in OnBase Foundation EP5.

General Requirements

For general requirement information that applies to Advanced Capture and other modules, see the sections on the following topics in the **Installation Requirements** manual:

- Database Requirements
- Supported Desktop Operating Systems
- Microsoft .NET Framework Requirements
- General C++ Requirements
- OCR Processing Workstation Hardware Requirements
- Miscellaneous Requirements

Licensing

See [Licensing on page 4](#) for licensing requirements.

Pre-Installation

In order to use Advanced Capture, your OnBase solution must be properly licensed and configured to import batches of image or PDF documents into OnBase scan queues.

A number of OnBase modules, such as Document Imaging, Disconnected Scanning, or Document Import Processor (DIP), can be used to import documents into batches via scan queues.

Hyland OCR Engine

Various OnBase modules (for example, Intelligent Capture for AP, Advanced Capture, Full-Page OCR, and Automated Redaction) use the Hyland OCR Engine as their recognition package.

The Hyland OCR Engine (version 21.00 or later) is required for these modules. This software is included with the purchase of these modules, and it is available from your OnBase solution provider. It does not require separate licensing or registration.

Tip: Even when older versions of the Hyland OCR Engine are supported, it is highly recommended that you use the most recent version with your solution.

When deploying the OCR engine on the Data Capture Server, the 64-bit version must be used.

With this in mind, note that the modules and functionalities that use the Hyland OCR Engine can reside in the following platforms:

- **OnBase Client only:** Advanced Capture and Automated Redaction.
- **Data Capture Server only:** Intelligent Capture for AP and Interactive Data Capture.
- **Either:** Full-Page OCR (batch or ad hoc).

Additional Information

- You must install the Hyland OCR Engine (64-bit) on each server that will perform OCR processing on the Data Capture Server.
- Multi-threaded access (that is, multiple documents can be processed at once) is supported by the Hyland OCR Engine (64-bit) on the Data Capture Server.
- In order for the Hyland OCR Engine to recognize fonts correctly, Windows ClearType must be disabled. If Windows ClearType is enabled when OCR is being performed, the OCR engine will disable it automatically.

Bar Code Processing

Bar code processing requires the Hyland Bar Code Recognition for OnBase software to be installed on the workstation that performs Advanced Capture.

For more information, see the **Bar Code Process** module reference guide or help files.

Installation

With the exception of the installation of the Hyland OCR Engine, no additional installation is required for the Advanced Capture module.

To install the Hyland OCR Engine:

Obtain the **Hyland OCR Engine [<version>] for DC Server x64** installer from Technical Support and copy it to each machine that will be performing OCR processing:

Note: The 64-bit OCR engine can only be deployed on 64-bit systems.

1. Double-click the Hyland OCR Engine [<version>] executable to launch the installer. The setup wizard's **Welcome** screen is displayed.
2. Click **Next**. The **Destination Folder** screen is displayed.

3. On the **Destination Folder** screen, you must specify the installation location on the OCR processing workstation.
The following default installation location is automatically selected:
C:\Program Files\Hyland\OCR
To select a different location, manually enter the path in the **Destination Folder** field or click **Change** to browse to the location.
When the destination folder has been selected, click **Next**.
4. The **Ready to install Hyland OCR Engine [<version>]** screen is displayed. Click **Install** to begin the installation.
5. Once the installation is complete, click **Finish**.

Command Line Switches

The following command line switches can be used in conjunction with Advanced Capture:

- **-SCHED**. This switch is used if the workstation will be performing scheduled Advanced Capture processes.

Auto-Processing Switches

-SCANAUTOINDEX

When this switch is used, the workstation will poll the **Awaiting Advanced Capture** queue and automatically process batches of documents found within it. Batches are processed in the order in which they were brought into the system.

-SCANAUTOQUEUE

This switch is used in combination with any of the auto-processing command line switches to direct the processing workstation to only automatically process batches from a specific scan queue and/or automatically process documents from the corresponding ad hoc processing scan queue (e.g., **Awaiting Ad Hoc Advanced Capture** or **Awaiting Ad Hoc OCR**). The batch scan queue(s) to be processed are identified by their Scan Queue Number, while the corresponding ad hoc processing scan queue to be processed is identified by a **0**.

Consider the following examples:

- To automatically perform OCR processing on batches scanned using scan queues #101 and #102, as well as on documents residing in the **Awaiting Ad Hoc OCR** queue, you would add the following switches, in this order, to the Client module's shortcut:
-SCANAUTOOCR -SCANAUTOQUEUE:0,101,102
- To automatically perform OCR processing on batches scanned using scan queues #101 and #102 only (i.e., without performing OCR processing on documents residing in the **Awaiting Ad Hoc OCR** queue), you would add the following switches, in this order, to the Client module's shortcut:
-SCANAUTOOCR -SCANAUTOQUEUE:101,102
- To automatically perform OCR processing on documents residing in the **Awaiting Ad Hoc OCR** queue only (i.e., without performing OCR processing on any batches), you would add the following switches, in this order, to the Client module's shortcut:
-SCANAUTOOCR -SCANAUTOQUEUE:0

Tip: The Scan Queue Number of a scan queue is displayed when the scan queue is selected in the **Scan Queue Configuration** dialog box (click **Import | Scan Queues** from the OnBase Configuration module).

Note: If you use both the **-SCANAUTOQUEUE** and **-SCANAUTOQUEUEEXCLUDE** switches in combination with an auto-processing command line switch, the **-SCANAUTOQUEUEEXCLUDE** switch overrides the **-SCANAUTOQUEUE** switch.

AutoProcessDelay INI Setting

A timer is used to poll the batch status queues affected by these switches to search for new batches to process. By default, this timer is set to 300 seconds (five minutes), but this value can be changed by modifying the **AutoProcessDelay** value in the onbase32.ini file.

For more information on this setting, see the information on the **AutoProcessDelay** setting in the **INI Settings** section below.

INI Options

The following INI settings are related to Advanced Capture:

System

The **[System]** section contains the following INI settings:

OCRLocation

The value for this entry is the path to the OCR engine's installation files. When it is absent, the OCR engine must be installed in the same location as the OnBase executable.

OCRRestartLimit

This setting controls the number of documents that the OCR engine will process before it is automatically restarted. This allows for more efficient processing of large batches of documents and is considered a best practice by the OCR engine's developer.

The OCR engine is automatically stopped and restarted without user interaction. It is not noticeable to the user and will not affect any OCR processing.

Note: The default setting is based on the OCR engine developer's recommendations. It should not be changed unless you are instructed to do so by Technical Support.

Scan

The **[Scan]** section contains the following INI setting:

AutoProcessDelay

The **AutoProcessDelay** setting pertains to workstations running with the auto-processing command line switches (for example, **-SCANAUTOBARCODE**, **-SCANAUTOCOMMIT**, or **-SCANAUTOCUSTOMPROCESS**). These switches are used to poll the batch status queues in the Document Imaging window where batch processing actions take place and automatically process batches of documents found within them.

The **AutoProcessDelay** setting in the onbase32.ini file specifies (in seconds) how often these batch status queues are polled. By default, this setting is set to **300**, meaning the service polls and processes all batches within the batch status queues every 300 seconds (five minutes). You can change this setting to modify the frequency in which the batch status queues are polled and the batches found within are then processed. The minimum value for this setting is **5** (the batch status queues are polled and the batches found within them every five seconds).

AutomatedIndex

The **[AutomatedIndex]** section contains the following INI settings:

Diagnostic

When enabled (**Diagnostic=1**), this option will add the information from the **Result Verification** panel of the **Advanced Capture Configuration** window to the Verification Report.

When this option is not enabled (**Diagnostic=0**), the Result Verification information is not included in the Verification Report.

AnalysisMin

This option sets the minimum number of document-to-form matches needed to perform form order analysis. By default, it is set to **100**. (**AnalysisMin=100**).

Note: If your solution does not contain more documents than the minimum specified by the **AnalysisMin** setting, you will not be able to perform form order analysis.

AnalysisMax

This option sets the maximum number of document-to-form matches used when performing form order analysis. By default, it is set to **30,000**. (**AnalysisMax=30000**).

Note: If your solution contains more documents than the maximum number of allowed documents, you are able to perform form order analysis; however, the analysis is only performed on the number of documents specified by the **AnalysisMax** setting (the excess documents are not included in the analysis).

MinProcessHeight

This option sets the minimum height (in whole inches) a page must be to be searched for Form Identification Zones during Advanced Capture processing. By default, it is set to **0** (**MinProcessHeight=0**), meaning that there is no minimum height. If this option is set to a positive integer value, and a page being processed for Form Identification Zones has a height less than this value, then the Advanced Capture engine automatically classifies the page as unidentified.

INI File

INI files (initialization files) are plain-text files that contain configuration information. These files are used by Windows and Windows-based applications to save and access information about your preferences and operating environment. OnBase uses an initialization file named onbase32.ini. If users do not have rights to access the onbase32.ini file, they will be unable to use the Client or the Configuration module.

The onbase32.ini file is primarily used to store settings specified in the Client or the Configuration module. For example, when a user selects a default data source in the OnBase Client's Workstation Options dialog box, this selection is saved to the onbase32.ini file. The onbase32.ini file is also used to make modifications to OnBase modules that cannot be made through the module's interface.

Location

For modern Windows operating systems, the default location of the onbase32.ini file is **C:\ProgramData\Hyland Software**. For previous versions of OnBase running on older operating systems, the default location of the onbase32.ini file was **C:\Documents and Settings\All Users\Application Data\Hyland Software**.

Note: To maintain backwards compatibility with previous versions of OnBase, OnBase checks the workstation's **C:\Windows** folder for the OnBase INI file if it is not found in the folder specified above. If the OnBase INI file is found in the **C:\Windows** folder, OnBase copies the file to the new location. The previously existing version of the OnBase INI file remains in the **C:\Windows** folder, but it is no longer used by OnBase.

Your onbase32.ini file may reside in a different location, if that location is specified by the following command line switch on the OnBase Client shortcut target:

-INIFILE= "full path\filename", where **full path** and **filename** are replaced by the specific path and file name.

If this command line switch is not used and you move or rename your onbase32.ini file, OnBase recreates the file in the default folder and ignores the newly created file.

INI Considerations in a Citrix and Microsoft Windows Remote Desktop Environment

In Remote Desktop environments, a remote session is established in which the user is running applications that are not installed locally. This presents a challenge when an application, such as OnBase, requires a user-specific INI file to establish unique settings. In a Remote Desktop environment, you must ensure that each user has a single, unique INI file to make sure any user-specific settings are consistent for that user.

Note: The default location of the OnBase INI file is not unique in a Remote Desktop environment.

To ensure that the INI file is accessible by OnBase and unique to each user in a Remote Desktop environment, the **-INIFILE** command line switch must be applied to the OnBase Client and Configuration shortcuts and be set to a unique location for the INI file.

Note: Additional details regarding the deployment of OnBase in a remote desktop environment is discussed in detail in the **Citrix and Microsoft Windows Remote Desktop Environment Deployment Guide**, available from your first line of support.

Editing the INI File

Users with the **Configuration** product right can open the onbase32.ini file from the OnBase Client by selecting **Admin | Utilities | Edit INI File**. When multiple onbase32.ini files exist, opening the onbase32.ini file from the OnBase Client ensures that a user is editing the correct onbase32.ini file instance. In most cases, this will be the onbase32.ini file residing in the default directory described above. If an alternate location for the onbase32.ini file is specified by the **-INIFILE** command line switch, the file in the specified location will be opened.

Backup/Recovery

Backup

Configuration

The Advanced Capture configuration is stored in the OnBase database. A proper backup of the database will contain all Advanced Capture configuration information, including all Advanced Capture forms.

Registry Settings

No registry settings apply to Advanced Capture.

External Files

Ensure that the Hyland OCR Engine software has been properly backed up.

You must also backup your onbase32.ini file.

Module-Related INI Options

Use the following chart to track the current settings of all related INI options for Advanced Capture.

Section	Setting	Current Value
System	OCRLocation	
	OCRRestartLimit	
AutomatedIndex	Diagnostic	
	AnalysisMin	
	AnalysisMax	

Additional Steps

There are no additional steps required to backup Advanced Capture.

Recovery

Configuration

All Advanced Capture configuration, including all Advanced Capture forms, is stored in the OnBase database. Restoring the database will restore any Advanced Capture configuration.

Registry Settings

No registry settings apply to Advanced Capture.

External Files

Ensure that your onbase32.ini file is properly restored to the correct location by the backup copy.

The onbase32.ini file can be restored from the backup if the recovery machine is intended to be used for exactly the same purpose as the original machine. If this machine will be used for other modules, thorough testing may be required to ensure that there are no unintended consequences from adding/modifying INI settings on the existing workstation.

Restore the Hyland OCR Engine by re-installing it on the processing workstation using the **OCRSetup** executable.

Module-Related INI Options

The onbase32.ini file can be restored from the backup if the recovery machine is intended to be used for exactly the same purpose as the original machine. If the recovery machine will be used for other modules, you may need to recover only the listed INI setting(s) from the table above.

Registration

Migrate the registration of Advanced Capture from the original workstation to the recovery workstation. The registration may need to be revoked from the original machine and then added to the recovery machine.

Additional Steps

No additional recovery steps are required for Advanced Capture.

Upgrade Considerations

The following upgrade considerations have been compiled by OnBase subject matter experts. These upgrade considerations are general and applicable to most OnBase solutions and network environments and should be considered each time an upgrade is performed.

Carefully consider the impact of making any changes, including those listed below, prior to implementing them in a production environment.

For additional general information about upgrading OnBase, refer to the Upgrade Guidelines reference manual, and visit the Hyland Community at: <https://www.hyland.com/community>.

Advanced Capture Upgrade Considerations

The following information should be considered or noted when upgrading Advanced Capture deployments. Read this information prior to upgrading your version of OnBase.

General Deployment Considerations — The following should be considered with regard to general deployments:

- As of OnBase 17, multiple users with the appropriate rights can configure Advanced Capture forms concurrently in the same system. While a user is making a configuration change, however, the form or setting (and any related, child settings) being modified remains locked until the user saves or cancels the changes. If any user making configuration changes is using a build prior to 16.1.0.148, no other user can make configuration changes to any forms in the system until the former user saves or cancels the changes.

Additional Licensing Considerations — The following should be considered with regard to licensing:

- The Interactive Data Capture license is included with the Advanced Capture license. If you are using Interactive Data Capture functionality in your Advanced Capture solution, you may have additional upgrade considerations to review. For more information, see the **Interactive Data Capture** module reference guide.
- To perform Advanced Capture processing on Japanese, Korean, or Chinese languages, your solution must be licensed for Asian Language OCR. This license may include additional upgrade considerations to review. For more information, see the **Full-Page OCR** module reference guide.

Troubleshooting

Common Issues

When I attempt to extract Keyword Values from an image using Point and Shoot Indexing, the text values that are returned are consistently incorrect and/or scrambled.

Most likely, the cause of this problem is the quality and/or resolution of your image. Try performing Point and Shoot Indexing on a “cleaner” image file.

It is considered a best practice to ensure that an image has a resolution of at least 300 dpi in order for acceptable results to be returned when performing Advanced Capture or Point and Shoot Indexing upon it.

I have been having trouble with the Advanced Capture engine incorrectly mistaking some alphabetic characters for numeric characters (i.e., reading INVO1CE instead of INVOICE). In order to prevent these indexing errors, I enabled the Allowed character types filtering option to prevent numeric characters from being recognized in fields that are only supposed to contain alphabetic characters. However, instead of routing the documents for manual indexing as I expected, the Advanced Capture engine now correctly reads the value.

When the **Allowed character types** filter options are used, the Advanced Capture engine is better able to “guess” what character SHOULD be read.

For example: you are using Advanced Capture to read the value **INVOICE** from a document, but the engine occasionally has trouble distinguishing the **I** from a **1**, and returns the value as **INVO1CE**. If the filter options are used to prevent Advanced Capture from reading numeric characters for that value, the engine will favor the **I** character over the **1** character, and will be more likely to return the value correctly as **INVOICE**.

Note: If the Advanced Capture engine, however, is confident that the value is, in fact, a **1** and not an **I**, then the **1** character will be replaced with a tilde (~) and the value will be marked as suspect.

But, if you are not using the filter options to prevent Advanced Capture from reading numeric characters, the engine will have no reason to favor the **I** character over the **1** character, other than its normal character recognition logic. Only if the engine is unsure of, and gives a high suspect level to, both the **I** and the **1** characters will the value itself be marked as suspect.

How can I configure the Date Format for Date Keyword Values?

The Date Format of any captured Date Keyword Values will match the workstation’s regional format.

For example, if your processing workstation is set to use **English (United States)** as its workstation locale, then all captured dates must be in the MM/DD/YYYY format.

If you are attempting to capture Date Keyword Values in a format that does not match the workstation’s regional format, then you must configure one of the following:

- a VB script for the Data Field Zone configured to parse the value into the correct format
- a regular expression match for the Data Field Zone configured to apply the correct format to the value with the **Applied data format** setting

Batches processed through Advanced Capture are not respecting the zoom and highlight features during the indexing process.

This issue can occur if both the Image Window Workstation Option is set to resize and no zoomed area has been saved on the image. To resolve this issue, do one of the following:

- In the OnBase Client, select **User | Workstation Options** to open the **Workstation Options** dialog box. On the Document tab, in the Image Window section, select **Default Window**.
- During indexing, right-click on the image in the Working window, and select **Image Zooming | Save Zoomed Area**.

For the changes to take effect, either re-launch the OnBase Client or return to the batch status queue list and re-enter indexing for the batch.

After being processed, documents exhibit poor image quality.

This issue can occur if documents are scanned or swept in a format other than JPEG compression and then undergo a type of processing through which the documents are recreated (e.g., Advanced Capture processing). If the **Use JPEG compressed TIFF as default color image format** option is enabled in the **Global Client Settings**, then the documents will be recreated using JPEG compression when they are processed.

While JPEG compression reduces file size, it also reduces image quality. If you repeatedly compress and decompress the same documents, they will progressively lose quality. To prevent the image quality from deteriorating, consider disabling the **Use JPEG compressed TIFF as default color image format** option in the **Global Client Settings**. For more information on this option, see the **System Administration** module reference guide.

Process Locks

Another administrator is configuring Advanced Capture forms for a scan queue, and I am trying to configure Advanced Capture templates for a different scan queue. We have sufficient Advanced Capture licenses available, but I keep getting a message that I am locked out.

I am attempting to configure Advanced Capture forms, but I keep getting a locked message even though no other users are currently logged on.

Only one OnBase administrator can access the Advanced Capture configuration interface at a time, regardless of the number of Advanced Capture licenses that are available. If another administrator is currently using the Advanced Capture configuration interface, then you will not be able to access it until that user has exited it.

You can check to see if another user is currently configuring Advanced Capture by viewing the Process Locks in the OnBase Client by clicking **Admin | Utilities | Process Lock Administration**.

If you are getting the locked message and you are sure that no other users could be using the Advanced Capture configuration interface, it is possible that a process lock was not removed when a user exited the Advanced Capture configuration interface.

For information on process locks, including information on deleting unnecessary process locks, see the **System Administration** module reference guide or help file.

Advanced Capture & AutoFill Keyword Sets

I am attempting to store AutoFill Keyword Set data from documents indexed via Advanced Capture process using the scan queue's Store Keyword Set Data option, but the AutoFill Keyword Set data is saved inconsistently.

Instances of AutoFill Keyword Sets can automatically be created using data applied to a document from a batch Advanced Capture process if the associated scan queue has the **Store Keyword Set Data** option enabled. AutoFill Keyword Set data is only stored for documents that are indexed correctly and marked as fully-indexed; AutoFill Keyword Set data will not be stored for any document that is unidentified or has any values marked as suspect.

AutoFill Keyword Set information cannot be stored for any document indexed via an ad hoc Advanced Capture process because the document does not have a connection to a scan queue and, therefore, the **Store Keyword Set Data** scan queue option.

I am using Advanced Capture to identify a Keyword Value that is used as the primary Keyword Value to trigger an AutoFill Keyword Set. However, when the primary Keyword Value is associated with more than one instance of an AutoFill Keyword Set, I'm not given the option to select the instance of the AutoFill Keyword Set(s) I want.

If you are using ad hoc Advanced Capture to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:

- If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document). This is the default behavior.
- If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

Regardless of the setting of the **Expand All Matching Instances** option you are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document.

If you are using batch Advanced Capture to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:

- If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, the batch is routed to either the **Awaiting Index** or **Index in Progress** batch status queue so that you can select the instance, or instances, of the AutoFill Keyword Set to assign to the document. This is the default behavior.
- If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

I am using Point and Shoot Indexing to extract a Keyword Value that is used as the primary Keyword Value to trigger an AutoFill Keyword Set. However, after I obtain the Keyword Value, I have to tab to the next Keyword field to trigger the Autofill Keyword Set expansion.

When using Point and Shoot Indexing, be aware that the focus remains on the field containing the just-extracted Keyword Value.

If you are using Point and Shoot Indexing to extract a Keyword Value that is the primary value for an AutoFill Keyword Set, the AutoFill Keyword Set is not triggered until you set the focus to another Keyword field.

Contacting Support

When contacting your solution provider, please provide the following information:

- The OnBase module where the issue was encountered.
- The OnBase version and build.
- The type and version of the connected database, such as Microsoft SQL Server 2014 or Oracle 12c, and any Service Pack that has been installed.
- The operating system that the workstation is running on, such as Windows 10 or Windows Server 2012 R2, and any Service Pack that has been installed. Check the supported operating systems for this module to ensure that the operating system is supported.
- The name and version of any application related to the issue.
- The version of Internet Explorer and any Service Pack that has been installed, if applicable.
- A complete description of the problem, including actions leading up to the issue.
- Screenshots of any error messages.

Supplied with the above information, your solution provider can better assist you in correcting the issue.



Advanced Capture

Administration Guide

Configuration Overview

The following steps are necessary for configuring Advanced Capture:

1. Grant User Groups the rights to configure and process Advanced Capture forms. See [Granting User Groups Rights for Configuration and Processing on page 26](#) for more information.
2. Configure Advanced Capture forms. See [Configuring Advanced Capture Forms on page 27](#) for more information.
3. Save the Advanced Capture configuration before processing changes or exiting the Advanced Capture configuration interface. See [Exiting the Advanced Capture Form Configuration Interface on page 289](#) for more information.

The following steps are optional for configuring Advanced Capture:

- Configure OCR formats. See [OCR and Document Type Configuration on page 22](#) for more information.
- Configure AutoFill Keyword Sets. See [Working With AutoFill Keyword Sets in Advanced Capture on page 22](#) for more information.
- Configure Advanced Capture forms for institutional databases. See [Configuring Advanced Capture Forms for Institutional Databases on page 24](#) for more information.
- Import/export Advanced Capture forms into/from your solution. See [Importing/Exporting Advanced Capture Forms on page 253](#) for more information.
- Perform a form order analysis to improve processing efficiency. See [Advanced Capture Form Order Analysis on page 263](#) for more information.
- Configure XML renditions of extracted Keyword Values. See [Configuring XML Renditions on page 274](#) for more information.
- Test the Advanced Capture forms. See [Testing Advanced Capture Forms on page 274](#) for more information.
- Configure Advanced Capture processes to use intelligent character recognition. See [Configuring Advanced Capture to Use Intelligent Character Recognition on page 278](#) for more information.
- Configure scan queues for Advanced Capture. When configuring batch Advanced Capture, this step is required. See [Configuring a Scan Queue for Advanced Capture on page 282](#) for more information.
- Configure Advanced Capture batches to process automatically. See [Configuring Automatic Advanced Capture Processes on page 286](#) for more information.

Note: Point and Shoot Indexing requires no additional configuration.

OCR and Document Type Configuration

Prior to configuring Advanced Capture, if desired, you can create a custom OCR format and assign it to the Document Types that will undergo processing. Otherwise, the **<Default>** OCR format will be used for Advanced Capture processing.

If you do create a custom OCR format, be aware that only the **Languages** options and certain **Preprocessor** options are respected by Advanced Capture.

For more information on creating an OCR format, see the **Full-Page OCR** module reference guide or help files.

Working With AutoFill Keyword Sets in Advanced Capture

Give additional consideration when using Advanced Capture to index documents that are associated with an AutoFill Keyword Set. For information on configuring AutoFill Keyword Sets, see the **AutoFill Keyword Sets** module reference guide or help files.

Ad Hoc Advanced Capture

If you are using ad hoc Advanced Capture to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:

- If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document). This is the default behavior.
- If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

Regardless of the setting of the **Expand All Matching Instances** option you are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document.

Batch Advanced Capture

If you are using batch Advanced Capture to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:

- If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, the batch is routed to either the **Awaiting Index** or **Index in Progress** batch status queue so that you can select the instance, or instances, of the AutoFill Keyword Set to assign to the document. This is the default behavior.
- If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

Point and Shoot Indexing

When using Point and Shoot Indexing, be aware that the focus remains on the field containing the just-extracted Keyword Value.

If you are using Point and Shoot Indexing to extract a Keyword Value that is the primary value for an AutoFill Keyword Set, the AutoFill Keyword Set is not triggered until you set the focus to another Keyword Type field.

Creating AutoFill Keyword Sets via Advanced Capture

Instances of AutoFill Keyword Sets can automatically be created using data applied to a document from a batch Advanced Capture process if the associated scan queue has the **Store Keyword Set Data** option enabled. AutoFill Keyword Set data is only stored for documents that are indexed correctly and marked as fully-indexed; AutoFill Keyword Set data is not stored for any document that is unidentified or has any values marked as suspect.

AutoFill Keyword Set information cannot be stored for any document indexed via an ad hoc Advanced Capture process because the document does not have a connection to a scan queue and, therefore, the **Store Keyword Set Data** scan queue option.

Configuring Point and Shoot Indexing

Point and Shoot Indexing can be performed on image documents displayed in the Document Imaging Working window. It does not require any special configuration.

A Note About Single and Double-Byte Characters

If your solution is configured to identify Latin characters and/or Arabic numerals (i.e., single-byte characters such as those found in English, French, German, etc.) and you wish to use Point and Shoot Indexing to identify Asian characters on an image, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the pointer to draw a box around the value to be extracted.

Likewise, if your solution is configured to identify Asian characters (i.e., double-byte characters such as those found in Chinese and Japanese) and you wish to use Point and Shoot Indexing to identify Latin characters and/or Arabic numerals, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the point to draw a box around the value to be extracted.

Configuring Advanced Capture

Note: Advanced Capture forms can be configured either from the OnBase Client or from within the Document Imaging interface. For information on importing batches of documents into a scan queue, consult the module reference guide or help files of the module you are using to import the documents into OnBase.

Configuring Advanced Capture Forms for Institutional Databases

If your OnBase solution is configured to use Institutional Databases, Advanced Capture forms can be created at the institution level or at a global level.

- Institution-level forms are available only to the institution that they were created for.
- Global forms are available to all institutions that are part of your OnBase solution.

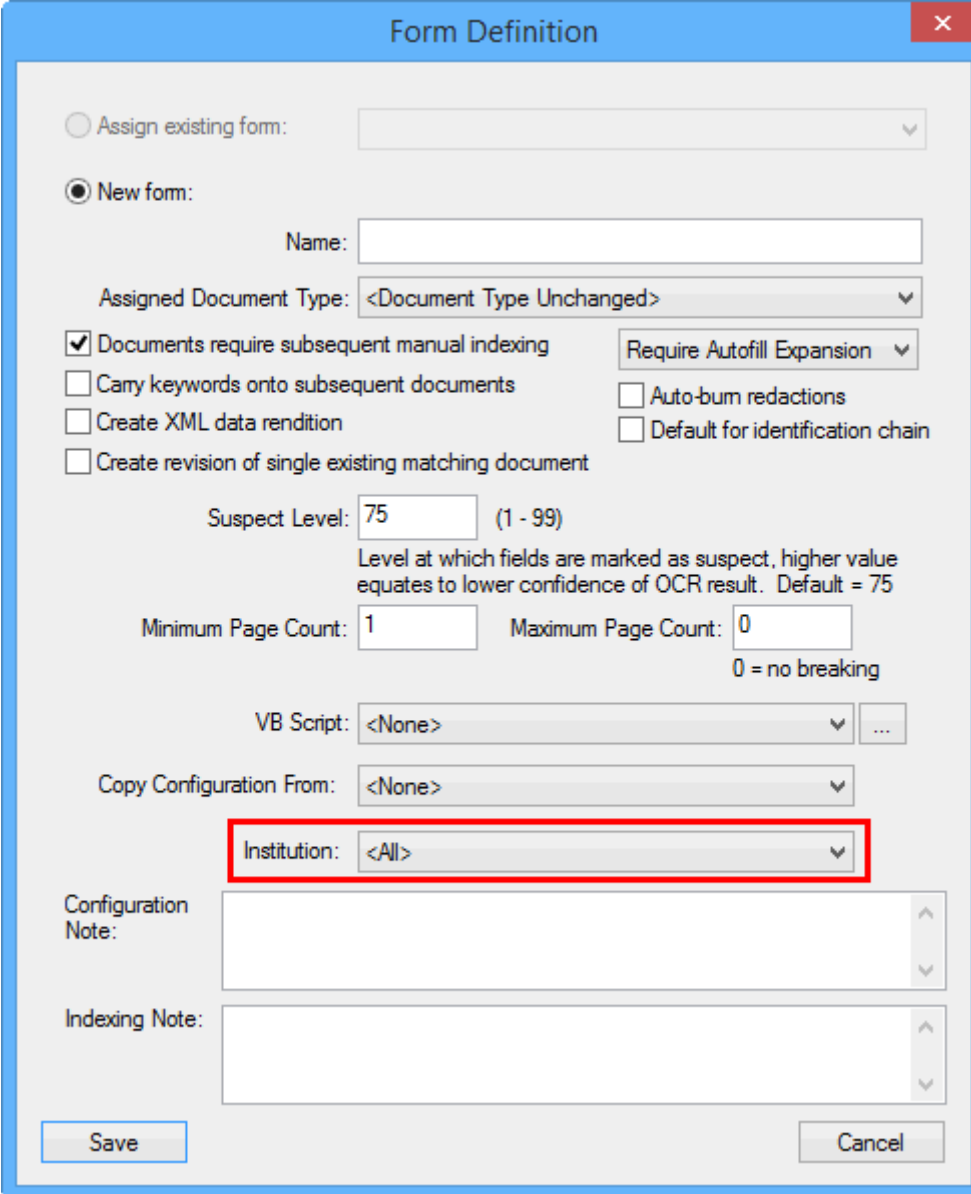
Non-super users can only create and configure Advanced Capture forms for their own institutions.

Super users can configure Advanced Capture forms for any institution, provided that the institutions using the forms have access to the Document Types and Keyword Types associated with the form. Super users may also configure global Advanced Capture forms.

Tip: It is considered a best practice for super users to create only global Advanced Capture forms. Non-super users should create all institution-level Advanced Capture forms.

Advanced Capture forms are assigned to institutions using the **Institution** drop-down list on the **Form Definition** dialog box.

- For non-super users, the current institution is pre-selected in the **Institution** drop-down list and the **Institution** drop-down list is disabled.
- For super users, the **Institution** drop-down list is enabled, allowing them to select any institution or the **<All>** option (used to create global Advanced Capture forms).

The image shows a 'Form Definition' dialog box with a blue title bar and a red close button. It contains several sections for configuring a form. The 'New form' section is active, showing fields for Name, Assigned Document Type, and various checkboxes for document processing. The 'Institution' dropdown is highlighted with a red rectangle and set to '<All>'. Below are text areas for Configuration Note and Indexing Note, and 'Save' and 'Cancel' buttons at the bottom.

Form Definition

☐ Assign existing form:

☒ New form:

Name:

Assigned Document Type: <Document Type Unchanged>

☒ Documents require subsequent manual indexing Require Autofill Expansion

☐ Carry keywords onto subsequent documents ☐ Auto-bum redactions

☐ Create XML data rendition ☐ Default for identification chain

☐ Create revision of single existing matching document

Suspect Level: 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: 1 Maximum Page Count: 0
0 = no breaking

VB Script: <None> ...

Copy Configuration From: <None>

Institution: <All>

Configuration Note:

Indexing Note:

Save Cancel

Using Advanced Capture Forms with Institutional Databases

Be aware that documents that are to undergo Advanced Capture must be pre-indexed with their **Institution #** Keyword Value.

Note: Documents cannot be swept into OnBase in pre-index mode. If you are sweeping in batches of documents that you want to undergo Advanced Capture, the **Institution #** Keyword Value must be added to the documents after they are swept but before they undergo Advanced Capture (e.g., via a batch bar code process using the Bar Code Recognition Server, via the **Add/Modify Keywords** dialog box before the batch undergoes Advanced Capture processing, etc.).

Testing Advanced Capture Forms

When testing Advanced Capture form configuration in the **Advanced Capture Configuration** window, the documents being processed are treated as if they belong to the current institution, even if they have not yet been assigned to that institution.

Tip: When a super user is performing a test Advanced Capture process (i.e., after a user has clicked **Process Current Document** or **Process Batch** in the **Tools** panel), ensure that the current institution is the institution you want to be testing the form(s) for.

During the test Advanced Capture process, information about the institution is displayed in the Results Verification window.

Performing Advanced Capture Processes

When a batch undergoes Advanced Capture, the OCR engine only attempts to match documents in the batch against Advanced Capture forms available to the documents' assigned Document Type and Advanced Capture forms available to the current institution.

During processing, information about the current institution is displayed in the Indexing Status window of the **Document Imaging** window.

Granting User Groups Rights for Configuration and Processing

In order to perform Advanced Capture processing on a document or batch of documents, or to route documents to the **Awaiting Ad Hoc Advanced Capture** batch status queue when a user's workstation is not registered, a user must belong to a User Group that has been given the Advanced/Intelligent Capture product right.

In order to create, modify, or delete Advanced Capture forms and associate Advanced Capture forms to existing documents, a user must belong to a User Group that has been given both the Advanced/Intelligent Capture product right and the Document Imaging administrative processing privilege.

Caution: Careful consideration should be taken when granting or restricting privileges or product rights to/from a User Group. For more information on User Group configuration, see the **System Administration** module reference guide or help file.

To assign the necessary rights for Advanced Capture configuration and processing to a User Group:

1. From the OnBase Configuration module, click **Users | User Groups /Rights**. The **User Groups & Rights** dialog box is displayed.
2. Select the User Group that is to be granted the rights from the User Group Name list and click **Product Rights**. The **Assigning Product Rights for <User Group Name> Group** dialog box is displayed.
3. Ensure that the following options are selected:
 - **Advanced/Intelligent Capture** in the Registered Processing Products section.
 - **Document Imaging** in the Administrative Processing Privileges section.
4. Click **Save**.

Configuring Advanced Capture Forms

Advanced Capture forms are the pre-configured templates that are matched to incoming documents to determine how the documents will be classified and indexed.

In order to configure Advanced Capture forms (e.g., create, modify, or delete Advanced Capture forms or manually assign Advanced Capture forms to documents in a batch):

- You must belong to a User Group with the Advanced/Intelligent Capture product right and the Document Imaging administrative processing privilege (see [Granting User Groups Rights for Configuration and Processing on page 26](#) for more information).
- Your OnBase solution and workstation must be licensed and registered for Advanced Capture or Ad Hoc Advanced Capture.

Note: As of OnBase 17, multiple users with the appropriate rights can configure Advanced Capture forms concurrently in the same system. While a user is making a configuration change, however, the form or setting (and any related, child settings) being modified remains locked until the user saves or cancels the changes. If any user making configuration changes is using a build prior to 16.1.0.148, no other user can make configuration changes to any forms in the system until the former user saves or cancels the changes.

Accessing the Advanced Capture Configuration Window

Advanced Capture forms are created and configured in the **Advanced Capture Configuration** window.

The **Advanced Capture Configuration** window can be accessed in two ways:

- **From the Client.** The Advanced Capture Configuration window can be opened from the Client's main menu (outside of the Document Imaging window). This allows you to create Advanced Capture forms from documents currently stored by your Advanced Capture solution.
- **From the Document Imaging Interface.** The Advanced Capture Configuration window can be opened from the Document Imaging window. This allows users to create forms from documents that have been scanned and currently reside in the batch status queue.

Note: For information on importing batches of documents into a scan queue, consult the module reference guide or help files of the module you are using to import the documents into OnBase.

Opening the Advanced Capture Configuration Window from the Client

This option exists to allow you to create Advanced Capture forms from documents that are currently available in the Client. Prior to creating Advanced Capture forms using this method, ensure that you have one or more documents that can be used as a basis for creating Advanced Capture forms.

1. From the Client, click **Admin | Advanced Capture Configuration**. The **Advanced Capture Configuration Documents** dialog box is displayed.

2. Using the Document Type Group and Document Type lists, select the Document Type that the document you want to use as a basis for the Advanced Capture form belongs to.

Once a Document Type is selected, the Keyword Type fields associated with the selected Document Type are displayed in the Keywords section.

3. Using the Date Range fields (i.e., **From** and **To**) and the Keyword Type fields, enter a date range and/or Keyword Values to retrieve the document you want to use to create the Advanced Capture form.

Tip: To limit the documents returned by the search to only those documents that have already been processed and were not successfully matched to any Advanced Capture form, select the **Limit document search results to only those document(s) that failed Advanced Capture recognition** check box.

4. To configure a maximum number of documents that can be returned by the search, select the **Maximum number of document(s) returned in result set** check box and enter the maximum number of documents allowed to be returned in the associated field.
5. As an alternative to the above search options, you can use the **Scan Queue** and **Batch** drop-down lists in the **Find Active Scan Batch** section to search the documents in a specific batch. This allows you to search documents in uncommitted batches belonging to custom capture processes.
 - a. From the **Scan Queue** drop-down list, select the scan queue associated with the batch you wish to search. Note that all scan queues you have rights to, including those configured for custom capture processes, are available for selection.

Note: Once you select a scan queue, all of the above options are deselected and disabled. To re-enable these options for selection, select **<None>** from the **Scan Queue** drop-down list.

- b. From the **Batch** drop-down list, select the batch you wish to search.

Note: The **Batch** drop-down list is only enabled when a scan queue is selected in the **Scan Queue** drop-down list.

6. Click **Find** to perform the search.
All documents returned by the search can be used to configure Advanced Capture forms in the **Advanced Capture Configuration** window.

Opening the Advanced Capture Configuration Window from the Document Imaging Window

This option exists to allow you to create Advanced Capture forms from documents in a batch. Prior to creating Advanced Capture forms using this method, ensure that you have scanned a batch of documents that can be used as a basis for creating Advanced Capture forms.

1. From the **Document Imaging** window, in the Queue List window, select the batch status queue (e.g., **Awaiting Advanced Capture**, **Awaiting Commit**, **Committed**, etc.) containing the batch of documents that are to be used as a basis for creating Advanced Capture forms.
2. All batches residing in the batch status queue are displayed in the Working window. Right-click on the desired batch in the Working window and select **Configure Advanced Capture**.
All documents in the batch are available for Advanced Capture form configuration in the **Advanced Capture Configuration** window.

The Advanced Capture Configuration Window

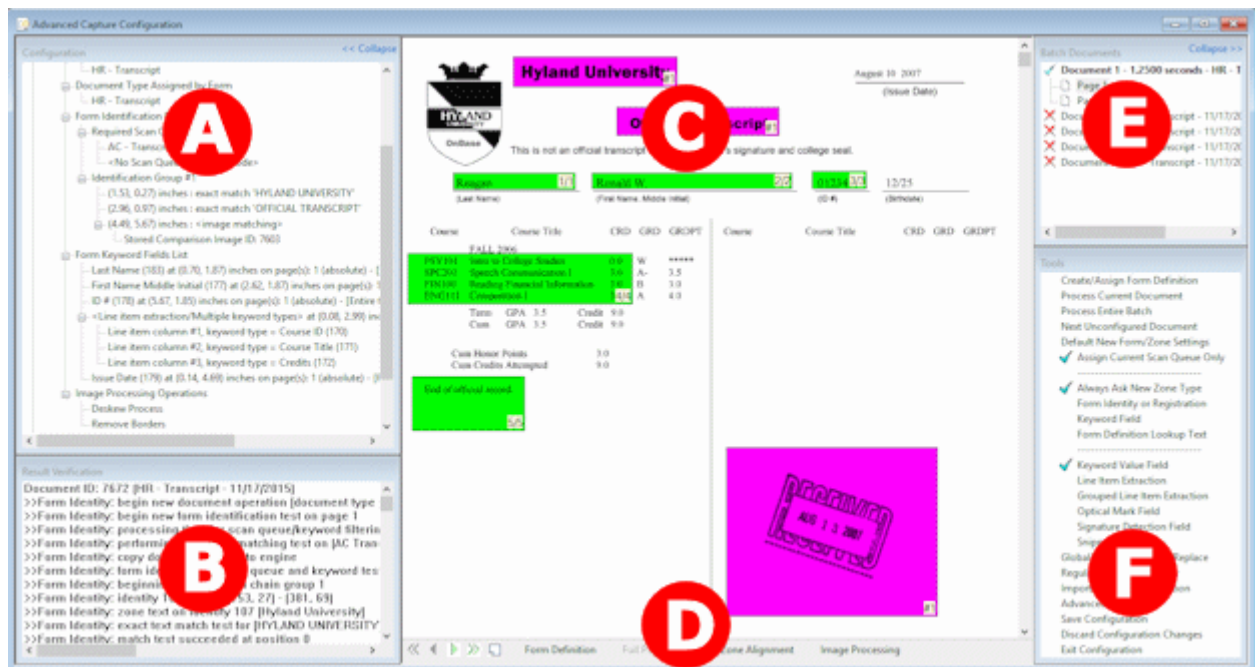
Before the **Advanced Capture Configuration** window is displayed, you are prompted with a message asking if you would like to pre-process the documents in the selected batch to attempt to match them to an existing Advanced Capture form.

- Click **Yes** to pre-process the documents.
 - If pre-processing results in documents in the batch being matched to Advanced Capture forms, then you will have the chance to modify the forms from within the **Advanced Capture Configuration** window.
 - If pre-processing fails to match Advanced Capture forms to documents in the batch, if pre-processing is cancelled before a document in the batch is processed, or if pre-processing is skipped altogether, then the documents in the batch are unclassified and you will be able to configure new forms for the documents' layouts.
- Click **No** to skip pre-processing.

Note: If pre-processing is skipped or cancelled before a document is pre-processed, that document is as unclassified, even if a matching Advanced Capture form exists, because the OCR engine was not able to attempt to match the document to the matching Advanced Capture form.

Tip: Documents can always be processed, or re-processed, later from within the **Advanced Capture Configuration** window by clicking **Process Entire Batch** or **Process Current Document** in the **Tools** pane.

The **Advanced Capture Configuration** window is displayed.



Allowing you to create, configure, and process documents using Advanced Capture forms, the **Advanced Capture Configuration** window contains the following unique sections. For more information on these sections, click the following section name links:

- **A: The Form Configuration Pane**
- **B: The Result Verification Pane**
- **C: The Document Viewer**
- **D: The Forms Toolbar**
- **E: The Batch Documents Pane**
- **F: The Tools Pane**

While the preceding diagram displays the default view of the **Advanced Capture Configuration** window, you can collapse the left and/or right sides of the window to expand the Document Viewer by clicking **Collapse** in the Form Configuration pane and the **Batch Documents** pane, respectively.

Once collapsed, you can re-expand the left and/or right side of the window by clicking the **Show Configuration** and **Show Tools** side tabs, respectively.

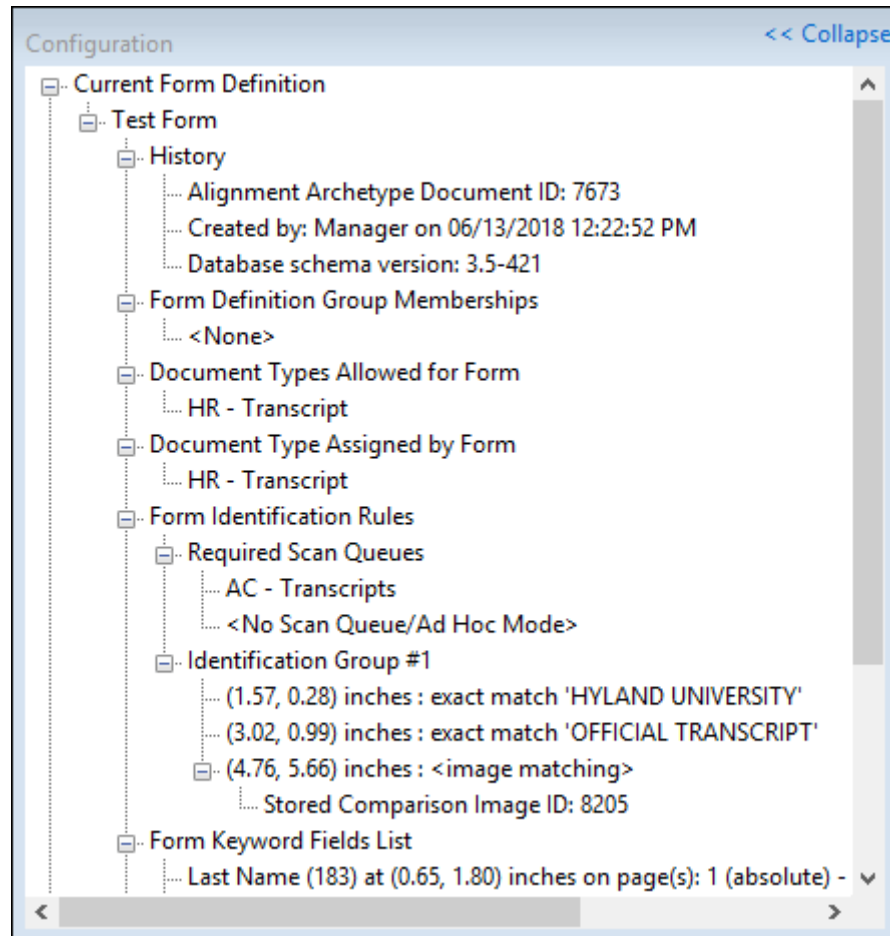
Note: When the right side of the window is collapsed, you can click the **Set Default New Zone Type** side tab to view and select the same **Default New Form/Zone Settings** you could otherwise select from the expanded **Tools** pane. For more information on these settings, see [The Tools Pane on page 41](#).

Once the **Advanced Capture Configuration** window is displayed, the first page (depending on how you accessed the **Advanced Capture Configuration** window, either the first page in the batch or the first page of the first document returned by your query) is displayed in the Document Viewer and the **Form Definition** dialog box is automatically displayed.

- To select a different page or document, either click the proper page/document from the **Batch Documents** pane or use the following shortcut keys:
 - **Page Up**: selects the previous page of a multi-page document
 - **Page Down**: selects the next page of a multi-page document
 - **Ctrl + Page Up**: selects the previous document in the batch
 - **Ctrl + Page Down**: selects the next document in the batch
- If the **Form Definition** dialog box is not automatically displayed, click **Create New Form Definition** in the **Tools** pane.

The Form Configuration Pane

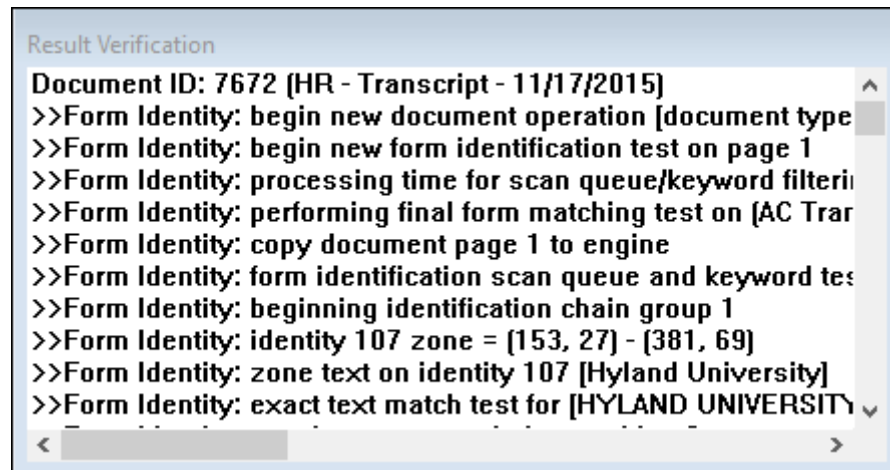
The Form Configuration pane displays all form configuration information in a tree view.



Tip: Double-clicking on a Document Type in the Form Configuration pane displays additional configuration options for that Document Type. For more information, see [Modifying Form Configuration Options for a Document Type](#) on page 99.

The Result Verification Pane

The **Result Verification** pane displays information on the fields processed and the values read by the OCR engine.



Tip: Double-clicking on the **Result Verification** pane exports the contents of the **Result Verification** pane to a new text document. This does not export the preview image displayed when configuring a Form Identification Zone or Data Field Zone. If the **Result Verification** pane is empty, a new text document is not opened.

The Document Viewer

The Document Viewer displays the currently selected document with the Advanced Capture form applied to it.

Hyland University #1

Official Transcript #1

August 10, 2007
(Issue Date)

This is not an official transcript without Registrar's signature and college seal.

Reagan 1/1 (Last Name) **Ronald W** 2/2 (First Name, Middle Initial) **01234** 3/3 (ID #) **12/25** (Birthdate)

Course	Course Title	CRD	GRD	GRDPT
FALL 2006				
PSY101	Intro to College Studies	0.0	W	*****
SPC201	Speech Communication I	3.0	A-	3.5
FIN100	Reading Financial Information	3.0	B	3.0
ENG111	Composition I	4/4	A	4.0
Term	GPA	3.5	Credit	9.0
Cum	GPA	3.5	Credit	9.0
Cum Honor Points		3.0		
Cum Credits Attempted		9.0		

End of official record. 5/5

RECEIVED
AUG 13 2007
#1

The different types of zones are highlighted in the following colors:

- Form Identification Zones are highlighted in magenta (as in the preceding image).
- Keyword Data Field Zones are highlighted either in green (as in the preceding image).

- Snippet Data Field Zones are highlighted in hatched red (as in the preceding image).
- Page Registration Zones are highlighted in yellow (as in the following image).

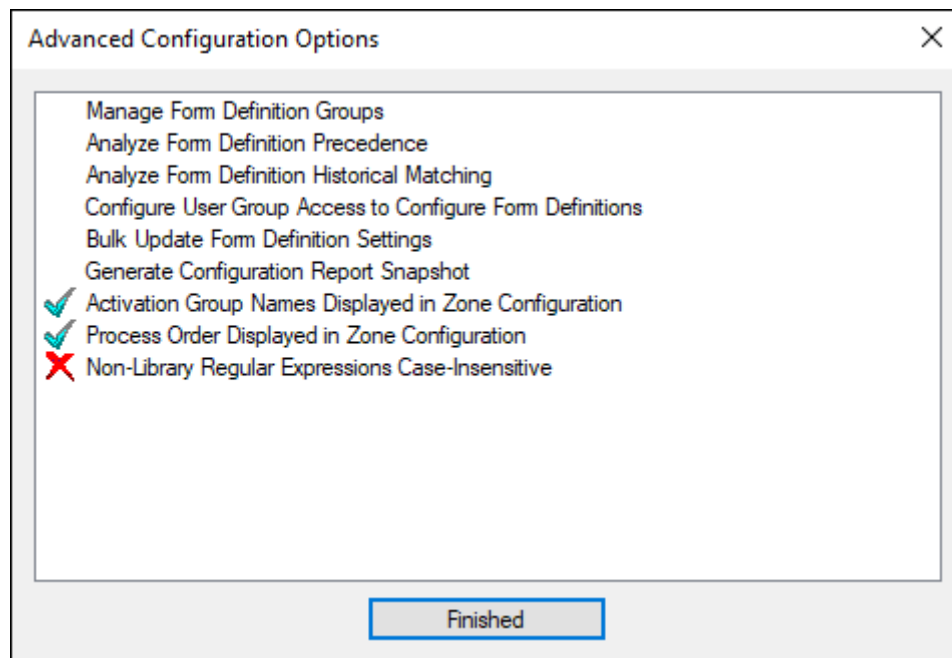
If any zones are configured as part of an activation group, by default, the name of the activation group is displayed in the upper-left corner of the zone (as in the following image).



Also by default, the ordering of identification groups and extraction zones is displayed in the lower-right corners of the Form Identification and Data Field Zones, respectively (as in the first image of this section).

To toggle either of these default settings on or off, do the following:

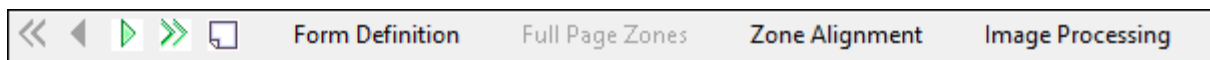
1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.



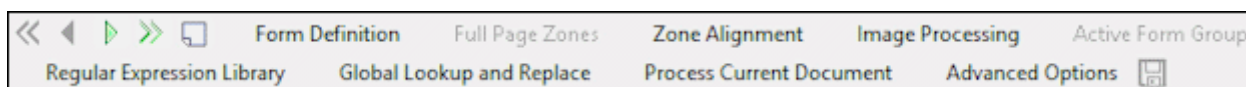
2. To turn the activation group names on (indicated by a teal checkmark) or off (indicated by a red X), click **Activation Group Names Displayed in Zone Configuration**.
3. To turn the zone process order numbers on or off, click **Process Order Displayed in Zone Configuration**.

The Forms Toolbar

The Forms toolbar contains various buttons that can be clicked to display information about the current page or assigned Advanced Capture form or to open tools for configuring the form. When the right side of the **Advanced Capture Configuration** window is expanded, the default buttons are displayed in the toolbar, as shown in the following image.








When the right side of the **Advanced Capture Configuration** window is collapsed, several additional buttons are also displayed in the toolbar, as shown in the following image.




Depending on your configuration and current selections in the window, some of the toolbar options might not be available.

Click the appropriate button or option in the toolbar to take the actions described in the following table.

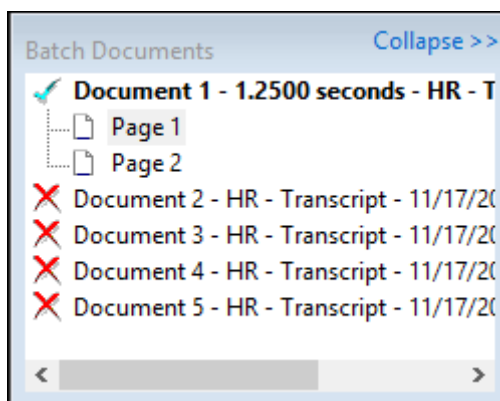
Button/Option	Description
Previous Document 	Selects the first page of the previous document in the batch.
Previous Page 	Selects the previous page of the current document.
Next Page 	Selects the next page of the current document.
Next Document 	Selects the first page of the next document in the batch.

Button/Option	Description
Page Information 	<p>Displays the following pieces of information about the current page:</p> <ul style="list-style-type: none"> • Document ID • Document name • Document Type • Batch ID • Current page • File format • Page dimensions • DPI • Color depth
Form Definition	Displays information about the current form and options to modify or delete the form or to create a Keyword Data Zone that covers the entire page.
Full Page Zones	Displays a menu of entire page zones that have been configured for the form. Selecting an entire page zone from the menu displays information about the entire page zone and options to modify, delete, or copy the zone or to show detailed extraction results.
Zone Alignment	<hr/> <p>Note: The Zone Alignment button is only available if the current page contains Form Identification or Page Registration Zones that apply registration points.</p> <hr/> <p>Instructs the Advanced Capture engine to calculate the offset/scaling for all registration zones on the current page and to shift the physical positions of all extraction zones to match these calculations.</p> <p>While this zone alignment feature is activated, a window is also displayed to show the offset/scaling information about the registration points as the extraction zones are being shifted. The extraction zones continue to be adjusted until you click elsewhere in the interface to deactivate the alignment feature and close the informational window.</p>
Image Processing	<hr/> <p>Note: The Image Processing button is only available if you are licensed for Bar Code Recognition Server or Barcode Recognition for Advanced Capture.</p> <hr/> <p>Displays a menu of options to show individual or all image processes configured for the form, hide all image processes, or edit the configured image processes. For more information, see Showing or Hiding Image Processing Operations on page 97.</p>

Button/Option	Description
Active Form Group	Displays a menu of any form definition groups that have been configured for the form. To simulate how the Advanced Capture engine will process the current document or batch with a form definition group activated, select that group from the menu before clicking the Process Current Document button.
Regular Expression Library	Opens the Regular Expression Library, allowing you to add, edit, or delete the regular expressions to be used in your system. For more information on configuring regular expressions, see The Regular Expression Library on page 247 .
Global Lookup and Replace	Opens the Keyword Lookup and Replace dialog box, allowing you to configure global Keyword Lookup/Replace dictionaries. For more information, see Configuring Keyword Lookup/Replace Dictionaries on page 73 .
Process Current Document	Runs a test Advanced Capture process on the current document.
Advanced Options	Opens the Advanced Configuration Options dialog box. For more information on the options available, see the advanced options information in The Tools Pane on page 41 .
Save Configuration 	Saves any changes made to the current configuration.

The Batch Documents Pane

The **Batch Documents** pane displays a list of all documents available for Advanced Capture form configuration (that is, depending on how the **Advanced Capture Configuration** window was accessed, either the documents returned by the retrieval operation or the documents in the current batch).

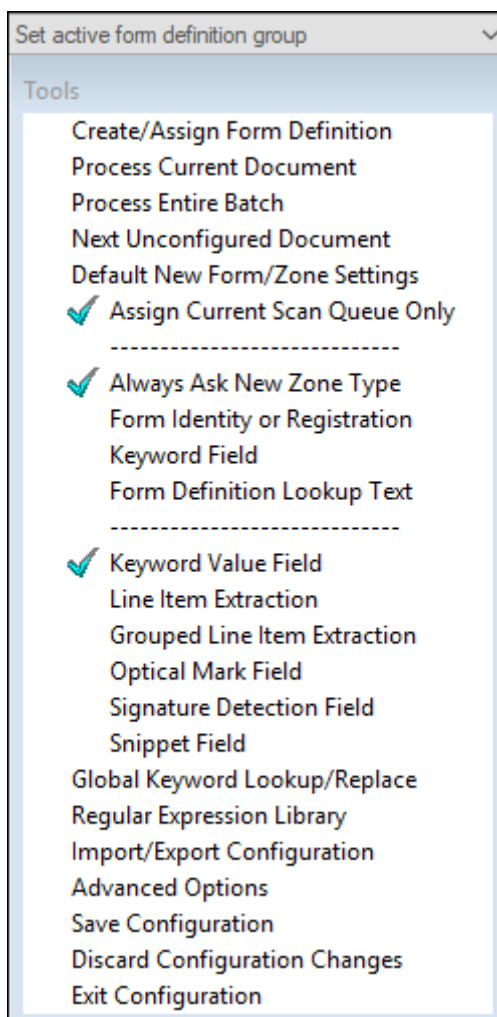


A green check mark is displayed next to documents that have been matched to Advanced Capture forms and a red X is displayed next to documents that are unclassified.

Selecting a document opens it in the Document Viewer.

The Tools Pane

The **Tools** pane contains links to all of the tools necessary to configure Advanced Capture forms.



The following tools are available:

- **Create a New Form Definition.** Click to open the **Form Definition** dialog box.
- **Process Current Document.** Click to run a test Advanced Capture process on the document currently displayed in the Document Viewer.

- **Process Entire Batch.** Click to run a test Advanced Capture process on all documents listed in the **Batch Documents** pane.

Note: If any form definition groups have been configured for the form, a drop-down list is displayed above the **Tools** pane. To simulate how the Advanced Capture engine will process the current document or batch with a form definition group activated, select that group from the drop-down list before clicking **Process Current Document** or **Process Entire Batch**.

- **Next Unconfigured Document.** Click to display the next document that has not yet been matched to a form.
- **Default New Template/Zone Settings.** Click one of the **Default Zone Type** sub-options to set the default behavior when a new Advanced Capture form zone is created.
 - **Assign Current Scan Queue Only.** This option is selected by default. When selected, this option makes any Advanced Capture forms you create available for only the scan queue that the currently displayed document is associated with and for ad hoc Advanced Capture processes.
 - When not selected, any Advanced Capture forms you create are automatically made available to all scan queues.
 - **Always Ask New Zone Type.** Select this option to always prompt the user to select the type of zone being created.
 - **Form Identity or Registration.** Select this option to automatically set each newly created zone as a Form Identity or Registration Zone.
 - **Keyword Field.** Select this option to automatically set each newly created zone as a Data Field Zone. When this option is selected, you can also specify the type of Data Field Zone being configured by selecting one of the following options: **Keyword Value Field, Line Item Extraction, Group Line Item Extraction, Optical Mark Field, or Signature Detection Field.**
 - **Form Definition Lookup Text.** Select this option to automatically set each newly created zone as a Form Definition Lookup Text Zone.
- **Global Keyword Lookup/Replace.** Click to configure a global Keyword Lookup/Replace dictionary to replace Keyword Values identified by the Advanced Capture process with different values.
 For example, if you require a **Term** Keyword Value to contain a Season and a Four Digit Year (for example, **Fall 2010**) and some incoming documents display a Season and a Two Digit Year (for example, **Fall 10**), you could configure a lookup/replace action to automatically replace all Keywords Values identified as **Fall 10** with **Fall 2010**.
- **Regular Expression Library.** Click to edit your system's default regular expressions or to configure custom regular expressions to be used in your Advanced Capture solution.
- **Import/Export Configuration.** Click to import or export Advanced Capture forms from/to another Advanced Capture solution.

- **Advanced Options.** Click to open a dialog box containing the following options:
 - **Manage Form Definition Groups:** Allows you to group forms together into named groups, which you can later assign to specific instances of Advanced Capture processing when configuring a custom capture process in the Capture Process Designer
 - **Analyze Form Definition Precedence:** Allows you to perform an analysis of how frequently your forms are being matched to documents
 - **Analyze For Definition Historical Matching:** Allows you to perform an analysis of historical matches made between forms and documents.
 - **Configure User Group Access to Configure:** Allows you to grant Advanced Capture configuration access to users who do not have the **Document Imaging** administrative processing privilege
 - **Bulk Update Form Definition Settings:** Allows you to apply modifications to multiple form definitions simultaneously.
 - **Generate Configuration Report Snapshot:** Allows you to generate a configuration report to show all form definitions and related information in an easy-to-read format.
 - **Activation Group Names Displayed in Zone Configuration:** Allows you to display the activation group a zone belongs to in the upper-left corner of the zone. By default, this option is enabled (indicated by a teal checkmark).
 - **Process Order Displayed in Zone Configuration:** Allows you to display the ordering of identification groups and extractions zones in the lower-right corners of the Form Identification and Data Field Zones, respectively. By default, this option is enabled (indicated by a teal checkmark).
 - **Non-Library Regular Expressions Case-Insensitive:** Allows you to set case sensitivity as a consideration when matching extracted values to non-library regular expressions (for example, regular expressions defined when configuring Data Field Zone Keyword Field Options). By default, this option is disabled (indicated by a red X), and non-library regular expressions will be case sensitive.

Note: Case sensitivity considerations for regular expressions in the Regular Expression Library are unaffected by the **Non-Library Regular Expressions Case-Insensitive** option.

- **Save Configuration.** Click to save the current configuration, including any configured Advanced Capture forms. A confirmation message is displayed with the following options:
 - **Yes:** Save your configuration changes.
 - **Yes, don't remind me again this session:** Save your configuration changes, and prevent the confirmation message from being displayed again during this session.
 - **No:** Cancel the save.

If you save your configuration changes, a second confirmation message is displayed with the following options:

- **Generate Report:** Generate a configuration report with all form definitions and related information.
- **Continue:** Continue working without generating a configuration report.
- **Continue, don't ask me again this session:** Continue working without generating a configuration report, and prevent the confirmation message from being displayed again during this session.
- **Discard Configuration Changes.** Click to discard any unsaved Advanced Capture form changes and revert to the last saved form configuration.
- **Exit Configuration.** Click to close the **Advanced Capture Configuration** window and return to the **Document Imaging** window.

Defining Advanced Capture Forms

Advanced Capture forms are defined by setting the options in the **Form Definition** dialog box, which can be accessed in different ways when doing any of the following:

- Creating a new form
- Assigning an existing form to documents
- Modifying an existing form

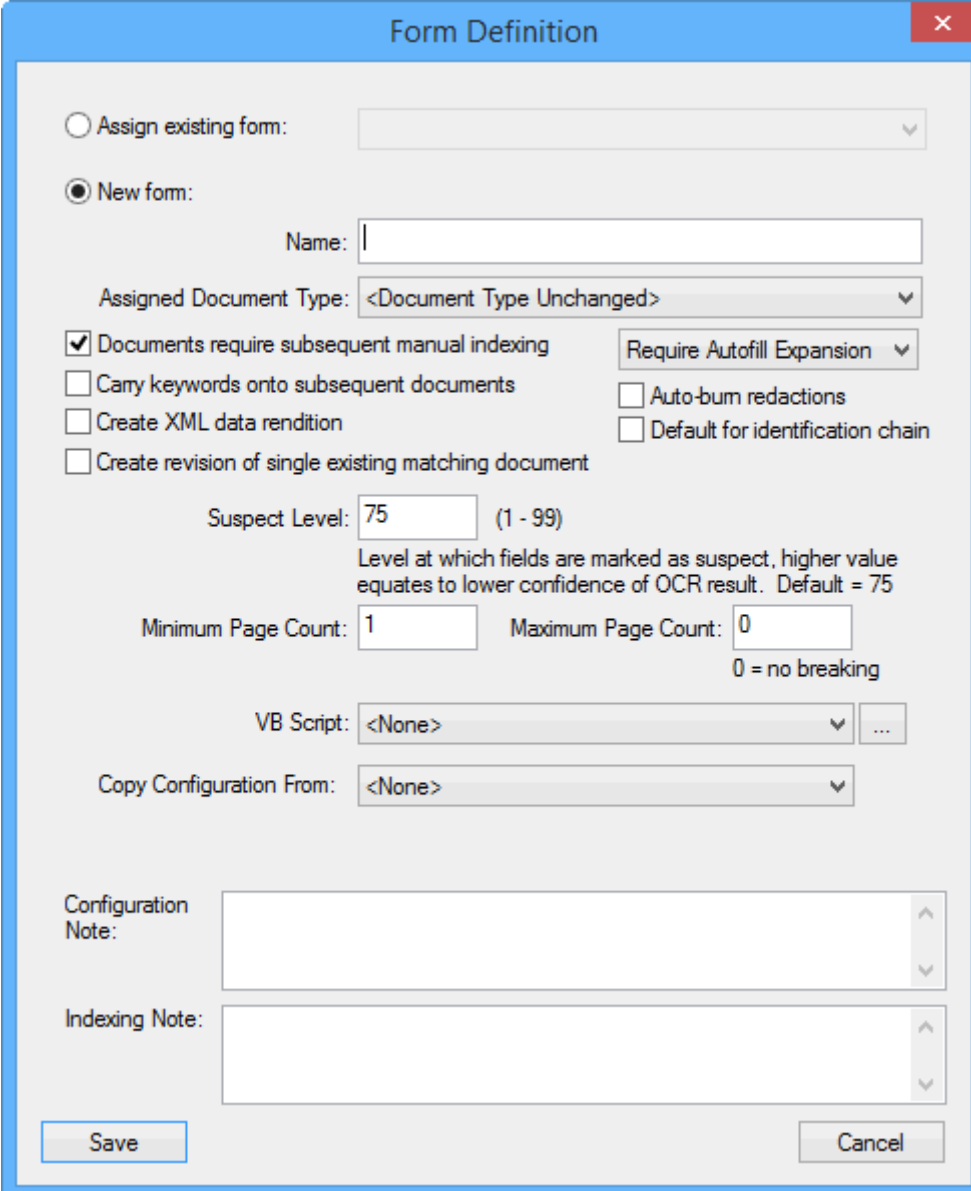
Depending on how the **Form Definition** dialog box is accessed, different options may be available. For more information on these options and their availability, see [Form Definition Options on page 53](#).

Creating a New Advanced Capture Form

To create a new Advanced Capture form:

1. From the **Advanced Capture Configuration** window, ensure that the document the Advanced Capture form is being matched to is displayed in the Document Viewer. If needed, select the correct document from the **Batch Documents** panel.
2. Open the **Form Definition** dialog box.
 - When accessing the **Advanced Capture Configuration** window, if pre-processing fails, or is skipped or canceled, the **Form Definition** dialog box is automatically displayed.
 - If the **Form Definition** dialog box is not automatically displayed, click **Create/Assign Form Definition** in the **Tools** panel.

Note: If you click **Create/Assign Form Definition** on an interior page of the document, a confirmation message is displayed. To treat this interior page as the beginning of a new sub-document (i.e., a sub-document assigned a different Advanced Capture form than the first page of the main document), click **Start new document here**. Conversely, to apply to the existing first page of the document the Advanced Capture form you are about to assign to the interior page, click **Apply form definition to existing document**.

The image shows a 'Form Definition' dialog box with a blue title bar and a red close button. It contains two radio buttons: 'Assign existing form:' (unselected) and 'New form:' (selected). Below 'New form:' is a 'Name:' text field. Underneath is an 'Assigned Document Type:' dropdown menu showing '<Document Type Unchanged>'. There are four checkboxes: 'Documents require subsequent manual indexing' (checked), 'Carry keywords onto subsequent documents' (unchecked), 'Create XML data rendition' (unchecked), and 'Create revision of single existing matching document' (unchecked). To the right of these are two more checkboxes: 'Require Autofill Expansion' (checked) and 'Auto-bum redactions' (unchecked), with 'Default for identification chain' (unchecked) below them. Below the checkboxes is a 'Suspect Level:' field with the value '75' and a range '(1 - 99)'. A note below this field states: 'Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75'. Below this are 'Minimum Page Count:' (1) and 'Maximum Page Count:' (0) fields, with a note '0 = no breaking' below the maximum field. There is a 'VB Script:' dropdown menu showing '<None>' and a button with three dots. Below that is a 'Copy Configuration From:' dropdown menu also showing '<None>'. At the bottom are two text areas: 'Configuration Note:' and 'Indexing Note:'. At the very bottom are 'Save' and 'Cancel' buttons.

Form Definition

☐ Assign existing form:

☒ New form:

Name:

Assigned Document Type: <Document Type Unchanged>

☒ Documents require subsequent manual indexing ☒ Require Autofill Expansion

☐ Carry keywords onto subsequent documents ☐ Auto-bum redactions

☐ Create XML data rendition ☐ Default for identification chain

☐ Create revision of single existing matching document

Suspect Level: 75 (1 - 99)

Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: 1 Maximum Page Count: 0

0 = no breaking

VB Script: <None>

Copy Configuration From: <None>

Configuration Note:

Indexing Note:

Save Cancel

3. From the **Form Definition** dialog box, click the **New Form** radio button to enable the new Advanced Capture form options on the dialog box.
4. Set the configuration options for the Advanced Capture form being created. See [Form Definition Options on page 53](#) for more information.

5. Click **Save** to save the form and close the **Form Definition** dialog box.

Tip: At any time, click **Save Configuration** in the **Tools** panel to save the Advanced Capture form. It is considered a best practice to save your Advanced Capture form after the form is created and after any configuration change is made (e.g., any Form Identification, Registration, or Data Field Zones are added, deleted, modified, moved, etc.).

To discard all changes since the last time the form was saved, click **Discard Configuration Changes** in the **Tools** panel.

Assigning Existing Advanced Capture Forms to Documents in the Batch

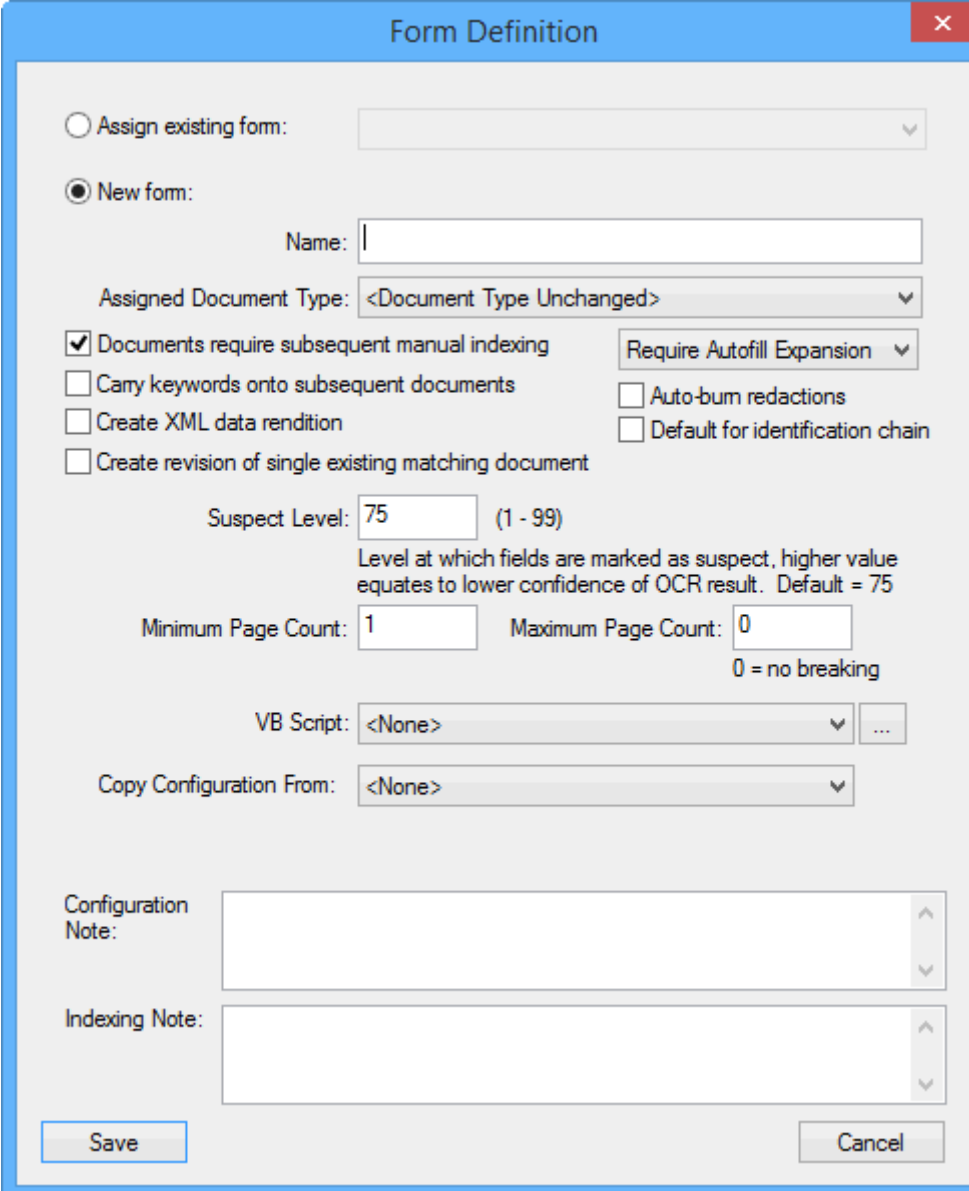
When a batch is processed in the **Awaiting Advanced Capture** batch status queue, the OCR engine automatically evaluates the documents in the batch against all available Advanced Capture forms.

Sometimes, however, it is necessary to manually assign a form to a document in the **Advanced Capture Configuration** window (such as when the text or image used as a form identifier is of poor quality for a scanned document, or if a document's layout has been changed and its Advanced Capture form needs to be updated to match it).

To manually assign a form to a document:

1. From the **Advanced Capture Configuration** window, ensure that the document the Advanced Capture form is being matched to is displayed in the Document Viewer. If needed, select the correct document from the **Batch Documents** panel.
2. Open the **Form Definition** dialog box.
 - When accessing the **Advanced Capture Configuration** window, if pre-processing fails, or is skipped or canceled, the **Form Definition** dialog box is automatically displayed.
 - If the **Form Definition** dialog box is not automatically displayed, click **Create/Assign Form Definition** in the **Tools** panel.

Note: If you click **Create/Assign Form Definition** on an interior page of the document, a confirmation message is displayed. To treat this interior page as the beginning of a new sub-document (i.e., a sub-document assigned a different Advanced Capture form than the first page of the main document), click **Start new document here**. Conversely, to apply to the existing first page of the document the Advanced Capture form you are about to assign to the interior page, click **Apply form definition to existing document**.



The image shows a 'Form Definition' dialog box with a blue title bar and a red close button. It contains several sections for configuring a form. The 'New form' section is active, showing a 'Name' field, an 'Assigned Document Type' dropdown, and several checkboxes for document processing options. There are also input fields for 'Suspect Level', 'Minimum Page Count', and 'Maximum Page Count'. At the bottom, there are 'Configuration Note' and 'Indexing Note' text areas, and 'Save' and 'Cancel' buttons.

Form Definition

☐ Assign existing form:

☒ New form:

Name:

Assigned Document Type: <Document Type Unchanged>

☒ Documents require subsequent manual indexing Require Autofill Expansion

☐ Carry keywords onto subsequent documents ☐ Auto-bum redactions

☐ Create XML data rendition ☐ Default for identification chain

☐ Create revision of single existing matching document

Suspect Level: (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: Maximum Page Count:
0 = no breaking

VB Script: <None> ...

Copy Configuration From: <None>

Configuration Note:

Indexing Note:

Save Cancel

3. Select the **Assign existing form** radio button. Once selected, use the drop-down to select the Advanced Capture form to be associated with the document.

Note: The **Assign existing form** radio button and the associated drop-down menu are only enabled when at least one Advanced Capture form is configured for your Advanced Capture solution.

4. Click **Save**. The Advanced Capture form is applied to the document in the Document Viewer. Form Identification, Page Registration, and Data Field Zones are highlighted on the displayed document.

Tip: At any time, click **Save Configuration** in the **Tools** panel to save the Advanced Capture form. It is considered a best practice to save your Advanced Capture form after the form is created and after any configuration change is made (e.g., any Form Identification, Registration, or Data Field Zones are added, deleted, modified, moved, etc.).




To discard all changes since the last time the form was saved, click **Discard Configuration Changes** in the **Tools** panel.

Now that the form is associated with the currently displayed document, it can be used as-is, modified, or deleted.

Modifying or Deleting a Form

An existing Advanced Capture form can be modified or deleted by right-clicking on the document it has been applied to in the Document Viewer.

Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**
 **Delete Form**
 **Create Entire Page Zone**

Information about the current Advanced Capture form is displayed above the following options:

- **Modify Form.** Click to display the **Form Definition** dialog box to modify the Advanced Capture form's configuration settings. For more information, see [Modifying a Form on page 52](#).

From **Form Definition** dialog box, you can also specify the Document Types that the Advanced Capture form is available to be applied to. For more information, see [Specifying Advanced Capture Form Availability by Document Type on page 62](#).

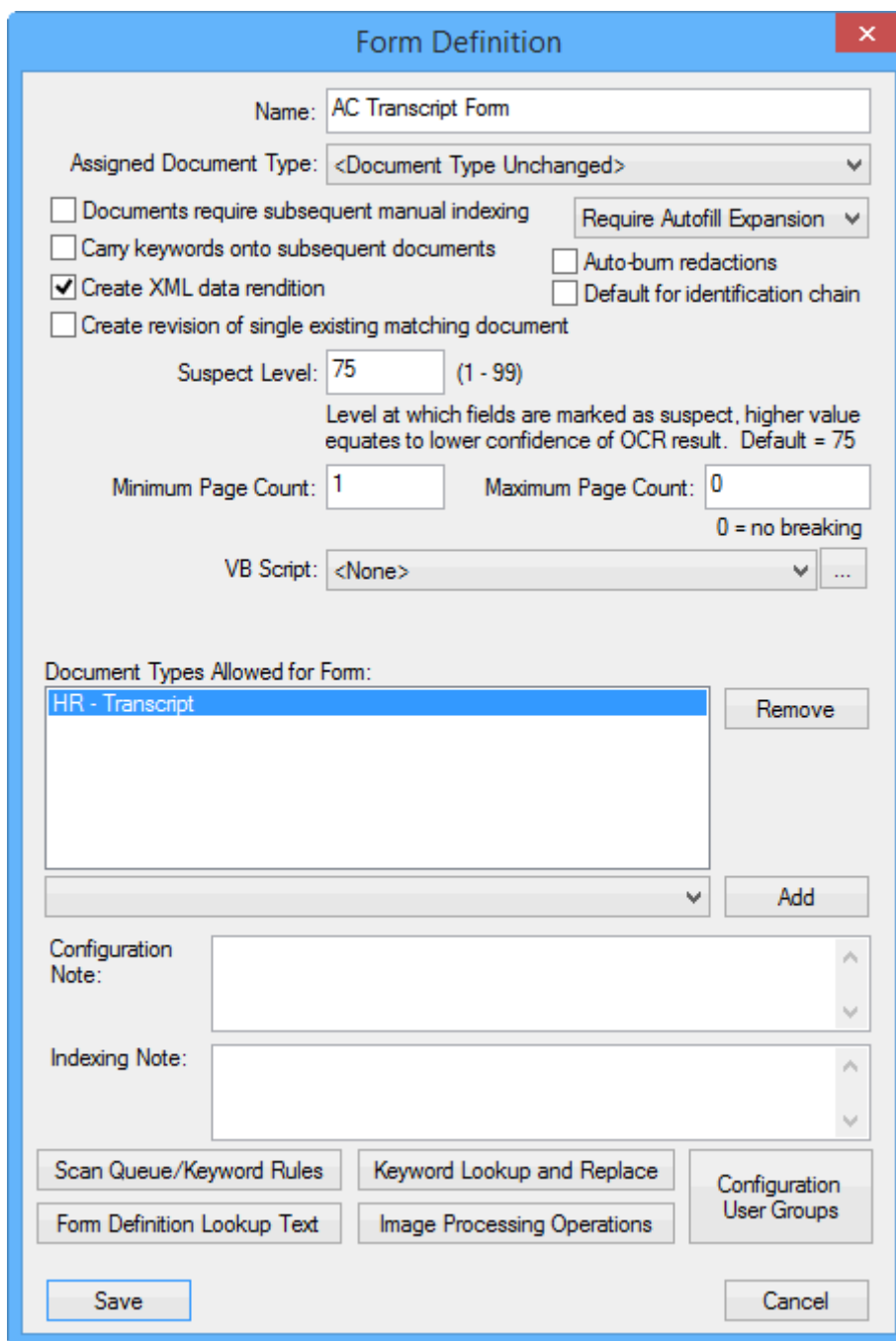
- **Delete Form.** Click to delete the selected Advanced Capture form. All associated Form Identification, Registration, and Data Field Zones are deleted.

Tip: You can also delete other Advanced Capture forms that are not currently selected by right-clicking on the form's name beneath the **All Form Definitions** node in the Form Configuration panel and selecting **Delete**.

- **Create Entire Page Keyword Zone.** Click to create a Keyword Data Zone that covers the entire page.

Modifying a Form

The **Form Definition** dialog box is displayed:



The **Form Definition** dialog box is shown with the following settings:

- Name:** AC Transcript Form
- Assigned Document Type:** <Document Type Unchanged>
- ☐ Documents require subsequent manual indexing
- ☐ Carry keywords onto subsequent documents
- ☒ Create XML data rendition
- ☐ Create revision of single existing matching document
- Require Autofill Expansion:** (dropdown menu)
- ☐ Auto-burn redactions
- ☐ Default for identification chain
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None> (dropdown menu with ellipsis button)
- Document Types Allowed for Form:**
 - HR - Transcript (selected)
 - Remove button
 - Add button
- Configuration Note:** (text area with up/down arrows)
- Indexing Note:** (text area with up/down arrows)
- Buttons:**
 - Scan Queue/Keyword Rules
 - Keyword Lookup and Replace
 - Configuration User Groups
 - Form Definition Lookup Text
 - Image Processing Operations
 - Save
 - Cancel

Use the configuration options (as needed) to modify the current Advanced Capture form. For more information, see [Form Definition Options](#) on page 53 and [Additional Information on Form Definition Options](#) on page 60.

Once the form has been modified, click **Save** to save your changes and close the **Form Definition** dialog box.

Form Definition Options

Depending on how the **Form Definition** dialog box is accessed, different options may be available.

Common Options

The following options are available regardless of how the **Form Definition** dialog box is accessed:

Option	Description
Name	Enter the name of the Advanced Capture form that you are creating.
Assigned Document Type	<p>Select the Document Type that documents will be assigned to once they are associated with this Advanced Capture form.</p> <p>To leave the Document Type of documents undergoing Advanced Capture processing unchanged (e.g., to ensure that the document remains unindexed or that the document still belongs to the Document Type it was assigned to at the time of import), select <Document Type Unchanged>.</p> <hr/> <p>Tip: If the Document Type is set at the scan queue level and does not need to be modified, it is considered a best practice to use the default <Document Type Unchanged> for this setting.</p> <hr/>
Documents require subsequent manual indexing	<hr/> <p>Note: When you are creating a new Advanced Capture form, this option is selected by default.</p> <hr/> <p>Select this check box to automatically route documents matched to this Advanced Capture form for index verification and review after processing.</p>

Option	Description
Carry keywords onto subsequent documents	<p>Select this check box to apply Keyword Values identified on the source document (i.e., the document currently selected in the Document Viewer) to subsequent documents in the same batch or ad hoc processing group.</p> <p>When using this option, note the following:</p> <ul style="list-style-type: none"> • If a subsequent document is assigned a different Document Type, only values from the common Keyword Types are applied. • If a subsequent document has not been assigned a Document Type, or has no Keyword Types in common with the source document, none of the carried Keyword Values is applied. • The carried Keyword Values are applied to subsequent documents until either a new source document is encountered or the end of the batch or ad hoc processing group is reached. If a new source document is encountered, the process begins again with the new source document's set of Keyword Values being carried forward to subsequent documents. If the end of the batch/processing group is reached, the next batch/processing group starts with an empty set of carry-forward Keywords until the first source document of the batch/processing group is encountered, in which case the process starts over. • The carried Keyword Values are applied to subsequent documents that have been assigned a Document Type, regardless of whether the subsequent documents have been matched to Advanced Capture forms.
Create XML data rendition	<hr/> <p>Note: In order for this setting to be respected, the assigned Document Type must be configured to allow renditions.</p> <hr/> <p>Select this check box to create XML renditions of the documents matched to this Advanced Capture form.</p> <p>The XML renditions contain indexing values (e.g., the Document Date and Keyword Values) identified during the Advanced Capture process. The structure of the XML file (i.e., the element that the Keyword Value is contained in) is defined when configuring the Data Field Zone for that Keyword Value.</p>

Option	Description
Create revision of single existing matching document	<hr/> <p>Note: This option is the form-level equivalent of the batch-level Check for Previous Revisions option, and it can be used to create revisions of matching documents when performing ad hoc Advanced Capture. The Create revision of single existing matching document option overrides the Check for Previous Revisions option if the former is selected at the form level and the latter is deselected at the batch level. For more information on the Check for Previous Revisions option, see the Document Imaging module reference guide.</p> <hr/> <p>Select this check box to tell OnBase to check for existing versions of the document being indexed. OnBase recognizes documents with the same file format, Document Type, and Keyword Values as potential revisions or renditions.</p> <p>If an existing document is found, the revision settings associated with the Document Type control the revision prompt behavior. For example, if a document with the same Document Type and Keyword Values is found, and the Document Type's Revision Settings are Always assume new revision and Force comment, upon indexing, the document is assumed to be a revision and the user must enter a revision comment.</p> <hr/> <p>Note: A revision of a document cannot be created while the document is locked.</p> <hr/> <p>Note: When a document indexed via Advanced Capture is determined to be a revision of an existing document and a revision comment is required for its Document Type, a revision comment is automatically generated indicating that the revision was identified via Advanced Capture.</p> <hr/> <p>If more than one existing document of the same Document Type with matching Keyword Values is found, the Document Revisions dialog box is displayed and the user is prompted to select the existing document to use.</p> <p>For more information on the Document Revisions dialog box, see the EDM Services module reference guide or help file.</p>

Option	Description
Autofill Expansion	<p>Select one of the following options to determine how the expansion of AutoFill Keyword Sets will be handled:</p> <ul style="list-style-type: none"> • Require Autofill Expansion: Select this option to ensure that batches are sent to either the Awaiting Index or Index in Progress batch status queue for manual index verification whenever an AutoFill Keyword Set on a document in a batch fails to expand during Advanced Capture processing. This is the default setting. • Allow Autofill Expansion: Select this option to allow AutoFill Keyword Sets to be expanded but also to prevent batches from automatically being sent for manual index verification whenever an AutoFill Keyword Set fails to expand during Advanced Capture processing. • Prevent Autofill Expansion: Select this option to prevent AutoFill Keyword Sets from being expanded on any primary Keyword Type during Advanced Capture processing. <hr/> <p>Note: Batches containing documents with missing required Keywords or suspect Keywords are still sent for manual index verification without respect to this setting.</p> <hr/>
Auto-burn redactions	<hr/> <p>Note: This option is only available when the system performing Advanced Capture is licensed for Automated Redaction.</p> <hr/> <p>Select this check box to immediately burn redactions on the areas of batch documents that have been configured for redactions during Advanced Capture processing. This automatic burning of the redactions prevents the need to send the documents for Automated Redaction review and approval after Advanced Capture processing is complete.</p> <hr/> <p>Note: If any values captured in the areas of the batch documents configured for redactions are marked as suspect during Advanced Capture processing, the redactions are not burned and the documents are sent to the Automated Redaction Review window for review and approval.</p> <hr/> <p>Deselect this check box to simply mark these configured areas for redaction. Once Advanced Capture processing is complete, the marked documents are sent to the Automated Redaction Review window for review and approval.</p> <p>For more information on Automated Redaction review and approval, see the Automated Redaction documentation.</p>

Option	Description
Default for identification chain	<p>Select this check box to set this Advanced Capture form as the default form. If a document's page is not matched to an existing Advanced Capture form during pre-processing, the document is automatically assigned to the highest-level default form.</p> <p>Deselect this check box to leave a document as unclassified when the document's page is not matched to an existing Advanced Capture form during pre-processing.</p> <hr/> <p>Note: Because a given Advanced Capture form might be included in different Document Type identification chains and in different positions of preference relative to the other forms included in each chain, you can select this option for multiple Advanced Capture forms. If you mark multiple forms as default forms, the form with the highest level of precedence in the current Document Type's identification chain is assigned to an unmatched document.</p> <hr/>

Option	Description
Suspect Level	<p>Enter the Suspect Level threshold, 1-99, in this field.</p> <p>The Suspect Level is the level of confidence placed in the data values captured using this form.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold causes the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Suspect Levels can also be set at the zone level for individual zones if more or less stringent standards are needed for their values.</p> <hr/> <p>Note: Setting the Suspect Level in the Form Definition dialog box sets the default Suspect Level for all Form Identification, Data Field, and Page Registration Zones configured for the form.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p> <hr/>
Minimum Page Count	<p>The Minimum Page Count field contains the minimum number of pages that can compose a document identified by this Advanced Capture form.</p> <p>This setting must be set to at least a value of 1.</p>
Maximum Page Count	<p>The Maximum Page Count field contains the maximum number of pages that can compose a document identified by this Advanced Capture form.</p> <p>Enter 0 if the Advanced Capture process should not attempt to break documents (i.e., Advanced Capture respects the pre-existing document breaks).</p>

Option	Description
VB Script	<p>Use the VB Script drop-down list to select a VB script to associate with the identification and processing of this Advanced Capture form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>
Configuration Note	<p>Enter a configuration note of up to 255 characters to be displayed for documents matched to this Advanced Capture form.</p> <p>This note is displayed in the Advanced Capture Configuration window only, under the form's name in the Form Configuration panel and within the Modify/Delete Form dialog box in the Document Viewer.</p> <hr/> <p>Note: To ensure readability, only the first 35 characters of a configuration note are displayed in the Form Configuration panel.</p> <hr/>
Indexing Note	<p>Enter an indexing note of up to 250 characters to be displayed for documents matched to this Advanced Capture form.</p> <p>This note is displayed to users when a document matched to this Advanced Capture form is displayed in the Working window during index verification and review.</p>

Create/Assign Form Options

The following options are only available when the **Form Definition** dialog box is accessed automatically (i.e., when pre-processing fails, or is skipped or canceled) or by clicking **Create/Assign Form Definition** in the **Tools** panel:

Option	Description
Assign existing form	<p>Select the Assign existing form radio button, and then select an existing Advanced Capture form from the drop-down menu to assign that form to the document currently displayed in the Document Viewer.</p> <p>Once selected, the remaining options in the dialog box are disabled.</p>
New form	<p>Select the New form radio button to create a new Advanced Capture form for the document currently displayed in the Document Viewer.</p> <p>Once selected, the remaining options in the dialog box, other than the Assign existing form option, are enabled.</p>
Copy Configuration From	<p>If you would like to copy all of the configured zones from an existing Advanced Capture form, select the existing form from this drop-down list.</p>

Modify Form Options

The following options are only available when the **Form Definition** dialog box is accessed by right-clicking on a document in the Document Viewer and selecting **Modify Form**:

Option	Description
Document Types Allowed for Form	This option allows you to specify the Document Types that this Advanced Capture form is available to (i.e., the allowable incoming Document Types for this form).
Scan Queue/Keyword Rules	Click this button to configure scan queue or Keyword Value rules for this Advanced Capture form.
Keyword Lookup and Replace	Click this button to configure a form-level Keyword Lookup/Replace dictionary for this Advanced Capture form. A Keyword Lookup/Replace dictionary is used to replace Keyword Values identified by the Advanced Capture process with different values. For example, if you require a Term Keyword Value to contain a Season and a Four Digit Year (e.g., Fall 2010) and some documents display a Season and a Two Digit Year (e.g., Fall 10), you could configure a lookup/replace action to automatically replace Term Keywords Values identified as Fall 10 with Fall 2010 .
Form Definition Lookup Text	Click this button to configure lookup text for this Advanced Capture form.
Image Processing Operations	Click this button to configure image processing or cleanup operations on documents matched to this Advanced Capture form. Note: This button is only available if you are licensed for Bar Code Recognition Server or Barcode Recognition for Advanced Capture.
Configuration User Groups	Click this button to configure which User Groups will have the rights to modify this Advanced Capture form.

Additional Information on Form Definition Options

Refer to the sections below for additional information on the following **Form Definition** dialog box options:

- **Minimum/Maximum Page Count** (see [Configuring Document Breaks on page 61](#))
- **Document Types Allowed for Form** (see [Specifying Advanced Capture Form Availability by Document Type on page 62](#))
- **Scan Queue/Keyword Rules** (see [Configuring Scan Queue or Keyword Value Rules for a Form on page 65](#))
- **Keyword Lookup and Replace** (see [Configuring Keyword Lookup/Replace Dictionaries on page 73](#))

- **Form Definition Lookup Text** (see [Configuring Lookup Text for a Form on page 83](#))
- **Image Processing Operations** (see [Configuring Image Processing Operations on page 87](#))
- **Configuration User Groups** (see [Restricting Form Modification Rights By User Group on page 94](#))

Configuring Document Breaks

If the documents matched to the Advanced Capture form consist of multiple pages, you can configure the form to automatically identify the beginning and end of a document (i.e., document breaks) in the batch based on the minimum and maximum number of pages allowed for a document matched to that form.

Note: If document breaks are specified by the scan format you are using to scan documents, then this step is not required.

Note: When document breaks are applied to PDF documents through Advanced Capture, the renditions created are image documents, not PDFs. When document breaks are applied to image documents, however, the Advanced Capture engine respects the original image format and file extension.

To specify the minimum and maximum number of pages allowed for a document matched to a form, use the **Minimum Page Count** and **Maximum Page Count** fields on the **Form Definition** dialog box when creating a new form or modifying an existing form.

- **Minimum Page Count.** Set the **Minimum Page Count** field to the minimum number of pages that can compose a document identified by this Advanced Capture form. This setting must be set to at least a value of 1.
- **Maximum Page Count.** Set the **Maximum Page Count** field to the maximum number of pages that can compose a document identified by this Advanced Capture form. Enter 0 if the OCR engine should not attempt to break documents (i.e., Advanced Capture should respect the pre-existing document breaks).

The following process describes how Advanced Capture attempts to determine how to break documents:

1. When the OCR engine encounters a new document (i.e., the first page of the first document in the batch or the first unclassified page after the end of the previously identified document), it processes the page and attempts to match it to a form.
2. Once the page is matched to a form, the OCR engine determines the minimum number of pages required for documents matched to that form and appends a number of pages to the current page to create a document containing the minimum number of pages.
3. Once that document is created, the OCR engine moves to the next page in the batch to attempt to match it to a form.
4. If this page is not matched to a form to create a new document, the page is appended to the previous document.

5. Steps 3 and 4 are repeated until a page can be matched to a form to start a new document or until the page count of the document reaches the **Maximum Page Count** configured for the Advanced Capture form.
6. Once the **Maximum Page Count** is reached, the next page in the batch is automatically treated as a new document. If it is not matched to a form, this document is considered unclassified and the OCR engine moves to the next page in the batch and attempts to match it to a form.

For example:

A 20-page batch is scanned and the first page in the batch (Page 1) is matched to a form with a **Minimum Page Count** of 5 and a **Maximum Page Count** of 8.

The OCR engine automatically appends the next four pages in the batch (Pages 2-5) to Page 1 to create a 5-page document and processes the next page in the batch (Page 6).

- If Page 6 can be matched to a form, it is treated as the first page in a new document.
- If Page 6 cannot be matched to a form, then it is appended to the previous document because it would now only consist of six pages, not the maximum of eight.

If Pages 6, 7, and 8 all could not be matched to an Advanced Capture form and are appended to the first document, then Page 9 is automatically treated as the first page of a new document because the maximum number of pages was reached for the first document.

If Page 9 cannot be matched to a form, it is treated as an unclassified document, and the OCR engine processes Page 10 to attempt to match it to an Advanced Capture form.

Specifying Advanced Capture Form Availability by Document Type

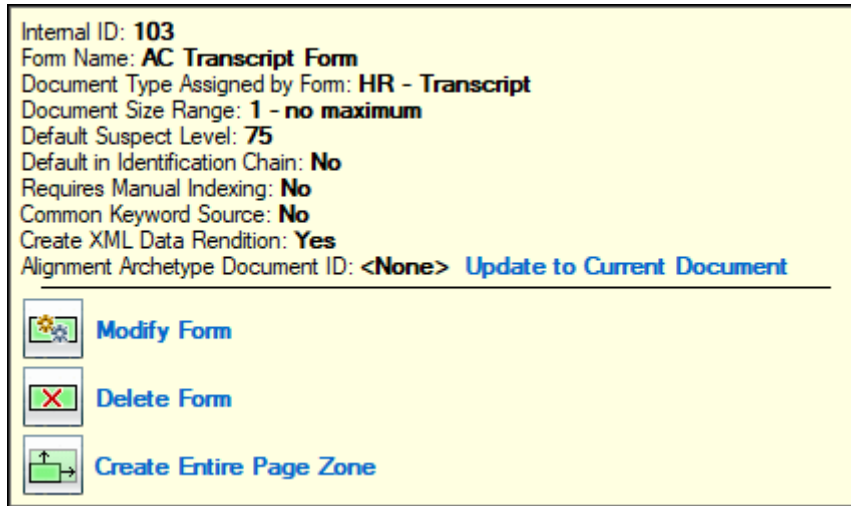
By default, a form is available only to the Document Type of the document it was associated with when it was created (i.e., only documents belonging to the same Document Type as the document used to create the Advanced Capture form can be matched to the Advanced Capture form).

However, you can configure a form to be available to additional Document Types (including unindexed documents) during the Advanced Capture process.


Tip: In addition to using this option to restrict Advanced Capture form availability by Document Type, you may also restrict Advanced Capture form availability by scan queue or the presence of a Keyword Value. These restrictions can be used separately or in conjunction with one another. Because a limitless number of Advanced Capture forms can exist in your Advanced Capture solution, it is considered a best practice to limit the number of Advanced Capture forms available as much as possible (by Document Type, Keyword Value, and/or scan queue) in order to save time and processing resources.


To specify a form's availability by Document Type:


1. Create the Advanced Capture form as usual.
2. Once the form has been created, from the **Advanced Capture Configuration** window, right-click on the document associated with the form in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.



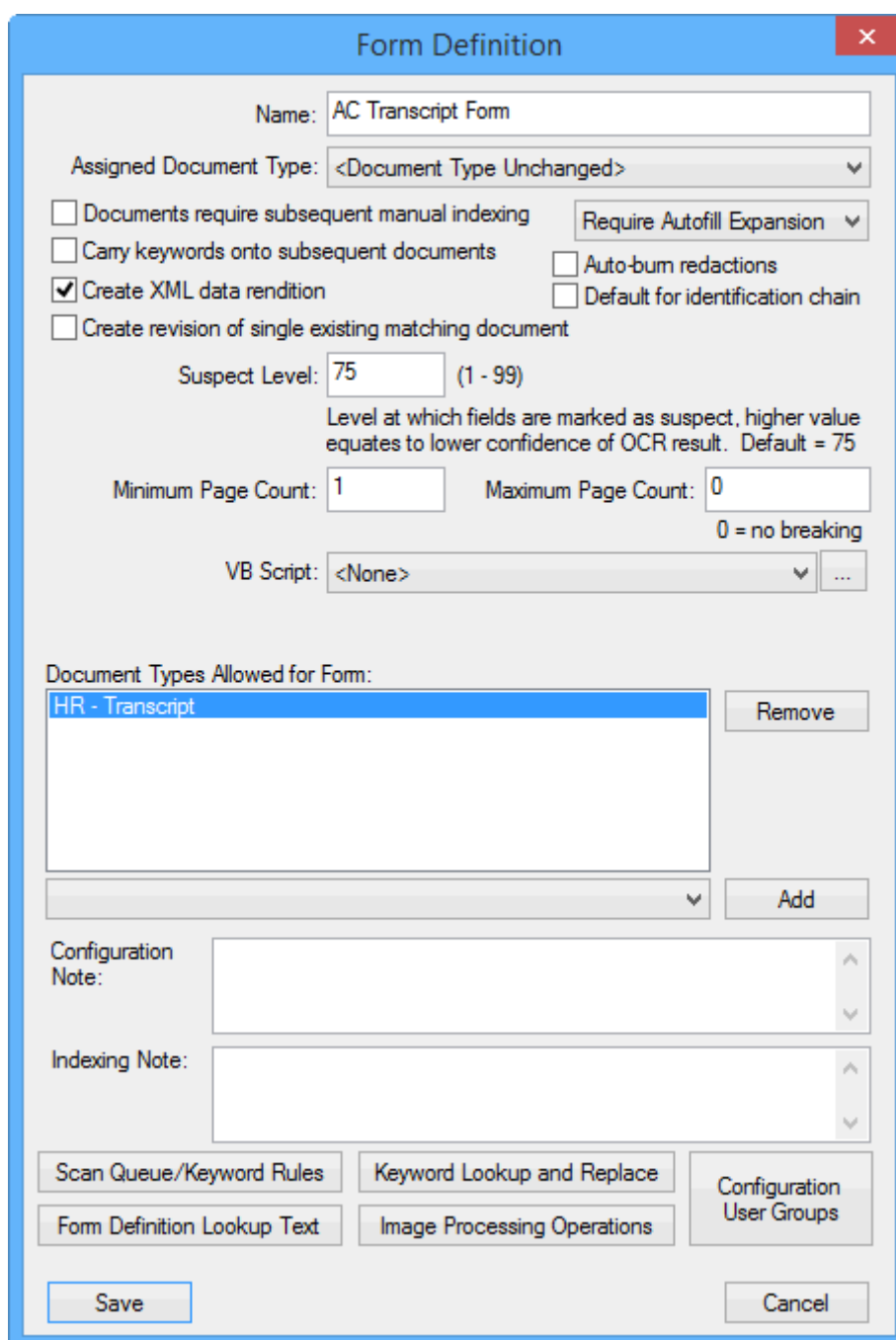
Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**

 **Delete Form**

 **Create Entire Page Zone**

3. Click **Modify Form**. The **Form Definition** dialog box is displayed.

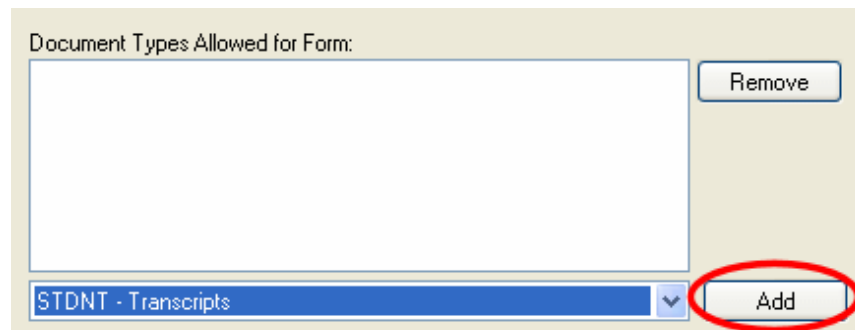


The **Form Definition** dialog box is shown with the following settings:

- Name:** AC Transcript Form
- Assigned Document Type:** <Document Type Unchanged>
- ☐ Documents require subsequent manual indexing
- ☐ Carry keywords onto subsequent documents
- ☒ Create XML data rendition
- ☐ Create revision of single existing matching document
- Require Autofill Expansion:** (dropdown menu)
- ☐ Auto-burn redactions
- ☐ Default for identification chain
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None> (dropdown menu)
- Document Types Allowed for Form:**
 - HR - Transcript (selected)
- Remove** (button)
- Add** (button)
- Configuration Note:** (text area)
- Indexing Note:** (text area)
- Scan Queue/Keyword Rules** (button)
- Keyword Lookup and Replace** (button)
- Form Definition Lookup Text** (button)
- Image Processing Operations** (button)
- Configuration User Groups** (button)
- Save** (button)
- Cancel** (button)

Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Advanced Capture form.

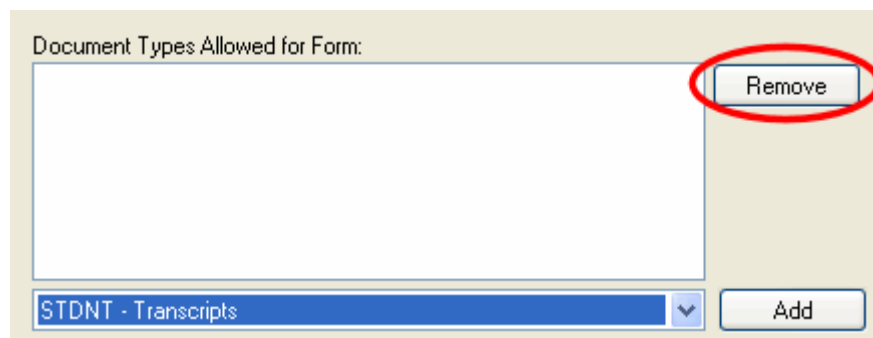
4. To add a Document Type allowed for the form (i.e., to allow documents from a Document Type to be matched to the Advanced Capture form), select the Document Type you would like to add from the drop-down below the Document Types Allowed for Form list and click **Add**.



Note: Only Document Types assigned to the scan queue are available in the drop-down.

Documents of this Document Type can now be matched to this Advanced Capture form. Depending on the Advanced Capture form's configuration, after the document is matched to the Advanced Capture form, its Document Type may be changed.

5. To remove a Document Type allowed for the form (i.e., to prevent documents from a Document Type from being matched to the Advanced Capture form), select the Document Type you would like to remove from the Document Types Allowed for Form list and click **Remove**.



Documents of this Document Type now cannot be matched to this Advanced Capture form. During processing, documents assigned to this Document Type are not compared to this Advanced Capture form.

6. Once you set the Document Types correctly for the Advanced Capture form, click **Save**.

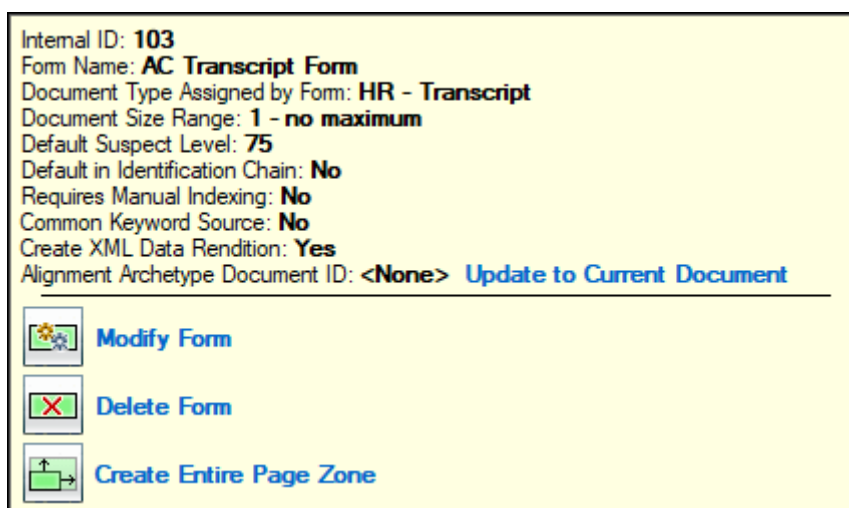
Configuring Scan Queue or Keyword Value Rules for a Form

You may configure scan queue or Keyword Value requirements for associating documents with a form (i.e., in order to be associated with this Advanced Capture form, documents must have been scanned using a specified scan queue or must have one or more specified Keyword Values).


Additionally, you can assign default Keyword Values for documents associated with a form (e.g., all documents associated with the **HYL-Hyland University** Advanced Capture form are assigned a Keyword Value of **RECEIVED** for the **Status** Keyword Type) and, depending on your form's configuration, default XML nodes for these default Keyword Values to be contained in when XML renditions of the documents are created (e.g., the default value of **RECEIVED** for the **Status** Keyword Type is listed within the assigned node under the XML rendition's DOCUMENTDATA section).


To configure scan queue or Keyword Value rules for a form:


1. Create the Advanced Capture form as usual.
2. Once the form has been created, from the **Advanced Capture Configuration** window, right-click on the document displayed in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.



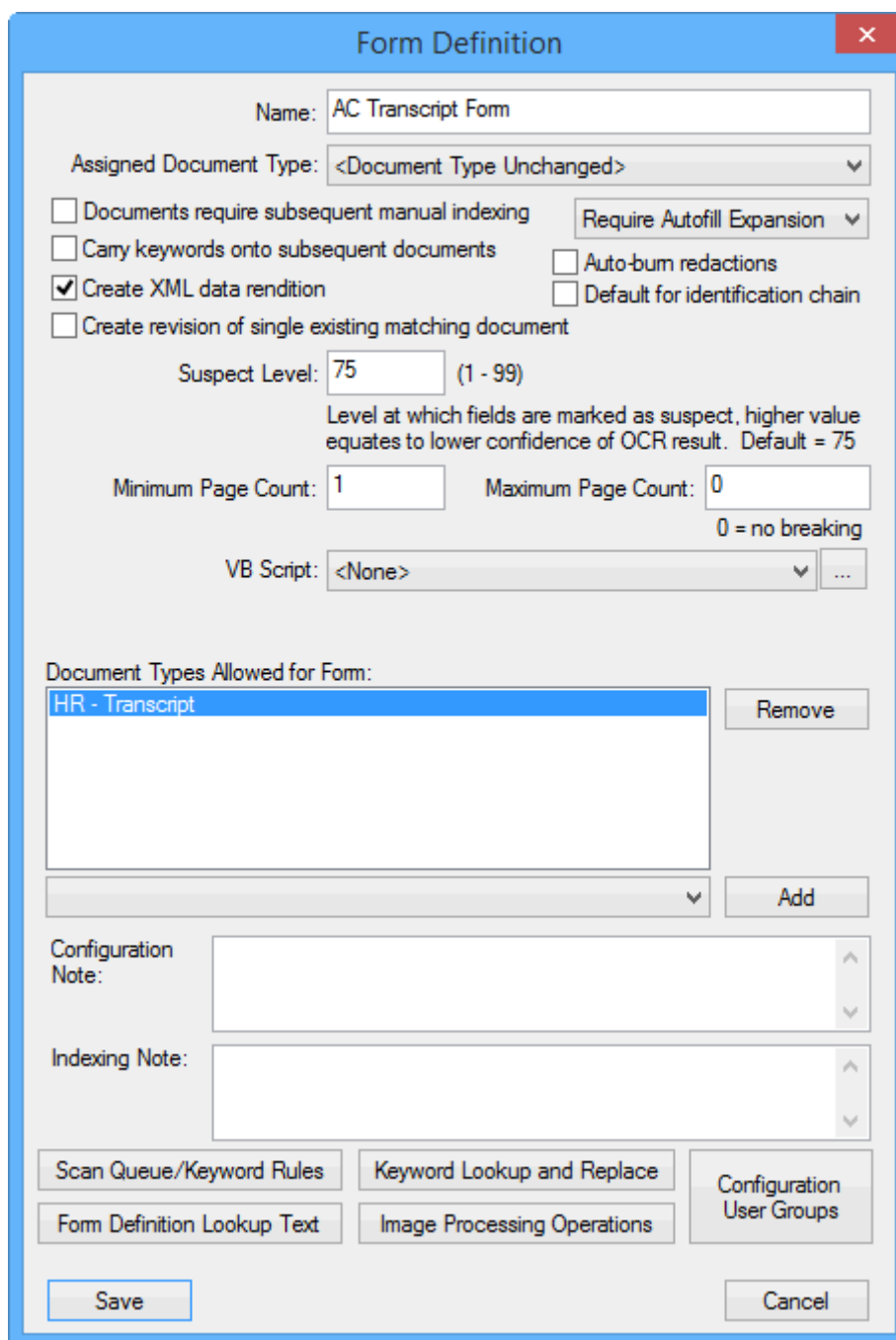
Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**

 **Delete Form**

 **Create Entire Page Zone**

3. Click **Modify Form**. The **Form Definition** dialog box is displayed.

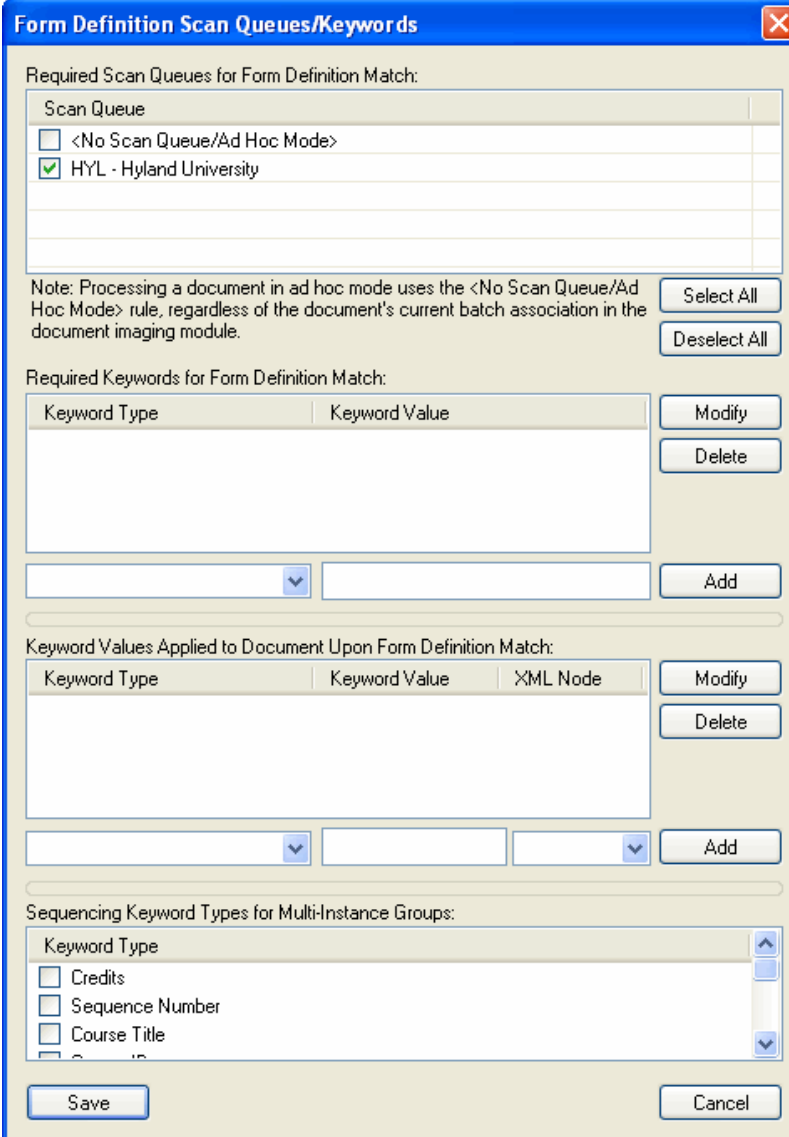


The **Form Definition** dialog box is shown with the following settings:

- Name:** AC Transcript Form
- Assigned Document Type:** <Document Type Unchanged>
- ☐ Documents require subsequent manual indexing
- ☐ Carry keywords onto subsequent documents
- ☒ Create XML data rendition
- ☐ Create revision of single existing matching document
- Require Autofill Expansion** (dropdown)
- ☐ Auto-burn redactions
- ☐ Default for identification chain
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None>
- Document Types Allowed for Form:**
 - HR - Transcript (selected)
- Remove** button
- Add** button
- Configuration Note:** (text area)
- Indexing Note:** (text area)
- Buttons:** Scan Queue/Keyword Rules, Keyword Lookup and Replace, Configuration User Groups, Form Definition Lookup Text, Image Processing Operations
- Save** and **Cancel** buttons

Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Advanced Capture form.

- Click **Scan Queue/Keyword Rules**. The **Form Definition Scan Queues/Keywords** dialog box is displayed.



Form Definition Scan Queues/Keywords

Required Scan Queues for Form Definition Match:

Scan Queue
<input type="checkbox"/> <No Scan Queue/Ad Hoc Mode>
<input checked="" type="checkbox"/> HYL - Hyland University

Note: Processing a document in ad hoc mode uses the <No Scan Queue/Ad Hoc Mode> rule, regardless of the document's current batch association in the document imaging module.

Select All
Deselect All

Required Keywords for Form Definition Match:

Keyword Type	Keyword Value
--------------	---------------

Modify
Delete

▼

Add

Keyword Values Applied to Document Upon Form Definition Match:

Keyword Type	Keyword Value	XML Node
--------------	---------------	----------

Modify
Delete

▼

Add

Sequencing Keyword Types for Multi-Instance Groups:

Keyword Type
<input type="checkbox"/> Credits
<input type="checkbox"/> Sequence Number
<input type="checkbox"/> Course Title

Save Cancel

5. Configure the scan queue and/or Keyword Value rules as needed:
 - **To specify the scan queues that the Advanced Capture form is available to (i.e., only documents scanned from the specified scan queues can be matched to the current Advanced Capture form):**

Tip: In addition to using this option to restrict Advanced Capture form availability by scan queue, you may also restrict Advanced Capture form availability by Document Type or the presence of a Keyword Value. These restrictions can be used separately or in conjunction with one another. Because a limitless number of Advanced Capture forms can exist in your Advanced Capture solution, it is considered a best practice to limit the number of Advanced Capture forms available as much as possible (by Document Type, Keyword Value, and/or scan queue) in order to save time and processing resources.

By default, an Advanced Capture form is available to only the scan queue that the document used to create for the form was scanned with and to documents that are not associated with a scan queue.

Note: Depending on your configuration, a form may be available to all scan queues once it is created. This default behavior is controlled by the **Assign Current Scan Queue** option in the **Tools** panel of the **Advanced Capture Configuration** window.

To make the Advanced Capture form available to additional scan queues (or restrict it from additional scan queues):

In the Required Scan Queues for Form Definition Match section, select the check boxes next to the scan queues that the Advanced Capture form is available to. To select all scan queues, click **Select All**. To deselect all scan queues, click **Deselect All**.

To make the Advanced Capture form available to documents not associated with a scan queue or those documents that undergo ad-hoc Advanced Capture (where a scan queue association would not be known), select the **<No Scan Queue/Ad Hoc Mode>** check box.

Required Scan Queues for Form Definition Match:

Scan Queue	
<input type="checkbox"/> <No Scan Queue/Ad Hoc Mode>	
<input checked="" type="checkbox"/> HYL- Hyland University	

Note: Processing a document in ad hoc mode uses the <No Scan Queue/Ad Hoc Mode> rule, regardless of the document's current batch association in the document imaging module.

Select All
Deselect All

In the example illustrated above, only documents imported via the **HYL-Hyland University** scan queue can be matched to the Advanced Capture form.

- **To only allow documents with a specified Keyword Value to be matched to the Advanced Capture form:**

Tip: In addition to using this option to restrict Advanced Capture form availability by the presence of a Keyword Value, you may also restrict Advanced Capture form availability by Document Type or scan queue. These restrictions can be used separately or in conjunction with one another. Because a limitless number of Advanced Capture forms can exist in your Advanced Capture solution, it is considered a best practice to limit the number of Advanced Capture forms available as much as possible (by Document Type, Keyword Value, and/or scan queue) in order to save time and processing resources.

In the Required Keywords for Form Definition Match section, use the **Keyword Type** drop-down to select the Keyword Type, enter the Keyword Value in the **Keyword Value** field that documents must have in order to be associated with this Advanced Capture form, and then click **Add**.

Keyword Type	Keyword Value
Course ID	PSY111

Modify
Delete

Course ID Add

Note: Only Keyword Types common to all Document Types specified in the Document Types Allowed for Form list for the Advanced Capture form can be selected from the **Keyword Type** drop-down. If only **Unindexed Document** is specified in the Document Types Allowed for Form list, no Keyword Types are available to be selected (i.e., no Keyword Value restrictions can be configured).

In the example illustrated above, only documents with a value of **PSY111** for the **Course ID** Keyword Type can be matched to the Advanced Capture form.

Repeat this process for as many Keyword Value restriction rules as are needed for the Advanced Capture form. To modify a default Keyword Value rule, double-click it in the Keyword Type/Keyword Value list or select it from the list and click **Modify**. To delete a default Keyword Value rule, select it from the Keyword Type/Keyword Value list and click **Delete**.

- **To add a default Keyword Value to all documents matched to the Advanced Capture form:**

In the Keyword Values Applied to Document Upon Form Definition Match section, use the **Keyword Type** drop-down to select the Keyword Type and enter the Keyword Value in the **Keyword Value** field that is to be automatically assigned as a default Keyword Value to documents matched to this Advanced Capture form, and then click **Add**.

Keyword Type	Keyword Value
Status	RECEIVED

Buttons: Modify, Delete

Form fields: Status (dropdown), [Empty text field], Add button

Note: Only Keyword Types assigned to the Document Type specified as the **Assigned Document Type** for the Advanced Capture form can be selected from the **Keyword Type** drop-down. If only **Unindexed Document** is specified in the Document Types Allowed for Form list and the Assigned Document Type is set to **<Document Type Unchanged>**, then no Keyword Types are available to be selected (i.e., no default Keyword Values can be configured)

In the example illustrated above, a value of **RECEIVED** is assigned to the **Status** Keyword Type for all documents that are matched to this Advanced Capture form. Repeat for as many default Keyword Values as you would like to add to documents matched to this Advanced Capture form. To modify a default Keyword Value rule, double-click it in the Keyword Type/Keyword Value list or select it from the list and click **Modify**. To delete a default Keyword Value rule, select it from the Keyword Type/Keyword Value list and click **Delete**.

- To assign a default XML node to a default Keyword Value:

Note: If you selected **Create XML data rendition** in the **Form Definition** dialog box, you can assign a default XML node to a default Keyword Value, even if the assigned Document Type is not configured to allow renditions. However, in order for XML renditions to actually be created during Advanced Capture processing, the assigned Document Type must be configured to allow renditions.

In the Keyword Values Applied to Document Upon Form Definition Match section, use the **Keyword Type** drop-down to select the Keyword Type, enter the Keyword Value in the **Keyword Value** field that is to be automatically assigned as a default Keyword Value to documents matched to this Advanced Capture form, enter the name of the XML node in which the default Keyword Value will be contained in the documents' XML data renditions, and then click **Add**.

Keyword Type	Keyword Value	XML Node
Status	RECEIVED	DocNote

Status ▼ ▼

Note: While XML node names may include alphanumeric characters, underscores (_), hyphens (-), or periods (.), they must begin with either a letter or an underscore.

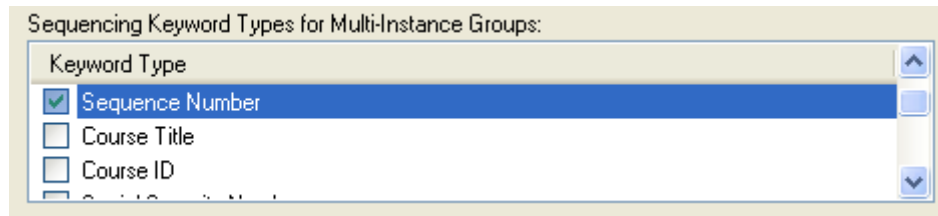
Note: Once XML node names are added and the configuration is saved, the names will be added to the **XML Node** drop-down list for future selection.

In the example above, the XML node name **DocNote** is assigned to the default Keyword Value **RECEIVED** for the **Status** Keyword Type. This means that the **RECEIVED** value will be listed within the **DocNote** node in the XML rendition's DOCUMENTDATA section.

Repeat for as many default XML nodes as you would like to assign to default Keyword Values for documents matched to this Advanced Capture form. To modify a default XML node rule, double-click it in the Keyword Type/Keyword Value/XML Node list or select it from the list and click **Modify**. To delete a default XML node rule, select it from the Keyword Type/Keyword Value/XML Node list and click **Delete**.

- **To assign a Keyword Type to keep track of the order of line items on documents matched to the Advanced Capture form:**

In the Sequencing Keyword Types for Multi-Instance Groups section, select a Keyword Type that is to be assigned a numerical value representing the order in which the instance of a Multi-Instance Keyword Type Group appears on the document.



Note: Only Keyword Types assigned to the Document Type specified as the **Assigned Document Type** for the Advanced Capture form can be selected from the Keyword Type list. Only selected Keyword Types that belong to an MIKG configured for Line Item Extraction Data Field Zones will be assigned the sequential value when the document is processed.

In the example illustrated above, the **Sequence Number** Keyword Type is configured to be assigned the sequential value of the instance of the MIKG when the document is processed. The value will correspond to the line number on which the instance of the MIKG appears.

If you have multiple MIKGs configured for the Document Type assigned to the Advanced Capture form, you can select a different sequencing Keyword Type for each MIKG. If you select multiple sequencing Keyword Types for the same MIKG, only the first Keyword Type is assigned the sequential value.

- When all scan queue and/or Keyword Value rules have been configured, click **Save**.

Configuring Keyword Lookup/Replace Dictionaries

Your Advanced Capture solution can be configured to use a Keyword Lookup/Replace dictionary to replace Keyword Values extracted from documents via the Advanced Capture process with other values.

For example:

Your institution stores **Term** Keyword Values as a season and a four-digit year (e.g., **Fall 2008**, **Spring 2009**, **Summer 2009**, etc.). However, some institutions that you receive transcripts from identify terms using a season and a two-digit year (e.g., **Fall 08**, **Spring 09**, **Summer 09**).

You can configure a Keyword Lookup/Replace dictionary to automatically replace the undesirably-formatted Keyword Values extracted from the transcripts with your preferred values.

Keyword Lookup/Replace dictionaries can be configured at either of two different levels:

- **At the Global Level.** When a global Keyword Lookup/Replace dictionary is configured, a Keyword Value extracted using any Advanced Capture form in your solution can be replaced by an entry in the dictionary.

Entries in the Global Keyword Lookup/Replace dictionary are configured by clicking **Global Keyword Lookup/Replace** in the **Tools** panel of the **Advanced Capture Configuration** window.

For example: You have several dozen Advanced Capture forms configured for your solution, and your global Keyword Lookup/Replace dictionary specifies that **Fall 08** should be replaced by **Fall 2008** for the **Term** Keyword Value. As a result, all instances where the **Term** Keyword Value is identified as **Fall 08** by the Advanced Capture process are replaced by **Fall 2008**, regardless of which Advanced Capture form was matched to the document the Keyword Value was extracted from.

- **At the Advanced Capture Form Level.** When a form-level Keyword Lookup/Replace dictionary is configured, a Keyword Value is only replaced if it is specified in the Keyword Lookup/Replace dictionary for the form used to identify it. Form-level Keyword Lookup/Replace dictionaries can be configured by clicking **Keyword Lookup and Replace** in the **Form Definition** dialog box when modifying an existing Advanced Capture form.

The image shows a 'Form Definition' dialog box for a form named 'AC Transcript Form'. The dialog has a blue title bar with a close button. Inside, there are several sections:

- Name:** A text field containing 'AC Transcript Form'.
- Assigned Document Type:** A dropdown menu showing '<Document Type Unchanged>'.
- Options:** A group of checkboxes:
 - ☐ Documents require subsequent manual indexing
 - ☐ Carry keywords onto subsequent documents
 - ☒ Create XML data rendition
 - ☐ Create revision of single existing matching document
 - ☐ Require Autofill Expansion
 - ☐ Auto-bum redactions
 - ☐ Default for identification chain
- Suspect Level:** A text field with '75' and a range '(1 - 99)'. Below it is a note: 'Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75'.
- Minimum Page Count:** A text field with '1'.
- Maximum Page Count:** A text field with '0'. Below it is a note: '0 = no breaking'.
- VB Script:** A dropdown menu showing '<None>' and a button with three dots.
- Document Types Allowed for Form:** A list box containing 'HR - Transcript' (highlighted) and 'Unindexed Document'. To the right are 'Remove' and 'Add' buttons.
- Configuration Note:** A text area with up and down arrow buttons.
- Indexing Note:** A text area with up and down arrow buttons.
- Buttons:** At the bottom, there are buttons for 'Scan Queue/Keyword Rules', 'Form Definition Lookup Text', 'Image Processing Operations', 'Configuration User Groups', 'Save', and 'Cancel'. The 'Keyword Lookup and Replace' button is highlighted with a red rectangle.

For example: You have two Advanced Capture forms in your solution (**HYL-Hyland**

University and **WSU-Westlake State University**), and a Keyword Lookup/Replace dictionary is configured for the Hyland University form (but not the Westlake State University form) to replace **Term** Keyword Value with **Fall 2008** when it is identified by the Advanced Capture process as **Fall 08**. As a result, all instances of the **Term** Keyword set to **Fall 08** are replaced with **Fall 2008** for documents matched to Hyland University form, but **Fall 08** remains set as the **Term** Keyword Value for all documents matched to the Westlake State University form.

If both a global and a form-level Keyword Lookup/Replace dictionary entry are configured for the same Keyword Value, then the form-level entry will override the global entry.

Note: A Keyword Value can only be replaced once. If a Keyword Value is replaced by another value, and the replacement Keyword Value also has an entry in the Keyword Lookup/Replace dictionary, the replacement value is not replaced.

Creating a Keyword Lookup/Replace Dictionary Entry

1. Determine whether you would like to create an entry in the global Keyword Lookup/Replace dictionary or a form-level Keyword Lookup/Replace dictionary:
 - To create an entry in the global Keyword Lookup/Replace dictionary: click **Global Keyword Lookup/Replace** in the Tools panel of the **Advanced Capture Configuration** window.

The **Keyword Lookup and Replace** dialog box is displayed.

- Click **Add**. The **Edit Keyword Lookup and Replace Entry** dialog box is displayed.

Keyword Type: Last Name

Description:

When Document Type is:

Compare Value:
 (Literal text value enclosed in quote marks, or regular expression not in quotes)

☐ Only match if value is marked as suspect ☐ Comparison is case-sensitive

Replacement Value:

☒ Replace with static text ☐ Replace using capture groups
☐ Replace with <blank> ☐ Leave value unchanged

☒ Suspect Level Remains Unchanged
☐ Mark Value as Suspect
☐ Mark Value as Not Suspect

Display Color Automatic Suspect Color Automatic
 Automatic Automatic Automatic Automatic

Save Cancel

- In the **Description** field, enter a description of the entry (up to 255 characters).
- From the **When Document Type is** drop-down list, select the Document Type for which Keyword Lookup/Replace actions will be performed.
 - If you are creating an entry in a form-level Keyword/Lookup Replace dictionary for a form that has been configured to assign documents to a specific Document Type, then this Document Type is automatically selected.
 - If you are creating an entry in a global Keyword/Lookup Replace dictionary, or if you are creating an entry in a form-level Keyword/Lookup Replace dictionary for a form that has *not* been configured to assign documents to a specific Document Type, then **<All>** is selected by default. When this option is selected, the Keyword Lookup/Replace actions will be performed for all Document Types with the specified Keyword Type.

6. In the **Compare Value** field, enter the Keyword Value that is to be replaced if it is found. This value can be a literal text string (i.e., specific text, such as **Fall 08**) or a regular expression (i.e., text that matches a certain pattern, such as a value containing four letters, a space, and two digits).

Literal values must be enclosed in quotation marks to differentiate them from regular expressions. Multiple values (literal values, regular expression, and/or both) can be configured by separating them by a space in the **Compare Value** field. Literal compare values are not case-sensitive.

Note: To access the Regular Expression Library, click in the field and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

- If you wish to only replace values that are marked as suspect, select the **Only match if value is marked as suspect** check box.
 - If you wish to restrict the values that will be replaced according to case-sensitivity matches, select the **Comparison is case-sensitive** check box. The comparison between the value in this field and the Keyword Values extracted from the document will now be case-sensitive, whether the value is a literal text string or a regular expression. If the value is a regular expression, however, be aware that case-sensitivity may not be respected if the regular expression does not imply case one way or the other (e.g., **\D**, **\w**, etc.). For more information on regular expressions, see [Regular Expressions on page 350](#).
7. Select the radio button that describes how you would like the compare value to be treated once it is identified:
 - **Replace with static text.** Select this option if you would like to replace the compare value with a static text string. Enter the static text you would like to replace the compare value with in the **Replacement Value** field.
For example: You would like to replace all **Term** Keywords set to **Fall 08** with **Fall 2008**. Set the compare value to **"Fall 08"** (a literal text value), select the **Replace with static text** radio button, and specify the replacement value as **Fall 2008**.

- **Replace with <blank>**. Select this option if you would like to replace the compare value with a blank value. Once this option is selected, the **Replace Value** field is disabled.

For example: You would like to discard all **Term** Keywords set to **Fall 08** and have a user manually enter a new value during Index Verification and Review. Set the compare value to **"Fall 08"** (a literal text value) and select the **Replace with <blank>** radio button.

Replace using captured groups. Select this option if you would like to replace the compare value with a new value composed of sections of the compare value.

Note: This feature is only available when specifying a regular expression for the compare value.

By enclosing sections of your regular expression compare value in parentheses, you can specify the portions of the compare value (i.e., the building blocks) that are used to create the replacement value. Then, specify the sections of the compare value (e.g., **[1] [2] [4]**) you would like to use to compose the new Keyword Value in the **Replacement Value** field.

For example: Your institution stores Keyword Values for the **Term** Keyword using a **<Season> <Four-Digit Year>** format (e.g., **Fall 2008, Spring 2009, Summer 2009**). However, some transcripts received from other institutions use a **<Season> Semester <Four-Digit Year>** format for **Term** Keyword Values (e.g., **Fall Semester 2008, Spring Semester 2009, Summer Semester 2009**).

Create the following regular expression to identify the compare value: **(\S{4,6}\s)(SEMESTER)\s(\d{4})**. As you can see, parentheses are used to divide the value into three sections: the season **(\S{4,6}\s)**, Semester **(SEMESTER)**, and the year **(\d{4})**.

To drop the Semester "building block" (section #2) from the compare value and use only the season and the year "building blocks" (sections #1 and #3) to create the new value, enter **[1] [3]** in the Replacement Value field. **Term** Keyword Values matching the **<Season> Semester <Four-Digit Year>** format are replaced with Keyword Values matching the **<Season> <Four-Digit Year>** format.

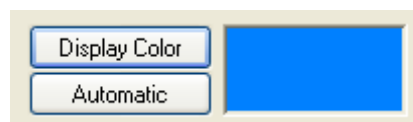
- **Leave value unchanged.** Select this option if you would like to leave the compare value as-is. Once this option is selected, the **Replace Value** field is disabled.

Tip: This option is typically used in a form-level Keyword Lookup/Replace dictionary entry to override an entry in the global dictionary.

For example: You have an entry in the global Keyword Lookup/Replace dictionary that specifies that **Term** Keywords set to **Fall 08** should be changed to **Fall 2008**. However, you have one non-transcript document for which you would like to leave the **Term** Keyword Value as-is.

Leave the global dictionary entry in place and configure a form-level dictionary entry for the Advanced Capture form that is matched to the non-transcript document where the compare value is set to **"Fall 08"** (a literal text value) and select the **Leave Value Unchanged** radio button. Because the form-level dictionary overrides the global dictionary, Fall 08 is changed to Fall 2008 for **Term** Keywords on all documents indexed via Advanced Capture except for the one non-transcript document you configured the form-level dictionary entry for.

8. Select the radio button that determines how the replacement value should be treated in terms of Suspect Level.
 - **Suspect Level Remains Unchanged.** The replacement value is given the Suspect Level score given to the compare value when it was identified by the OCR engine. The replacement value's score is compared to the Suspect Level threshold configured for the zone (or the form itself) to determine if it should be considered suspect.
 - **Mark Value as Suspect.** The replacement value is automatically marked as suspect, regardless of the Suspect Level score given to the compare value.
 - **Mark Value as Not Suspect.** The replacement value is automatically marked as not suspect, regardless of the Suspect Level score given to the compare value.
9. If you wish to set all instances of this Keyword Lookup/Replace dictionary entry to a specific color in the Indexing panel when the entry is matched in Advanced Capture processing, click **Display Color** to display your machine's color palette. Then select a color from the palette and click **OK**. The selected color is displayed in the box next to the **Display Color** button.

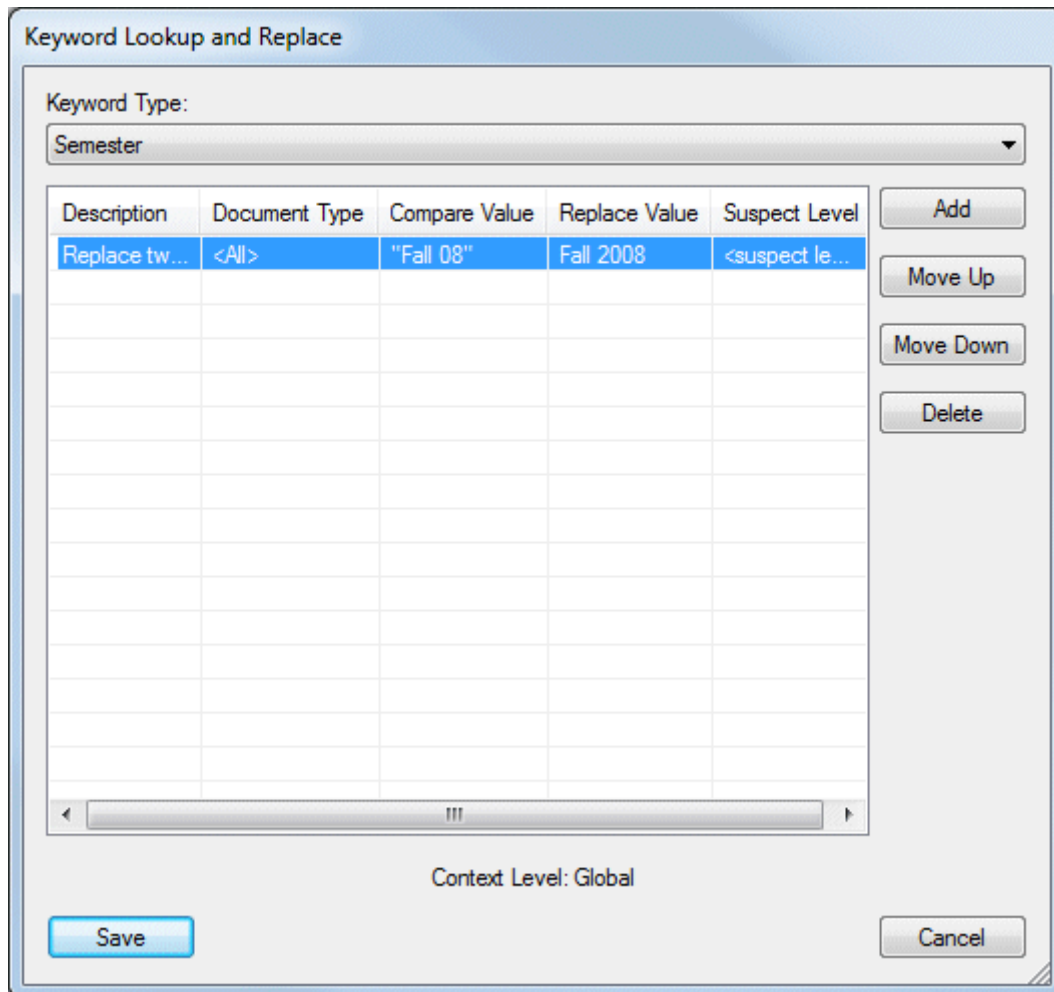


Note: A selected display color overrides any red highlighting for suspect Keyword Values.

If you wish to forgo display colors and retain the default coloring scheme (i.e., with red highlights for suspect Keyword Values and default coloring otherwise), click **Automatic**. This is the default setting.

Tip: When configuring a display color, select a color that will appropriately contrast adjacent colors and foreground/background text.

10. Click **Save**. The **Edit Keyword Lookup and Replace Entry** dialog box is closed and you are returned to the **Keyword Lookup and Replace** dialog box. The entry you created in Steps 2-9 is added to the dictionary.



11. Repeat Steps 2-10 to add additional entries to the Keyword Lookup/Replace dictionary.
- To modify the order of the entries in the dictionary, select an entry and click **Move Up** or **Move Down**.

Tip: If multiple entries in the Keyword Lookup/Replace dictionary affect the same Keyword Value, only the first entry in the dictionary will modify its value (remember -- a Keyword Value can only be replaced once). Use the **Move Up** and **Move Down** buttons to ensure the order of the entries will correctly identify and replace Keyword Values in your solution.

- To delete an entry, select it and click **Delete**.

12. Once the Keyword Lookup/Replace dictionary is complete, click **Save**. The **Keyword Lookup and Replace** dialog box is closed.

Adding or Modifying/Deleting a Keyword Lookup/Replace Dictionary Entry

New entries can be created and existing entries can be modified or deleted at any time for either the global or a form-level Keyword Lookup/Replace dictionary by accessing the **Keyword Lookup and Replace** dialog box.

- To access the global Keyword Lookup/Replace dictionary: click **Global Keyword Lookup/Replace** in the **Tools** panel of the **Advanced Capture Configuration** window.
- To access a form-level Keyword Lookup/Replace dictionary: click **Keyword Lookup and Replace** in the **Form Definition** dialog box when modifying an existing Advanced Capture form.

Note: If you select **<None>** from the **Keyword Type** drop-down list, Keyword Lookup/Replace entries that were imported into your system under the **<None>** Keyword Type are displayed. These entries can be edited or deleted, but they cannot be added to a Keyword Lookup/Replace dictionary.

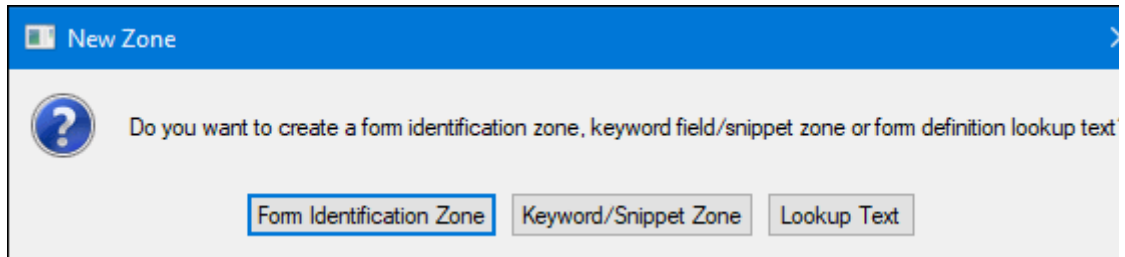
Configuring Lookup Text for a Form

You may configure specific, first-page text values to search for when associating documents with a form. If a document contains the configured text values on its first page, the document is associated with this Advanced Capture form. If a document does not contain the configured text values on its first page, the Advanced Capture engine continues to search for form matches not related to lookup text (i.e., documents matching the configured Document Types, scan queues, Keyword Values, etc.).

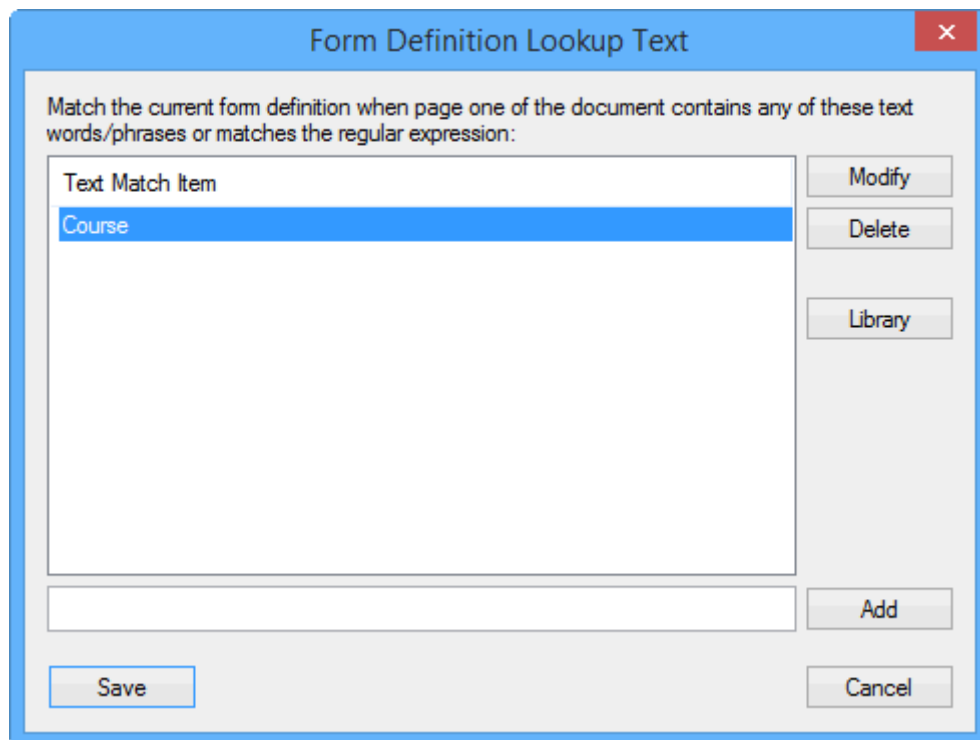
To configure lookup text for a form:

1. Create the Advanced Capture form as usual.
2. Once the form has been created, from the **Advanced Capture Configuration** window, do one of the following:
 - To configure a Lookup Text Zone on the first page of the document, proceed to the next step.
 - To configure lookup text according to a manually entered value, proceed to step 5.
3. To create a Lookup Text Zone on the first page of the document, click and hold the left mouse button and draw a box around the text that is to be used as the lookup text value.

4. Release the left mouse button to display the **Form Definition Lookup Text** dialog box. Depending on the setting of the **Default Zone Type** option in the **Tools** panel, the **New Zone** dialog box might be displayed.




If the **New Zone** dialog box is displayed, click **Lookup Text** to continue to the **Form Definition Lookup Text** dialog box.





Once you have accessed the **Form Definition Lookup Text** dialog box, proceed to step 8.

5. To configure lookup text according to a manually entered value, right-click on the document displayed in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.

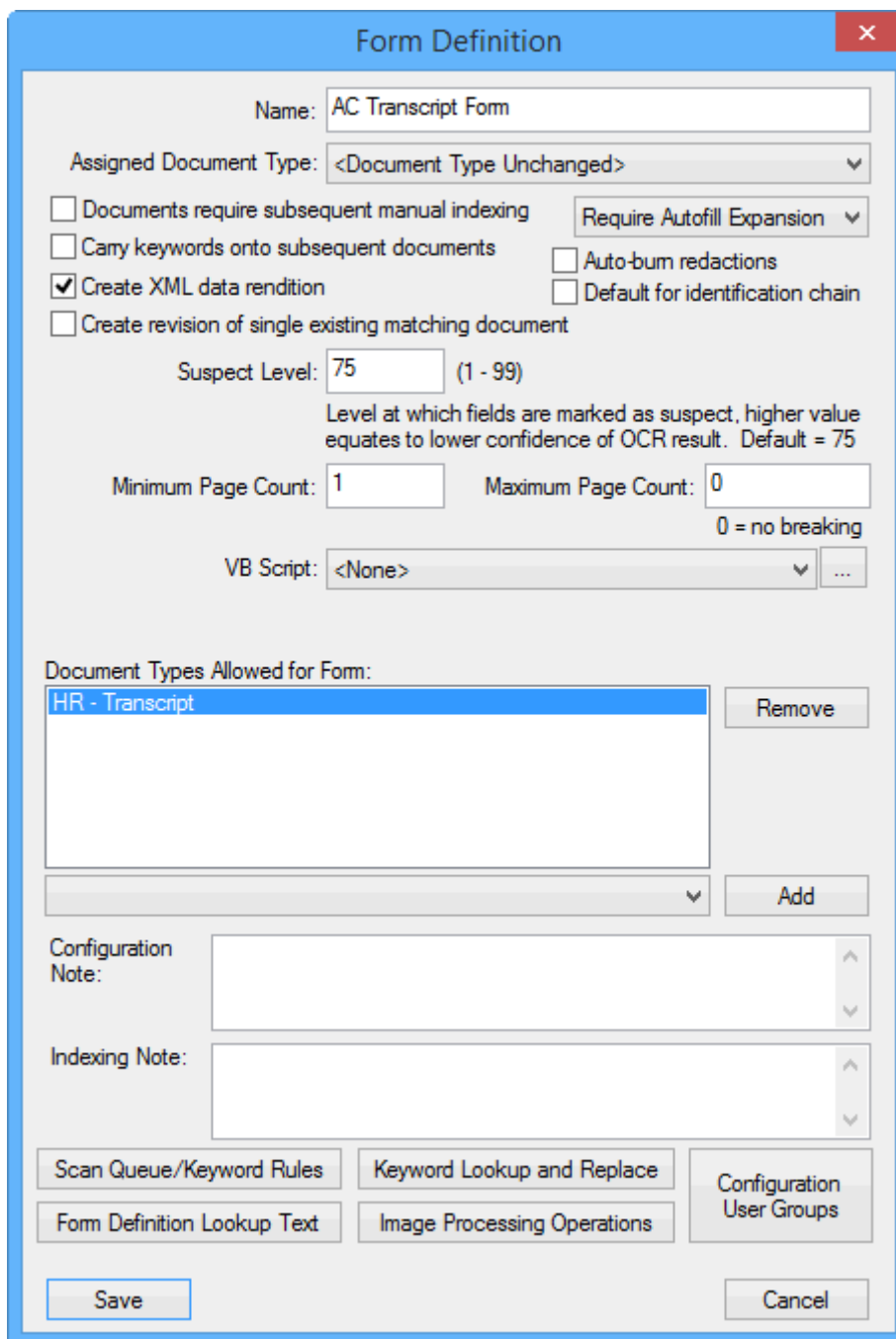
Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**

 **Delete Form**

 **Create Entire Page Zone**

- Click **Modify Form**. The **Form Definition** dialog box is displayed.



The **Form Definition** dialog box is shown with the following settings:

- Name:** AC Transcript Form
- Assigned Document Type:** <Document Type Unchanged>
- ☐ Documents require subsequent manual indexing
- ☐ Carry keywords onto subsequent documents
- ☒ Create XML data rendition
- ☐ Create revision of single existing matching document
- Require Autofill Expansion:** (dropdown menu)
- ☐ Auto-burn redactions
- ☐ Default for identification chain
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None> (dropdown menu)
- Document Types Allowed for Form:**
 - HR - Transcript (selected)
- Remove** (button)
- Add** (button)
- Configuration Note:** (text area)
- Indexing Note:** (text area)
- Scan Queue/Keyword Rules** (button)
- Keyword Lookup and Replace** (button)
- Configuration User Groups** (button)
- Form Definition Lookup Text** (button)
- Image Processing Operations** (button)
- Save** (button)
- Cancel** (button)

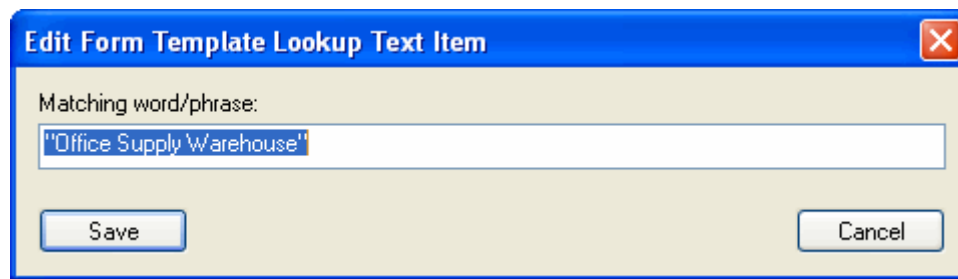
Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Advanced Capture form.

- Click **Form Definition Lookup Text**. The **Form Definition Lookup Text** dialog box is displayed.

8. If you created a Lookup Text Zone on the first page of the document, the text value captured from that zone is displayed in the **Text Match Item** list.
If you are manually entering a lookup text value, in the text field, enter the value you wish to associate with the form. This value can be a literal text string (i.e., specific text, such as **Office Supply Warehouse**) or a regular expression (i.e., text that matches a certain pattern, such as a value containing four letters, a space, and two digits).
Literal values must be enclosed in quotation marks to differentiate them from regular expressions.

Note: To access the Regular Expression Library, either click **Library** or click in the text field and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

9. Once you have entered the text value in the field, click **Add**. The value is added to the **Text Match Item** list.
10. To modify an existing text value:
 - a. Select the value in the **Text Match Item** list and click **Modify**. The **Edit Form Template Lookup Text Item** dialog box is displayed.



- b. In the **Matching word/phrase** field, edit the text value as desired.

Note: To access the Regular Expression Library, click in the field and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

- c. Click **Save**. You are returned to the **Form Definition Lookup Text** dialog box.
11. To delete an existing text value, select the value in the **Text Match Item** list and click **Delete**. A confirmation message is displayed.
12. When you have finished configuring lookup text values, click **Save**. The **Form Definition Lookup Text** dialog box is closed.


Configuring Image Processing Operations


You may configure various image processing operations to be performed on documents matched to the Advanced Capture form. Once the documents are matched to a form, the image processing operations are performed in the configured order in an attempt to improve the accuracy of the OCR extracted results. The effects of these image processing operations exist only on the in-memory copies of the images that are used for OCR processing; the effects are not permanently applied to the documents.


To configure image processing operations for documents:

1. Create the Advanced Capture form as usual.
2. Once the form has been created, from the **Advanced Capture Configuration** window, right-click on the document displayed in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.

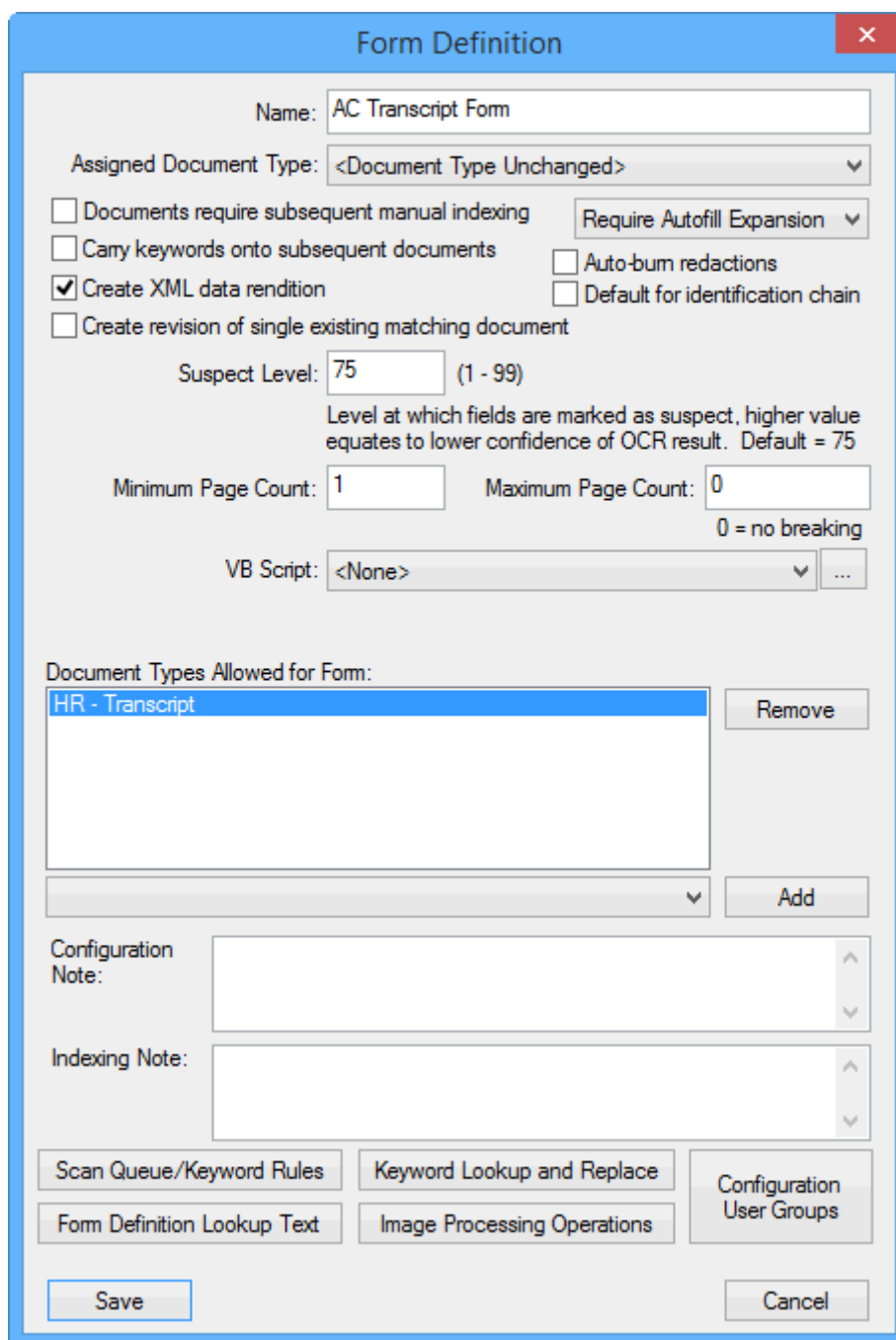
Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**

 **Delete Form**

 **Create Entire Page Zone**

3. Click **Modify Form**. The **Form Definition** dialog box is displayed.

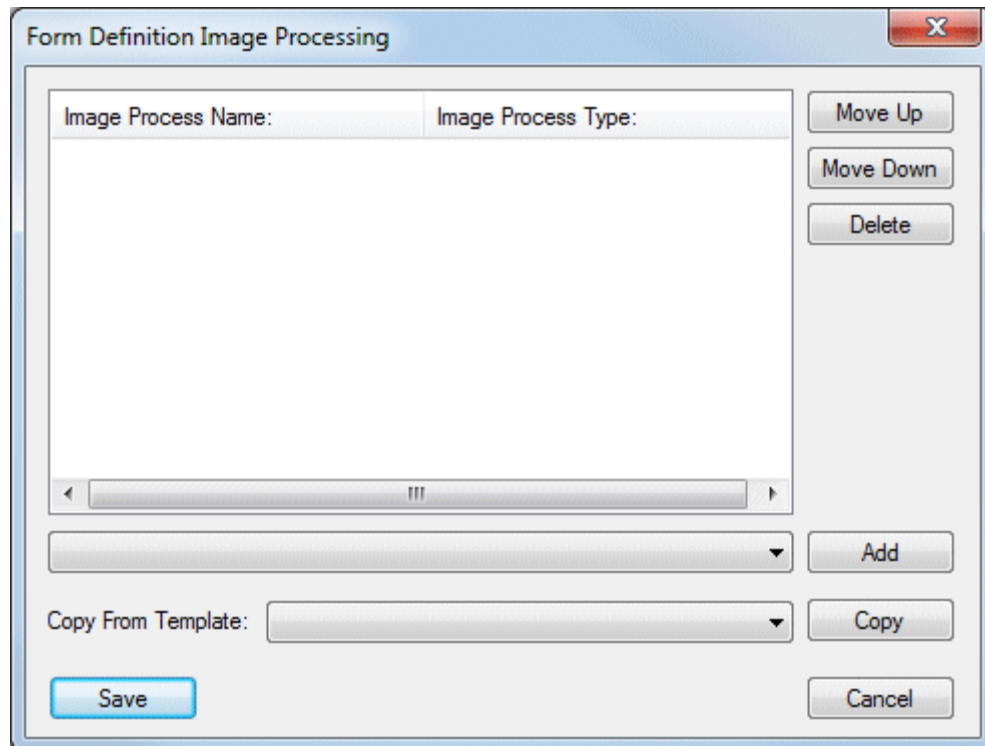


The **Form Definition** dialog box is shown with the following settings:

- Name:** AC Transcript Form
- Assigned Document Type:** <Document Type Unchanged>
- ☐ Documents require subsequent manual indexing
- ☐ Carry keywords onto subsequent documents
- ☒ Create XML data rendition
- ☐ Create revision of single existing matching document
- Require Autofill Expansion:** (dropdown menu)
- ☐ Auto-burn redactions
- ☐ Default for identification chain
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None> (dropdown menu)
- Document Types Allowed for Form:**
 - HR - Transcript (selected)
- Remove** (button)
- Add** (button)
- Configuration Note:** (text area)
- Indexing Note:** (text area)
- Scan Queue/Keyword Rules** (button)
- Keyword Lookup and Replace** (button)
- Configuration User Groups** (button)
- Form Definition Lookup Text** (button)
- Image Processing Operations** (button)
- Save** (button)
- Cancel** (button)

Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Advanced Capture form.

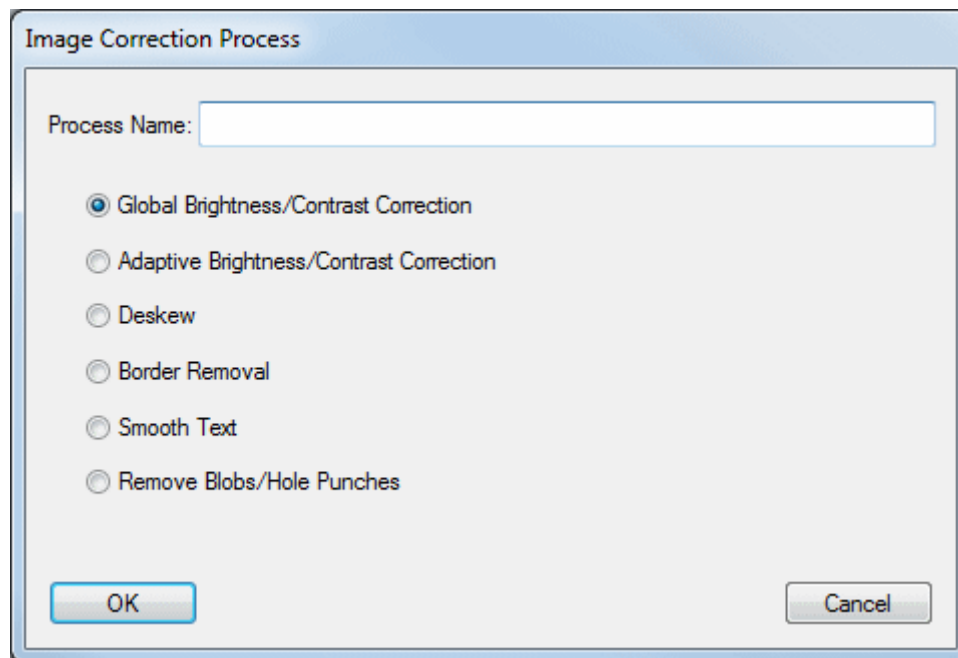
4. Click **Image Processing Operations**. The **Form Definition Image Processing** dialog box is displayed.



5. Using the **Image Process Type** drop-down list, select the type of image processing operation to be performed:
 - **Cleanup operations**
 - **Color removal**
 - **Line removal**
6. Click **Add**. A dialog box with options relating to your selection is displayed. For more information on these options, see the sections below.
7. Once you have named and configured the image processing operation, it is added to the list of operations at the top of the **Form Definition Image Processing** dialog box.
 - To reorder the listed operations, select the appropriate operation and use the **Move Up** and **Move Down** buttons. When processed, the operations will be performed in the order listed here.
 - To delete an operation from the list, select the operation and click **Delete**.
8. If you wish to copy a pre-configured list of image processing operations from another Advanced Capture form and apply it to the current form, use the **Copy From Template** drop-down list to select the appropriate form and click **Copy**. If you already added operations to the list for the current form, you will be prompted to confirm the removal of these operations before copying the full list from the other form.
9. When finished configuring image processing operations, click **Save**.

Cleanup Operations

When configuring image processing operations, if you select the **Cleanup operations** image process type and click **Add**, the **Image Correction Process** dialog box is displayed.



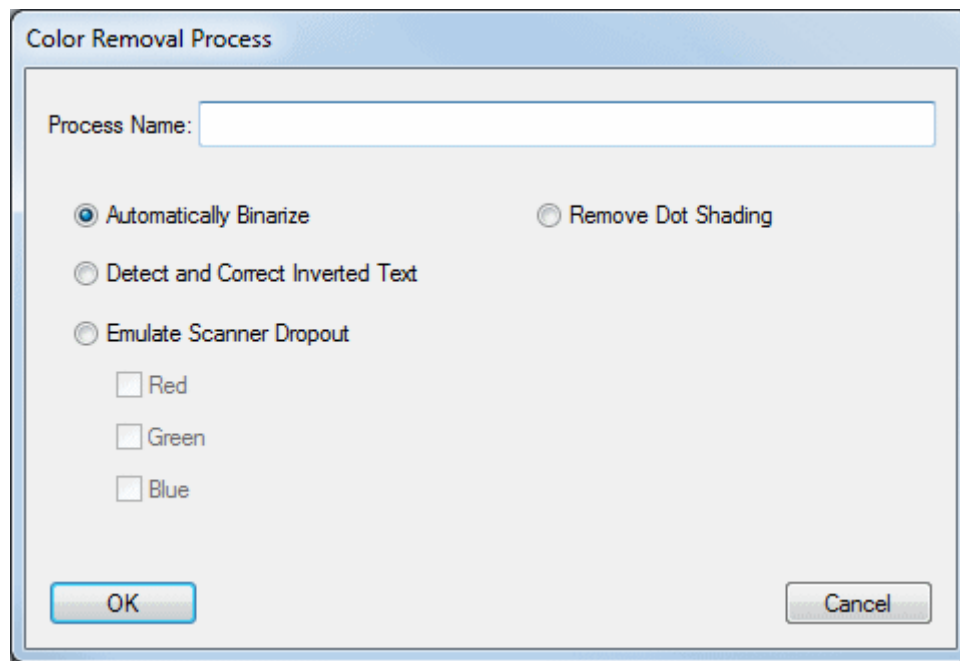
Type a descriptive name for the process in the **Process Name** field, and select the desired process:

- **Global Brightness/Contrast Correction:** Corrects brightness/contrast uniformly to help bring out faint text. This process is useful for images with low background variance.
- **Adaptive Brightness/Contrast Correction:** Corrects brightness/contrast adaptively to help bring out faint text. This process is useful for images with high background variance.
- **Deskew:** Straightens and scales misaligned images.
- **Border Removal:** Removes black borders from images.
- **Smooth Text:** Corrects text with pitted or eroded characters.
- **Remove Blobs/Hole Punches:** Removes larger marks from images.

When finished, click **OK** to save this process and add it to the list of configured operations in the **Form Definition Image Processing** dialog box.

Color Removal

When configuring image processing operations, if you select the **Color removal** image process type and click **Add**, the **Color Removal Process** dialog box is displayed.

The image shows a dialog box titled "Color Removal Process". It has a "Process Name:" label followed by a text input field. Below this, there are four radio button options: "Automatically Binarize" (which is selected), "Detect and Correct Inverted Text", "Emulate Scanner Dropout", and "Remove Dot Shading". Under the "Emulate Scanner Dropout" option, there are three checkboxes: "Red", "Green", and "Blue", all of which are currently unchecked. At the bottom of the dialog box, there are two buttons: "OK" and "Cancel".

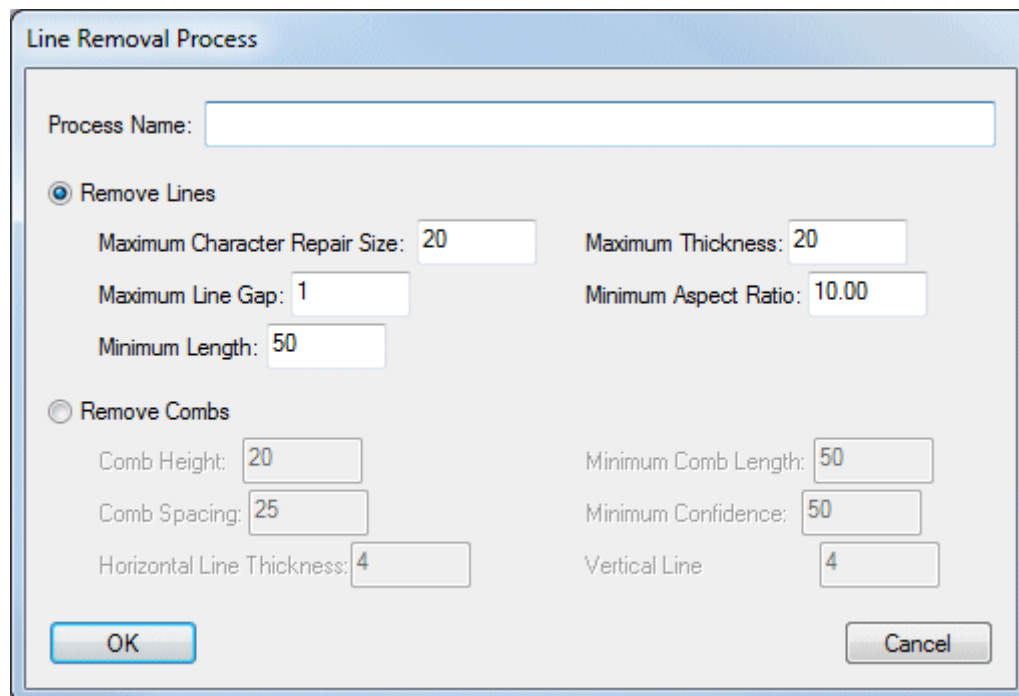
Type a descriptive name for the process in the **Process Name** field, and select the desired process:

- **Automatically Binarize:** Converts the image to black and white with adaptive thresholding.
- **Detect and Correct Inverted Text:** Detects white-on-black text and converts it to black-on-white text.
- **Emulate Scanner Dropout:** Removes the selected color(s) (that is, **Red**, **Green**, and/or **Blue**) from the image by converting the selected color(s) to white. Other colors in the image are converted to shades of gray or black.
- **Remove Dot Shading:** Removes dot shading areas behind text.

When finished, click **OK** to save this process and add it to the list of configured operations in the **Form Definition Image Processing** dialog box.

Line Removal

When configuring image processing operations, if you select the **Line removal** image process type and click **Add**, the **Line Removal Process** dialog box is displayed.



The dialog box is titled "Line Removal Process". It contains a "Process Name:" label followed by a text input field. Below this, there are two radio buttons: "Remove Lines" (which is selected) and "Remove Combs". Under "Remove Lines", there are four input fields: "Maximum Character Repair Size:" (20), "Maximum Thickness:" (20), "Maximum Line Gap:" (1), and "Minimum Length:" (50). Under "Remove Combs", there are four input fields: "Comb Height:" (20), "Minimum Comb Length:" (50), "Comb Spacing:" (25), and "Horizontal Line Thickness:" (4). There are also two input fields for "Minimum Aspect Ratio:" (10.00) and "Vertical Line" (4). At the bottom left is an "OK" button and at the bottom right is a "Cancel" button.

Option	Parameter	Value
Remove Lines (Selected)	Maximum Character Repair Size:	20
	Maximum Thickness:	20
	Maximum Line Gap:	1
	Minimum Length:	50
Remove Combs	Comb Height:	20
	Minimum Comb Length:	50
	Comb Spacing:	25
	Horizontal Line Thickness:	4
Other Parameters	Minimum Aspect Ratio:	10.00
	Vertical Line	4

Type a descriptive name for the process in the **Process Name** field, and select the desired process:

- **Remove Lines:** Removes lines from the image and attempts to reconstruct characters that intersect the removed lines. The default values for the following sub-options are recommended for general use.
 - **Maximum Character Repair Size:** Sets the maximum size (**0-100**) for character repair after line removal. The default value is **20**.
 - **Maximum Line Gap:** Sets the maximum gap (**0-20**) allowed in a line for the line to still be considered one object. The default value is **1**.
 - **Minimum Length:** Sets the minimum length (**0-2000** pixels) a line must be to meet the removal criteria. The default value is **50**.
 - **Maximum Thickness:** Sets the maximum thickness (**0-50** pixels) allowed for a line to meet the removal criteria. The default value is **20**.
 - **Minimum Aspect Ratio:** Sets the minimum ratio of the line length to the line width (**1.0-100.0**) required to meet the removal criteria. The default value is **10.0**.
- **Remove Combs:** Removes combs from the image and attempts to reconstruct characters that intersect the removed combs. The default values for the following sub-options are recommended for general use.
 - **Comb Height:** Sets the typical height (**4-1000** pixels) of a comb mark (i.e., the distance measured from the top of the horizontal base line to the top of the mark). The default value is **20**.
 - **Comb Spacing:** Sets the typical distance (**10-1000** pixels) between comb marks. The default value is **25**.
 - **Horizontal Line Thickness:** Sets the typical thickness (**1-20** pixels) of the horizontal lines that comprise the comb. The default value is **4**.
 - **Minimum Comb Length:** Sets the minimum distance (**10-5000** pixels) that a comb must extend (i.e., from the leftmost to the rightmost comb mark) to meet the removal criteria. The default value is **50**.
 - **Minimum Confidence:** Sets the minimum confidence (**0-100**) an individual comb must have to meet the removal criteria. The default value is **50**.
 - **Vertical Line:** Sets the typical thickness (**1-20** pixels) of the vertical lines that comprise the comb. The default value is **4**.

When finished, click **OK** to save this process and add it to the list of configured operations in the **Form Definition Image Processing** dialog box.

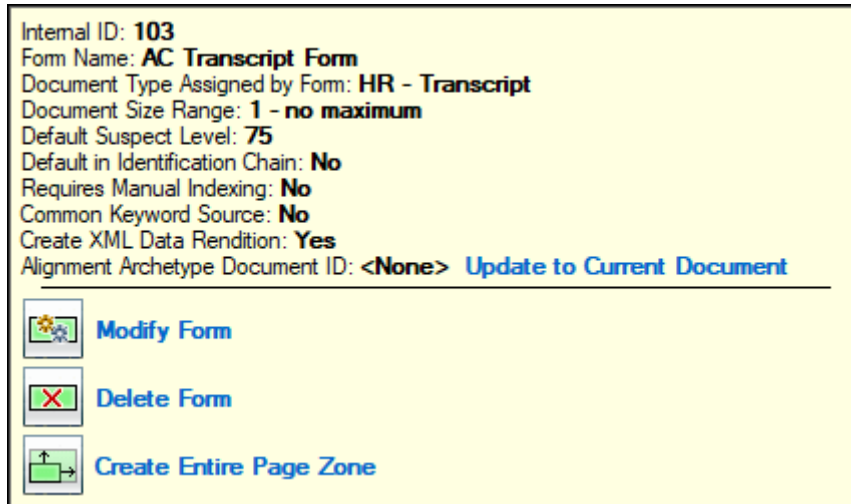
Restricting Form Modification Rights By User Group

By default, all users who have been granted the necessary rights to access the **Advanced Capture Configuration** window can modify an Advanced Capture form. However, you may restrict who can modify a particular Advanced Capture form by assigning specific User Groups to it.


Note: While assigning User Groups to a particular Advanced Capture form prevents members of the unassigned User Groups from modifying the form, the members of the unassigned groups can still view the form and initiate Advanced Capture processing.


To restrict form modification rights by User Group:


1. Create the Advanced Capture form as usual.
2. Once the form has been created, from the **Advanced Capture Configuration** window, right-click on the document displayed in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.



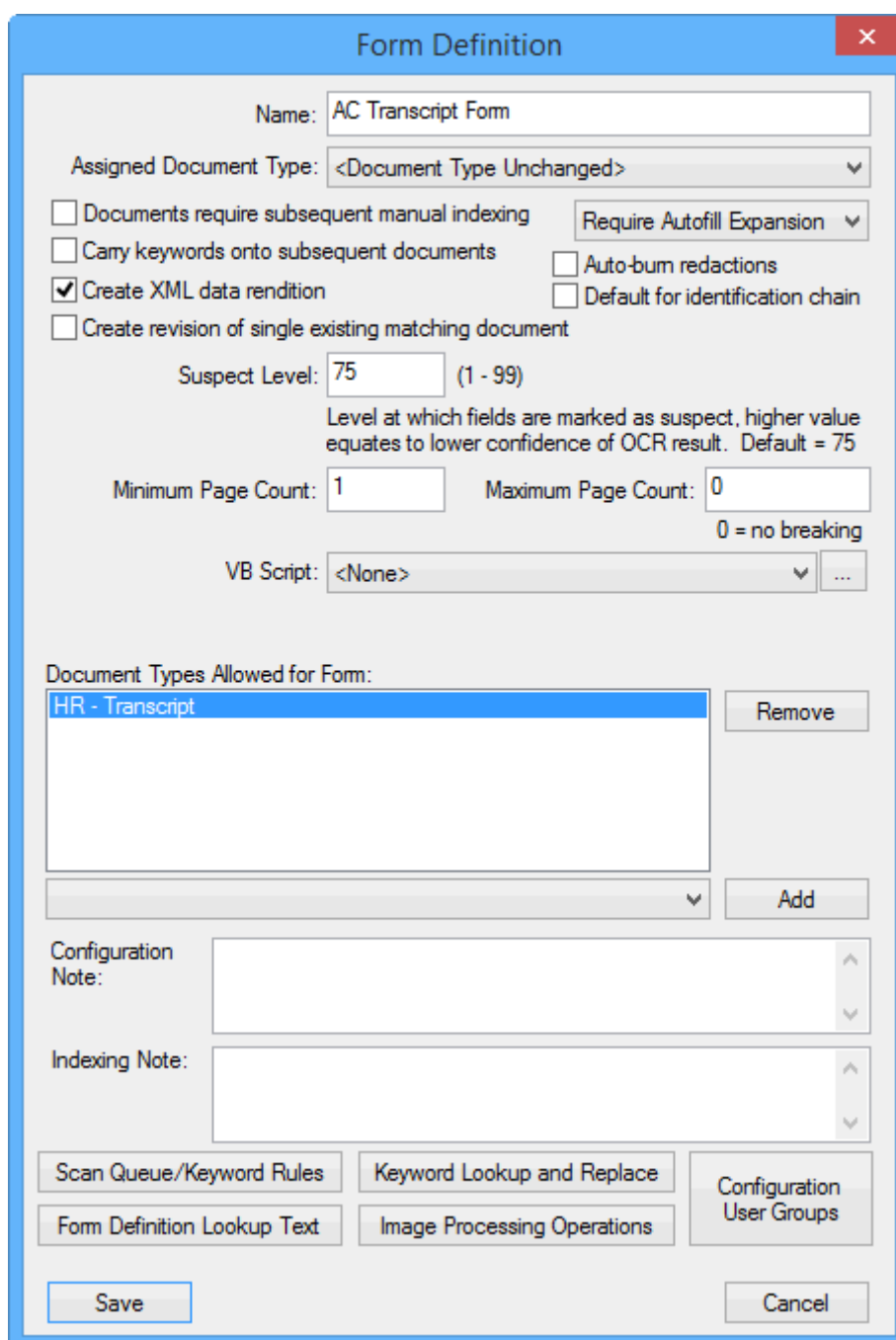
Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**

 **Delete Form**

 **Create Entire Page Zone**

- Click **Modify Form**. The **Form Definition** dialog box is displayed.

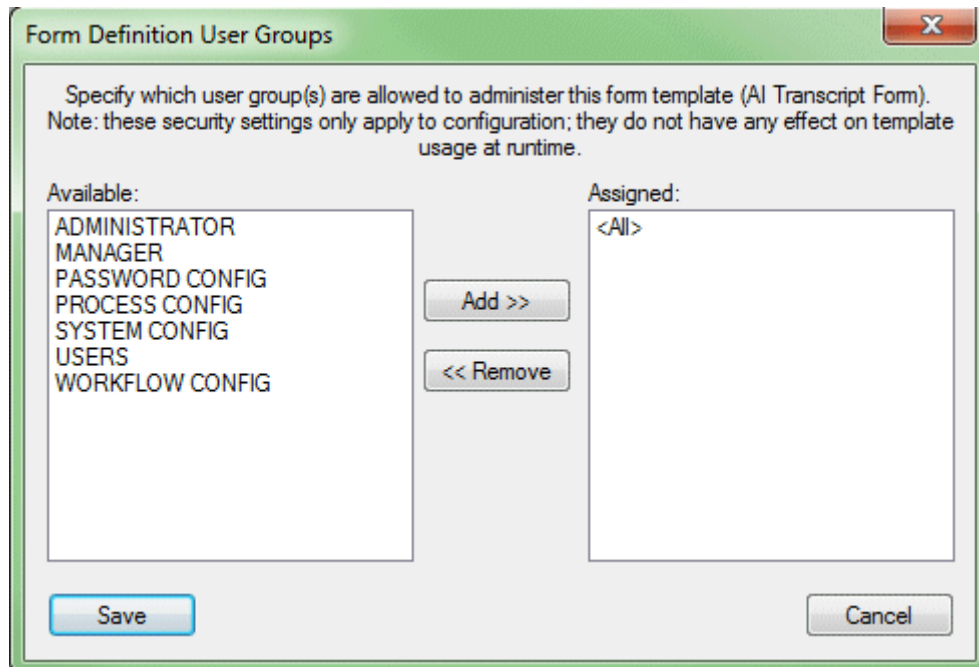


The **Form Definition** dialog box is shown with the following settings:

- Name:** AC Transcript Form
- Assigned Document Type:** <Document Type Unchanged>
- ☐ Documents require subsequent manual indexing
- ☐ Carry keywords onto subsequent documents
- ☒ Create XML data rendition
- ☐ Create revision of single existing matching document
- Require Autofill Expansion:** (dropdown menu)
- ☐ Auto-burn redactions
- ☐ Default for identification chain
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None> (dropdown menu)
- Document Types Allowed for Form:**
 - HR - Transcript (selected)
- Remove** (button)
- Add** (button)
- Configuration Note:** (text area)
- Indexing Note:** (text area)
- Buttons:**
 - Scan Queue/Keyword Rules
 - Keyword Lookup and Replace
 - Form Definition Lookup Text
 - Image Processing Operations
 - Configuration User Groups
- Save** (button)
- Cancel** (button)

Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Advanced Capture form.

4. Click **Configuration User Groups**. The **Form Definition User Groups** dialog box is displayed.



5. To assign a User Group to the form, do one of the following:
 - In the **Available** list, select the User Group and click **Add >>**. The User Group is moved to the **Assigned** list.
 - In the **Available** list, double-click the User Group. The User Group is moved to the **Assigned** list.
6. To remove an assigned User Group, do one of the following:
 - In the **Assigned** list, select the User Group and click **<< Remove**. The User Group is moved to the **Available** list.
 - In the **Assigned** list, double-click the User Group. The User Group is moved to the **Available** list.
7. When you have finished assigning User Groups to the form, click **Save**. The **Form Definition User Groups** dialog box is closed.

Showing or Hiding Image Processing Operations

Once you have configured image processing operations to be performed on documents matched to the Advanced Capture form, you can configure which image processing operations are displayed on the current document in the Document Viewer.

To show or hide different image processing operations on the current document, click the **Image Processing** button in the Forms toolbar and then select any of the following options from the menu:

Option	Description
Show Image Processing	Shows all image processing operations.
Hide Image Processing	Hides all image processing operations.
Edit Image Processing	Opens the Form Definition Image Processing dialog box, which allows you to configure the image processing operations for the form. For more information, see Configuring Image Processing Operations on page 87 .
Show Individual Image Process Only	Displays the individual image processing operations that have been configured for the form. Select the single image processing operation you want to show.

Updating the Archetype Document

The Advanced Capture form maintains a reference to the document that was used to align the form's configured zones. This document is called the form's archetype document. For a new Advanced Capture form, the archetype document is automatically set to the current document.

When a form's archetype document is set to another document, you can manually update the form's archetype to the current document by right-clicking on the document in the Document Viewer and then clicking **Update to Current Document**.

Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** **Update to Current Document**

Modify Form
Delete Form
Create Entire Page Zone

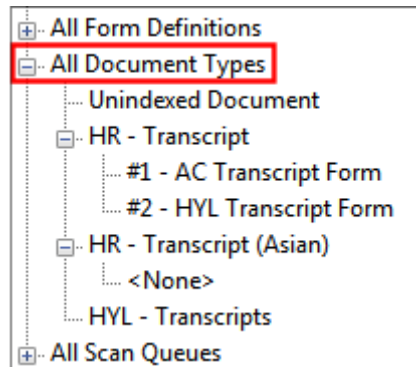
When viewing a form's configuration information in the Form Configuration pane, you can double-click on the archetype document ID to view the archetype document on top of the current document in the Document Viewer for diagnostics purposes (for example, to visually compare the original archetype document to see if the alignment zones are valid).

Modifying Form Configuration Options for a Document Type

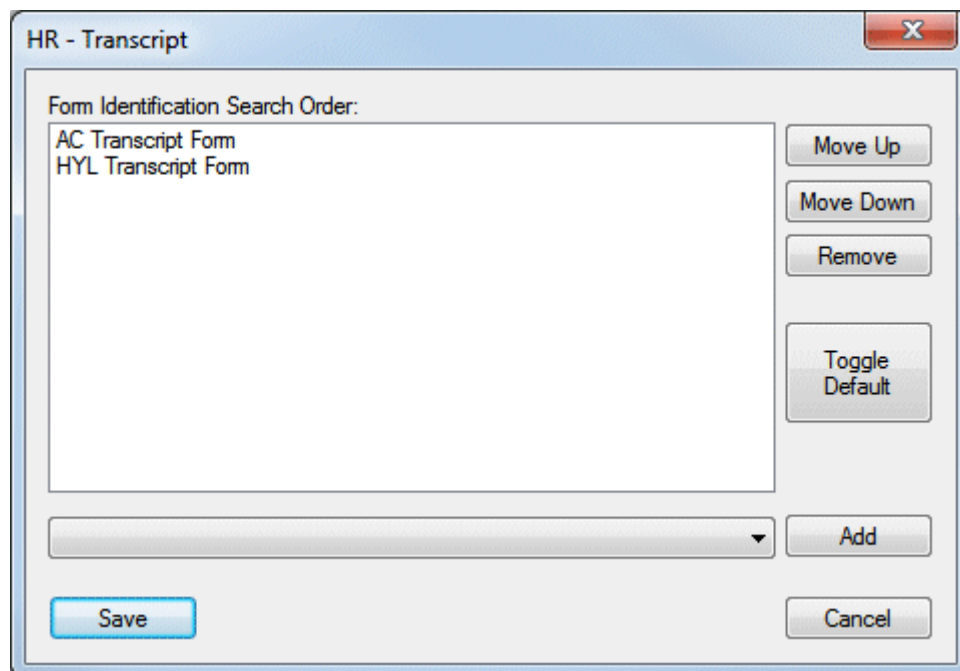
While reviewing information in the Form Configuration panel, you can access and modify the existing configuration options set for your Advanced Capture forms according to Document Type.

To modify the form configuration options set for a Document Type:

1. In the Form Configuration panel, expand the All Document Types list in the tree view.



2. Double-click the Document Type whose configuration options you wish to modify. The **<Document Type>** dialog box is displayed, with the Advanced Capture forms associated with the Document Type displayed in order of preference in the **Form Identification Search Order** list.



3. To add an available form to the list, select this form from the drop-down list and click **Add**.

4. To remove a form from the list, select this form in the list and click **Remove**.
5. To change a form's precedence order (i.e., to change the order in which the form is compared to a document in search of a match, with respect to the other configured forms in the list), select this form in the list and click **Move Up** or **Move Down**, as appropriate.
6. To change a form's default status, select this form in the list and click **Toggle Default**. When a form is configured as a default form for matching documents belonging to the associated Document Type, **[DEFAULT]** is appended to the form's name in the list.
7. To save your configuration modifications to the forms associated with the Document Type, click **Save**.
8. To cancel your configuration modifications, click **Cancel**.

For more information on configuring Advanced Capture forms, see [Defining Advanced Capture Forms on page 44](#).

Configuring a Form Identification Zone

Once the general Advanced Capture form options have been set in the **Form Definition** dialog box, a Form Identification Zone must be configured for the form.

A Form Identification Zone is used to detect the images or text values on a document that are unique to that document layout. Advanced Capture uses these images or text values to determine if the Advanced Capture form is a match to a document (i.e., if the image or text value is detected in the Form Identification Zone, then the document is determined to be a match to that Advanced Capture form).

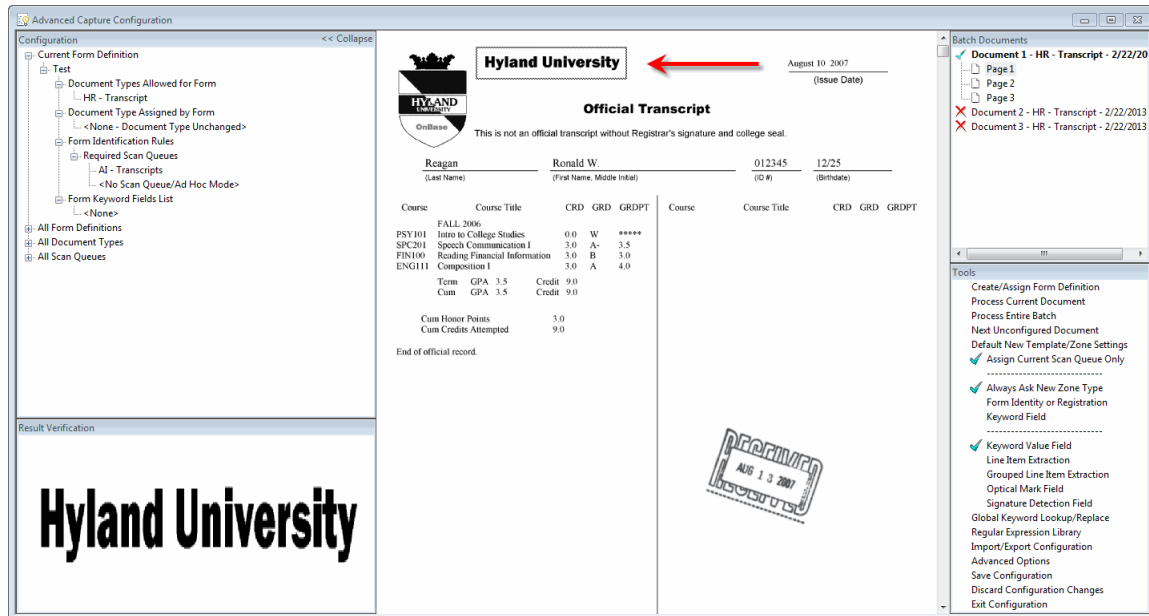
Tip: Multiple Form Identification Zones can be configured for a single Advanced Capture form. However, configuring multiple zones across different documents for the same form can reduce the accuracy of indexing results. Therefore, if you use multiple Form Identification Zones for the same form, it is recommended that you configure these zones on the same document.

Once a document is matched to a form, data values can be extracted from the document via the Advanced Capture process.

To create a Form Identification Zone:

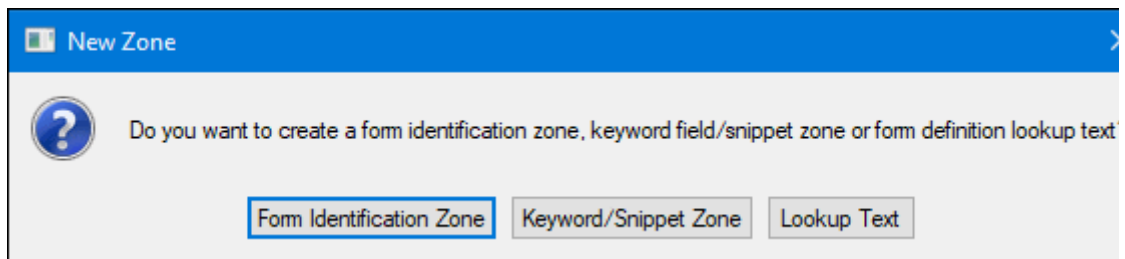
Note: Prior to creating the Form Identification Zone, ensure that the **Default Zone Type** option in the **Tools** panel is set to either **Always Ask** or **Form identity or registration**.

1. With the document that the Advanced Capture form is being created from displayed in the Document Viewer, click and hold the left mouse button and use the pointer to draw a box around the text or image that is to be used as the Form Identification value.

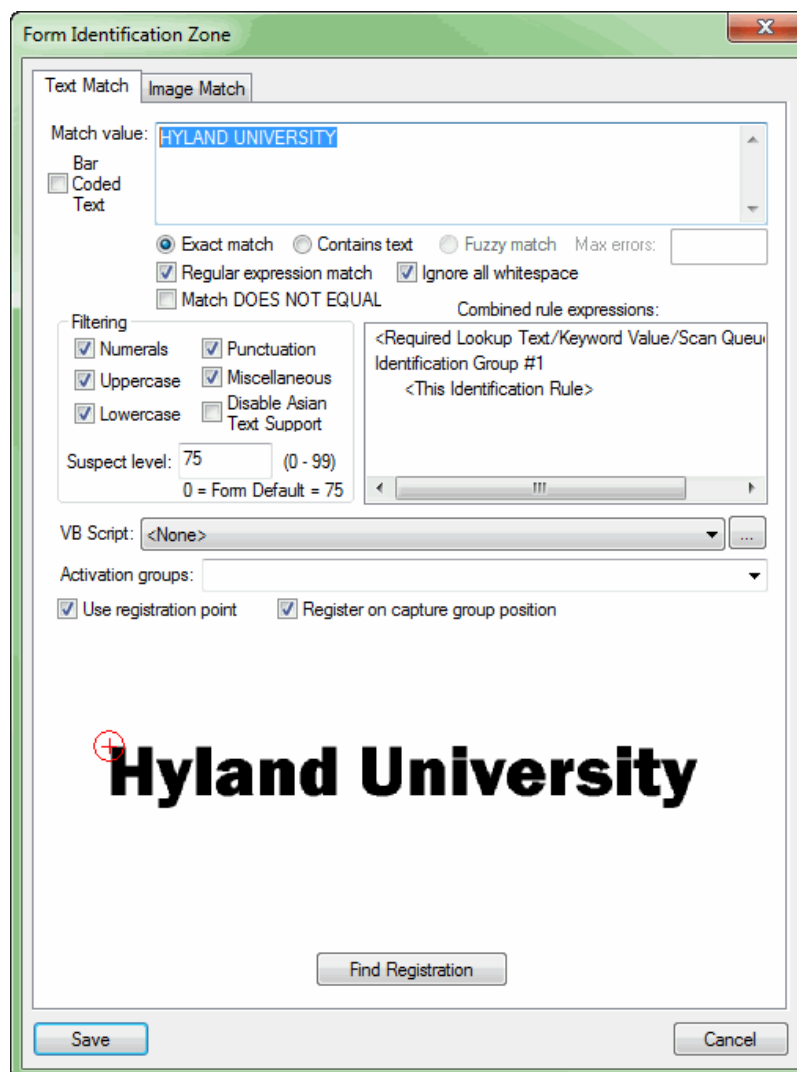


Tip: Note that the area of the document that you selected is displayed in the **Result Verification** panel.

2. Release the left mouse button to display the **Form Identification Zone** dialog box. Depending on the setting of the **Default Zone Type** option in the **Tools** panel, the **New Zone** dialog box may be displayed.



If the **New Zone** dialog box is displayed, click **Form Identification Zone** to continue to the **Form Identification Zone** dialog box.



3. Select the type of identifier you are using to match the Advanced Capture form to incoming documents.
 - If you are using a text value as an identifier to match the Advanced Capture form to a document, select the Text Match tab.
 - If you are using an image, such as a logo, as an identifier to match the Advanced Capture form to a document, select the Image Match tab.
4. Configure the Form Identification Zone using the **Form Identification Zone** dialog box.
5. Once you have set all necessary options on the **Form Identification Zone** dialog box, click **Save**. The **Form Identification Zone** dialog box is closed and the Form Identification Zone is highlighted in magenta on the document in the Document Viewer.



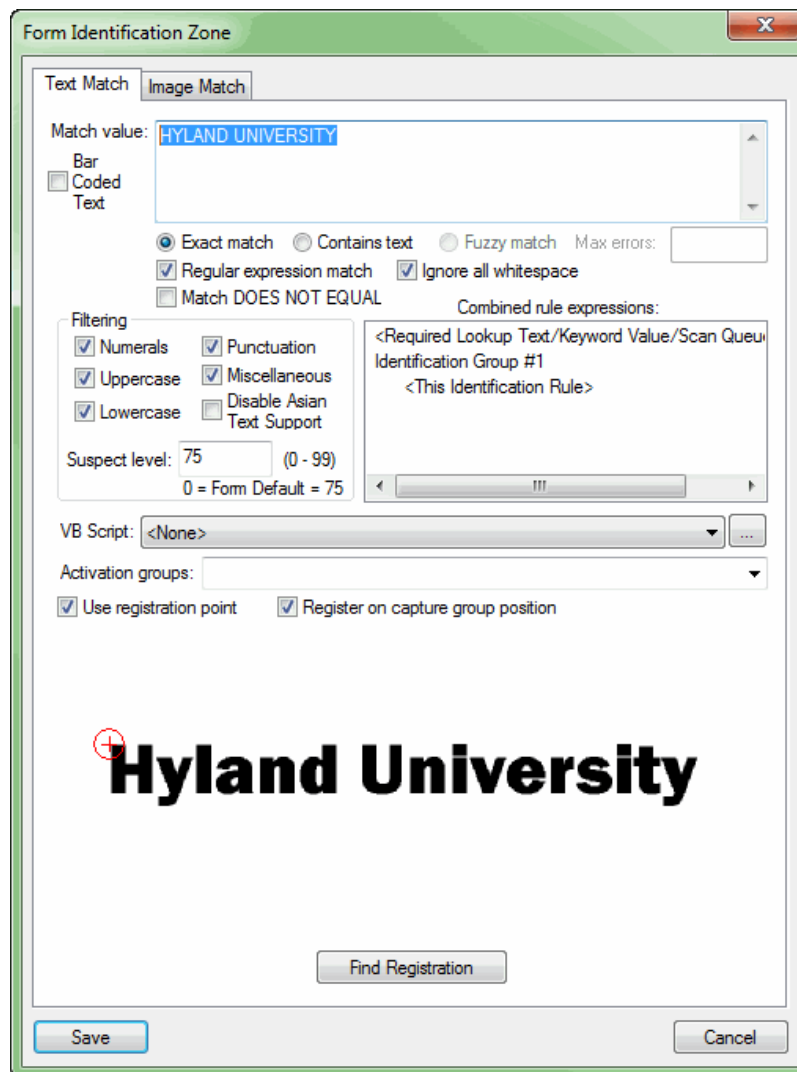
6. Repeat Steps 1-5 for each Form Identification Zone that you wish to create for the currently-displayed Advanced Capture form.

Tip: At any time, click **Save Configuration** in the **Tools** panel to save the form to the database. It is considered a best practice to save your Advanced Capture form configuration after the creation and configuration of every Advanced Capture form and each form's Form Identification, Registration, and Data Field Zones.

To discard all configuration changes (including any newly-created Form Identification Zones) made since the last time the form was saved, click **Discard Configuration Changes** in the **Tools** panel.

Configuring a Text Value Form Identifier

A text value form identifier is configured using the options on the Text Match tab of the **Form Identification Zone** dialog box.



The image shows the 'Form Identification Zone' dialog box with the 'Text Match' tab selected. The 'Match value' field contains 'HYLAND UNIVERSITY'. The 'Bar' checkbox is checked, while 'Coded' and 'Text' are unchecked. Under 'Exact match', 'Contains text', and 'Fuzzy match', the 'Exact match' radio button is selected. The 'Max errors' field is empty. The 'Regular expression match' checkbox is checked, and the 'Ignore all whitespace' checkbox is also checked. The 'Match DOES NOT EQUAL' checkbox is unchecked. In the 'Filtering' section, 'Numerals', 'Uppercase', and 'Lowercase' are checked, while 'Punctuation', 'Miscellaneous', and 'Disable Asian Text Support' are unchecked. The 'Suspect level' is set to 75, with a note '(0 - 99)' and '0 = Form Default = 75'. The 'VB Script' dropdown is set to '<None>'. The 'Activation groups' dropdown is empty. The 'Use registration point' and 'Register on capture group position' checkboxes are both checked. The 'Combined rule expressions' field contains the text: '<Required Lookup Text/Keyword Value/Scan Queue Identification Group #1>' and '<This Identification Rule>'. At the bottom, there is a large preview area showing the text 'Hyland University' with a red registration point icon (a circle with a crosshair) at the top left of the 'H'. Below the preview is a 'Find Registration' button. At the very bottom of the dialog are 'Save' and 'Cancel' buttons.

Form Identification Zone

Text Match Image Match

Match value: HYLAND UNIVERSITY

☒ Bar
☐ Coded
☐ Text

☒ Exact match ☐ Contains text ☐ Fuzzy match Max errors:

☒ Regular expression match ☒ Ignore all whitespace

☐ Match DOES NOT EQUAL

Filtering

☒ Numerals ☒ Punctuation
☒ Uppercase ☒ Miscellaneous
☒ Lowercase ☐ Disable Asian Text Support

Suspect level: 75 (0 - 99)
0 = Form Default = 75

VB Script: <None>

Activation groups:

☒ Use registration point ☒ Register on capture group position

Combined rule expressions:
<Required Lookup Text/Keyword Value/Scan Queue Identification Group #1>
<This Identification Rule>

Hyland University

Find Registration

Save Cancel

Form Identification Zone Text Match Options	Description
Match Value	<p>This field displays the text value returned from the OCR engine's evaluation of the Form Identification Zone.</p> <hr/> <p>Note: The Match Value is also displayed in the Result Verification panel below the enlarged image of the Form Identification Zone.</p> <hr/> <p>You may select the match rules for the Form Identification Zone by selecting one of the match rule radio buttons.</p> <hr/> <p>Tip: If the position of the text that is being compared to the Match Value shifts from document to document to the extent that a portion of this text might fall outside the Form Identification Zone, or different text might fall inside the zone, select the Contains text option to increase the chances of a desired match.</p> <hr/> <ul style="list-style-type: none"> • Exact match. Select this radio button if the value identified by the OCR engine when a document is processed must exactly match the Match Value in order to match the document to the Advanced Capture form. • Contains text. Select this radio button if the Match Value can be part of a longer text string identified by the OCR engine for the document being processed in order for Advanced Capture to match the document to the Advanced Capture form. <hr/> <p>Note: When using this option, only the Match Value portion of the text string is taken into consideration to determine the Suspect Level of the value.</p> <hr/> <ul style="list-style-type: none"> • Fuzzy match. Select this radio button if the Match Value should match the value detected by the OCR engine for the document, but some margin of error is allowed between the two values. <p>Enter a value in the Max Errors field to set the number of discrepancies allowed between the Match Value and the value read from the document being processed.</p> <p>For example, if your Match Value is HYLAND UNIVERSITY and the value detected by the OCR engine from the processed document is HYL4ND UNIVERS1TY, the document will still be matched to the Advanced Capture form if the Max Errors value is set to 2 or more.</p>

Form Identification Zone Text Match Options	Description
Bar Coded Text	<hr/> <p>Note: This option is only available if your solution is licensed for the Bar Code Recognition Server.</p> <hr/> <p>Select this check box if the Form Identification Zone contains a bar code. The bar code will be converted into text and then identified according to the zone's configured match rules.</p> <hr/> <p>Note: Depending on your configuration, the Advanced Capture engine will only search for certain bar code types in this zone. See Bar Code Types for more information.</p> <hr/>
Bar Code Types	<hr/> <p>Note: This option is only available when the Bar Coded Text option is selected.</p> <hr/> <p>Click this button to access the Select Bar Code Types dialog box.</p> <div data-bbox="628 894 1390 1413" data-label="Image"> </div> <p>Here you can select the types of bar codes for which the Advanced Capture engine will search when processing the Form Identification Zone. The engine will ignore any deselected bar code types when processing this zone. By default, all bar code types are selected.</p> <hr/> <p>Tip: To expedite Advanced Capture processing, select only the desired bar code types.</p> <hr/> <p>To save changes to your bar code type selections, click Save.</p>

Form Identification Zone Text Match Options	Description
Regular expression match	<p>Select this check box if the Match Value should match the form of a regular expression.</p> <p>For example: the regular expression rule for a Social Security Number is <code>\d{3}-\d{2}-\d{4}</code> (i.e., 3 digits, dash, 2 digits, dash, 4 digits). If the OCR engine identifies text matching this pattern in the zone, that text is returned as a Match Value.</p> <hr/> <p>Note: All regular expressions must be ECMA compliant.</p> <hr/>
Ignore all whitespace	<p>Select this check box if you would like to ignore all whitespace (e.g., spaces, paragraph returns, etc.) in the Match Value after it is read by the OCR engine.</p> <p>For example, if the Match Value read by the OCR engine is ENG 313 and the Ignore all whitespace check box is selected, the Match Value would be read as ENG313.</p>
Match DOES NOT EQUAL	<p>Select this check box if the document should be matched to the Advanced Capture form only when the Match Value does not appear in the Form Identification Zone.</p>

Form Identification Zone Text Match Options	Description
Allowed character types	<p>These options are available to assist the OCR engine in determining if the value it reads is correct. Select the check box next to each character type that is allowed in the Match Value. The available options are:</p> <ul style="list-style-type: none"> • Numerals. Numeric characters, 0-9. • Uppercase. Uppercase alphabetic characters. • Lowercase. Lowercase alphabetic characters. • Punctuation. Punctuation marks (i.e., . ! ?). • Miscellaneous. Other ASCII characters that do not fall into one of the above categories (i.e., # \$ * @). <p>When a character is recognized by the OCR engine that is not part of an allowable character set, the character is replaced by a tilde (~) and the value is automatically marked as suspect.</p> <hr/> <p>Tip: Using the Allowed character types options can sometimes help the OCR engine more easily determine the correct value by eliminating characters that are obviously not correct (e.g., an I is correctly identified instead of a 1 because numeric characters are filtered and prevented from being recognized as part of the value).</p> <hr/> <p>Select the Disable Asian Text Support check box to instruct the OCR engine to skip Asian (i.e., double-byte) characters.</p> <hr/> <p>Note: The Disable Asian Text Support check box is only available if the OCR format assigned to the Document Type is configured for an Asian language (e.g., Japanese, Korean, etc.).</p> <hr/> <p>Tip: Selecting the Disable Asian Text Support check box allows you to identify numeric data (i.e., Date Keyword Values, Currency Keyword Values, and Document Dates) when performing OCR on documents configured to contain Asian (i.e., double-byte) characters.</p> <hr/>

Form Identification Zone Text Match Options	Description
Suspect Level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured using this field. The default Suspect Level for the Advanced Capture form and the actual Suspect Level detected for the selected Form Identification Zone are displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p> <hr/> <p>Note: If the Form Identification Zone is marked suspect, it will not be used to determine if the form matches the document. If no other Form Identification Zones are configured for the form, or if all of the zones are marked suspect, the form will not be able to be matched to the document.</p> <hr/>

Form Identification Zone Text Match Options	Description
<p>Combined rule expressions</p>	<p>In the case where multiple Form Identification Zones are needed to match a document to a form, you can configure how these zones are processed using the Combined rule expressions option.</p> <p>Rules can be dragged-and-dropped into the proper order in this field and can be combined/evaluated using the AND or OR boolean operators.</p> <ul style="list-style-type: none"> Rules placed within the same Identification Group are combined using an AND operator. Different Identification Groups are evaluated using an OR operator. <hr/> <p>Note: Entire Identification Groups cannot be copied. However, individual rules can be copied to new or existing groups by right-clicking on the rule name and dragging it within the desired Identification Group.</p> <hr/> <p>A document is matched to a form when all of the rules within an Identification Group (i.e., an entire set of AND rules) are true. For example:</p> <div data-bbox="644 1016 1390 1310" data-label="Diagram"> <pre> graph TD subgraph IG1 [Identification Group #1] R1["(1.42, 0.23) inches : exact match 'HYLAND L'"] R2["(2.98, 0.93) inches : exact match 'OFFICIAL'"] R1 -- AND --> R2 end subgraph IG2 [Identification Group #2] R3["<This Identification Rule>"] R4["(2.98, 0.93) inches : exact match 'OFFICIAL'"] R3 -- AND --> R4 end subgraph IG3 [Identification Group #3] R5["(4.55, 5.68) inches : <image matching>"] R6["(2.98, 0.93) inches : exact match 'OFFICIAL'"] R5 -- AND --> R6 end IG1 -- OR --> IG2 IG2 -- OR --> IG3 </pre> </div> <p>In the example above, each Identification Group consists of two rules. For the document to be matched to the form, both rules within any one of the three Identification Groups must be true. If any rule in Identification Group #1 is false, the rules in Identification Group #2 are evaluated, and so on.</p>
<p>VB Script</p>	<p>Use the VB Script drop-down list to select a VB script to associate with the identification of this Advanced Capture form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>

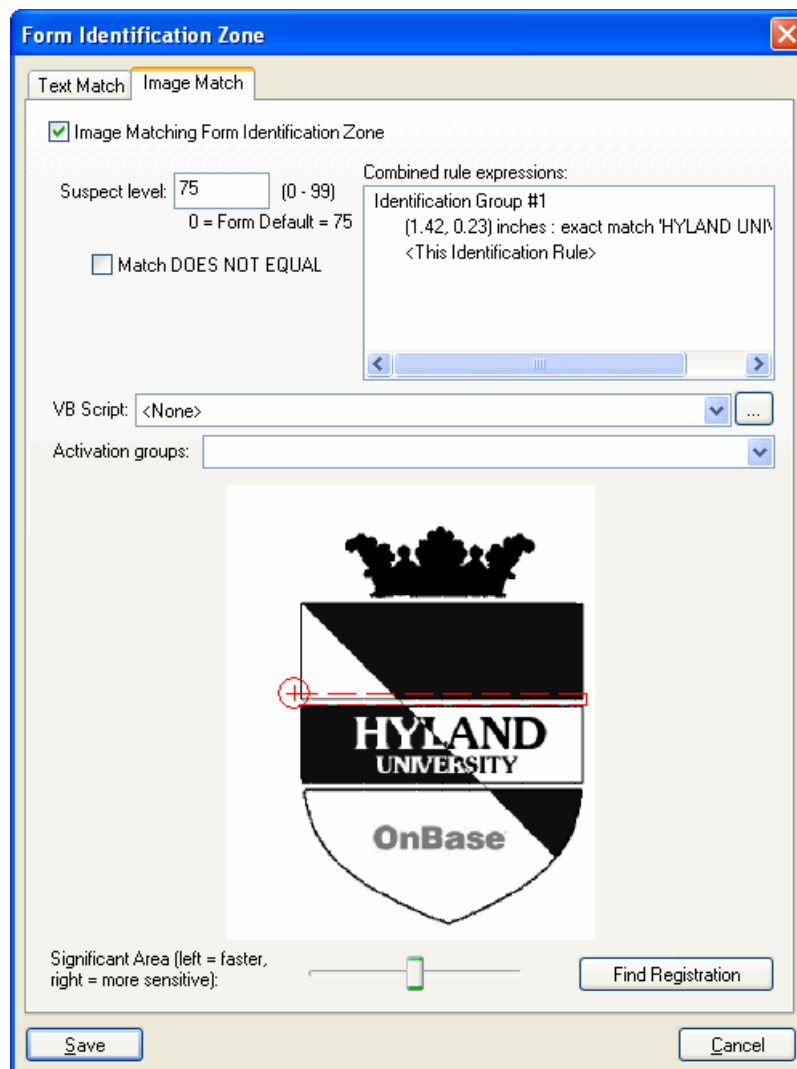
Form Identification Zone Text Match Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Form Identification Zone Text Match Options	Description
Use registration point	<p>Select this check box to enable the Registration Point feature.</p> <hr/> <p>Note: The Registration Point is the starting point (upper-left corner) of the value detected in the Form Identification Zone. Click Find Registration to automatically set the Registration Point at the starting point of the value in the Form Identification Zone, or double-click a location in the field below the check box to manually set the Registration Point at that position.</p> <hr/> <p>The position of the Registration Point on the Advanced Capture form is compared to the same position on the document being processed to determine the offset (i.e., the skew or rotation) of any imported documents. The position of the configured Form Identification Zone is adjusted on the document being processed to account for the detected offset to ensure that the Advanced Capture process is able to process the document properly.</p> <ul style="list-style-type: none">• If you selected the Regular expression match check box for the Match Value, the Register on capture group position check box will be enabled. Select this option and click Find Registration to set the Registration Point to the starting point of the first capture group contained within the Match Value's regular expression. This option is useful when the Match Value's first capture group is not the first value detected within the Form Identification Zone.

Configuring an Image Match Form Identifier

An image identifier is configured using the options on the Image Match tab of the **Form Identification Zone** dialog box.

Note: On a scaled document, the location of the Registration Point can become skewed when using an image match form identifier. As a result, the document might not be matched to the form. Therefore, when configuring a form to which scaled documents will be matched, it is recommended that you limit the use of image match form identifiers and that you configure at least one Identification Group that uses only text and/or regular expression Form Identification Zones with Registration Points set.



Form Identification Zone Image Match Options	Description
Image Matching Form Identification Zone	<p>Select this check box to configure this Form Identification Zone to use an image identifier. By default, Form Identification Zones are configured to use text values as identifiers.</p> <p>In order to match a document to this Advanced Capture form, the OCR engine must find the image displayed in this Form Identification Zone in the same location on the document being processed.</p> <p>This check box must be selected in order to enable the other options on the Image Match tab of the Form Identification Zone dialog box.</p>

Form Identification Zone Image Match Options	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured using this field. The default Suspect Level for the Advanced Capture form is displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p> <hr/> <p>Note: If the Form Identification Zone is marked suspect, it will not be used to determine if the form matches the document. If no other Form Identification Zones are configured for the form, or if all of the zones are marked suspect, the form will not be able to be matched to the document.</p> <hr/>
Match DOES NOT EQUAL	<p>Select this check box if the document should be matched to the Advanced Capture form only when the image displayed does not appear in the Form Identification Zone.</p>

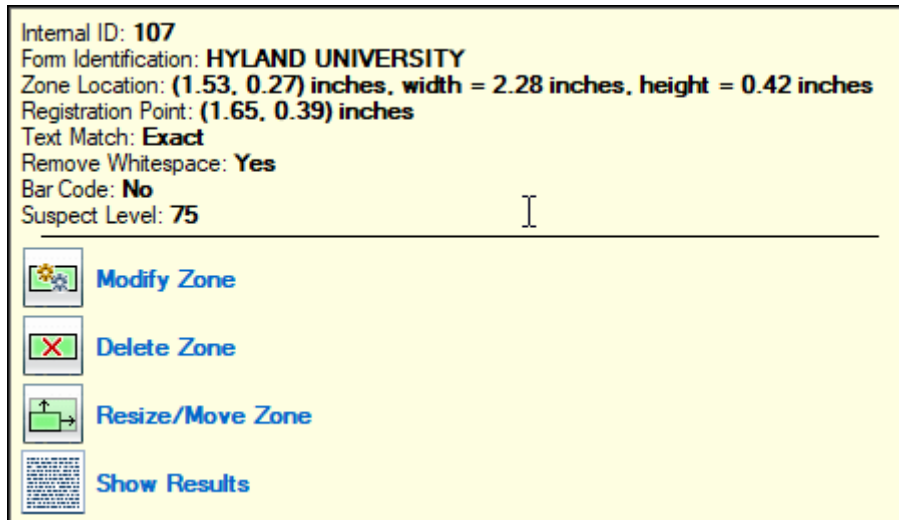
Form Identification Zone Image Match Options	Description
<p>Combined rule expressions</p>	<p>In the case where multiple Form Identification Zones are needed to match a document to a form, you can configure how these zones are processed using the Combined rule expressions option.</p> <p>Rules can be dragged-and-dropped into the proper order in this field and can be combined/evaluated using the AND or OR boolean operators.</p> <ul style="list-style-type: none"> Rules placed within the same Identification Group are combined using an AND operator. Different Identification Groups are evaluated using an OR operator. <hr/> <p>Note: Entire Identification Groups cannot be copied. However, individual rules can be copied to new or existing groups by right-clicking on the rule name and dragging it within the desired Identification Group.</p> <hr/> <p>A document is matched to a form when all of the rules within an Identification Group (i.e., an entire set of AND rules) are true. For example:</p> <div data-bbox="644 1016 1390 1310" data-label="Diagram"> <pre> graph TD subgraph IG1 [Identification Group #1] R1["(1.42, 0.23) inches : exact match 'HYLAND L'"] R2["(2.98, 0.93) inches : exact match 'OFFICIAL'"] R1 -- AND --> R2 end subgraph IG2 [Identification Group #2] R3["<This Identification Rule>"] R4["(2.98, 0.93) inches : exact match 'OFFICIAL'"] R3 -- AND --> R4 end subgraph IG3 [Identification Group #3] R5["(4.55, 5.68) inches : <image matching>"] R6["(2.98, 0.93) inches : exact match 'OFFICIAL'"] R5 -- AND --> R6 end IG1 -- OR --> IG2 IG2 -- OR --> IG3 </pre> </div> <p>In the example above, each Identification Group consists of two rules. For the document to be matched to the form, both rules within any one of the three Identification Groups must be true. If any rule in Identification Group #1 is false, the rules in Identification Group #2 are evaluated, and so on.</p>
<p>VB Script</p>	<p>Use the VB Script drop-down list to select a VB script to associate with the identification of this Advanced Capture form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>

Form Identification Zone Image Match Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).





Form Identification Zone Image Match Options	Description
Significant Area	<p>Use the slider to control the size of the significant processing area. The significant processing area is the red box displayed on the image snippet.</p> <p>Increasing the size of the significant processing area (i.e., moving the slider to the right) will increase the accuracy of the image match process, but will decrease the speed of the process.</p> <p>Decreasing the size of the significant processing area (i.e., moving the slider to the left) will increase the speed of the image match process, but will decrease the accuracy of the process.</p>
Find Registration	<p>The Registration Point feature is required when using an image identifier.</p> <p>The Registration Point is the starting point (upper-left corner) of the image's significant processing area (the red box displayed on the image snippet). It is used to determine the page's offset (i.e., any skew or rotation) that might have occurred during scanning.</p> <p>The position of the Registration Point on the Advanced Capture form is compared to the same position on the document being processed. The position of the configured Form Identification Zone is adjusted on the document being processed to account for the detected offset to ensure that the Advanced Capture process is able to process the document properly.</p> <p>Click Find Registration to automatically set the Registration Point, or double-click a location on the image to manually set the registration point at that position.</p>

Modifying or Analyzing a Form Identification Zone

An existing Form Identification Zone can be modified, deleted, or resized/moved by right-clicking on it in the Document Viewer. If multiple zones overlap, the **Previous** and **Next** links can be used to select the desired zone.



Internal ID: **107**
 Form Identification: **HYLAND UNIVERSITY**
 Zone Location: **(1.53, 0.27) inches, width = 2.28 inches, height = 0.42 inches**
 Registration Point: **(1.65, 0.39) inches**
 Text Match: **Exact**
 Remove Whitespace: **Yes**
 Bar Code: **No**
 Suspect Level: **75**

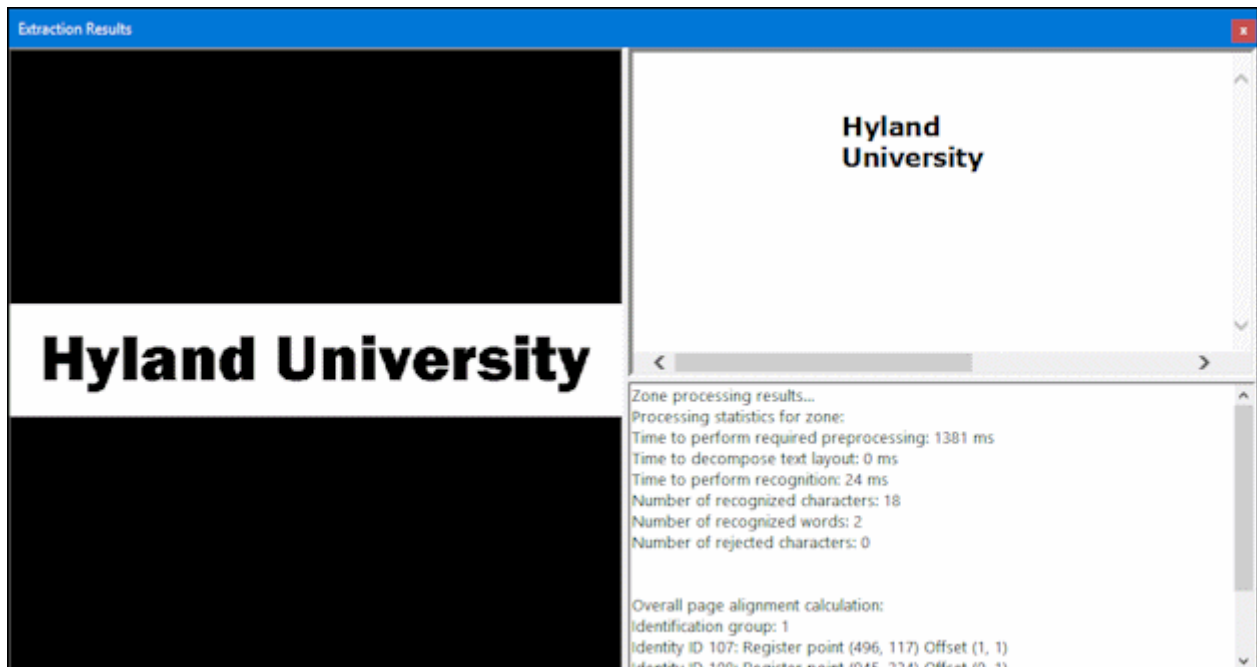
 **Modify Zone**
 **Delete Zone**
 **Resize/Move Zone**
 **Show Results**

Information about the selected zone is displayed above the following options:

- **Modify Zone.** Click to display the **Form Identification Zone** dialog box to modify any of the Form Identification Zone's configuration settings.
- **Delete Zone.** Click to delete the selected Form Identification Zone.
- **Resize/Move Zone.** Click to make the Form Identification Zone editable in the Document Viewer. It can be resized and/or moved without changing any of its existing configuration settings using either the pointer or the following shortcut keys:

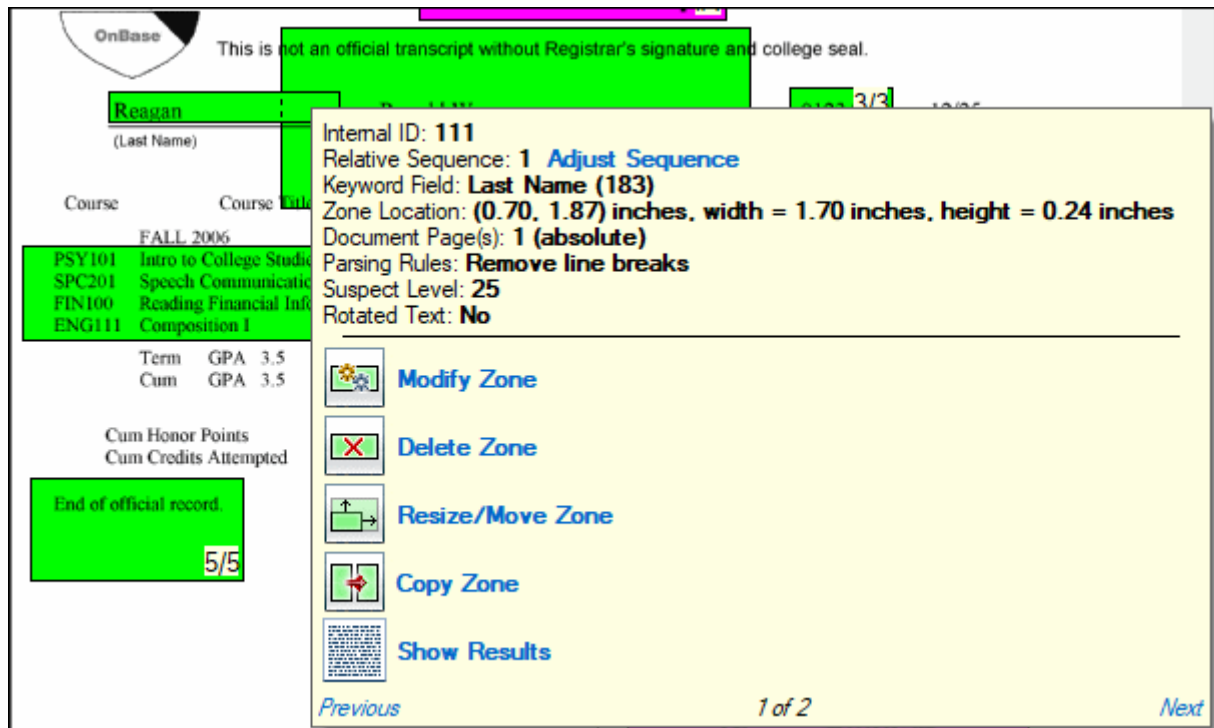
Shortcut Key	Action
Up / Down / Left / Right	Moves the zone one pixel in the specified direction.
Shift + Up / Down / Left / Right	Expands the zone by one pixel in the specified direction.
Ctrl + Up / Down / Left / Right	Shrinks the zone by one pixel in the specified direction.
Space / Enter	Saves the changes made to the size and position of the zone.
Esc	Cancels the changes made to the size and position of the zone.

- **Show Results.** Click to view the **Extraction Results** window, which displays the extraction area of the Form Identification Zone on the left, the text or image extracted from the zone in the upper right, and diagnostics information about the extraction process in the lower right.



Modifying or Analyzing Overlapping Zones

Depending on your document layout, zones may overlap. Overlapping zone borders are displayed with a dotted line.



When right-clicking on an area of the document with multiple zones, the name of the currently selected zone is displayed at the top of the dialog box. To switch between zones, click the **Previous** and **Next** links; once the desired zone is selected, it can be modified, deleted, resized/moved, copied, or analyzed.

Configuring A Data Field Zone

Data Field Zones are used to identify, extract, and assign Keyword Values to the document.

Data Field Zones can identify Keyword Values in a number of ways:

- By extracting text from the document and assigning that text as a Keyword Value.
- By performing OMR or signature detection and assigning pre-configured Keyword Values based on whether a mark or a signature is present.

To create a Data Field Zone:

Note: Prior to creating a Data Field Zone, ensure that the **Default Zone Type** option in the **Tools** panel is set to either **Always Ask** or **Keyword field**.

1. With the document that the Advanced Capture form is being created from displayed in the Document Viewer, click and hold the left mouse button and use the pointer to draw a box around the value that is to be extracted/evaluated.

Hyland University

August 10, 2007
(Issue Date)

Official Transcript

This is not an official transcript without Registrar's signature and college seal.

Reagan Ronald W.
(Last Name) (First Name, Middle Initial)

012345 12/25
(ID #) (Birthdate)

Course	Course Title	CRD	GRD	GRDPT
FALL 2006				
PSY101	Intro to College Studies	0.0	W	*****
SPC201	Speech Communication I	3.0	A-	3.5
FIN100	Reading Financial Information	3.0	B	3.0
ENG111	Composition I	3.0	A	4.0
Term GPA: 2.5 Credit: 0.0				

Tip: Note that the area of the document that you selected is displayed in the **Result Verification** panel.

2. Release the left mouse button to display the **Data Field Zone** dialog box.
Depending on the setting of the **Default Zone Type** option in the **Tools** panel, the **New Zone** dialog box may be displayed.

New Zone

Do you want to create a form identification zone, keyword field/snippet zone or form definition lookup text?

Form Identification Zone Keyword/Snippet Zone Lookup Text

3. If the **New Zone** dialog box is displayed, click **Keyword/Snippet Zone** to continue to the **Data Field Zone** dialog box.

Data Field Zone

Optical Mark Field Signature Detection Field Snippet

Text Field Line Item Extraction Grouped Line Item Extraction

Keyword type: <None> XML Node:

Data Application

- ☒ Entire zone ☐ Remove line breaks
- ☐ Specific line(s):
- ☐ Table style/vertical list
- ☐ Find by tag/expression
- ☐ Person name extraction
- ☐ Address extraction
- ☐ Remove all whitespace from result

Page Location(s)

- ☒ Absolute page: 1
 - Field only appears on a single page; always the same absolute page number (one-based)
- ☐ Relative page(s):
 - ☐ First page
 - ☐ First page (not only page)
 - ☐ Interior pages
 - ☐ Last page
 - ☐ Last page (not only page)

☐ Delete this page from document

☐ Redact zone from document after keyword assignment

Filtering

- ☒ Numerals ☒ Punctuation
- ☒ Uppercase ☒ Miscellaneous
- ☒ Lowercase

Advanced filtering:

VB script: <None> ...

Suspect level: 25 (0-99) 0 = Form Default (25), Actual = 20

- ☐ Add implied decimal point
- ☐ Remove text clipped by zone boundaries
- ☐ Find/validate in keyword dataset

Specific allowed characters:

Advanced recognition: Machine print text/standard OCR

Keyword association: Activation groups: Colors

Save Cancel

4. Select the type of data the OCR engine needs to process in order to identify, extract, and assign Keyword Values to the document being processed.
 - If the OCR engine is reading a text value from the image document and assigning that value as a Keyword Value, select the Text Field tab.
 - If the OCR engine is going to be reading multiple Keyword Values in a table format, select the Line Item Extraction tab.
 - If the OCR engine is going to be reading multiple Keyword Values from a single zone when line item data is separated into multiple groups on the document, select the Grouped Line Item Extraction tab.

- If the OCR engine is going to be performing optical mark recognition and assigning a pre-configured Keyword Value based on the presence or lack of an optical mark, select the Optical Mark tab.
 - If the OCR engine is going to be performing signature detection and assigning a pre-configured Keyword Value based on the presence or lack of a signature, select the Signature Detection tab.
5. Configure the Data Field Zone using the **Data Field Zone** dialog box.
 6. Once you have set all necessary options on the **Data Field Zone** dialog box, click **Save**. The **Data Field Zone** dialog box is closed, and the Data Field Zone is temporarily saved to the Advanced Capture form.

The Data Field Zone is highlighted either in green (for a Keyword zone, as in the following image) or in hatched red (for a snippet zone) on the document in the Document Viewer.

Hyland University

August 10, 2007
(Issue Date)

Official Transcript

This is not an official transcript without Registrar's signature and college seal.

Reagan Ronald W. 012345 12/25
(Last Name) (First Name, Middle Initial) (ID #) (Birthdate)

Course	Course Title	CRD	GRD	GRDPT
FALL 2006				
PSY101	Intro to College Studies	0.0	W	0.0000
SPC201	Speech Communication I	3.0	A-	3.5
FIN100	Reading Financial Information	3.0	B	3.0
ENG111	Composition I	3.0	A	4.0
Term: CDA 2.5 Credit: 0.0				

7. Repeat steps 1-5 for each Data Field Zone that you wish to create for the currently-displayed Advanced Capture form.

Tip: At any time, click **Save Configuration** in the **Tools** panel to save the form to the database. It is considered a best practice to save your Advanced Capture form configuration after the creation and configuration of every Advanced Capture form and of each form's Form Identification, Page Registration, and Data Field Zones.

To discard all configuration changes (including all Data Field Zones created since the last time the form was saved), click **Discard Configuration Changes** in the **Tools** panel.

Configuring a Text Data Field Zone

A Text Data Field Zone is configured using the options on the Text Field tab of the **Data Field Zone** dialog box.

The screenshot shows the 'Data Field Zone' dialog box with the 'Text Field' tab selected. The dialog is divided into several sections for configuration.

Tab Selection: Optical Mark Field, Signature Detection Field, Snippet, **Text Field**, Line Item Extraction, Grouped Line Item Extraction.

Keyword type: <None> (dropdown) **XML Node:** (dropdown)

Data Application:

- ☒ Entire zone ☐ Remove line breaks
- ☐ Specific line(s): (text input)
- ☐ Table style/vertical list
- ☐ Find by tag/expression
- ☐ Person name extraction
- ☐ Address extraction
- ☐ Remove all whitespace from result

Page Location(s):

- ☒ Absolute page: (text input: 1)
 - Field only appears on a single page; always the same absolute page number (one-based)
- ☐ Relative page(s):
 - ☐ First page
 - ☐ First page (not only page)
 - ☐ Interior pages
 - ☐ Last page
 - ☐ Last page (not only page)

☐ Delete this page from document

☐ Redact zone from document after keyword assignment

Filtering:

- ☒ Numerals ☒ Punctuation
- ☒ Uppercase ☒ Miscellaneous
- ☒ Lowercase

Advanced filtering:

VB script: <None> (dropdown) ...

Suspect level: 25 (0-99) 0 = Form Default (25), Actual = 20

- ☐ Add implied decimal point
- ☐ Remove text clipped by zone boundaries
- ☐ Find/validate in keyword dataset

Specific allowed characters: (text input)

Advanced recognition: Machine print text/standard OCR (dropdown)

Keyword association: (dropdown) **Activation groups:** (dropdown) **Colors** (button)

Buttons: Save, Cancel

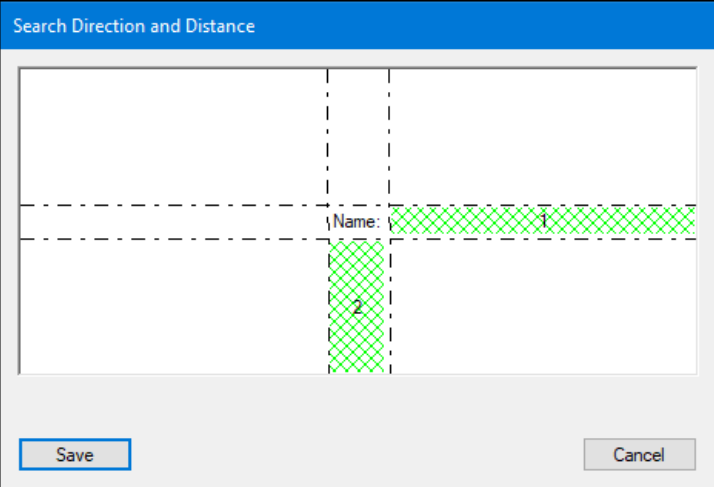
Data Field Zone Keyword Field Options	Description
Keyword type	<p>Using this drop-down list, select the Keyword Type that values extracted from the Data Field Zone are to be assigned to.</p> <p>If the currently-displayed form has an assigned Document Type, only the Keyword Types associated with the assigned Document Type are available in the drop-down list. If no Document Type is assigned to the currently-displayed form, only the <None> and >>Document Date options are available in the drop-down list.</p> <p>Some notes about selecting a Keyword Type:</p> <ul style="list-style-type: none"> • If the Keyword Value identified by the OCR engine exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is truncated to fit this length. • If you are configuring a Data Field Zone to capture a Date Keyword Value, the date format of the value you are attempting to capture must match the workstation's regional format. <p>If the value's date format does not match the workstation's regional format, then you must configure a VB script for the Data Field Zone to parse the value into the correct format.</p> <ul style="list-style-type: none"> • If you are configuring a Data Field Zone to capture a Date Keyword Value, the year portion of the value must be less than 100 or greater than 1900 to be considered valid (e.g., 01/01/11 or 01/01/2011). If the year does not fall within these ranges, the Date Keyword Value will be marked as suspect. <hr/> <p>Note: When the year portion of a Date Keyword Value consists of two digits, the Two Digit Year Settings configured for your solution override the workstation's regional settings for the pivot year. For more information on configuring the Two Digit Year Settings, see the System Administration module reference guide or help file.</p> <hr/> <ul style="list-style-type: none"> • If you are configuring a Data Field Zone to capture a Currency Keyword Value, the Add implied decimal point check box is displayed below the Suspect Level field. Select this check box to automatically insert a decimal point into a Currency Keyword Value if one is not found when the zone is processed. <hr/> <p>Note: When the Person name extraction Data Application option is selected, the Keyword type drop-down is replaced by the Last Name Keyword drop-down.</p> <hr/>

Data Field Zone Keyword Field Options	Description
XML Node	<p>Note: This field is only displayed if the Advanced Capture form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Advanced Capture form).</p> <p>Enter the name of the XML node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p> <p>Note: While the XML Node name may include alphanumeric characters, underscores (_), hyphens (-), or periods (.), it must begin with either a letter or an underscore.</p>

Data Field Zone Keyword Field Options	Description
Data Application	<p>Determine how the Data Field Zone is processed by selecting one of the Data Application radio buttons.</p> <p>Entire zone value: Select this option if all text detected in the Data Field Zone should be processed and assigned to the Keyword Value.</p> <p>To allow for the removal of line breaks in the Keyword Value, which might be present if multiple lines of text are displayed in the Data Field Zone, the Remove line breaks check box is selected by default. To prevent line breaks from being removed in the Keyword Value, deselect this option.</p> <p>Specific line(s): Select this option if multiple lines of text may be displayed in the Data Field Zone, but only the text from specific detected lines should be assigned to the Keyword Value. Enter the lines from which text is to be extracted into the field next to the option. Multiple line numbers and line ranges can be included in the field. Individual line numbers and line ranges must be separated by commas, and line ranges are denoted by a dash between the first included line and the last included line. For example, entering 1, 3-5, 7 into the field next to the option will assign values extracted from lines 1, 3, 4, 5, and 7 to the Keyword Value.</p> <p>Note the following:</p> <ul style="list-style-type: none"> • If Specific line(s) is selected and the Remove line breaks check box is selected, any values extracted from the specified lines will be concatenated into a single value and assigned to one instance of the Keyword. If the resulting Keyword Value exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is truncated to fit this length. • If Specific line(s) is selected and Remove line breaks is not selected, each line will be assigned as a separate instance of the Keyword. <p>Table style/vertical list: Select this option if multiple lines of text should be read as a table. Each line will be assigned as a separate instance of the Keyword.</p> <p>Multi-Instance Keyword Type Groups can be used with values read as a table to group related values together accordingly. Each Multi-Instance Keyword Type Group is constructed with Keyword Values in the order they were detected.</p> <p>For example: You have an invoice for multiple line items. You would like to extract indexing information using three Keyword Types (Course ID, Course Name, and Credits) for each individual line item, and you would like to keep the Keyword Values for each line item grouped together using Multi-Instance Keyword Type Groups.</p>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>In order to correctly identify and group the Keyword Values, you must create a separate Data Field Zone for each of the Keyword Types, and each zone must be configured to read each zone as a table. Each recognized Keyword Value will be associated with the others that compose the Multi-Instance Keyword Type Group in the order they were read (i.e., the first Class ID Keyword Value is grouped with the first Class Name and Credits Keyword Values, the second Class ID Keyword Value is grouped with the second Class Name and Credits Keyword Values, etc.)</p> <hr/> <p>Caution: In order to read Data Field Zones as a table to create Multi-Instance Keyword Type Groups, a value must exist for every Keyword Value in the Multi-Instance Keyword Type Group. If a value is missing, the subsequent values for the Keyword Type may be unintentionally assigned to the wrong Multi-Instance Keyword Type Group.</p> <hr/> <p>When the Entire zone value, Specific line(s), or Table style option is selected, the Remove text clipped by zone boundaries option is displayed in the Filtering section below. Select this option if you want to exclude from the Keyword Value any characters that straddle a boundary of the Data Field Zone.</p> <div data-bbox="778 1016 1241 1262"> </div> <hr/> <p>Note: The Remove text clipped by zone boundaries option only excludes characters that actually straddle boundaries of the zone. All whole characters in the zone, including those that stray into the zone from nearby text, will be included in the Keyword Value.</p> <hr/>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>Find by tag/expression: Select this option if you want to identify information to be stored as Keyword Values based on a specific tag or regular expression.</p> <hr/> <p>Note: When extracting Keyword Values using text tags or regular expressions, you may configure the entire page as a Data Field Zone instead of configuring individual Data Field Zones. For more information, see Creating an Entire Page Zone for Find by Tag/Expression on page 147.</p> <hr/> <p>When this option is selected, the Find near tag/expression and Find matching reg. expression options are displayed in the Filtering section (as in the following image), allowing you to define a text tag or regular expression to identify the Keyword Value to be extracted.</p> <div data-bbox="783 789 1236 1012" data-label="Form"> </div> <p>Select either the Find near tag/expression or the Find matching reg. expression radio button and enter the tag or the regular expression rule in the associated text field.</p> <hr/> <p>Note: To access the Regular Expression Library, click in the field and press F2. See The Regular Expression Library on page 247 for more information.</p> <hr/> <ul style="list-style-type: none"> • Find near tag/expression: The OCR engine searches the zone for a tag and extracts the next logical value (by default, either to the right of or underneath the tag) as a Keyword Value. Tags can be literal (that is, the text specified in the text box field is the tag) or regular expressions (that is, the text in the text box field defines a regular expression, and text matching that pattern is the tag). Literal text tags must be enclosed in quotation marks. The tag is case insensitive. However, if the case of the tag and document value match, the suspect level of the Keyword Value is lower than it would be if they did not match.

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>To change the default search direction for the tag, click Search Direction. The Search Direction and Distance dialog box is displayed.</p>  <p>The eight directional search areas are separated by dotted lines, with the tag in the center. The order of area search preference is indicated by ascending numerals, from 1 up to 4. When an area is selected to be considered in the search preference order, it is filled in hatched green with the ordinal numeral in the center. Click any area to select/deselect it.</p> <p>Tip: To reorder the preference for the selected areas, first deselect all of them and then select up to four areas in the appropriate order.</p> <p>Note: The area order preference only applies when the search identifies multiple hits roughly the same distance from the tag; otherwise, proximity to the tag automatically takes precedence.</p> <p>For entire page zones, you can also enter a maximum search distance value (0.25 to 20.00 inches) in the Maximum search distance from tag upper left corner field.</p> <ul style="list-style-type: none"> • Find matching reg. expression: The OCR engine searches the zone for text that matches the regular expression rule. If text matching the rule is found, it is extracted as a Keyword Value. <p>For example: the regular expression rule for a Social Security Number is <code>\d{3}-\d{2}-\d{4}</code> (that is, 3 digits, dash, 2 digits, dash, 4 digits). If the OCR engine identifies text matching this pattern in the zone, that text is stored as a Keyword Value.</p> <p>Note: All regular expressions must be ECMA compliant.</p>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>Select the Applied data format check box if you would like to use a regular expression to identify the Keyword Value, but you would like to apply a specific format to the extracted data. Enter the format in the text field, enclosing capture groups within brackets.</p> <p>For example:</p> <p>A document displays a date without separators (for example, Date: 01 01 2011). To capture this value and apply the appropriate date format (for example, DD/MM/YYYY), you would first enter the following regular expression in the Find matching reg. expression field, with parentheses enclosing the three capture groups: (\d\d)\s+(\d\d)\s+(\d\d\d\d). Then you would enter the date format in the Applied data format field, with brackets enclosing the capture groups: [1]/[2]/[3]. The Keyword Value can then be stored in the desired format (for example, 01/01/2011).</p> <hr/> <p>Note: When the Applied data format field is blank, the entire value captured in the Data Field Zone is shown in highlights and image snippets (that is, when indexing). When you specify capture groups in this field, however, only the matching capture groups in the Data Field Zone are shown in the highlights and image snippets.</p> <hr/> <p>Once you have entered the tag or regular expression rule in the associated text field, you can create variations or synonyms of this value to search for when the entered value is not found. Click the ... button to open the Search Tag Synonyms dialog box.</p> <div data-bbox="647 1224 1370 1835"> </div>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>To create a new search tag synonym, type the value in the text field and click Add. The synonym is added to the list in the top field. Once you have added multiple synonyms to the list, you can select them and then click the Move Up and Move Down buttons to reorder them for search precedence.</p> <p>Select the Find multiple occurrences check box if you would like to extract multiple Keyword Values from the zone (for example, multiple values in the zone meet the regular expression rule or the tag appears multiple times in the zone and you would like to extract each instance as a Keyword Value).</p> <p>Person name extraction: Select this option if you want the OCR engine to identify and extract any text in the zone that is formatted like a name.</p> <p>When this option is selected, the Person Name options are displayed in the Filtering section below, allowing you to define how the name should be identified and extracted as a Keyword Value.</p> <div data-bbox="616 884 1402 1098"> </div> <p>The following name formats are supported by default.</p> <ul style="list-style-type: none"> • Last, First [Middle] • LAST, FIRST [MIDDLE] • First [Middle] Last • FIRST [MIDDLE] LAST <hr/> <p>Note: A middle name value is optional.</p> <hr/> <p>If the name being extracted is displayed in the Last First [Middle] format, select the Format is Last First [Middle] with or without comma check box to distinguish it from other supported formats.</p>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>While the entire name is captured in the zone, the text is parsed into separate values (first name, middle name, and last name) and each value can be stored as an individual Keyword Value.</p> <ul style="list-style-type: none"> • Use the Last Name Keyword drop-down list to select the Keyword Type that the last name is assigned to. • Use the First Name Keyword drop-down list to select the Keyword Type that the first name is assigned to. • Use the Middle Name Keyword drop-down list to select the Keyword Type that the middle name is assigned to. <p>Once you have entered the tag or regular expression rule in the associated text field, you can create variations or synonyms of this value to search for when the entered value is not found. Click the ... button to open the Search Tag Synonyms dialog box. From here, you can add multiple synonyms and order them by search preference.</p> <p>To ignore the case of the text value being extracted as a Keyword Value, select the Match against any letter case check box.</p> <hr/> <p>Tip: The Match against any letter case option should be used when the zone contains only the name and is not surrounded by other text.</p> <hr/> <p>Address extraction: Select this option if you want the OCR engine to identify and extract any text in the zone that is formatted like an address.</p> <p>When this option is selected, the Address options are displayed in the Filtering section below, allowing you to define how the address should be identified and extracted as a Keyword Value.</p> <div data-bbox="616 1249 1401 1465" data-label="Form"> </div> <p>The following address formats are supported by default:</p> <ul style="list-style-type: none"> • Street Address 1 [Street Address 2] City, State/Province Code, Postal Code[-Secondary Postal Code] • Street Address 1, [Street Address 2,] City, State/Province Code, Postal Code[-Secondary Postal Code] <hr/> <p>Note: State/Province Code values must be two characters in length. Street Address 2 and Secondary Postal Code values are optional.</p>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>While the entire address is captured in the zone, the text is parsed into separate values (street address, city, state/province, and postal code) and each value can be stored as an individual Keyword Value.</p> <ul style="list-style-type: none"> • Use the Street Address Keyword drop-down list to select the Keyword Type that the street address is assigned to. • Use the City Keyword drop-down list to select the Keyword Type that the city is assigned to. • Use the State/Province Keyword drop-down list to select the Keyword Type that the state/province is assigned to. • Use the Postal Code Keyword drop-down list to select the Keyword Type that the postal code is assigned to. <hr/> <p>Note: While all four of the above values must be captured within the zone for address extraction to be processed, you do not have to assign each of the values to a Keyword Type if one or more of the values are not needed.</p> <hr/> <p>You may also identify a text tag or a regular expression to find the name values within the zone. When a text tag is configured, the OCR engine searches the zone for the specified text and extracts the next logical value (either to the right of or underneath the tag) as a Keyword Value. When a regular expression is configured, the text matching the pattern of the regular expression is identified as a name value.</p> <p>To configure a tag or regular expression, enter the text tag or regular expression in the Find after tag/expression: field. Text tags must be enclosed in quotation marks so that they are not treated as regular expressions. Multiple tags and/or regular expressions can be configured; they must be separated by a space in the Find after tag/expression field.</p> <p>Once you have entered the tag or regular expression rule in the associated text field, you can create variations or synonyms of this value to search for when the entered value is not found. Click the ... button to open the Search Tag Synonyms dialog box. From here, you can add multiple synonyms and order them by search preference.</p>

Data Field Zone Keyword Field Options	Description
Remove all whitespace from result	<p>Note: This option is not displayed if the Person name extraction radio button is selected in the Data Application section.</p> <p>Select this check box if you would like to remove all whitespace (for example, spaces or paragraph returns) from the Keyword Value after it is read by the OCR engine.</p> <p>For example, if the Keyword Value read by the OCR engine is PSY 103 and the Remove all whitespace from result check box is selected, the Keyword Value is modified to PSY103.</p>
Delete this page from document	<p>Select this check box if you would like the page to be deleted from the document if a value is successfully extracted for the Data Field Zone and this value is not suspect.</p> <p>If a value is not extracted for the Data Field Zone, or if the extracted value is suspect, the page will not be deleted.</p>
Redact zone from document after keyword assignment	<p>Note: This option is only available when the system performing Advanced Capture is licensed for Automated Redaction.</p> <p>Select this check box if you would like to redact a Data Field Zone from a document once the Keyword Value has been extracted from the zone and assigned.</p> <p>Tip: If you would like to redact a Data Field Zone from a document without capturing and assigning a Keyword Value, select <None> from the Keyword type drop-down list.</p>

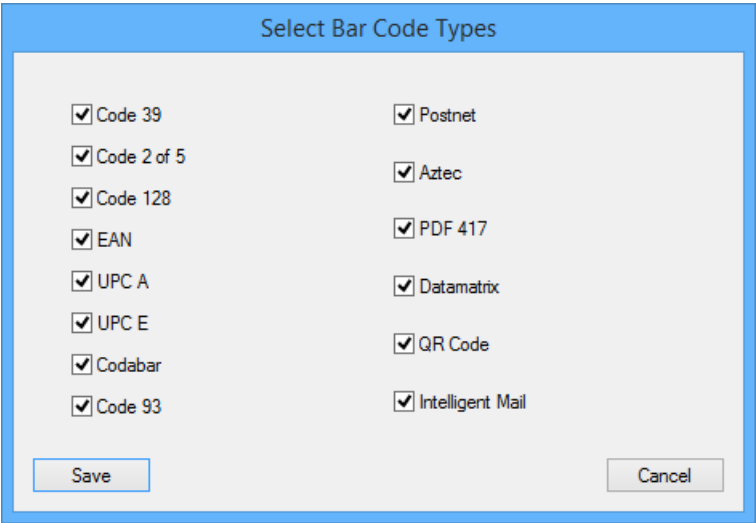
Data Field Zone Keyword Field Options	Description
Character Search Hints	<p>These options are available to assist the OCR engine in determining if the value it reads is correct. Select the check box next to each character type that is allowed to be included in the Keyword Value. The available options are:</p> <ul style="list-style-type: none"> • Numerals. Numeric characters, 0-9. • Uppercase. Uppercase alphabetic characters. • Lowercase. Lowercase alphabetic characters. • Punctuation. Punctuation marks (e.g., . ! ?). • Miscellaneous. Other ASCII characters that do not fall into one of the above categories (e.g., # \$ * @). <p>When a character is recognized by the OCR engine that is not part of an allowable character set, the character is replaced by a tilde (~) and the value is automatically marked as suspect.</p> <hr/> <p>Tip: Using the allowed character type options can sometimes help the OCR engine more easily determine the correct value by eliminating characters that are obviously not correct (e.g., an I is correctly identified instead of a 1 because numeric characters are filtered and prevented from being recognized as part of the value).</p> <hr/> <p>Select the Disable Asian Text Support check box to instruct the OCR engine to skip Asian (i.e., double-byte) characters.</p> <hr/> <p>Note: The Disable Asian Text Support check box is only available if the OCR format assigned to the Document Type is configured for an Asian language (e.g., Japanese, Korean, etc.).</p> <hr/> <p>Tip: Selecting the Disable Asian Text Support check box allows you to identify numeric data (e.g., Date Keyword Values, Currency Keyword Values, and Document Dates) when performing OCR on documents configured to contain Asian (i.e., double-byte) characters.</p> <hr/>

Data Field Zone Keyword Field Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the Keyword Value being read by the Data Field Zone is only displayed on one page and is always displayed on the same page (for example, the Keyword Value is always displayed on page 1). Enter the page number that the Keyword Value is located on in the associated field. • Select the Relative page radio button if the Keyword Value may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) the Data Field Zone may be located on. <ul style="list-style-type: none"> Select the First page check box if the Data Field Zone is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes. Select the First page (not only page) check box if the Data Field Zone is located on the first page and other pages in the document. Select the Interior pages check box if the Data Field Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes. Select the Last page check box if the Data Field Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes. Select the Last page (not only page) check box if the Data Field Zone is located on the last page and other pages in the document.
Specific allowed characters	<p>The Specific allowed characters field allows you to configure the OCR engine to identify only the characters that you specify in the field.</p> <p>For example, if you entered ABC in the Specific allowed characters field and the OCR engine identified the value as ABCDEFGF, the DEFG characters would be stripped from the value and the Keyword Value associated with the document would be ABC.</p>

Data Field Zone Keyword Field Options	Description
Advanced Recognition	<p>Note: Options involving ICR processing below are only enabled if your solution is licensed for Intelligent Character Recognition (ICR).</p> <hr/> <p>Note: The OCR engine does not support Asian characters when reading dot matrix-printed text.</p> <hr/> <p>Using this drop-down list, select the type of processing you would like to perform on this Data Field Zone.</p> <ul style="list-style-type: none"> • Machine print text/OCR. The OCR engine will perform optical character recognition to read machine-printed text. This option is selected by default. • Machine print text/detect structure. The OCR engine will attempt to determine whether the value consists of machine-printed text, dot matrix-printed text, or hand-written text, and then use the appropriate type of character recognition to read the text. • Machine print text/flowing text. The OCR engine will attempt to read machine-printed text within the Data Field Zone as if the text flowed like a paragraph with similar character sizes, spacing, and fonts. If this option is not selected, by default, the OCR engine will attempt to read machine-printed text using the automatic zoning recognition mode, which simply attempts to determine what the character sizes, spacing, and fonts look like on their own (i.e., independent of their positioning within a paragraph or section of flowing text). • Machine print text/labeled form field. The OCR engine will attempt to identify and remove form tag data from the extracted value. If there are two or more consecutive characters at the beginning or end of the extracted value that are smaller in font height than the middle text of the line, then the smaller characters are removed from the value. If the image is processed in color, and some of the characters within the Data Field Zone are red, then these characters are removed from the value. However, if your OCR format is configured for black and white processing, color characters may not be removed. Additionally, if one set of characters appears on a different line than another set of characters, the first set of characters may be removed from the value. For example, if State OH is extracted from the Data Field Zone, and State has a smaller font height, appears on a different line, or has a different color than OH, then State will be identified as a tag from a form and thus be removed from the value.

Data Field Zone Keyword Field Options	Description
Advanced Recognition (cont.)	<ul style="list-style-type: none"> • Machine print text/Rotated 90 degrees counter-clockwise. The OCR engine will perform optical character recognition to read machine-printed text that has been rotated 90 degrees counter-clockwise. • Machine print text/Rotated 90 degrees clockwise. The OCR engine will perform optical character recognition to read machine-printed text that has been rotated 90 degrees clockwise. • Machine print text/dot matrix printer. The OCR engine will perform optical character recognition to read dot matrix-printed text. • Machine print text/dot matrix - rotated 90 degrees counter-clockwise. The OCR engine will perform optical character recognition to read dot matrix-printed text that has been rotated 90 degrees counter-clockwise. • Machine print text/dot matrix - rotated 90 degrees clockwise. The OCR engine will perform optical character recognition to read dot matrix-printed text that has been rotated 90 degrees clockwise. • Bar Code. The bar code found within the Data Field Zone will be converted into text and then read by the OCR engine. <hr/> <p>Note: This option is only available if your solution is licensed for the Bar Code Recognition Server.</p> <hr/> <p>Note: Depending on your configuration, the Advanced Capture engine will only search for certain bar code types in this zone. See Bar Code Types for more information.</p> <hr/> <ul style="list-style-type: none"> • MICR Font. The OCR engine will perform optical character recognition to read MICR line data from check documents. • Handwriting/ICR - numerals/grouping punctuation (North American style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the North American style (e.g., the 7 character is not crossed). • Handwriting/ICR - numerals/grouping punctuation (European style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the European style (e.g., the 7 character is crossed).

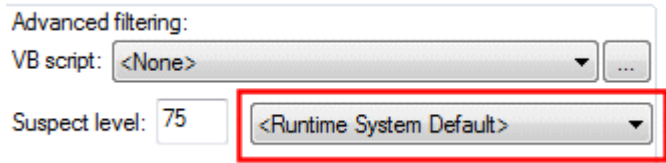
Data Field Zone Keyword Field Options	Description
Advanced Recognition (cont.)	<ul style="list-style-type: none"> • Handwriting/ICR - alphanumeric/punctuation. The OCR engine will perform intelligent character recognition to read hand-written text. This option will attempt to recognize all alphanumeric characters. • Auto-detect ICR/OCR - Default to ICR/Alphanumeric. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for alphanumeric characters will be used by default. • Auto-detect ICR/OCR - Default to ICR/Numeric North American. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, North American-style characters (e.g., the 7 character is not crossed) will be used by default. • Auto-detect ICR/OCR - Default to ICR/Numeric European. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, European-style characters (e.g., the 7 character is crossed) will be used by default. • Auto-detect ICR/OCR - Default to OCR. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, optical character recognition will be used by default. <hr/> <p>Note: The Auto-detect ICR/OCR options may not work properly if the Data Field Zone contains less than 25 characters.</p> <hr/> <p>Tip: Of the two Auto-detect ICR/OCR options, Default to ICR is more likely to produce the best results when the type of text cannot be determined. This is because the OCR engine's intelligent character recognition generally reads machine-printed text more accurately than the engine's optical character recognition reads hand-written text.</p> <hr/> <p>If one of the Handwriting/ICR options is selected, the Remove vertical/horizontal lines from zone before processing check box is displayed. Select this option if you would like the OCR engine to attempt to strip vertical or horizontal lines (i.e., constraint boxes) from the zone before it is processed.</p>

Data Field Zone Keyword Field Options	Description
Bar Code Types	<p>Note: This option is only available when the Bar Code option is selected in the Advanced recognition drop-down list.</p> <p>Click this button to access the Select Bar Code Types dialog box.</p>  <p>Here you can select the types of bar codes for which the Advanced Capture engine will search when processing the Data Field Zone. The engine will ignore any deselected bar code types when processing this zone.</p> <p>By default, all bar code types are selected.</p> <p>Tip: To expedite Advanced Capture processing, select only the desired bar code types.</p> <p>To save changes to your bar code type selections, click Save.</p>
VB script	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>
Keyword association	<p>You can logically group Data Field Zones using Keyword association. These groupings can be used to identify Keywords that belong to a Multi-Instance Keyword Type Group.</p> <p>Use the Keyword association field to enter or select a logical group name for the Data Field Zones that identify Keywords belonging to an MIKG. The Advanced Capture engine will attempt to maintain this Keyword grouping on the resulting document.</p>

Data Field Zone Keyword Field Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (for example, +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (that is, through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (for example, when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Data Field Zone Keyword Field Options	Description
Activation groups (cont.)	<p>Alternatively, you can assign a form definition group as the Data Field Zone's activation group to activate the zone for processing. Form definition groups can be used to extract only specific types of information (for example, header data vs. detail data) during processing. In the Activation groups drop-down list, form definition groups are enclosed in brackets (for example, [Group1]).</p>
Colors	<p>If you would like to assign specific colors to the Keyword Type configured for the Data Field Zone, click the Colors button. The Display Colors dialog box is displayed.</p> <div data-bbox="667 657 1348 961" data-label="Image"> </div> <p>Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once Advanced Capture processing has taken place.</p> <ul style="list-style-type: none"> • To change the display color for regular values, click Display Color to open your machine's color palette, select a color, and click OK. • To change the display color for suspect values, click Suspect Color to open your machine's color palette, select a color, and click OK. • To revert back to the default display color for regular or suspect values, click the Automatic button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values). <hr/> <p>Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.</p>

Data Field Zone Keyword Field Options	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Advanced Capture form and the actual Suspect Level detected for the selected value are displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that suspect Keyword Values are being consistently identified.</p> <hr/>

Data Field Zone Keyword Field Options	Description
Override Default Regional Setting	<p>When a Date, Date & Time, or Currency Keyword Type has been selected for the Data Field Zone, the Override Default Regional Setting drop-down list is displayed to the right of the Suspect level field.</p> <hr/> <p>Note: If Person name extraction or Address extraction has been selected for the Data Application, the Override Default Regional Setting drop-down list is not displayed.</p> <hr/>  <p>This option allows you to override your system's default regional settings when extracting values for the applicable Keyword Types in the Data Field Zone.</p> <p>Select a regional language from the drop-down list to parse the applicable Keyword Values using the selected region's formatting rules and language-specific names for days, months, etc.</p> <p>Select <Runtime System Default> to maintain your system's default regional settings when parsing these values.</p> <hr/> <p>Note: This option does not affect which languages the OCR format is configured to recognize. For information on configuring OCR formats, see the Full-Page OCR module reference guide or help files.</p> <hr/>

Data Field Zone Keyword Field Options	Description
Find/validate in keyword dataset	<p>Note: This option is not enabled if the Person name extraction option is selected in the Data Application section. Additionally, it is only enabled if the Keyword Type selected in the Keyword type drop-down list is configured to use a Keyword Data Set (but not a Cascading Data Set).</p> <p>Select this option if you wish to validate the value identified by the OCR engine using values specified for the Keyword Type's Data Set.</p> <p>If the extracted value is an exact or fuzzy match to a value found in the Data Set (i.e., the number of mismatched characters in the extracted value is less than 10% of the number of characters in the extracted value), then the value found in the Data Set is assigned as the Keyword Value.</p> <p>Additionally, if the extracted value is a fuzzy match to a value found in the Data Set, then the assigned Keyword Value (i.e., the value fuzzy-matched in the Data Set) is marked as suspect.</p> <p>Note: Fuzzy matches are disabled for values of five characters or less. If the extracted value is not an exact match to a value found in the Data Set, then the extracted value is assigned as the Keyword Value and is marked as suspect.</p> <p>If the Find/validate in keyword dataset check box is not selected, the OCR engine does not attempt to validate the extracted value against any values in the Keyword Data Set, and the extracted value is assigned as the Keyword Value.</p>
Add implied decimal point	<p>Note: This check box is only enabled if a Currency Keyword Type is selected in the Keyword Type drop-down list.</p> <p>Select this check box if you would like the OCR engine to automatically insert a decimal point in the extracted value if one is not detected.</p> <p>For example, if this option is selected and the extracted value for the Total Amount Due Keyword is \$12345, the OCR engine would automatically modify the value to \$123.45.</p> <p>If this option is selected and a decimal point is detected in the value, or if this option is not selected, then the extracted value is left as-is.</p>

Creating an Entire Page Zone for Find by Tag/Expression


Note: The **Create Entire Page Keyword Zone** should only be used in using text tags or regular expressions to identify and extract Keyword Values.


When using a tag or regular expression to identify Keyword Values, you may elect to search the entire page instead of a specific Data Field Zone.


To configure an entire page zone:

1. To display the **Modify/Delete Form** options, in the Document Viewer, right-click on the document that the Advanced Capture form is being created from.

Internal ID: **103**
Form Name: **AC Transcript Form**
Document Type Assigned by Form: **HR - Transcript**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Default in Identification Chain: **No**
Requires Manual Indexing: **No**
Common Keyword Source: **No**
Create XML Data Rendition: **Yes**
Alignment Archetype Document ID: **<None>** [Update to Current Document](#)

 **Modify Form**

 **Delete Form**

 **Create Entire Page Zone**

2. Click **Create Entire Page Keyword Zone**. The **Data Field Zone** dialog box is displayed.

Data Field Zone

Optical Mark Field Signature Detection Field Snippet

Text Field Line Item Extraction Grouped Line Item Extraction

Keyword type: <None> XML Node:

Data Application

☒ Entire zone ☐ Remove line breaks

☐ Specific line(s):

☐ Table style/vertical list

☐ Find by tag/expression

☐ Person name extraction

☐ Address extraction

☐ Remove all whitespace from result

Page Location(s)

☒ Absolute page: 1 ☐ Relative page(s):

Field only appears on a single page; always the same absolute page number (one-based)

☐ First page

☐ First page (not only page)

☐ Interior pages

☐ Last page

☐ Last page (not only page)

☐ Delete this page from document

☐ Redact zone from document after keyword assignment

Filtering

☒ Numerals ☒ Punctuation

☒ Uppercase ☒ Miscellaneous

☒ Lowercase

Advanced filtering:

VB script: <None> ...

Suspect level: 25 (0-99) 0 = Form Default (25), Actual = 20

☐ Add implied decimal point

☐ Remove text clipped by zone boundaries

☐ Find/validate in keyword dataset

Specific allowed characters:

Advanced recognition: Machine print text/standard OCR

Keyword association: Activation groups: Colors

Save **Cancel**

3. Ensure that the Text Field tab is displayed.
4. Create a Text Field Data Zone, ensuring that the you are using the **Find by tag/expression** option in the Data Application section.
5. When finished, click **Save**.

Configuring a Line Item Extraction Data Field Zone

A Line Item Extraction Field Zone allows you to extract multiple Keyword Values from a single zone when the values to be extracted are arranged in line item format using rows and columns. Values in each column of the table are assigned to the same Keyword Type.

Tip: Line Item Extraction Field Zones are intended to be used with Multi-Instance Keyword Type Groups in order to capture line item data while maintaining the Keyword Values' relationship to one another.

Note: Keyword Values can also be extracted from line-item-formatted data into a Multiple Instance Keyword Type Group using multiple Text Data Field Zones (one for each column of data) set to use the **Table style** Data Application option.

Line Item Extraction Zones have two modes:

- **Automatic Table Decomposition.** The OCR engine attempts to automatically detect the location of individual table columns within the Data Field Zone.
- **Manual Table Decomposition.** Individual table columns are manually defined when configuring the Data Field Zone.

Both Line Item Extraction modes are configured using the options on the Line Item Extraction tab of the **Data Field Zone** dialog box. Once the table decomposition mode-specific options are set, you are able to set other, non-mode-specific configuration options.

Data Field Zone [X]

Optical Mark Field Signature Detection Field Snippet

Text Field Line Item Extraction Grouped Line Item Extraction

☒ Use line item style data extraction XML Node (Parent): [v]

Column	Keyword Type	Req	Expression	XML Node

[Delete] [Move Up] [Move Down]

[v] ☐ Press F2 in this field [v] [Add]

Begin table: [Press F2 in this field] End table: [Press F2 in this field] Invalid row: [Press F2 in this field]

☐ Inclusive ☐ Inclusive

Page Location(s)

☒ Absolute page: ☐ Relative page(s):

Field only appears on a single page; always the same absolute page number (one-based)

☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)

Options

☐ Manual table decomposition mode
☐ Automatic continued-row merging

VB script: [None] [v] [...]

Suspect level: [75]
 (0-99) 0 = Form Default (75), Actual = 30

Advanced recognition: [Machine print text/standard OCR] [v]

Activation groups: [v]

[Save] [Cancel]

Configuring Line Item Extraction in Automatic Table Decomposition Mode

1. Select the **Use line item style data extraction** check box. The remaining options on the tab are enabled.
2. Use the **Keyword type** drop-down list (located below the Column and Keyword Type columns) to select the Keyword Type that values in the first column of the table are assigned to as Keyword Values.
To ignore a column or capture it as XML data only, select **<None>** from the drop-down list.
3. If this Keyword Value is required, select the **Required** check box (located to the right of the Keyword type drop-down list). If a Keyword Value that is marked as required is missing, the data from the entire row is discarded.
4. Optional: To specify a regular expression rule for the extracted text, enter the rule in the **Expression** field (located next to the **Required** check box). The OCR engine will compare the extracted text to the defined regular expression rule; if the text is a match, the value is stored as a Keyword Value. If the text does not match, it is discarded.

For example, if you specify the following regular expression rule:

[[::upper:]]::lower:]]::digit:]]::space:]]+

Any value containing a character that is not a letter, number, or space is discarded.

Note: To access the Regular Expression Library, click in the field and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

Note: All regular expressions must be ECMA compliant.

Tip: Regular expressions can be used to discard column data that consists of entirely unwanted characters (e.g., a **123***456**, where you want to capture **123** and **456** as separate Keyword Values and discard the asterisk separator column). However, if the unwanted characters are located in the middle of valid data (e.g., **123***456**, where you want to capture **123456** as one Keyword Value), you can configure a Keyword Lookup/Replace dictionary entry to replace the data.

5. Click **Add**.
6. Repeat Steps 2-5 for all columns in the table. Columns must be added to the list as they appear in the document from left-to-right in order for Keyword Values to be extracted correctly.
Select a configured column and use the **Move Up** and **Move Down** buttons to re-arrange it as needed. Select a configured column and click **Delete** to delete the column configuration.
7. Click **Save**.

Configuring Line Item Extraction in Manual Table Decomposition Mode

1. Select the **Use line item style data extraction** check box. The remaining options on the tab are enabled.
2. In the Options section, select the **Manual table decomposition mode** check box. A preview of the Data Field Zone is displayed to the right of the **Data Field Zone** dialog box.
3. Optional: Select the **Automatic continued-row merging** check box.
When this option is selected, the OCR engine compares the average spacing between rows. If a row is closer than average to the row that precedes it and the row does not have data in at least half of its processed columns, the OCR engine recognizes it as a continuation of the previous row and merges the extracted values.

4. In the preview of the Data Field Zone displayed to the right of the **Data Field Zone** dialog box, click and hold the left mouse button and use the pointer to draw a box around the first column of data to be extracted.

The **Modify Line Item Column** dialog box is displayed.

Depending on the type of Data Field Zone you are configuring, you will have different settings available to you.

- For information on configuring a text extraction zone, see [Configuring a Text Extraction Column on page 155](#).
 - For information on configuring a signature detection zone, see [Configuring a Signature Detection Column on page 159](#).
5. Click **Save**. The Column Extraction Zone is highlighted in the Data Field Zone preview.
 6. If you wish to edit a configured column, double-click on its entry in the table to re-open the **Modify Line Item Column** dialog box, change any desired settings, and click **Save**.
 7. Repeat step 4 and the steps in the sections below for any remaining columns in the Data Field Zone preview whose data you wish to capture.

Configuring a Text Extraction Column

To begin configuring a text extraction column for Line Item Extraction, ensure that the **Text Extraction** tab is selected in the **Modify Line Item Column** dialog box. Then follow the procedures below.

The screenshot shows the 'Modify Line Item Column' dialog box with the 'Text Extraction' tab selected. The dialog has three tabs: 'Text Extraction', 'Optical Mark Detection', and 'Signature Detection'. The 'Text Extraction' tab contains the following fields and options:

- Keyword Type:** A dropdown menu with 'Issue Date' selected.
- XML Node:** A dropdown menu.
- Colors:** A button next to the XML Node dropdown.
- Second Capture Group Keyword Type:** A dropdown menu with '<None>' selected.
- XML Node:** A dropdown menu.
- Colors:** A button next to the XML Node dropdown.
- Expression to Match:** A text input field.
- Extract capture group only:** A checkbox.
- Suspect Level:** A text input field with '0' entered, followed by the text '(0 - 99 : 0 = parent line item zone default)'.
- Value must be NULL text:** A checkbox.
- Column data required:** A checkbox.
- Enforce absolute column break on right:** A checkbox.
- Remove all whitespace from result:** A checkbox.
- Save:** A button at the bottom left.
- Cancel:** A button at the bottom right.

1. Use the **Keyword Type** drop-down list to select the Keyword Type that values in the selected column are assigned to as Keyword Values.

To ignore the data in the selected column or capture it as XML data only, select **<None>** from the **Keyword Type** drop-down list.

Note: You can configure the Advanced Capture engine to split each value in the selected column into two separate Keyword Values by specifying a regular expression rule for the extracted text, setting two capture groups within this regular expression rule, and assigning a Keyword Type to each of these capture groups. This may be useful when two different types of data are listed on a document as one value (e.g., a transcript course subject and course number need to be captured as two separate values, but they appear together on a document as **ENG101**). If you are configuring the selected column for two Keyword Values, the first of these values is assigned to your **Keyword Type** selection.

2. If the form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the form), the **XML Node** drop-down list is displayed next to the corresponding Keyword Type drop-down list.

Use the **XML Node** drop-down list to select the XML node (i.e., the element) that Keyword Values extracted from this column are contained in when the XML rendition of the document is created.

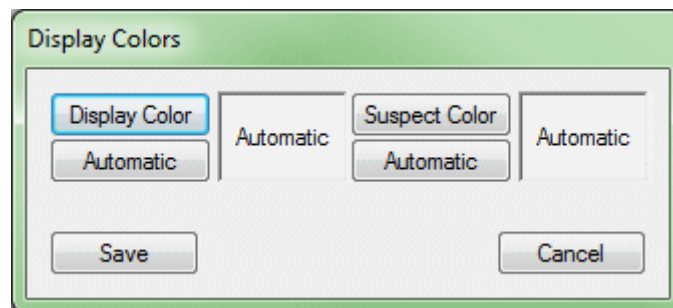
3. If you selected a **Date**, **Date & Time**, or **Currency** Keyword Type for the selected column, the **Override Default Regional Setting** drop-down list is displayed below this selected Keyword Type. If you have not selected one of these Keyword Types, the **Override Default Regional Setting** drop-down list is disabled.

When available, you can use the **Override Default Regional Setting** option to override your system's default regional settings when extracting values for the applicable Keyword Types in the Data Field Zone. To use this option, do one of the following:

- Select a regional language from the drop-down list to parse the applicable Keyword Values using the selected region's formatting rules and language-specific names for days, months, etc.
- Select **<Runtime System Default>** to maintain your system's default regional settings when parsing these values.

Note: This option does not affect which languages the OCR format is configured to recognize. For information on configuring OCR formats, see the **Full-Page OCR** module reference guide or help files.

4. If you would like to assign specific colors to the selected Keyword Type, click the **Colors** button. The **Display Colors** dialog box is displayed.



Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once processing has taken place.

- To change the display color for regular values, click **Display Color** to open your machine's color palette, select a color, and click **OK**.
- To change the display color for suspect values, click **Suspect Color** to open your machine's color palette, select a color, and click **OK**.

- To revert back to the default display color for regular or suspect values, click the **Automatic** button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values).

Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.

5. If you are configuring the selected column for two Keyword Values, use the **Second Capture Group Keyword Type** drop-down list to select the second Keyword Type that values in the selected column are assigned to as Keyword Values.
If you are configuring the selected column for just one Keyword Value, or if you wish to capture a second Keyword Value as XML data only, select **<None>** from the **Second Capture Group Keyword Type** drop-down list.
6. If you wish to specify a regular expression rule for the extracted text, enter the rule in the **Expression to Match** field. The OCR engine will compare the extracted text to the defined regular expression rule; if the text is a match, the value is stored as a Keyword Value. If the text does not match, it is discarded.
For example, if you specify the following regular expression rule:
[:upper:][:lower:][:digit:][:space:]]+
Any value containing a character that is not a letter, number, or space is discarded.

Note: To access the Regular Expression Library, click in the field and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

Note: All regular expressions must be ECMA compliant.

Note: If you are configuring the selected column for two Keyword Values, you must specify a regular expression rule for the extracted text and set two capture groups within the rule to align to these two values.

Tip: Regular expressions can be used to discard column data that consists of entirely unwanted characters (e.g., a **123***456**, where you want to capture **123** and **456** as separate Keyword Values and discard the asterisk separator column). However, if the unwanted characters are located in the middle of valid data (e.g., **123***456**, where you want to capture **123456** as one Keyword Value), you can configure a Keyword Lookup/Replace dictionary entry to replace the data.

7. If you entered a regular expression rule containing a capture group in the previous step, and you would like to have Keyword Values extracted for only this capture group, select the **Extract capture group only** check box. If you would like to have Keyword Values extracted for the entire regular expression (including the capture group), deselect the **Extract capture group only** check box.

Note: This option is not available if the **Second Capture Group Keyword Type** drop-down list is set to anything other than **<None>**, or if the **Value must be NULL text** check box is selected.

8. If you wish to set a different Suspect Level for the selected Keyword Type(s) than the parent Suspect Level set for the entire Line Item Extraction Zone, enter this value (**0** to **99**) in the **Suspect Level** field. By default, this value is set to **0**, which retains the parent Suspect Level.
9. If you would like to have Keyword Values extracted from the row only if no value exists in the specified column, select the **Value must be NULL text** check box. When this option is selected, the **Column Data Required** check box is selected automatically and the **Expression to Match** field is disabled.
10. If a Keyword Value is required, select the **Column Data Required** check box. If a Keyword Value that is marked as required is missing, the data from the entire row is discarded.
11. If you wish to restrict the Keyword Values that are extracted from the column to only those characters fully contained within the column's boundaries, select the **Enforce absolute column break on right** check box.
When this option is not selected, some carryover is allowed when characters on either side of the column's right boundary are very close together, assuming that these characters belong to the same Keyword Value. The characters just outside the right boundary will be included in the extracted value.
12. If you wish to remove all whitespace (i.e., spaces, paragraph returns, etc.) from a Keyword Value after it is read by the OCR engine, select the **Remove all whitespace from result** check box.
For example, if the Keyword Value read by the OCR engine is **PSY 103** and this option is selected, the Keyword Value is modified to **PSY103**.

Note: If you specified a regular expression rule for the extracted text in the **Expression to Match** field, the OCR engine will compare the extracted text without any whitespace to the defined regular expression rule.

Configuring a Signature Detection Column

To begin configuring a signature detection column for Line Item Extraction, ensure that the **Signature Detection** tab is selected in the **Modify Line Item Column** dialog box. Then follow the procedures below.

The screenshot shows the 'Modify Line Item Column' dialog box with the 'Signature Detection' tab selected. The 'Keyword type' dropdown is set to 'Last Name'. The 'Keyword Values' section has empty fields for 'Positive result value' and 'Negative result value'. The 'XML Node' dropdown is empty. The 'Sensitivity' section has radio buttons for 'Highest' (selected), 'Low', 'Lower', 'Lowest', and 'Custom' (with a value of 1-99). The 'Options' section has radio buttons for 'Signature is framed/underlined' (selected), 'Signature area is unframed', and a checkbox for 'Exclude machine print text'. A 'Colors' button is at the bottom right of the main content area. 'Save' and 'Cancel' buttons are at the bottom of the dialog.

1. Use the **Keyword Type** drop-down list to select the Keyword Type that values in the selected column are assigned to as Keyword Values.
To ignore the data in the selected column or capture it as XML data only, select **<None>** from the **Keyword Type** drop-down list.
2. If the form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the form), the **XML Node** drop-down list is displayed.
Use the **XML Node** drop-down list to select the XML node (i.e., the element) that Keyword Values extracted from this column are contained in when the XML rendition of the document is created.
3. In the **Keyword Values** section, do the following:
 - a. In the **Positive Result Value** field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is present.
 - b. In the **Negative Result Value** field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is not present.

For example: you are processing an application that requires a signature to indicate that the data has been verified. Depending on if a signature is present or not, a Keyword Value is assigned to the **Verified** Keyword Type.

- If a signature is detected, **Yes** is assigned as the **Verified** Keyword Value.
- If a signature is not detected, **No** is assigned as the **Verified** Keyword Value.

4. In the **Options** section, do the following:

- If the signature is located in a pre-defined space, such as a box or on a signature line, select **Signature is framed/underlined**.
- If the signature is located in an undefined or blank space, select **Signature area is unframed**.
- To ignore machine-printed text in the Data Field Zone (e.g., **Signature**, **Applicant Signature**, etc.), select the **Exclude machine print text** check box.

This prevents the signature detection results from being impacted by the presence of machine-printed text.

5. In the **Sensitivity** section, select a radio button to determine the Suspect Level for this Data Field Zone.

- **Highest**. Sets the Suspect Level for this Data Field Zone to 70.
- **Low**. Sets the Suspect Level for this Data Field Zone to 75.
- **Lower**. Sets the Suspect Level for this Data Field Zone to 80.
- **Lowest**. Sets the Suspect Level for this Data Field Zone to 85.
- **Custom**. Allows you to enter a value (from 1 to 99) for the Suspect Level for this Data Field Zone. If enter a value of 0 or leave the field blank, the value will reset to 50.

The Suspect Level is the level of confidence placed in the processing results for this field.

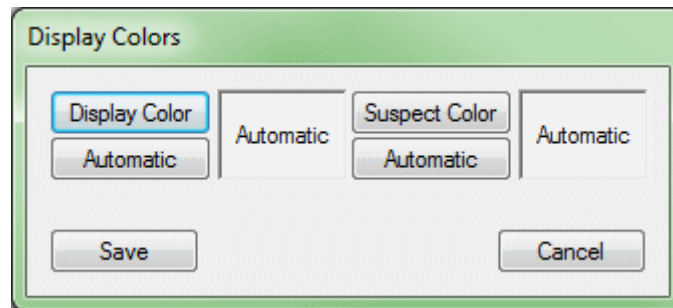
After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.

The Sensitivity level selected for this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.

For example, setting the Sensitivity to **Lowest** would indicate you have a fair amount of confidence in the result returned by the OCR engine because few higher scores could be returned and fewer results would be determined to be suspect.

Setting the Suspect Level to **Highest** would indicate you have less confidence in the result because a great number of lower scores could be returned and more results would be determined as suspect.

6. If you would like to assign specific colors to the Keyword Type configured for the Data Field Zone, click the **Colors** button. The **Display Colors** dialog box is displayed.



Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once processing has taken place.

- To change the display color for regular values, click **Display Color** to open your machine's color palette, select a color, and click **OK**.
- To change the display color for suspect values, click **Suspect Color** to open your machine's color palette, select a color, and click **OK**.
- To revert back to the default display color for regular or suspect values, click the **Automatic** button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values).

Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.

Additional Line Item Extraction Configuration Options

Once you have configured the Line Item Extraction Field Zone's table decomposition mode, you are able to set other, non-mode specific configuration options.

Data Field Zone Line Item Extraction Option	Description
XML Node (Parent)	<p>Note: This field is only displayed if the Advanced Capture form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Advanced Capture form).</p> <p>Enter the name of the XML parent node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p> <p>Note: While the XML Node (Parent) name may include alphanumeric characters, underscores (_), hyphens (-), or periods (.), it must begin with either a letter or an underscore.</p>
Begin table marker	<p>Use the Begin table marker field to specify the text that indicates the beginning of the information being extracted.</p> <p>The markers can be configured as literal text (i.e., the text specified in the Begin Table Marker field is the marker) or regular expressions (i.e., the text in the Begin Table Marker field defines a regular expression, and text matching that pattern is the marker).</p> <p>Literal markers must be enclosed in quotation marks to differentiate them from the regular expression markers.</p> <p>Multiple begin table markers (either literal, regular expression, or both) can be configured by separating the text by a space in the Begin table marker field.</p> <p>Note: To access the Regular Expression Library, click in the field and press F2. See The Regular Expression Library on page 247 for more information.</p> <p>By default, text used as a Begin table marker is not included in the Keyword Value(s) being extracted. To include this text in the Keyword Value, select the Inclusive check box beneath the Begin table marker field.</p>

Data Field Zone Line Item Extraction Option	Description
End table marker	<p>Use the End table marker field to specify the text that indicates the end of the information being extracted.</p> <p>The markers can be configured as literal text (i.e., the text specified in the End Table Marker field is the marker) or regular expressions (i.e., the text in the End Table Marker field defines a regular expression, and text matching that pattern is the marker). Literal markers must be enclosed in quotation marks to differentiate them from the regular expression markers.</p> <p>Multiple end table markers (either literal, regular expression, or both) can be configured by separating the text by a space in the End table marker field.</p> <hr/> <p>Note: To access the Regular Expression Library, click in the field and press F2. See The Regular Expression Library on page 247 for more information.</p> <hr/> <p>By default, text used as an End table marker is not included in the Keyword Value(s) being extracted. To include this text in the Keyword Vale, select the Inclusive check box beneath the End table marker field.</p>
Invalid Row	<p>Use the Invalid Row field to specify text that indicates the beginning of the row that should be discarded</p> <p>The markers can be configured as literal text (i.e., the text specified in the Invalid Row field is the marker) or regular expressions (i.e., the text in the Invalid Row field defines a regular expression, and text matching that pattern is the marker). Literal markers must be enclosed in quotation marks to differentiate them from the regular expression markers.</p> <p>Multiple invalid row markers (either literal, regular expression, or both) can be configured by separating the text by a space in the Invalid Row field.</p> <hr/> <p>Note: To access the Regular Expression Library, click in the field and press F2. See The Regular Expression Library on page 247 for more information.</p> <hr/>

Data Field Zone Line Item Extraction Option	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the table being read in the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the table is always displayed on page 1). Enter the page number that the table is located on in the associated field. • Select the Relative page radio button if the table may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the table may be located on. <p>Select the First page check box if the table is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the table is located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the table is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the table is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the table is located on the last page and other pages in the document.</p>
VB script	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>

Data Field Zone Line Item Extraction Option	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Advanced Capture form and the actual Suspect Level detected for the selected table are displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that suspect Keyword Values are being consistently identified.</p> <hr/>

Data Field Zone Line Item Extraction Option	Description
Advanced Recognition	<p>Note: Options involving ICR processing below are only enabled if your solution is licensed for Intelligent Character Recognition (ICR).</p> <hr/> <p>Note: The OCR engine does not support Asian characters when reading dot matrix-printed text.</p> <hr/> <p>Using this drop-down list, select the type of processing you would like to perform on this Data Field Zone.</p> <ul style="list-style-type: none"> • Machine print text/OCR. The OCR engine will perform optical character recognition to read machine-printed text. This option is selected by default. • Machine print text/detect structure. The OCR engine will attempt to determine whether the value consists of machine-printed text, dot matrix-printed text, or hand-written text, and then use the appropriate type of character recognition to read the text. • Machine print text/flowing text. The OCR engine will attempt to read machine-printed text within the Data Field Zone as if the text flowed like a paragraph with similar character sizes, spacing, and fonts. If this option is not selected, by default, the OCR engine will attempt to read machine-printed text using the automatic zoning recognition mode, which simply attempts to determine what the character sizes, spacing, and fonts look like on their own (i.e., independent of their positioning within a paragraph or section of flowing text). • Machine print text/dot matrix printer. The OCR engine will perform optical character recognition to read dot matrix-printed text. • Handwriting/ICR -numerals/grouping punctuation (North American style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the North American style (e.g., the 7 character is not crossed). • Handwriting/ICR -numerals/grouping punctuation (European style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the European style (e.g., the 7 character is crossed). • Handwriting/ICR - alphanumeric/punctuation. The OCR engine will perform intelligent character recognition in order to read hand-written text. This option will attempt to recognize all alphanumeric characters.

Data Field Zone Line Item Extraction Option	Description
Advanced Recognition (cont.)	<ul style="list-style-type: none"> • Auto-detect ICR/OCR - Default to ICR/Alphanumeric. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for alphanumeric characters will be used by default. • Auto-detect ICR/OCR - Default to ICR/Numeric North American. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, North American-style characters (e.g., the 7 character is not crossed) will be used by default. • Auto-detect ICR/OCR - Default to ICR/Numeric European. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, European-style characters (e.g., the 7 character is crossed) will be used by default. • Auto-detect ICR/OCR - Default to OCR. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, optical character recognition will be used by default. <hr/> <p>Note: The Auto-detect ICR/OCR options may not work properly if the Data Field Zone contains less than 25 characters.</p> <hr/> <p>Tip: Of the two Auto-detect ICR/OCR options, Default to ICR is more likely to produce the best results when the type of text cannot be determined. This is because the OCR engine's intelligent character recognition generally reads machine-printed text more accurately than the engine's optical character recognition reads hand-written text.</p> <hr/> <p>Note: The Line Item Extraction Data Field Zone cannot contain a mixture of OCR and ICR values. All values within the zone must use one type of recognition only.</p> <hr/>

Data Field Zone Line Item Extraction Option	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Data Field Zone Line Item Extraction Option	Description
Activation groups (cont.)	Alternatively, you can assign a form definition group as the Data Field Zone's activation group to activate the zone for processing. Form definition groups can be used to extract only specific types of information (e.g., header data vs. detail data) during processing. In the Activation groups drop-down list, form definition groups are enclosed in brackets (e.g., [Group1]).

Configuring a Grouped Line Item Extraction Data Field Zone

A Grouped Line Item Extraction Data Field Zone allows you to extract Keyword Values from multiple groups, or tables, within a single zone.

Keyword Values in each group can be identified by a tag (either a literal tag or a regular expression), regular expression, or by columns in line item data.

Tip: Grouped Line Item Extraction Data Field Zones are intended to be used with Multi-Instance Keyword Type Groups in order to capture data from multiple tables while maintaining the Keyword Values' relationship to one another.

For example:

Fall 2009			
BIO111	Biology 1	B	3.0
ENG212	Composition 1	A-	3.0
ECO211	Micro Econ 1	C	3.0
PHS111	Ice Skating	W	0.0
TOTAL			9.0

Spring 2010			
PSY 101	Psychology 1	B	3.0
ENG 312	Lit Analysis	A	3.0
MTH 151	Calculus 1	C	3.0
TOTAL			9.0

Here you have two groups of course data organized by term. For each course, you want to identify five pieces of information: the term the course was taken, the Course ID, the Course Description, Grade, and Credits). This course information is to be stored in a Multi-Instance Keyword Type Group; each instance represents information about one course the student has completed.

In this example, the **Term** Keyword Value is identified by a regular expression, and the remaining course information is identified by the column that it is displayed in.

A Grouped Line Item Data Field Zone is configured using the options on the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box.

The screenshot shows the 'Data Field Zone' dialog box with the 'Grouped Line Item Extraction' tab selected. The 'Use grouped line item style data extraction' checkbox is checked. The 'XML Node (Parent)' dropdown is empty. Below this is a table with columns: Column, Keyword Type, Req, Expression, and XML Node. To the right of this table is another table with columns: Keyword Type, Tag/Expression T..., Req, Position, and XML Node. To the right of this second table are buttons: New Tag, Modify, and Delete. Below the tables are sections for 'Page Location(s)' and 'Options'. The 'Page Location(s)' section has two radio buttons: 'Absolute page:' (selected) and 'Relative page(s):'. The 'Absolute page:' section has a text box with '1' and a description: 'Field only appears on a single page; always the same absolute page number (one-based)'. The 'Relative page(s):' section has five checkboxes: 'First page', 'First page (not only page)', 'Interior pages', 'Last page', and 'Last page (not only page)'. The 'Options' section has two checkboxes: 'Automatic continued-row merging' and 'Extract only between tags'. Below these is a 'VB script:' dropdown menu with '<None>' selected and a button with three dots. Below the dropdown is a 'Suspect level:' text box with '75' and a description: '(0-99) 0 = Form Default (75), Actual = 30'. At the bottom are three dropdown menus: 'Advanced recognition:' (set to 'Machine print text/standard OCR'), 'Activation groups:', and 'Context group:'. At the very bottom are 'Save' and 'Cancel' buttons.

Unlike Line Item Extraction Zones, Grouped Line Item Extraction Zones use only manual table decomposition.

To configure a Grouped Line Item Extraction Zone, select the **Use grouped line item style data extraction** check box. The remaining options on the tab are enabled and a preview of the Data Field Zone is displayed to the right of the **Data Field Zone** dialog box.

There are several steps required to configure a Grouped Line Item Extraction Zone:

1. **Specify How Line Item Data is to be Extracted as Keyword Values.** Each column in the group is mapped to a Keyword Type.
2. **Specify How Data is to be Extracted as Keyword Values Using Tags or Regular Expressions.** Data identified by a tag (tags can be literal text or regular expressions) or a regular expression can be assigned as Keyword Values to specified Keyword Types.
3. **Configure the Additional Zone Configuration Options.** Once you have specified how the OCR engine should address the data in the group, you can specify other configuration options for the zone (e.g., page locations, VB scripts, Suspect Level, etc.).

Identifying Keyword Values from Line Item Data

To identify Keyword Values from line item data:

1. In the preview of the Data Field Zone displayed to the right of the **Data Field Zone** dialog box, click and hold the left mouse button and use the pointer to draw a box around a column of data to be extracted as Keyword Values.

The **Modify Line Item Column** dialog box is displayed.

Modify Line Item Column

Text Extraction | Optical Mark Detection | Signature Detection

Keyword Type: Issue Date XML Node: <Runtime System Default>

Second Capture Group Keyword Type: <None> XML Node: <None>

Expression to Match:

☐ Extract capture group only

Suspect Level: 0 (0 - 99 : 0 = parent line item zone default)

☐ Value must be NULL text

☐ Column data required

☐ Enforce absolute column break on right

☐ Remove all whitespace from result

Save Cancel

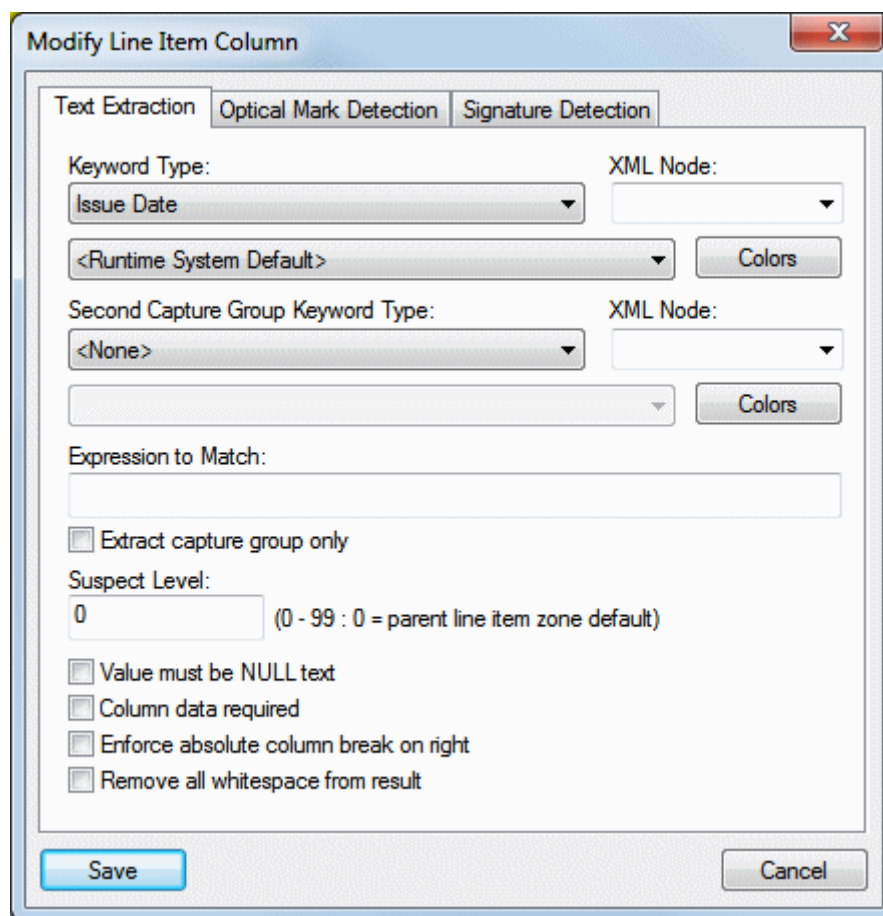
Depending on the type of Data Field Zone you are configuring, you will have different settings available to you.

- For information on configuring a text extraction zone, see [Configuring a Text Extraction Column on page 175](#).
 - For information on configuring a signature detection zone, see [Configuring a Signature Detection Column on page 179](#).
2. Click **Save**. The Column Extraction Zone is highlighted in the Data Field Zone preview.

3. If you wish to edit a configured column, double-click on its entry in the table to re-open the **Modify Keyword Result for Column** dialog box, change any desired settings, and click **Save**.
4. Repeat step 1 and the steps in the sections below for any remaining columns in the Data Field Zone preview whose data you wish to capture.

Configuring a Text Extraction Column

To begin configuring a text extraction column for Grouped Line Item Extraction, ensure that the **Text Extraction** tab is selected in the **Modify Line Item Column** dialog box. Then follow the procedures below.



1. Use the **Keyword Type** drop-down list to select the Keyword Type that values in the selected column are assigned to as Keyword Values.

To ignore the data in the selected column or capture it as XML data only, select **<None>** from the **Keyword Type** drop-down list.

Note: You can configure the Advanced Capture engine to split each value in the selected column into two separate Keyword Values by specifying a regular expression rule for the extracted text, setting two capture groups within this regular expression rule, and assigning a Keyword Type to each of these capture groups. This may be useful when two different types of data are listed on a document as one value (e.g., a transcript course subject and course number need to be captured as two separate values, but they appear together on a document as **ENG101**). If you are configuring the selected column for two Keyword Values, the first of these values (i.e., the value extracted from the first capture group) is assigned to your **Keyword Type** selection and the second of these values is assigned to your **Second Capture Group Keyword Type** selection. If you are configuring the selected column for just one Keyword Value,

however, as long as you have set multiple capture groups within the regular expression rule, only the value extracted from the first capture group is still assigned to your **Keyword Type** selection; the values extracted from the subsequent capture groups are discarded.

2. If the form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the form), the **XML Node** drop-down list is displayed next to the corresponding Keyword Type drop-down list.

Use the **XML Node** drop-down list to select the XML node (i.e., the element) that Keyword Values extracted from this column are contained in when the XML rendition of the document is created.

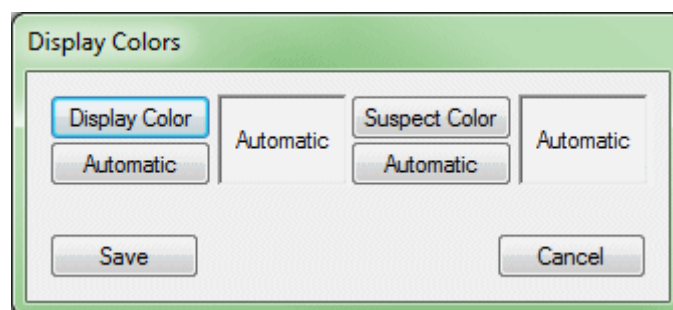
3. If you selected a **Date**, **Date & Time**, or **Currency** Keyword Type for the selected column, the **Override Default Regional Setting** drop-down list is displayed below this selected Keyword Type. If you have not selected one of these Keyword Types, the **Override Default Regional Setting** drop-down list is disabled.

When available, you can use the **Override Default Regional Setting** option to override your system's default regional settings when extracting values for the applicable Keyword Types in the Data Field Zone. To use this option, do one of the following:

- Select a regional language from the drop-down list to parse the applicable Keyword Values using the selected region's formatting rules and language-specific names for days, months, etc.
- Select **<Runtime System Default>** to maintain your system's default regional settings when parsing these values.

Note: This option does not affect which languages the OCR format is configured to recognize. For information on configuring OCR formats, see the **Full-Page OCR** module reference guide or help files.

4. If you would like to assign specific colors to the selected Keyword Type, click the **Colors** button. The **Display Colors** dialog box is displayed.



Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once processing has taken place.

- To change the display color for regular values, click **Display Color** to open your machine's color palette, select a color, and click **OK**.
- To change the display color for suspect values, click **Suspect Color** to open your machine's color palette, select a color, and click **OK**.

- To revert back to the default display color for regular or suspect values, click the **Automatic** button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values).

Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.

5. If you are configuring the selected column for two Keyword Values, use the **Second Capture Group Keyword Type** drop-down list to select the second Keyword Type that values in the selected column are assigned to as Keyword Values.
If you are configuring the selected column for just one Keyword Value, or if you wish to capture a second Keyword Value as XML data only, select **<None>** from the **Second Capture Group Keyword Type** drop-down list.

Note: If you are configuring the selected column for just one Keyword Value, and you have set capture groups within the regular expression rule, be aware that only the value extracted from the first capture group will be assigned to your **Keyword Type** drop-down selection.

6. If you wish to specify a regular expression rule for the extracted text, enter the rule in the **Expression to Match** field. The OCR engine will compare the extracted text to the defined regular expression rule; if the text is a match, the value is stored as a Keyword Value. If the text does not match, it is discarded.
For example, if you specify the following regular expression rule:
[:upper:][:lower:][:digit:][:space:]]+
Any value containing a character that is not a letter, number, or space is discarded.

Note: To access the Regular Expression Library, click in the field and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

Note: All regular expressions must be ECMA compliant.

Note: If you are configuring the selected column for two Keyword Values, you must specify a regular expression rule for the extracted text and set two capture groups within the rule to align to these two values.

Tip: Regular expressions can be used to discard column data that consists of entirely unwanted characters (e.g., a **123***456**, where you want to capture **123** and **456** as separate Keyword Values and discard the asterisk separator column). However, if the unwanted characters are located in the middle of valid data (e.g., **123***456**, where you want to capture **123456** as one Keyword Value), you can configure a Keyword Lookup/Replace dictionary entry to replace the data.

7. If you entered a regular expression rule containing a capture group in the previous step, and you would like to have Keyword Values extracted for only this capture group, select the **Extract capture group only** check box. If you would like to have Keyword Values extracted for the entire regular expression (including the capture group), deselect the **Extract capture group only** check box.

Note: This option is not available if the **Second Capture Group Keyword Type** drop-down list is set to anything other than **<None>**, or if the **Value must be NULL text** check box is selected.

8. If you wish to set a different Suspect Level for the selected Keyword Type(s) than the parent Suspect Level set for the entire Line Item Extraction Zone, enter this value (**0** to **99**) in the **Suspect Level** field. By default, this value is set to **0**, which retains the parent Suspect Level.
9. If you would like to have Keyword Values extracted from the row only if no value exists in the specified column, select the **Value must be NULL text** check box. When this option is selected, the **Column Data Required** check box is selected automatically and the **Expression to Match** field is disabled.
10. If a Keyword Value is required, select the **Column Data Required** check box. If a Keyword Value that is marked as required is missing, the data from the entire row is discarded.
11. If you wish to restrict the Keyword Values that are extracted from the column to only those characters fully contained within the column's boundaries, select the **Enforce absolute column break on right** check box.
When this option is not selected, some carryover is allowed when characters on either side of the column's right boundary are very close together, assuming that these characters belong to the same Keyword Value. The characters just outside the right boundary will be included in the extracted value.
12. If you wish to remove all whitespace (i.e., spaces, paragraph returns, etc.) from a Keyword Value after it is read by the OCR engine, select the **Remove all whitespace from result** check box.

For example, if the Keyword Value read by the OCR engine is **PSY 103** and this option is selected, the Keyword Value is modified to **PSY103**.

Note: If you specified a regular expression rule for the extracted text in the **Expression to Match** field, the OCR engine will compare the extracted text without any whitespace to the defined regular expression rule.

Configuring a Signature Detection Column

To begin configuring a signature detection column for Line Item Extraction, ensure that the **Signature Detection** tab is selected in the **Modify Line Item Column** dialog box. Then follow the procedures below.

The screenshot shows the 'Modify Line Item Column' dialog box with the 'Signature Detection' tab selected. The 'Keyword type:' dropdown is set to 'Last Name'. The 'Keyword Values' section has empty text boxes for 'Positive result value:' and 'Negative result value:'. The 'XML Node:' dropdown is empty. The 'Sensitivity' section has radio buttons for 'Highest' (selected), 'Low', 'Lower', 'Lowest', and 'Custom:'. Below 'Custom:' is a text box and the text '1-99 : greater value is more sensitive'. The 'Options' section has radio buttons for 'Signature is framed/underlined' (selected), 'Signature area is unframed', and a checkbox for 'Exclude machine print text'. A 'Colors' button is at the bottom right of the main area. 'Save' and 'Cancel' buttons are at the bottom of the dialog.

1. Use the **Keyword Type** drop-down list to select the Keyword Type that values in the selected column are assigned to as Keyword Values.
To ignore the data in the selected column or capture it as XML data only, select **<None>** from the **Keyword Type** drop-down list.
2. If the form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the form), the **XML Node** drop-down list is displayed.
Use the **XML Node** drop-down list to select the XML node (i.e., the element) that Keyword Values extracted from this column are contained in when the XML rendition of the document is created.
3. In the **Keyword Values** section, do the following:
 - a. In the **Positive Result Value** field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is present.
 - b. In the **Negative Result Value** field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is not present.

For example: you are processing an application that requires a signature to indicate that the data has been verified. Depending on if a signature is present or not, a Keyword Value is assigned to the **Verified** Keyword Type.

- If a signature is detected, **Yes** is assigned as the **Verified** Keyword Value.
- If a signature is not detected, **No** is assigned as the **Verified** Keyword Value.

4. In the **Options** section, do the following:

- If the signature is located in a pre-defined space, such as a box or on a signature line, select **Signature is framed/underlined**.
- If the signature is located in an undefined or blank space, select **Signature area is unframed**.
- To ignore machine-printed text in the Data Field Zone (e.g., **Signature, Applicant Signature**, etc.), select the **Exclude machine print text** check box.

This prevents the signature detection results from being impacted by the presence of machine-printed text.

5. In the **Sensitivity** section, select a radio button to determine the Suspect Level for this Data Field Zone.

- **Highest**. Sets the Suspect Level for this Data Field Zone to 70.
- **Low**. Sets the Suspect Level for this Data Field Zone to 75.
- **Lower**. Sets the Suspect Level for this Data Field Zone to 80.
- **Lowest**. Sets the Suspect Level for this Data Field Zone to 85.
- **Custom**. Allows you to enter a value (from 1 to 99) for the Suspect Level for this Data Field Zone. If enter a value of 0 or leave the field blank, the value will reset to 50.

The Suspect Level is the level of confidence placed in the processing results for this field.

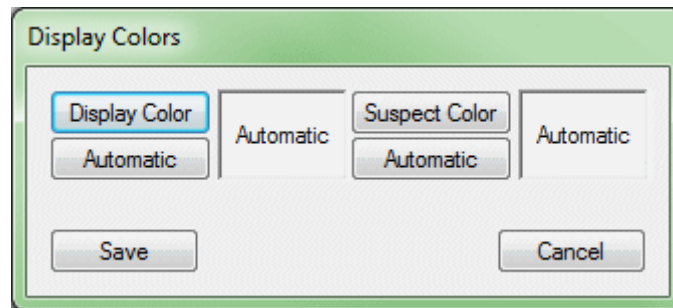
After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.

The Sensitivity level selected for this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.

For example, setting the Sensitivity to **Lowest** would indicate you have a fair amount of confidence in the result returned by the OCR engine because few higher scores could be returned and fewer results would be determined to be suspect.

Setting the Suspect Level to **Highest** would indicate you have less confidence in the result because a great number of lower scores could be returned and more results would be determined as suspect.

6. If you would like to assign specific colors to the Keyword Type configured for the Data Field Zone, click the **Colors** button. The **Display Colors** dialog box is displayed.



Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once processing has taken place.

- To change the display color for regular values, click **Display Color** to open your machine's color palette, select a color, and click **OK**.
- To change the display color for suspect values, click **Suspect Color** to open your machine's color palette, select a color, and click **OK**.
- To revert back to the default display color for regular or suspect values, click the **Automatic** button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values).

Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.

Identifying Keyword Values Using Tags or Regular Expressions

Note: To access the Regular Expression Library, click in any text field that will receive a regular expression and press **F2**. See [The Regular Expression Library on page 247](#) for more information.

To identify Keyword Values using tags (literal or regular expression) or regular expressions:

1. From the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box, click **New Tag** to create the text tag or regular expression to be used to identify the Keyword Value. The **Modify Line Item Tagged Value** dialog box is displayed.

Modify Line Item Tagged Value

Keyword Type:
Issue Date

Select the desired radio button on the left to activate the associated entry fields on the right:

Tag Text:
☒

Regular Expression Match:
☐
☐ Extract capture group only

Person Name Extraction:
☐
First Name: Middle Name:
☐ Case Insensitive ☐ Force LFM with or without comma

Meta Data Value:
☐

Line Item Totals Marker:
☐
☐ No tag value/always last row in group

☒ This item PRECEDES the associated line item group
☐ This item FOLLOWS the associated line item group
☐ Treat this row as a line item after reading this tagged value
☐ Treat this row as a line item/tag value starts new group
☐ The position of this item is not significant ☐ Carry Across Groups
☐ This entire line of data should be ignored

☐ Duplicate value does not start new logical group
☐ Force logical group to end regardless of future repeating values
☐ Data value required ☐ Remove all whitespace from result
☐ Ignore all whitespace in tag or regular expression search

<Runtime System Default>

XML

2. Using the **Keyword Type** drop-down list, select the Keyword Type that the value you are identifying is to be assigned to.
If you only want to capture this value as XML data (not as a Keyword Value), select **<None>**.
3. Select the radio button that describes how the OCR engine will identify the Keyword Value:
 - **Tag Text.** The OCR engine searches the Data Field Zone for a tag (e.g., **Name:**, **Date:**, etc.) and identifies the value to the right of the tag as the Keyword Value.

Note: When identifying a Keyword Value, a tag can only be used to identify the value to the right of the tag. A tag cannot be used to identify the value below the tag.

Enter the tag text in the **Tag Text** field. Tags can be literal (i.e., the text specified in the field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and the text matching that pattern is the tag).

Literal tags must be enclosed in quotation marks in order to differentiate them from regular expression tags. Regular expressions must be ECMA compliant.

- **Regular Expression Match.** The OCR engine searches the zone for text that matches the regular expression rule. If text matching the pattern set by the regular expression rule is found, it is identified as the Keyword Value.

Enter the regular expression in the **Regular Expression Match** field. All regular expressions must be ECMA compliant.

- Select the **Extract Capture Group Only** check box if you would like to use a regular expression to identify the Keyword Value, but you would only like to extract part of the value identified by the regular expression as a Keyword Value. The portion of the Keyword Value you wish to capture must be enclosed in parentheses in the regular expression.

For example: A document displays a student's Social Security Number (i.e., **SSN: 123-45-6789**), of which you only wish to capture the last four digits (**6789**) as a Keyword Value. To capture this value, you would select the **Extract Capture Group Only** check box and enter the following regular expression in the **Regular Expression Match** field: `\d{3}-\d{2}-(\d{4})`.

Additionally, you can use the **Extract Capture Group Only** option to capture only one value from a row of line item data. For example: a document contains a **Totals** line that contains tallied data for five columns, of which you only want to capture the Total value from the third column (i.e., **\$310.71**) as a Keyword Value.

TOTAL \$101.35 \$177.01 \$310.71 \$285.00 \$526.22

To capture this value, you would select the **Extract Capture Group Only** check box and enter the following regular expression in the **Regular Expression Match** field:

`TOTAL\s+\$\d{3}.\d{2}\s+\$\d{3}.\d{2}\s+(\$\d{3}.\d{2})\s+\$\d{3}.\d{2}\s+\$\d{3}.\d{2}`

- **Person Name Extraction.** Similar to **Tag Text** option, the OCR engine searches the Data Field Zone for a literal tag or regular expression tag; however, once the tag is identified, the value following the tag is stored as a name instead of a single Keyword Value.

When a name is stored, the extracted value is parsed into separate values (first name, middle name, and last name) and each value is stored separately as an individual Keyword Value.

The following are valid name formats (middle name values are not required):

- **Last, First [Middle]**
- **LAST, FIRST [MIDDLE]**
- **First [Middle] Last**
- **FIRST [MIDDLE] LAST**

To select the Keyword Types that the name values are assigned to:

- Use the **Last Name Keyword** drop-down list (renamed from **Keyword Type**) to select the Keyword Type that the last name is assigned to.
- Use the **First Name** drop-down list to select the Keyword Type that the first name is assigned to.
- Use the **Middle Name** drop-down list to select the Keyword Type that the middle name is assigned to.

Enter the tag text in the **Person Name Extraction** field. Tags can be literal (i.e., the

text specified in the field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and the text matching that pattern is the tag). Literal tags must be enclosed in quotation marks in order to differentiate them from regular expression tags. Regular expressions must be ECMA compliant.

- **Meta Data Value.** The OCR engine identifies the Keyword Value using data obtained when processing the document. Use the **Meta Data Value** drop-down list to select the data to be used as the Keyword Value:
 - **<Group Number>.** This option refers to the order of the groups in the zone. For example, if a zone contains three groups (e.g., **Fall 2009, Spring 2010, Fall 2010**), then Group Number for the first group (**Fall 2009**) is **1**, the Group Number for the second group (**Spring 2010**) is **2**, and the Group Number for the third group (**Fall 2010**) is **3**.
 - **<Line Number>.** This option refers to the line on which the data is displayed once the zone is processed. The first line at the top of the zone is 1, the second line is 2, etc.
 - **<None>.** This option is reserved for future use.
 - **<Page Number>.** This option refers to the page on which the data is displayed.
- **Line Item Totals Marker.** The OCR engine identifies the Keyword Value as the last line in the group (i.e., a “totals” line where column data is tallied). Enter the tag text in the **Line Item Totals** field. Tags can be literal (i.e., the text specified in the field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and the text matching that pattern is the tag). Select **No tag value/always last line in group** check box if the data is not denoted by a tag, but should captured because it is the last line in the group.

4. Select the radio button that describes the location of tag or regular expression in the group.
 - **This item PRECEDES the associated line item group.** Select this radio button if the Keyword Value is located before the line item data.
 - When this option is selected, the **Duplicate value does not start new logical group** check box is enabled. Select this check box to continue the same group when a duplicate tag value is found. If you do not select this option, a new logical group will be started when a duplicate tag value is found.
 - **This item FOLLOWS the associated line item group.** Select this radio button if the Keyword Value is located after the line item data.
 - When this option is selected, the **Force logical group to end regardless of future repeating values** check box is enabled. Select this check box to force the current group of data to end and ensure that a new group will be started once a new line item tag value is found for the next group. This prevents the next group from being appended to the current group, even if the first line of tagged values in the next group is identical.

This may be desirable in explanation of benefit forms (EOBs), for instance, when a single patient has multiple claims with different claim numbers, but the claim numbers do not appear in the first line of tags for the new group. In this case, even though the first line of tags will be identical, the new group should not be appended to the current group.

Note: If there is a page break in the middle of the associated line item group and the **FOLLOWS** tag appears on the second page, this tag does not capture any line item values from the associated group on the first page.

- **Treat this row as a line item after reading this tagged value.** Select this radio button if the line item containing the Keyword Value should also be treated as a row of line item data.

- **Treat this row as a line item/tag value starts new group.** Select this radio button if the line item containing the Keyword Value should be treated as a row of line item data and the Keyword Value that starts each new group is located in the leftmost position of the line item, concurrent with the line items for the first row of each group and not appearing again until the first row of the next group.

In the example below, the highlighted Keyword Values mark the beginning of each group:

TERM	DEPT	COURSE NUMBER	DESCRIPTIVE TITLE	SEM HRS	GRADE
85/FA	GEN	1000	FRESHMEN SURVEY	3.00	A
	ENG	1001	ENGLISH 101	3.00	A
	MAT	2001	COLLEGE MATH	3.00	B
	HIS	2002	U.S. HISTORY 2	3.00	A
	HIS	2650	INTERNATIONAL POLITICS	4.00	B
86/SP	HIS	2100	WORLD WAR II	3.00	A
	SCI	2235	BIOLOGY 2	3.00	B
	HUM	1250	INTRO TO THE ARTS	3.00	A
	HIS	2780	LATIN AMERICAN HISTORY	3.00	B
	HIS	3500	COLONIALISM	4.00	B
	HIS	3225	THE NUCLEAR AGE	4.00	A

- **The position of this item is not significant.** Select this radio button if the location of the Keyword Value is not significant in the zone.
 - When both this option and either the **Tag Text**, **Regular Expression Match**, or **Person Name Extraction** option is selected, the **Carry Across Groups** check box is enabled. Select this check box to apply the extracted value to the current group and all subsequent groups until a new value is read for this tag. When a new value is read, this loop starts over.

For example, an account number may be extracted as a top-level value. This account number applies to the current group and all subsequent groups, each of which has start and end markers, until a new account number is extracted. The pattern then repeats for the new account number.
 - **This entire line of data should be ignored.** Select this radio button if the entire line containing the matching tag or regular expression should be ignored.
 - This option is only available if the value is configured to be captured as XML data (i.e., if **<None>** is selected in the **Keyword Type** drop-down list).
5. Select the **Data Value required** check box if the Keyword Value is a required value. If this check box is selected and no value is found, then the entire group is discarded.

6. If you wish to remove all whitespace (i.e., spaces, paragraph returns, etc.) from a Keyword Value after it is read by the OCR engine, select the **Remove all whitespace from result** check box.

For example, if the Keyword Value read by the OCR engine is **PSY 103** and this option is selected, the Keyword Value is modified to **PSY103**.

Note: The **Remove all whitespace from result** check box is only available if you selected the **Tag Text** or **Regular Expression Match** option in step 3. Once the OCR engine finds a value following the specified tag or matching the specified regular expression, it removes all whitespace from the result.

7. If you wish to ignore all whitespace when searching for a text tag or potential regular expression match, select the **Ignore all whitespace in tag or regular expression search** check box.

Note: The **Ignore all whitespace in tag or regular expression search** check box is only available if you selected the **Tag Text**, **Regular Expression Match**, **Person Name Extraction**, or **Line Item Totals Marker** option in step 3. The OCR engine removes all whitespace from the text it is searching for before finding a match.

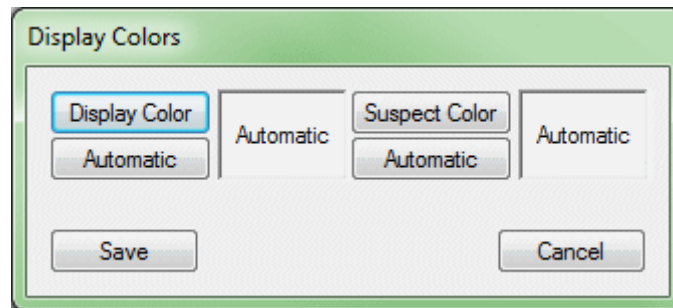
8. If you selected a **Date**, **Date & Time**, or **Currency** Keyword Type in step 2, the **Override Default Regional Setting** drop-down list is displayed above the **Save** button. If you have not selected one of these Keyword Types, or if you have selected the **Person Name Extraction**, **Meta Data Value**, or **Line Item Totals Marker** option, the **Override Default Regional Setting** drop-down list is disabled.

When available, you can use the **Override Default Regional Setting** option to override your system's default regional settings when extracting values for the applicable Keyword Types in the Data Field Zone. To use this option, do one of the following:

- Select a regional language from the drop-down list to parse the applicable Keyword Values using the selected region's formatting rules and language-specific names for days, months, etc.
- Select **<Runtime System Default>** to maintain your system's default regional settings when parsing these values.

Note: This option does not affect which languages the OCR format is configured to recognize. For information on configuring OCR formats, see the **Full-Page OCR** module reference guide or help files.

9. If you would like to assign specific colors to the Keyword Type configured for the Data Field Zone, click the **Colors** button. The **Display Colors** dialog box is displayed.



Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once processing has taken place.

- To change the display color for regular values, click **Display Color** to open your machine's color palette, select a color, and click **OK**.
- To change the display color for suspect values, click **Suspect Color** to open your machine's color palette, select a color, and click **OK**.
- To revert back to the default display color for regular or suspect values, click the **Automatic** button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values).

Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.

10. If the form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the form), the **XML Node** drop-down list is displayed.
- Use the **XML Node** drop-down list to select the XML node (i.e., the element) that Keyword Value is contained in when the XML rendition of the document is created.
11. Click **Save**. The **Modify Line Item Tagged Value** dialog box is closed and you are returned to the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box.
12. If you wish to edit a configured Keyword Value identification type, double-click on its entry in the table to re-open the **Modify Line Item Tagged Value** dialog box, change any desired settings, and click **Save**.

Note: If you hover over the list of configured tags/expressions in the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box, the list will be expanded to show more rows. When you stop hovering over the list, it will return to the default size.

Configuring Additional Grouped Line Item Extraction Options

The following configuration options for the zone can be configured in the Data Field Zone dialog box.

Data Field Zone Grouped Line Item Extraction Option	Description
XML Node (Parent)	<p>Note: This field is only displayed if the Advanced Capture form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Advanced Capture form).</p> <p>Enter the name of the XML parent node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p> <p>Note: While the XML Node (Parent) name may include alphanumeric characters, underscores (_), hyphens (-), or periods (.), it must begin with either a letter or an underscore.</p>

Data Field Zone Grouped Line Item Extraction Option	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the groups are only displayed on one page and are always displayed on the same page (e.g., the groups are always displayed on page 1). Enter the page number that the groups are located on in the associated field. • Select the Relative page radio button if the groups may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) the groups may be located on. <ul style="list-style-type: none"> Select the First page check box if the groups are located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes. Select the First page (not only page) check box if the groups are located on the first page and other pages in the document. Select the Interior pages check box if the groups are located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes. Select the Last page check box if the groups are located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes. Select the Last page (not only page) check box if the groups are located on the last page and other pages in the document.
Automatic continued-row merging	<p>Select the Automatic continued-row merging check box to compare the average spacing between rows.</p> <p>If a row is closer than average to the row that precedes it and the row does not have data in at least half of its processed columns, the OCR engine recognizes it as a continuation of the previous row and merges the extracted values.</p>
Extract only between tags	<p>Select the Extract only between tags check box if all Keyword Values to be extracted can be found between two tags (i.e., a start marker and an end marker).</p> <hr/> <p>Tip: This option should be selected when your groups contain “extra” data that can be ignored by the OCR engine because it does not contain Keyword Values.</p> <hr/>

Data Field Zone Grouped Line Item Extraction Option	Description
VB script	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Advanced Capture form is displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that suspect Keyword Values are being consistently identified.</p> <hr/>

Data Field Zone Grouped Line Item Extraction Option	Description
Advanced Recognition	<p>Note: Options involving ICR processing below are only enabled if your solution is licensed for Intelligent Character Recognition (ICR).</p> <hr/> <p>Note: The OCR engine does not support Asian characters when reading dot matrix-printed text.</p> <hr/> <p>Using this drop-down list, select the type of processing you would like to perform on this Data Field Zone.</p> <ul style="list-style-type: none"> • Machine print text/OCR. The OCR engine will perform optical character recognition to read machine-printed text. This option is selected by default. • Machine print text/detect structure. The OCR engine will attempt to determine whether the value consists of machine-printed text, dot matrix-printed text, or hand-written text, and then use the appropriate type of character recognition to read the text. • Machine print text/flowing text. The OCR engine will attempt to read machine-printed text within the Data Field Zone as if the text flowed like a paragraph with similar character sizes, spacing, and fonts. If this option is not selected, by default, the OCR engine will attempt to read machine-printed text using the automatic zoning recognition mode, which simply attempts to determine what the character sizes, spacing, and fonts look like on their own (i.e., independent of their positioning within a paragraph or section of flowing text). • Machine print text/dot matrix printer. The OCR engine will perform optical character recognition to read dot matrix-printed text. • Handwriting/ICR -numerals/grouping punctuation (North American style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the North American style (e.g., the 7 character is not crossed). • Handwriting/ICR -numerals/grouping punctuation (European style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the European style (e.g., the 7 character is crossed). • Handwriting/ICR - alphanumeric/punctuation. The OCR engine will perform intelligent character recognition in order to read hand-written text. This option will attempt to recognize all alphanumeric characters.

Data Field Zone Grouped Line Item Extraction Option	Description
Advanced Recognition (cont.)	<ul style="list-style-type: none"> • Auto-detect ICR/OCR - Default to ICR/Alphanumeric. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for alphanumeric characters will be used by default. • Auto-detect ICR/OCR - Default to ICR/Numeric North American. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, North American-style characters (e.g., the 7 character is not crossed) will be used by default. • Auto-detect ICR/OCR - Default to ICR/Numeric European. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, European-style characters (e.g., the 7 character is crossed) will be used by default. • Auto-detect ICR/OCR - Default to OCR. The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, optical character recognition will be used by default. <hr/> <p>Note: The Auto-detect ICR/OCR options may not work properly if the Data Field Zone contains less than 25 characters.</p> <hr/> <p>Tip: Of the two Auto-detect ICR/OCR options, Default to ICR is more likely to produce the best results when the type of text cannot be determined. This is because the OCR engine's intelligent character recognition generally reads machine-printed text more accurately than the engine's optical character recognition reads hand-written text.</p> <hr/> <p>Note: The Grouped Line Item Extraction Data Field Zone cannot contain a mixture of OCR and ICR values. All values within the zone must use one type of recognition only.</p> <hr/>

Data Field Zone Grouped Line Item Extraction Option	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Data Field Zone Grouped Line Item Extraction Option	Description
Activation groups (cont.)	Alternatively, you can assign a form definition group as the Data Field Zone's activation group to activate the zone for processing. Form definition groups can be used to extract only specific types of information (e.g., header data vs. detail data) during processing. In the Activation groups drop-down list, form definition groups are enclosed in brackets (e.g., [Group1]).
Context group	<p>You can logically group Data Field Zones using context groups. These groupings can be used to identify Keywords that belong to one Multi-Instance Keyword Type Group but are split across multiple Data Field Zones, for instance.</p> <p>Use the Context group field to enter or select a logical group name for the Data Field Zones that identify Keywords belonging to an MIKG (or other logical grouping). The Advanced Capture engine will attempt to maintain this Keyword grouping on the resulting document.</p>

Configuring an Optical Mark Data Field Zone

An OMR Data Field Zone is configured using the options on the Optical Mark Field tab of the **Data Field Zone** dialog box.

The screenshot shows the 'Data Field Zone' dialog box with the 'Optical Mark Field' tab selected. The 'Keyword type' is set to 'Course ID' and the 'XML Node' is empty. The 'OMR item group mode' checkbox is unchecked. Under 'Keyword Values', there are input fields for 'Positive result value' and 'Negative result value'. The 'Page Location(s)' section has 'Absolute page' selected with a value of '1', and a note stating 'Field only appears on a single page; always the same absolute page number (one-based)'. There are also options for 'Relative page(s)' with checkboxes for 'First page', 'First page (not only page)', 'Interior pages', 'Last page', and 'Last page (not only page)'. The 'Options' section includes 'Mark frame detection' with radio buttons for 'Framed', 'Unframed', and 'Auto-detect' (selected), and 'Sensitivity' with radio buttons for 'Highest' (selected), 'Low', 'Lower', and 'Lowest'. There is also a checkbox for 'Support filled-in-error' detection. A 'Test' button is present. The 'Advanced filtering' section has a 'VB script' dropdown set to '<None>' and a 'Colors' button. At the bottom, there are 'Keyword association' and 'Activation groups' dropdowns, and 'Save' and 'Cancel' buttons.

Data Field Zone

Text Field | **Line Item Extraction** | Grouped Line Item Extraction

Optical Mark Field | Signature Detection Field | Snippet

Keyword type: Course ID | XML Node:

☐ OMR item group mode

Keyword Values

Positive result value:

Negative result value:

Page Location(s)

☒ Absolute page: 1 | ☐ Relative page(s):

Field only appears on a single page; always the same absolute page number (one-based)

☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)

Options

Mark frame detection:

☐ Framed
☐ Unframed
☒ Auto-detect

Sensitivity:

☒ Highest ☐ Low ☐ Lower ☐ Lowest

☐ Support filled-in-error' detection

Test

Advanced filtering:

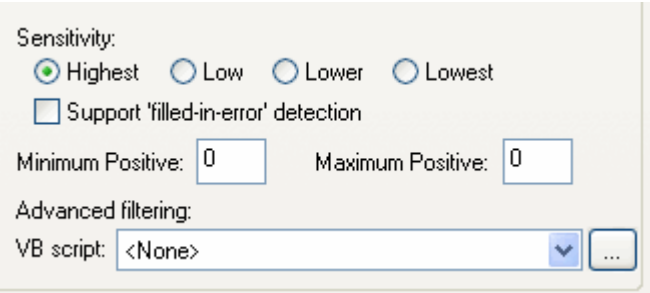
VB script: <None> ...

Keyword association: | Colors

Activation groups:

Save Cancel

Data Field Zone Optical Mark Field Options	Description
Keyword type	<p>Using this drop-down list, select the Keyword Type that this Data Field Zone will assign a Keyword Value to.</p> <p>If the currently-displayed Advanced Capture form has an assigned Document Type, only the Keyword Types associated with the assigned Document Type are available in the drop-down list. If no Document Type is assigned to the currently-displayed Advanced Capture form, only the <None> option is available in the drop-down list.</p> <hr/> <p>Note: If the Keyword Value identified by the OCR engine exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is truncated to fit this length.</p> <hr/>
XML Node	<hr/> <p>Note: This field is only displayed if the Advanced Capture form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Advanced Capture form).</p> <hr/> <p>Note: This field is disabled if the OMR item group mode check box is selected.</p> <hr/> <p>Enter the name of the XML node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p> <hr/> <p>Note: While the XML Node name may include alphanumeric characters, underscores (_), hyphens (-), or periods (.), it must begin with either a letter or an underscore.</p> <hr/>

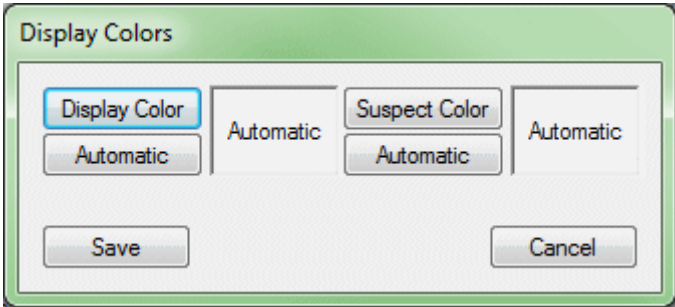
Data Field Zone Optical Mark Field Options	Description
OMR item group mode	<p>Note: To use this feature, multiple optical marks must be reside in the Data Field Zone.</p> <p>Select this check box to allow the OCR engine to process multiple optical marks inside the Data Field Zone. Each optical mark must share all of the same configuration options (e.g., Keyword Type, Framed/Unframed, Sensitivity level, etc.) except for their associated positive and negative Keyword Values.</p> <p>When this check box is selected, the Keyword Values section of the Data Field Zone dialog box is displayed as a list containing Item #, Positive Value, and Negative Value, and an image snippet of the area of the document selected for the Data Field Zone is displayed next to the Data Field Zone dialog box.</p> <p>From the image snippet, use the pointer to draw a box around the first optical mark in the Data Field Zone. The Modify Keyword Result for OMR Group Item dialog box is displayed, allowing you to specify the positive and negative Keyword Value for that optical mark. Click Save when finished. To re-open the dialog box and edit these values once they have been configured, double-click on the corresponding entry in the table, make the desired changes, and click Save.</p> <p>If the Advanced Capture form is configured to create XML renditions of the documents it is matched to, the XML Node field is displayed, giving you the opportunity to name the XML node that the Keyword Values are contained in when the XML rendition is created.</p> <p>When the OMR item group mode option is selected, the Minimum Positive and Maximum Positive options are displayed between the Sensitivity and Advanced filtering options.</p>  <p>In the Minimum Positive and Maximum Positive fields, you can specify the minimum and maximum number of optical marks that the OCR engine should detect within the Data Field Zone, respectively. If the number of optical marks detected is less than the specified minimum or greater than the specified maximum value, the values for the Data Field Zone will be automatically marked as suspect.</p>

Data Field Zone Optical Mark Field Options	Description
OMR item group mode (cont.)	<p>For example, if a form has two check boxes for gender, and you only expect one box to be checked (i.e., either male or female), you can specify both the Minimum Positive and Maximum Positive values to be 1. If neither box is checked, or if both boxes are checked, the values for the zone will be marked as suspect.</p> <p>By default, the Minimum Positive and Maximum Positive values are both set to 0, which keeps minimum/maximum optical mark validation disabled.</p> <p>When finished, click Save. The optical mark configuration zone is highlighted on the image snippet. Right-click on the optical mark configuration within the image snippet to move, resize, or delete the configuration information for that optical mark configuration zone.</p> <p>Repeat this process for each optical mark displayed in the Data Field Zone.</p>
Keyword Values	<p>In the Positive Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that an optical mark is present.</p> <p>In the Negative Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that an optical mark is not present.</p> <p>For example: you are processing an application and one of the questions asks the applicant to select a check box if he/she has previously applied to the university. Depending on the answer to this question, a Keyword Value is assigned to the Previously Applied Keyword Type.</p> <ul style="list-style-type: none"> • If the check box is selected, Yes is assigned as the Previously Applied Keyword Value. • If the check box is not selected, No is assigned as the Previously Applied Keyword Value.

Data Field Zone Optical Mark Field Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the data being read in the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the optical mark is always displayed on page 1). Enter the page number that the optical mark is located on in the associated field. • Select the Relative page radio button if the data may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) the optical mark may be located on. <ul style="list-style-type: none"> Select the First page check box if the optical mark is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes. Select the First page (not only page) check box if the optical mark is located on the first page and other pages in the document. Select the Interior pages check box if the optical mark is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes. Select the Last page check box if the optical mark is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes. Select the Last page (not only page) check box if the optical mark is located on the last page and other pages in the document.
Mark Frame Detection	<p>Select the radio button that describes the optical mark being read by the OCR engine.</p> <ul style="list-style-type: none"> • Framed. Select the Framed radio button if the optical mark is located in a pre-defined space, such as a check box or bubble. • Unframed. Select the Unframed radio button if the optical mark is located in an undefined or blank space. • Auto-detect. Select the Auto-detect radio button if the OCR engine should determine if the optical mark is framed or unframed. <hr/> <p>Tip: Selecting the Framed or Unframed selection, where appropriate, will help to increase the accuracy of the Advanced Capture process.</p> <hr/>

Data Field Zone Optical Mark Field Options	Description
Sensitivity	<p>Select a radio button to determine the Suspect Level for this Data Field Zone.</p> <ul style="list-style-type: none">• Highest. Sets the Suspect Level for this Data Field Zone to 70.• Low. Sets the Suspect Level for this Data Field Zone to 75.• Lower. Sets the Suspect Level for this Data Field Zone to 80.• Lowest. Sets the Suspect Level for this Data Field Zone to 85. <p>The Suspect Level is the level of confidence placed in the processing results for this field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The Sensitivity level selected for this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Sensitivity to Lowest would indicate you have a fair amount of confidence in the result returned by the OCR engine because few higher scores could be returned and fewer results would be determined to be suspect.</p> <p>Setting the Suspect Level to Highest would indicate you have less confidence in the result because a great number of lower scores could be returned and more results would be determined as suspect.</p>

Data Field Zone Optical Mark Field Options	Description
Support 'filled-in error' detection	<p>Note: This option is enabled only when the Framed or Auto-detect radio button is selected in the Options section.</p> <p>Select this check box to enable optical mark "filled in error" detection.</p> <p>When this option is selected, the OCR engine will attempt to determine if a user has "crossed out" an optical mark on the document in order to indicate it is not present.</p> <p>For example, users occasionally select a check box and later realize that they made an error (i.e., they meant to leave the check box unselected), so they attempt to cross it out. This option allows the OCR engine to attempt to distinguish between a selected check box and a crossed-out check box.</p> <p>If the OCR engine determines that a field has been marked in error (i.e., crossed out), the field is marked as suspect.</p> <p>If this option is not selected, the OCR engine will not attempt to discern the difference between a selected check box and a crossed-out check box.</p>
Advanced filtering	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>
Test	<p>Click the Test button to have the OCR engine perform a test process on the Data Field Zone and attempt to detect an optical mark using the options you have specified. If the Advanced Capture engine is configured to attempt to compensate for offset data, the resulting offset/scaling adjustments will take place during the test.</p> <p>Once the test process is complete, a dialog box is displayed indicating if an optical mark was detected and the Keyword Value that would be assigned based on this result. The offset/scaling information is also displayed in the dialog box.</p>
Keyword association	<p>You can logically group Data Field Zones using Keyword association. These groupings can be used to identify Keywords that belong to a Multi-Instance Keyword Type Group.</p> <p>Use the Keyword association field to enter or select a logical group name for the Data Field Zones that identify Keywords belonging to an MIKG. The Advanced Capture engine will attempt to maintain this Keyword grouping on the resulting document.</p>

Data Field Zone Optical Mark Field Options	Description
Colors	<p>If you would like to assign specific colors to the Keyword Type configured for the Data Field Zone, click the Colors button. The Display Colors dialog box is displayed.</p>  <p>Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once Advanced Capture processing has taken place.</p> <ul style="list-style-type: none"> • To change the display color for regular values, click Display Color to open your machine's color palette, select a color, and click OK. • To change the display color for suspect values, click Suspect Color to open your machine's color palette, select a color, and click OK. • To revert back to the default display color for regular or suspect values, click the Automatic button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values). <hr/> <p>Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.</p> <hr/>

Data Field Zone Optical Mark Field Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Data Field Zone Optical Mark Field Options	Description
Activation groups (cont.)	Alternatively, you can assign a form definition group as the Data Field Zone's activation group to activate the zone for processing. Form definition groups can be used to extract only specific types of information (e.g., header data vs. detail data) during processing. In the Activation groups drop-down list, form definition groups are enclosed in brackets (e.g., [Group1]).

Configuring a Signature Detection Data Field Zone

A Signature Detection Data Field Zone is configured using the options on the Signature Detection Field tab of the **Data Field Zone** dialog box.

The screenshot shows the 'Data Field Zone' dialog box with the 'Signature Detection Field' tab selected. The dialog has a blue title bar with a close button. Below the title bar are four tabs: 'Text Field', 'Line Item Extraction', 'Grouped Line Item Extraction', and 'Signature Detection Field' (which is active). Below these tabs are three sub-tabs: 'Optical Mark Field', 'Signature Detection Field', and 'Snippet'. The 'Signature Detection Field' sub-tab contains the following configuration options:

- Keyword type:** A dropdown menu set to 'Last Name'.
- XML Node:** A dropdown menu.
- Keyword Values:**
 - Positive result value:** A text input field.
 - Negative result value:** A text input field.
- Page Location(s):**
 - Absolute page:** A radio button (selected) and a text input field containing '1'. Below it, text reads: 'Field only appears on a single page; always the same absolute page number (one-based)'.
 - Relative page(s):** A radio button (unselected) with a list of checkboxes:
 - ☐ First page
 - ☐ First page (not only page)
 - ☐ Interior pages
 - ☐ Last page
 - ☐ Last page (not only page)
- Options:**
 - ☒ Signature is framed/underlined
 - ☐ Signature area is unframed
 - ☐ Exclude machine print text
 - Sensitivity:**
 - ☒ Highest
 - ☐ Low
 - ☐ Lower
 - ☐ Lowest
 - ☐ Custom: [text input field] 1-99 : greater value is more sensitive
- Advanced filtering:**
 - VB script:** A dropdown menu set to '<None>' with a '...' button next to it.
- Test:** A button.
- Keyword association:** A dropdown menu.
- Colors:** A button.
- Activation groups:** A dropdown menu.

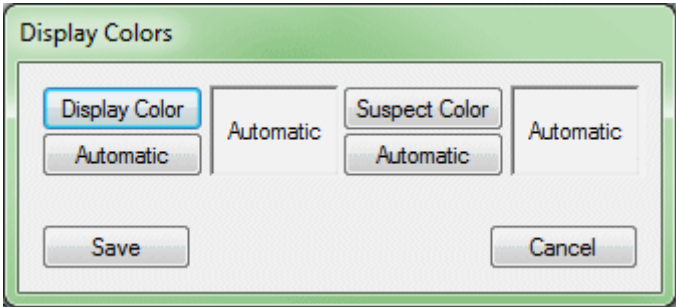
At the bottom of the dialog are two buttons: 'Save' and 'Cancel'.

Data Field Zone Signature Detection Field Options	Description
Keyword type	<p>Using this drop-down list, select the Keyword Type that this Data Field Zone will assign a Keyword Value to.</p> <p>If the currently-displayed Advanced Capture form has an assigned Document Type, only the Keyword Types associated with the assigned Document Type are available in the drop-down list. If no Document Type is assigned to the currently-displayed Advanced Capture form, only the <None> option is available in the drop-down list.</p> <hr/> <p>Note: If the Keyword Value identified by the OCR engine exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is truncated to fit this length.</p> <hr/>
XML Node	<hr/> <p>Note: This field is only displayed if the Advanced Capture form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Advanced Capture form).</p> <hr/> <p>Enter the name of the XML node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p> <hr/> <p>Note: While the XML Node (Parent) name may include alphanumeric characters, underscores (_), hyphens (-), or periods (.), it must begin with either a letter or an underscore.</p> <hr/>
Keyword Values	<p>In the Positive Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is present. In the Negative Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is not present.</p> <p>For example: you are processing an application that requires a signature to indicate that the data has been verified. Depending on if a signature is present or not, a Keyword Value is assigned to the Verified Keyword Type.</p> <ul style="list-style-type: none"> • If a signature is detected, Yes is assigned as the Verified Keyword Value. • If a signature is not detected, No is assigned as the Verified Keyword Value.

Data Field Zone Signature Detection Field Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the data being read in the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the signature is always displayed on page 1). Enter the page number that the optical mark is located on in the associated field. • Select the Relative page radio button if the data may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) the signature may be located on. <p>Select the First page check box if the signature is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the signature is located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the signature is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the signature is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the signature is located on the last page and other pages in the document.</p>
Options	<p>Select the radio button that describes the signature being read by the OCR engine.</p> <ul style="list-style-type: none"> • Signature is framed/underlined. Select this radio button if the signature is located in a pre-defined space, such as a box or on a signature line. • Signature area is unframed. Select this radio button if the signature is located in an undefined or blank space. <p>To ignore machine-printed text in the Data Field Zone (e.g., Signature, Applicant Signature, etc.), select the Exclude machine print text check box.</p> <p>This prevents the signature detection results from being impacted by the presence of machine-printed text.</p>

Data Field Zone Signature Detection Field Options	Description
Sensitivity	<p>Select a radio button to determine the Suspect Level for this Data Field Zone.</p> <ul style="list-style-type: none"> • Highest. Sets the Suspect Level for this Data Field Zone to 70. • Low. Sets the Suspect Level for this Data Field Zone to 75. • Lower. Sets the Suspect Level for this Data Field Zone to 80. • Lowest. Sets the Suspect Level for this Data Field Zone to 85. • Custom. Allows you to enter a value (from 1 to 99) for the Suspect Level for this Data Field Zone. If enter a value of 0 or leave the field blank, the value will reset to 50. <p>The Suspect Level is the level of confidence placed in the processing results for this field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The Sensitivity level selected for this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Sensitivity to Lowest would indicate you have a fair amount of confidence in the result returned by the OCR engine because few higher scores could be returned and fewer results would be determined to be suspect.</p> <p>Setting the Suspect Level to Highest would indicate you have less confidence in the result because a great number of lower scores could be returned and more results would be determined as suspect.</p>
Advanced filtering	<p>Use the VB script drop-down to select a VB script to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>

Data Field Zone Signature Detection Field Options	Description
Test	<p>Click the Test button to have the OCR engine perform a test process on the Data Field Zone and attempt to detect a signature using the options you have specified. If the Advanced Capture engine is configured to attempt to compensate for offset data, the resulting offset/scaling adjustments will take place during the test.</p> <p>Once the test process is complete, a dialog box is displayed indicating if a signature was detected and the Keyword Value that would be assigned based on this result. The offset/scaling information is also displayed in the dialog box.</p>
Keyword association	<p>You can logically group Data Field Zones using Keyword association. These groupings can be used to identify Keywords that belong to a Multi-Instance Keyword Type Group.</p> <p>Use the Keyword association field to enter or select a logical group name for the Data Field Zones that identify Keywords belonging to an MIKG. The Advanced Capture engine will attempt to maintain this Keyword grouping on the resulting document.</p>

Data Field Zone Signature Detection Field Options	Description
Colors	<p>If you would like to assign specific colors to the Keyword Type configured for the Data Field Zone, click the Colors button. The Display Colors dialog box is displayed.</p>  <p>Here you can change the colors in which any regular or suspect values for the corresponding Keyword Type are displayed in the Indexing panel once Advanced Capture processing has taken place.</p> <ul style="list-style-type: none"> • To change the display color for regular values, click Display Color to open your machine's color palette, select a color, and click OK. • To change the display color for suspect values, click Suspect Color to open your machine's color palette, select a color, and click OK. • To revert back to the default display color for regular or suspect values, click the Automatic button that corresponds to the desired type of values (i.e., the left button left for regular values, the right button for suspect values). <hr/> <p>Note: Any colors assigned here can be overridden by colors assigned through Keyword Lookup/Replace settings and/or VB scripting.</p> <hr/>

Data Field Zone Signature Detection Field Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Data Field Zone Signature Detection Field Options	Description
Activation groups (cont.)	Alternatively, you can assign a form definition group as the Data Field Zone's activation group to activate the zone for processing. Form definition groups can be used to extract only specific types of information (e.g., header data vs. detail data) during processing. In the Activation groups drop-down list, form definition groups are enclosed in brackets (e.g., [Group1]).

Configuring a Snippet Data Field Zone

A Snippet Data Field Zone allows you to capture specific regions of a page and combine these regions into a separate, snippet document that repurposes the extracted data into a condensed and easily readable format.

Depending on your configuration, the newly created snippet document(s) may be added to the same batch undergoing Advanced Capture processing or created as a separate document. When testing processing in the **Advanced Capture Configuration** window, any snippet documents are created locally as temporary files. The location of these temporary files is listed in the **Test Results Verification** window.

A Snippet Data Field Zone is configured using the options on the Snippet tab of the **Data Field Zone** dialog box.

Data Field Zone [X]

Text Field | Line Item Extraction | Grouped Line Item Extraction
Optical Mark Field | Signature Detection Field | **Snippet**

☒ Use image snippet style extraction

Document Type: [v]

Page Location(s)
☒ Absolute page: [1]
 Snippet only from single page; always the same absolute page number (one-based)
☐ Relative page(s):
☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)

Snippet Sizing
☒ Specified area only
☐ Fit width to page
☐ Fit height to page
☐ Entire Page

☐ Delete this entire page from source document after creating snippet
☐ Redact zone from source document
☐ Place snippet on blank page of original page size
☐ Combine with other snippets from this document into single target document (same document type)
☐ Use post-image processed page image for snippet source
☐ Copy common keywords from source to snippet document

Archive into:
☒ Non-batch document ☐ Same batch as source ☐ New batch in same scan queue

VB script: [<None>] [v] [...]

Activation groups: [v]

[Save] [Cancel]

To enable the following options, select **Use image snippet style extraction**.

Data Field Zone Snippet Options	Description
Document Type	Using this drop-down list, select the Document Type to which the new snippet document is to be assigned.

Data Field Zone Snippet Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the snippet being read by the Data Field Zone is only displayed on one page and is always displayed on the same page (for example, the snippet is always displayed on page 1). Enter in the associated field the page number on which the snippet is located. • Select the Relative page radio button if the snippet may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate on which page(s) the Data Field Zone may be located. <ul style="list-style-type: none"> Select the First page check box if the Data Field Zone is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes. Select the First page (not only page) check box if the Data Field Zone is located on the first page and other pages in the document. Select the Interior pages check box if the Data Field Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes. Select the Last page check box if the Data Field Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes. Select the Last page (not only page) check box if the Data Field Zone is located on the last page and other pages in the document.
Snippet Sizing	<p>The Snippet Sizing options control whether the snippet zone remains the size it was drawn or expands to fit the page's width, height, or entire area.</p> <ul style="list-style-type: none"> • Select Specified area only to keep the snippet zone the same size it was drawn. • Select Fit width to page to expand the snippet zone's boundaries to fit the page width. • Select Fit height to page to expand the snippet zone's boundaries to fit the page height. • Select Entire Page to expand the snippet zone to fit the dimensions of the entire page.
Delete this entire page from source document after creating snippet	<p>Select this option to delete the entire page from the source document once the snippet document is created.</p>

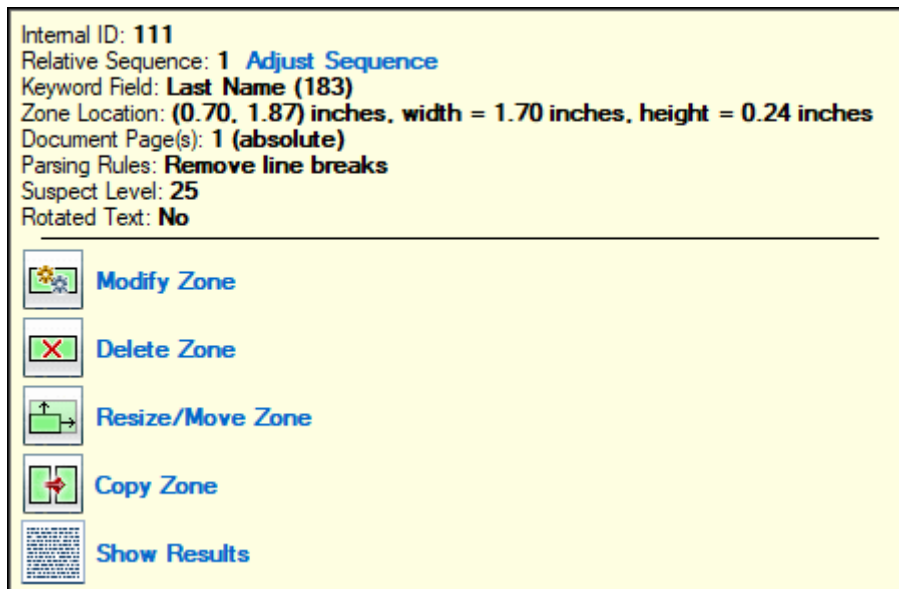
Data Field Zone Snippet Options	Description
Redact zone from source document	<p>Select this option to redact the snippet zone from the source document once the snippet document is created.</p> <hr/> <p>Note: This option is only respected if you are licensed for Automated Redaction.</p> <hr/>
Place snippet on blank page of original page size	<p>Select this option to place the snippet zone on a blank page with the same dimensions as the source page when creating the snippet document.</p> <p>If this option is not selected, the snippet document is only made as large as it needs to be to contain the snippet zone(s) from the source document.</p>
Combine with other snippets from this document into a single target document (same document type)	<p>Select this option to combine all snippet zones configured for the same Document Type and Archive into: option on the current document into a single snippet document.</p> <hr/> <p>Note: Snippet zones configured with different Archive into: options will not be combined. See Archive into: on page 217 for more information.</p> <hr/> <p>If this option is not selected, a separate snippet document is created for each snippet zone.</p>
Use post-image processed page image for snippet source	<p>Select this option to process the snippet zone image after image processing has been performed on the document.</p> <p>If this option is not selected, the snippet zone image is taken before any image processing is performed.</p>
Copy common keywords from source to snippet document	<p>Select this option to copy common Keyword Values from the source document to the snippet document once Advanced Capture processing has been performed.</p>
Archive into:	<p>You can select the batch into which a newly created snippet document is added.</p> <ul style="list-style-type: none"> • Select Non-batch document to create the snippet document outside of a batch. • Select Same batch as source to add the newly created snippet document to the same batch as the source document. • Select New batch in same scan queue to add the newly created snippet document to a new batch created in the same scan queue as the source batch.
VB script	<p>Use the VB script drop-down list to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>

Data Field Zone Snippet Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (for example, +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (that is, through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (for example, when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Data Field Zone Snippet Options	Description
Activation groups (cont.)	Alternatively, you can assign a form definition group as the Data Field Zone's activation group to activate the zone for processing. Form definition groups can be used to extract only specific types of information (for example, header data vs. detail data) during processing. In the Activation groups drop-down list, form definition groups are enclosed in brackets (for example, [Group1]).

Modifying or Analyzing a Data Field Zone

An existing Data Field Zone can be modified, deleted, resized/moved, copied, or analyzed by right-clicking on it in the Document Viewer. If multiple zones overlap, the **Previous** and **Next** links can be used to select the desired zone.



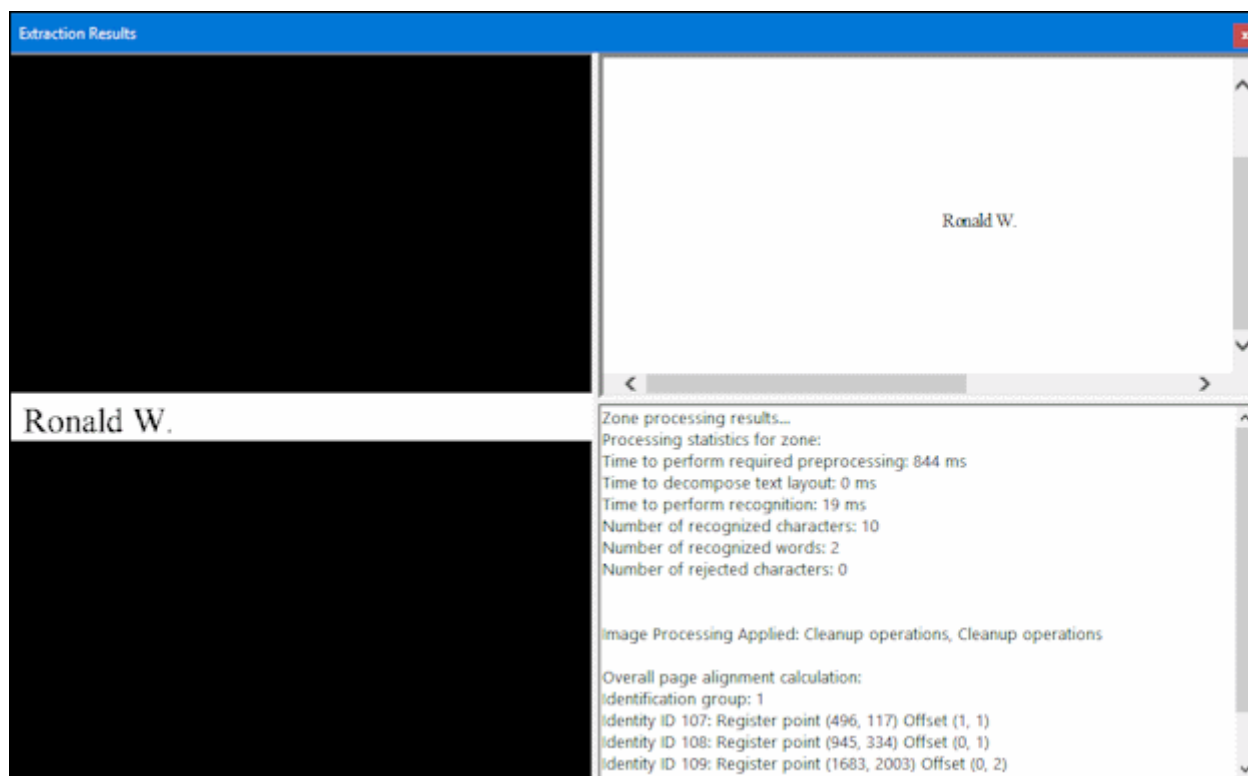
Information about the selected zone is displayed above the following options:

- **Modify Zone.** Click to display the **Data Field Zone** dialog box to modify any of the Data Field Zone's configuration settings.
- **Delete Zone.** Click to delete the selected Data Field Zone.
- **Resize/Move Zone.** Click to make the Data Field Zone editable in the Document Viewer. It can be resized and/or moved without changing any of its existing configuration settings using either the pointer or the following shortcut keys:

Shortcut Key	Action
Up / Down / Left / Right	Moves the zone one pixel in the specified direction.

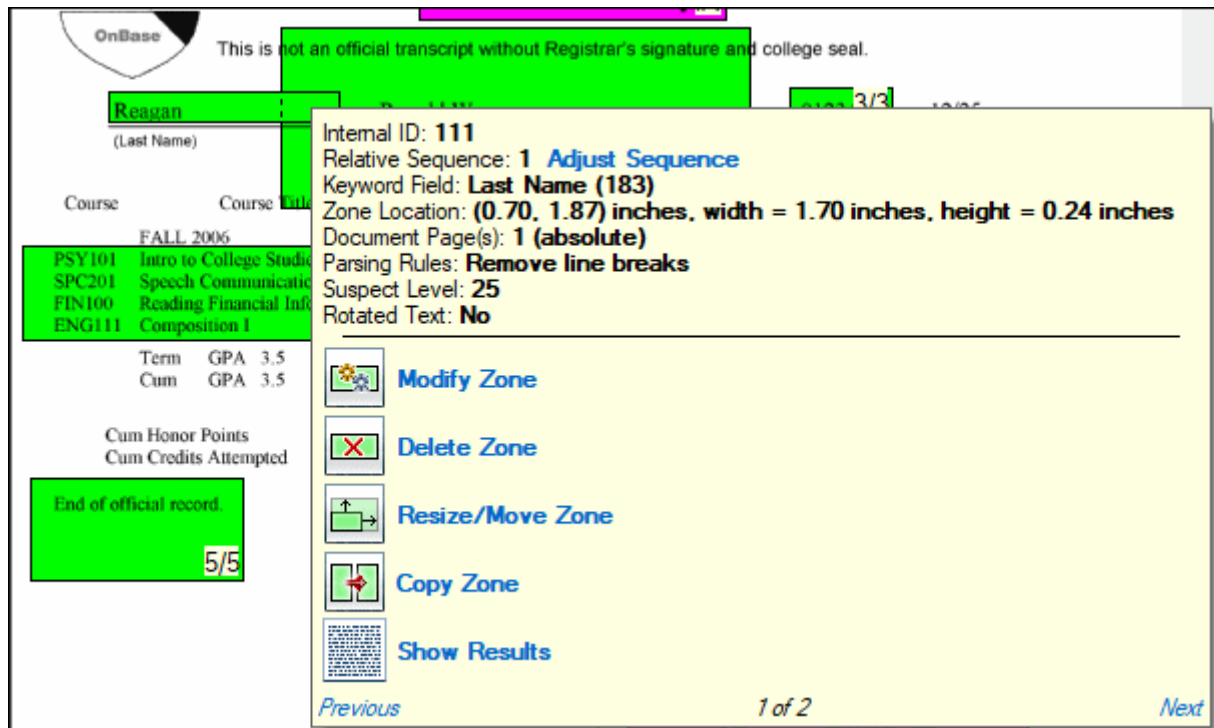
Shortcut Key	Action
Shift + Up / Down / Left / Right	Expands the zone by one pixel in the specified direction.
Ctrl + Up / Down / Left / Right	Shrinks the zone by one pixel in the specified direction.
Space / Enter	Saves the changes made to the size and position of the zone.
Esc	Cancels the changes made to the size and position of the zone.

- **Copy Zone.** Click to create a duplicate of the selected Data Field Zone in the Document Viewer. The new zone is created directly on top of the zone it was created from; it can be resized as needed and then dragged to a new location, while still inheriting all of the configuration settings of the zone it was created from.
- **Show Results.** Click to view the **Extraction Results** window, which displays the extraction area of the Data Field Zone on the left, the text extracted from the zone in the upper right, and diagnostics information about the extraction process in the lower right.



Modifying or Analyzing Overlapping Zones

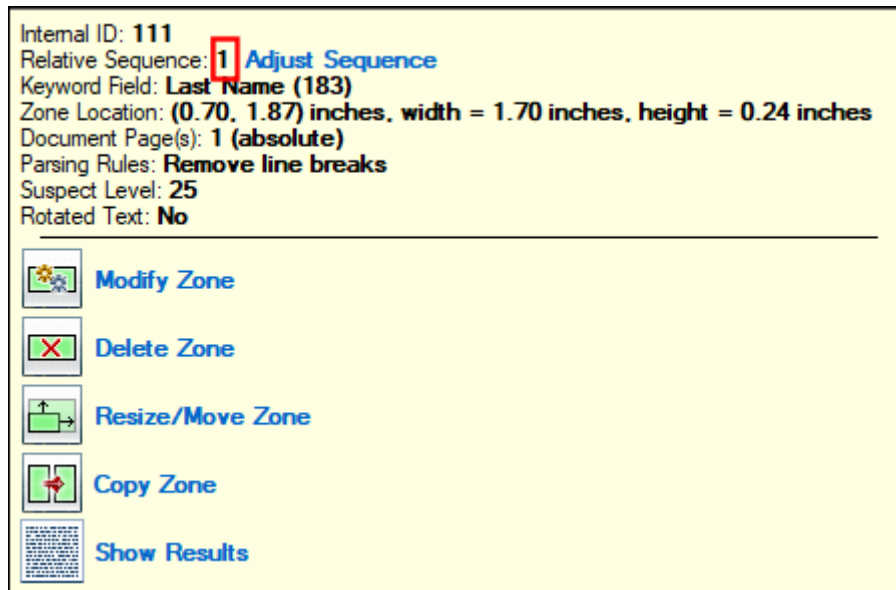
Depending on your document layout, zones may overlap. Overlapping zone borders are displayed with a dotted line.



When right-clicking on an area of the document with multiple zones, the name of the currently selected zone is displayed at the top of the dialog box. To switch between zones, click the **Previous** and **Next** links; once the desired zone is selected, it can be modified, deleted, resized/moved, copied, or analyzed.

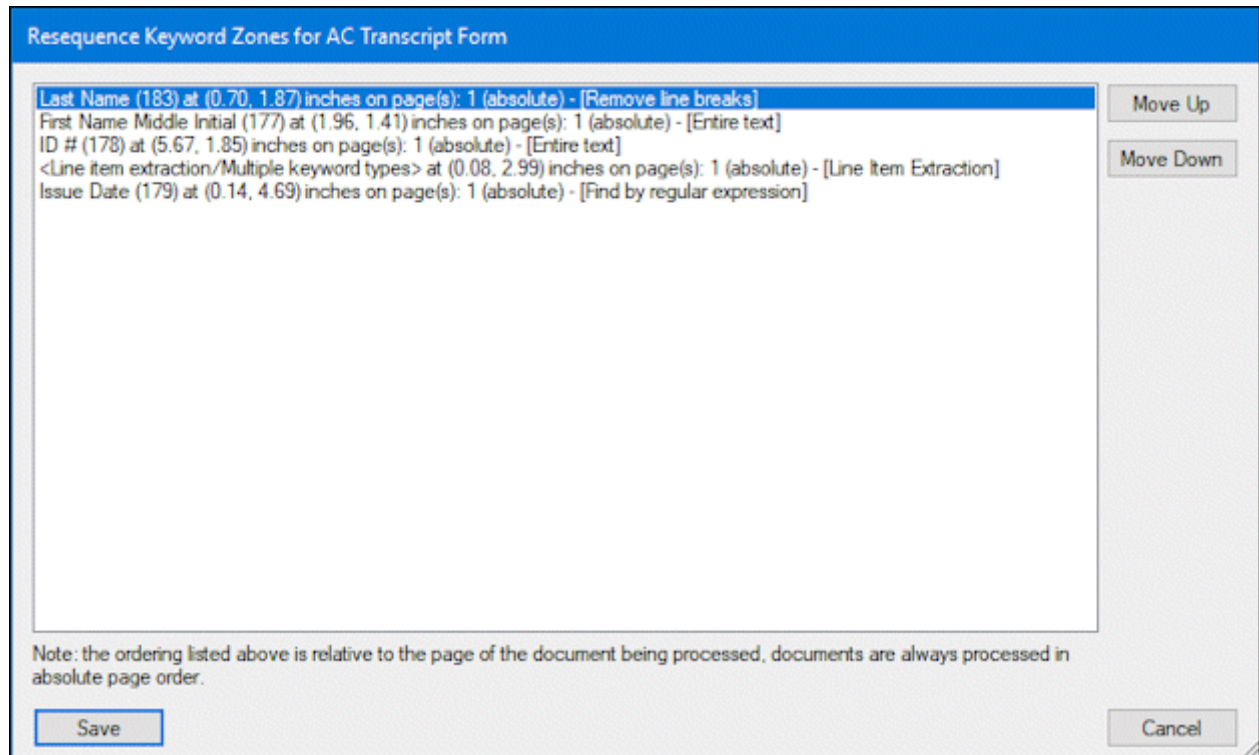
Changing the Data Field Zone Processing Order

The order in which Data Field Zones are processed can be adjusted by right-clicking on any of the Data Field Zones in the Document Viewer. The selected Data Field Zone's position in the processing order is displayed next to **Relative Sequence** in the **Modify/Delete Zone** dialog box.



To change the processing order of a form's Data Field Zones:

1. Right-click on any Data Field Zone in the Document Viewer. The **Modify/Delete Zone** dialog box is displayed.
2. Click **Adjust Sequence**. The **Resequence Keywords for <Form Name>** dialog box is displayed.



3. Select the Data Field Zone you wish to reorder.
4. Click **Move Up** or **Move Down** to change the Data Field Zone's position with respect to the other zones in the processing sequence.

Note: While Data Field Zones are processed according to their sequence on each page, the page order is taken into account first. All Data Field Zones on Page 1 are processed before the Data Field Zones on Page 2 are processed, and so on. For example: suppose you have a form with three Data Field Zones. Zones 1 and 3 appear on Page 1, while Zones 1 and 2 appear on Page 2. The zones will be processing in the following order: Zone 1 (Page 1), Zone 3 (Page 1), Zone 1 (Page 2), Zone 2 (Page 2).

5. Click **Save**.

Configuring a Page Registration Zone

If you are creating or modifying a form for a multi-page document, you can create a Page Registration Zone on pages of the document (other than the first page) to determine the offset (i.e., the skew or rotation) of the page so that it can be accounted for when performing the Advanced Capture process.

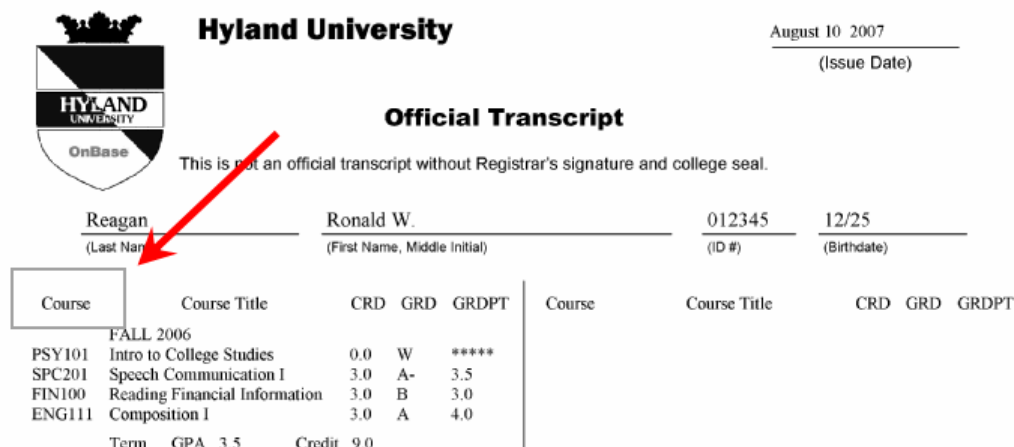
Note: A Page Registration Zone cannot be configured for the first page of a multi-page document. You can configure a Registration Point within the Form Identification Zone to determine and account for offset on the first page.

Tip: Multiple Page Registration Zones can be configured for an Advanced Capture form. However, configuring multiple zones across different documents for the same form can reduce the accuracy of indexing results. Therefore, if you use multiple Page Registration Zones for the same form, it is recommended that you configure these zones on the same document.

To create a Page Registration Zone:

Note: Prior to creating the Page Registration Zone, ensure that the **Default Zone Type** option in the **Tools** panel is set to either **Always Ask** or **Form Identity or Registration**.

1. With the page of the document that the Page Registration Zone is being created for displayed in the Document Viewer, click and hold the left mouse button and use the pointer to draw a box around the value that is to be used as the Page Registration value. Note that the area of the document that you selected is displayed in the **Result Verification** panel.



Hyland University August 10, 2007
(Issue Date)

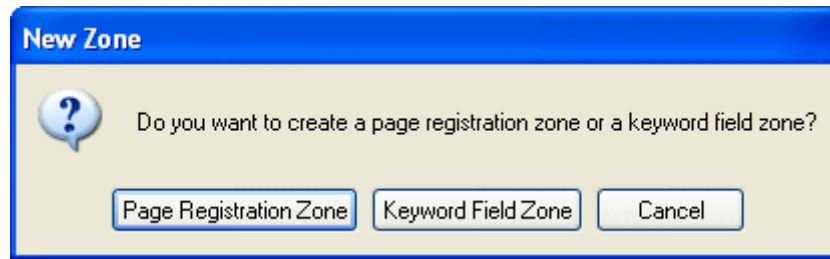
Official Transcript

This is not an official transcript without Registrar's signature and college seal.

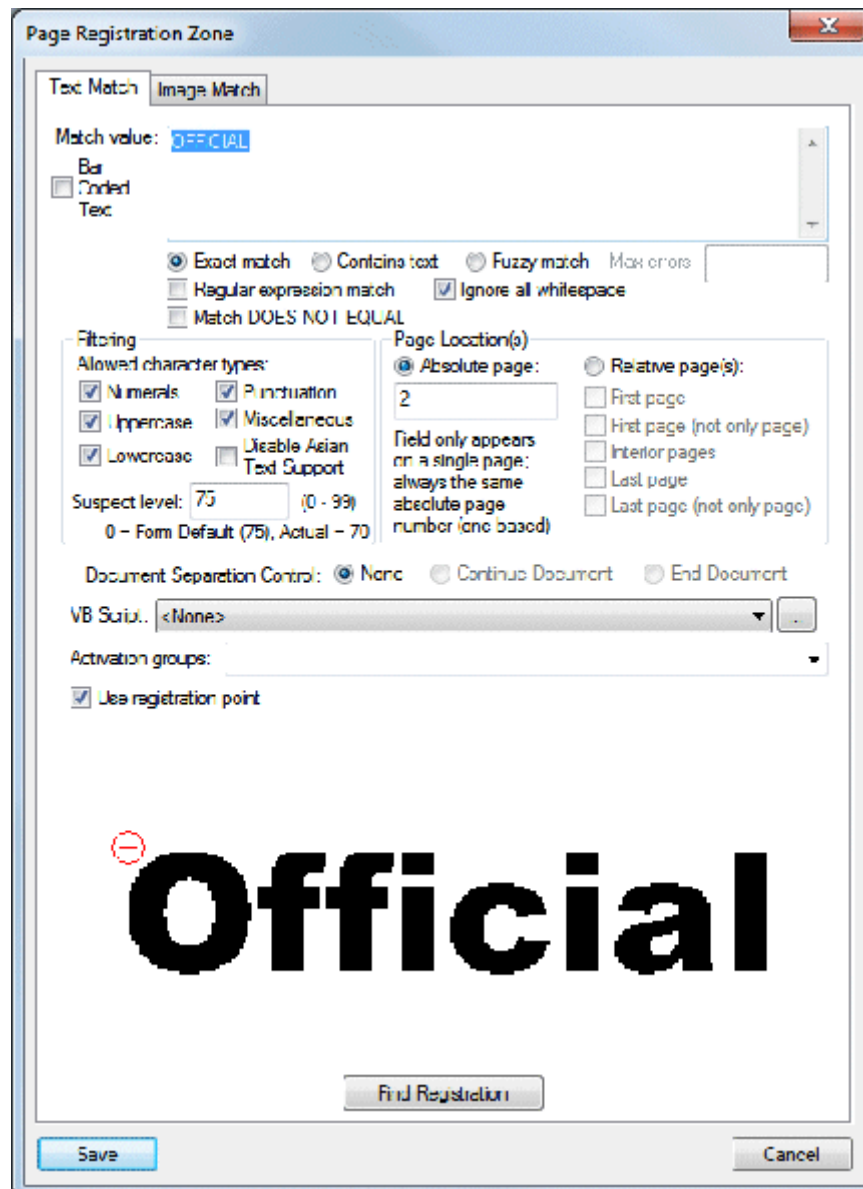
Reagan Ronald W. 012345 12/25
(Last Name) (First Name, Middle Initial) (ID #) (Birthdate)

Course	Course Title	CRD	GRD	GRDPT	Course	Course Title	CRD	GRD	GRDPT
FALL 2006									
PSY101	Intro to College Studies	0.0	W	*****					
SPC201	Speech Communication I	3.0	A-	3.5					
FIN100	Reading Financial Information	3.0	B	3.0					
ENG111	Composition I	3.0	A	4.0					
Term	GPA	3.5	Credit	9.0					

2. Release the left mouse button to display the **Page Registration Zone** dialog box. Depending on the setting of the **Default Zone Type** option in the **Tools** panel, the **New Zone** dialog box may be displayed.



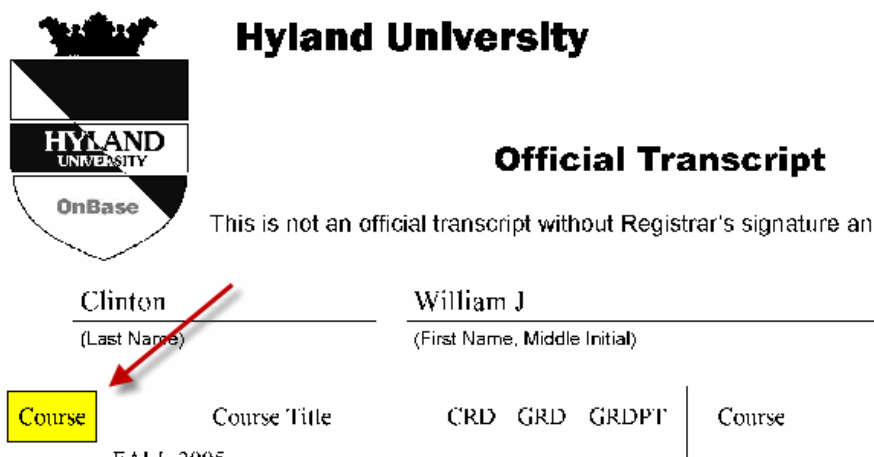
If the **New Zone** dialog box is displayed, click **Page Registration Zone** to continue to the **Page Registration Zone** dialog box.



3. Select if you would like to use a text value or an image to determine the page's registration.
 - If you are using a text value to determine the page's registration, select the Text Match tab of the **Page Registration Zone** dialog box.
 - If you are using an image to determine the page's registration, select the Image Match tab of the **Page Registration Zone** dialog box.
4. Configure the Page Registration Zone using the options on the **Page Registration Zone** dialog box.

- Once you have set all necessary options on the **Page Registration Zone** dialog box, click **Save**. The **Page Registration Zone** dialog box is closed, and the Page Registration Zone is temporarily saved to the Advanced Capture form.

The Page Registration Zone is highlighted in yellow on the document in the Document Viewer.



The image shows a form template for Hyland University. At the top left is the Hyland University logo, which includes a crest with a crown and the text 'HYLAND UNIVERSITY' and 'OnBase'. To the right of the logo is the text 'Hyland University' in a large, bold font. Below this is the text 'Official Transcript' in a bold font. A disclaimer line reads: 'This is not an official transcript without Registrar's signature an'. Below the disclaimer are two signature lines. The first line is labeled 'Clinton' and '(Last Name)'. The second line is labeled 'William J' and '(First Name, Middle Initial)'. Below the signature lines is a table with a yellow header row. The header row has a column labeled 'Course' (highlighted in yellow), a column labeled 'Course Title', and a column labeled 'CRD GRD GRDPT' (with a vertical line between 'GRD' and 'GRDPT'). Below the header row, the text 'FALL 2005' is visible under the 'Course' column, and 'Course' is visible under the 'CRD GRD GRDPT' column.

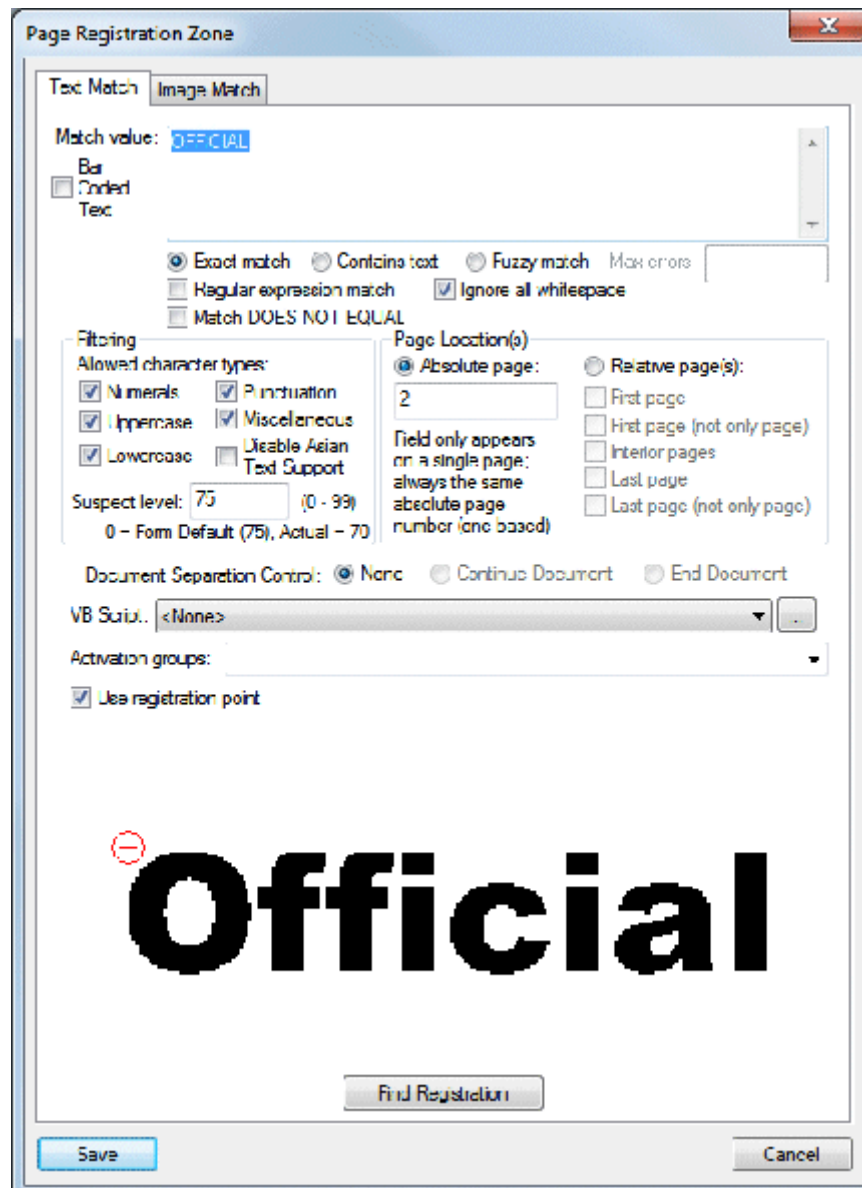
- Repeat Steps 1-6 for each Page Registration field you would like to create for the document.

Tip: At any time, click **Save Configuration** in the **Tools** panel to save the form to the database. It is considered a best practice to save your Advanced Capture form configuration after the creation and configuration of every Advanced Capture form and each form's Form Identification, Page Registration, and Data Field Zones.

To discard all configuration changes (including Page Registration Zones created since the last time the form was saved), click **Discard Configuration Changes** in the **Tools** panel.

Configuring a Page Registration Zone to Use a Text Value

A text value can be configured to determine a page's registration by using the options on the Text Match tab of the **Page Registration Zone** dialog box.



Page Registration Zone Text Match Options	Description
Match Value	<p>This field displays the text result returned from the OCR engine's evaluation of the Page Registration Zone.</p> <hr/> <p>Note: The Match Value is also displayed in the Result Verification panel below the enlarged image of the Page Registration Zone.</p> <hr/> <p>You may select the match rules for the Page Registration Zone by selecting one of the match rule radio buttons.</p> <hr/> <p>Tip: If the position of the text that is being compared to the Match Value shifts from document to document to the extent that a portion of this text might fall outside the Page Registration Zone, or different text might fall inside the zone, select the Contains text option to increase the chances of a desired match.</p> <hr/> <ul style="list-style-type: none"> • Exact match. Select this radio button if the OCR engine must recognize the Match Value exactly on a processed document in order for Advanced Capture to determine and account for offset using the Page Registration Zone. • Contains text. Select this radio button if the OCR engine needs only to identify the Match Value as part of a longer string on the document being processed in order to determine and account for offset using the Page Registration Zone. <hr/> <p>Note: When using this option, only the Match Value portion of the text string is taken into consideration to determine the Suspect Level of the value.</p> <hr/> <ul style="list-style-type: none"> • Fuzzy match. Select this radio button if the OCR engine needs to identify the Match Value on a processed document, but can allow for some margin for error between the Match Value and the value read from the document being processed in order to determine and account for offset using the Page Registration Zone. <p>Enter a value in the Max Errors field to set the number of discrepancies allowed between the Match Value and the value read from the document being processed.</p> <p>For example, if your Match Value is HYLAND UNIVERSITY and the value detected by the OCR engine from the processed document is HYL4ND UN1VERSITY, the Page Registration Zone can still be used to determine and account for offset if the Max Errors value is set to 2 or more.</p>

Page Registration Zone Text Match Options	Description
Bar Coded Text	<hr/> <p>Note: This option is only available if your solution is licensed for the Bar Code Recognition Server.</p> <hr/> <p>Select this check box if the Page Registration Zone contains a bar code. The bar code will be converted into text and then identified according to the zone's configured match rules.</p> <hr/> <p>Note: Depending on your configuration, the Advanced Capture engine will only search for certain bar code types in this zone. See Bar Code Types for more information.</p> <hr/>
Bar Code Types	<hr/> <p>Note: This option is only available when the Bar Coded Text option is selected.</p> <hr/> <p>Click this button to access the Select Bar Code Types dialog box.</p> <div data-bbox="628 894 1390 1413" data-label="Image"> </div> <p>Here you can select the types of bar codes for which the Advanced Capture engine will search when processing the Page Registration Zone. The engine will ignore any deselected bar code types when processing this zone.</p> <p>By default, all bar code types are selected.</p> <hr/> <p>Tip: To expedite Advanced Capture processing, select only the desired bar code types.</p> <hr/> <p>To save changes to your bar code type selections, click Save.</p>

Page Registration Zone Text Match Options	Description
Regular expression match	<p>Select this check box if the Match Value should match the form of a regular expression.</p> <p>For example: the regular expression rule for a Social Security Number is <code>\d{3}-\d{2}-\d{4}</code> (i.e., 3 digits, dash, 2 digits, dash, 4 digits). If the OCR engine identifies text matching this pattern in the zone, that text is returned as a Match Value.</p> <hr/> <p>Note: All regular expressions must be ECMA compliant.</p> <hr/>
Ignore all whitespace	<p>Select this check box if you would like to ignore all whitespace (e.g., spaces, paragraph returns, etc.) in the Match Value after it is read by the OCR engine.</p> <p>For example, if the Match Value read by the OCR engine is PSY 103, and the Ignore all whitespace check box is selected, the Match Value would be read as PSY103.</p>
Match DOES NOT EQUAL	<p>Select this check box if the page's registration should be determined only when the Match Value does not appear in the Form Identification Zone.</p>

Page Registration Zone Text Match Options	Description
Allowed character types	<p>These options are available to assist the OCR engine in determining if the value it reads is correct. Select the check box next to each character type that is allowed in the Match Value. The available options are:</p> <ul style="list-style-type: none"> • Numerals. Numeric characters, 0-9. • Uppercase. Uppercase alphabetic characters. • Lowercase. Lowercase alphabetic characters. • Punctuation. Punctuation marks (i.e., . ! ?). • Miscellaneous. Other ASCII characters that do not fall into one of the above categories (i.e., # \$ * @). <p>When a character is recognized by the OCR engine that is not part of an allowable character set, the character is replaced by a tilde (~) and the value is automatically marked as suspect.</p> <hr/> <p>Tip: Using the Allowed character types options can sometimes help the OCR engine more easily determine the correct value by eliminating characters that are obviously not correct (e.g., an I is correctly identified instead of a 1 because numeric characters are filtered and prevented from being recognized as part of the value).</p> <hr/> <p>Select the Disable Asian Text Support check box to instruct the OCR engine to skip Asian (i.e., double-byte) characters.</p> <hr/> <p>Note: The Disable Asian Text Support check box is only available if the OCR format assigned to the Document Type is configured for an Asian language (e.g., Japanese, Korean, etc.).</p> <hr/> <p>Tip: Selecting the Disable Asian Text Support check box allows you to identify numeric data (i.e., Date Keyword Values, Currency Keyword Values, and Document Dates) when performing OCR on documents configured to contain Asian (i.e., double-byte) characters.</p> <hr/>

Page Registration Zone Text Match Options	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Advanced Capture form and the actual Suspect Level detected for the selected value are displayed below the Suspect Level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that is returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured by the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that the page registration is being correctly determined.</p> <hr/>

Page Registration Zone Text Match Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Page Registration Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the Page Registration Zone is only displayed on one page and is always displayed on the same page (e.g., the Page Registration Zone is always on page 2). Enter the page number that the Page Registration Zone is located on in the associated field. <hr/> <p>Note: Valid Absolute page values are 2 or greater. If enter a value less than 2, this value will be changed to 2 automatically upon saving.</p> <hr/> <ul style="list-style-type: none"> • Select the Relative page radio button if the Page Registration Zone may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the Page Registration Zone may be located on. <hr/> <p>Note: The First page and First page (not only page) options are not available for Page Registration Zones.</p> <hr/> <p>Select the Interior pages check box if the Page Registration Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the Page Registration Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the Page Registration Zone is located on the last page and other pages in the document.</p>

Page Registration Zone Text Match Options	Description
Document Separation Control	<hr/> <p>Note: The Document Separation Control options are available only for Advanced Capture forms that are configured for document separation. For more information, see Configuring Document Breaks on page 61.</p> <hr/> <p>The Document Separation Control options control whether the presence of the Page Registration Zone on a page should force a continuation or a break of the document.</p> <ul style="list-style-type: none"> • Select None to use the form's default settings for document separation. • Select Continue Document to continue the document onto the next page automatically, without testing for matches to other Advanced Capture forms. • Select End Document to break the document on the current page. The next page will start a new document. <hr/> <p>Note: Irrespective of these settings, the document will still break when it reaches the Maximum Page Count that has been configured for the form.</p> <hr/>
VB Script	<p>Use the VB Script drop-down list to select a VB script to associate with the identification of this Advanced Capture form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>

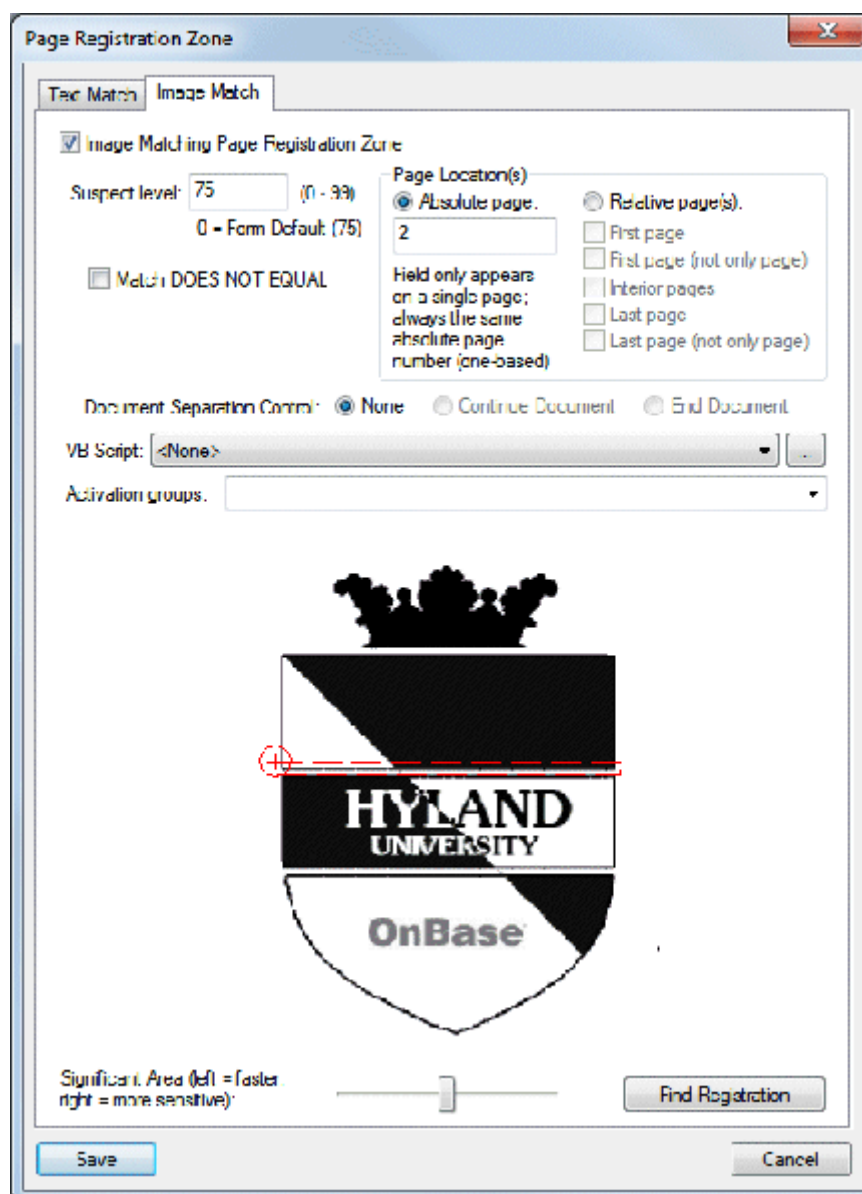
Page Registration Zone Text Match Options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Page Registration Zone Text Match Options	Description
Use registration point	<p>Select this check box to enable the Registration Point feature.</p> <hr/> <p>Note: The Registration Point is the starting point (upper-left corner) of the value detected in the Form Identification Zone. Click Find Registration to automatically set the Registration Point at the starting point of the value in the Form Identification Zone, or double-click a location in the field below the check box to manually set the Registration Point at that position.</p> <hr/> <p>The position of the Registration Point on the Advanced Capture form is compared to the same position on the document being processed to determine the offset (i.e., the skew or rotation) of any imported documents. The position of the configured Form Identification Zone is adjusted on the document being processed to account for the detected offset to ensure that the Advanced Capture process is able to process the document properly.</p> <ul style="list-style-type: none">• If you selected the Regular expression match check box for the Match Value, the Register on capture group position check box will be enabled. Select this option and click Find Registration to set the Registration Point to the starting point of the first capture group contained within the Match Value's regular expression. This option is useful when the Match Value's first capture group is not the first value detected within the Form Identification Zone.

Configuring a Page Registration Zone to Use an Image Match

An image can be configured to determine a page's registration by using the options on the Image Match tab of the **Page Registration Zone** dialog box.

Note: On a scaled document, the location of the Registration Point can become skewed when using an image to determine a page's registration. As a result, the document might not be matched to the form. Therefore, when configuring a form to which scaled documents will be matched, it is recommended that you limit the use of Page Registration Zones that use an image match and that you configure at least one Page Registration Zone that uses only text and/or regular expressions.



Page Registration Zone Image Match options	Description
Image Matching Form Identification Zone	<p>By default, Page Registration Zones are configured to use text values to determine the page's registration.</p> <p>Select this check box to configure this Page Registration Zone to use an image to determine the page's registration.</p> <p>This check box must be selected in order to enable the other options on the Image Match tab of the Page Registration Zone dialog box.</p>
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Advanced Capture form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Advanced Capture form is displayed below the Suspect Level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that is returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured by the zone to be marked as suspect. A score lower than the Suspect Level threshold indicates that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Setting the Suspect Level to 0 reverts back to the default threshold of 75.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that the page registration is being correctly determined.</p> <hr/>

Page Registration Zone Image Match options	Description
Match DOES NOT EQUAL	Select this check box if the page's registration should be determined only when the Match Value does not appear in the Page Registration Zone.
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Page Registration Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the Page Registration Zone is only displayed on one page and is always displayed on the same page (e.g., the Page Registration Zone is always on page 2). Enter the page number that the Page Registration Zone is located on in the associated field. <hr/> <p>Note: Valid Absolute page values are 2 or greater. If enter a value less than 2, this value will be changed to 2 automatically upon saving.</p> <hr/> <ul style="list-style-type: none"> • Select the Relative page radio button if the Page Registration Zone may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) the Page Registration Zone may be located on. <hr/> <p>Note: The First page and First page (not only page) options are not available for Page Registration Zones.</p> <hr/> <p>Select the Interior pages check box if the Page Registration Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the Page Registration Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the Page Registration Zone is located on the last page and other pages in the document.</p>

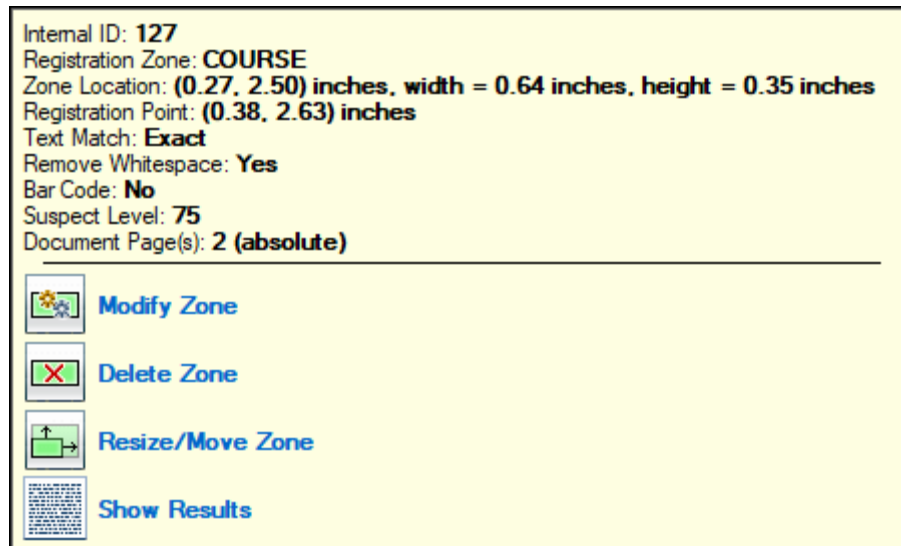
Page Registration Zone Image Match options	Description
Document Separation Control	<p>Note: The Document Separation Control options are available only for Advanced Capture forms that are configured for document separation. For more information, see Configuring Document Breaks on page 61.</p> <hr/> <p>The Document Separation Control options control whether the presence of the Page Registration Zone on a page should force a continuation or a break of the document.</p> <ul style="list-style-type: none"> • Select None to use the form's default settings for document separation. • Select Continue Document to continue the document onto the next page automatically, without testing for matches to other Advanced Capture forms. • Select End Document to break the document on the current page. The next page will start a new document. <hr/> <p>Note: Irrespective of these settings, the document will still break when it reaches the Maximum Page Count that has been configured for the form.</p> <hr/>
VB Script	<p>Use the VB Script drop-down list to select a VB script to associate with the identification of this Advanced Capture form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>

Page Registration Zone Image Match options	Description
Activation groups	<p>When you have configured multiple Form Identification Zones or Page Registration Zones for a document, you can assign individual Data Field Zones to a specific Form Identification or Page Registration Zone using activation groups. Activation groups allow you to activate only the Data Field Zones assigned to the Form Identification or Page Registration Zone that is used to match the document to an Advanced Capture form. Data Field Zones assigned to Form Identification or Page Registration Zones that are not used to match the document to a form will not be processed. Also, Data Field Zones present on pages other than the pages containing their assigned Form Identification or Page Registration Zones will not be processed, unless otherwise specified through the Page Location(s) setting or by adding a + to the front of the activation group name on the Form Identification or Page Registration Zone. This selective activation saves processing time and reduces the number of forms that need to be created for a Document Type.</p> <p>Use the Activation groups field to enter or select an activation group name. Add a + to the front of a group name (e.g., +Group1) on a Form Identification or Page Registration Zone to set all Data Field Zones assigned to this group to be processed. Use commas to separate multiple group names.</p> <ul style="list-style-type: none"> • When a Form Identification Zone or Page Registration Zone is matched to a form, all activation groups that have been configured for the zone will be activated. • Form Identification Zones are organized into Identification Groups (under Combined rule expressions), and only one Identification Group can be matched to a form. Once an Identification Group has been matched, any remaining Identification Groups on the document will be skipped. • Multiple Page Registration Zones can be matched to a form. Every Page Registration Zone on the document will be tested for a match. • If multiple activation groups have been configured for a Data Field Zone, the zone will be processed if any of these activation groups is activated. • If no activation groups have been configured for a Data Field Zone, the zone will be considered active and thus will be processed. • If a Data Field Zone is configured to only be searched for on certain pages (i.e., through the Page Location(s) setting), the zone can only be considered active on these pages. This overrides any conflicting settings that would otherwise activate the Data Field Zone (e.g., when a Data Field Zone is assigned to an activation group that is named with a + on the corresponding Form Identification or Page Registration Zone, or when a Data Field Zone is not assigned to any activation group).

Page Registration Zone Image Match options	Description
Significant Area	<p>Use the slider to control the size of the significant processing area. The significant processing area is the red box displayed on the image snippet.</p> <p>Increasing the size of the significant processing area (i.e., moving the slider to the right) will increase the accuracy of the image match process, but will decrease the speed of the process.</p> <p>Decreasing the size of the significant processing area (i.e., moving the slider to the left) will increase the speed of the image match process, but will decrease the accuracy of the process.</p>
Find Registration	<p>The Registration Point feature is required when using an image identifier.</p> <p>The Registration Point is the starting point (upper-left corner) of the image's significant processing area (the red box displayed on the image snippet). It is used to determine the page's offset (i.e., any skew or rotation) that might have occurred during scanning.</p> <p>The position of the Registration Point on the Advanced Capture form is compared to the same position on the document being processed. The position of the configured Page Registration Zone is adjusted on the document being processed to account for the detected offset to ensure that the Advanced Capture process is able to process the document properly.</p> <p>Click Find Registration to automatically set the Registration Point, or double-click a location on the image to manually set the Registration Point at that position.</p>

Modifying or Analyzing a Page Registration Zone

An existing Page Registration Zone can be modified, deleted, or resized/moved by right-clicking on it in the Document Viewer. If multiple zones overlap, the **Previous** and **Next** links can be used to choose the desired zone.

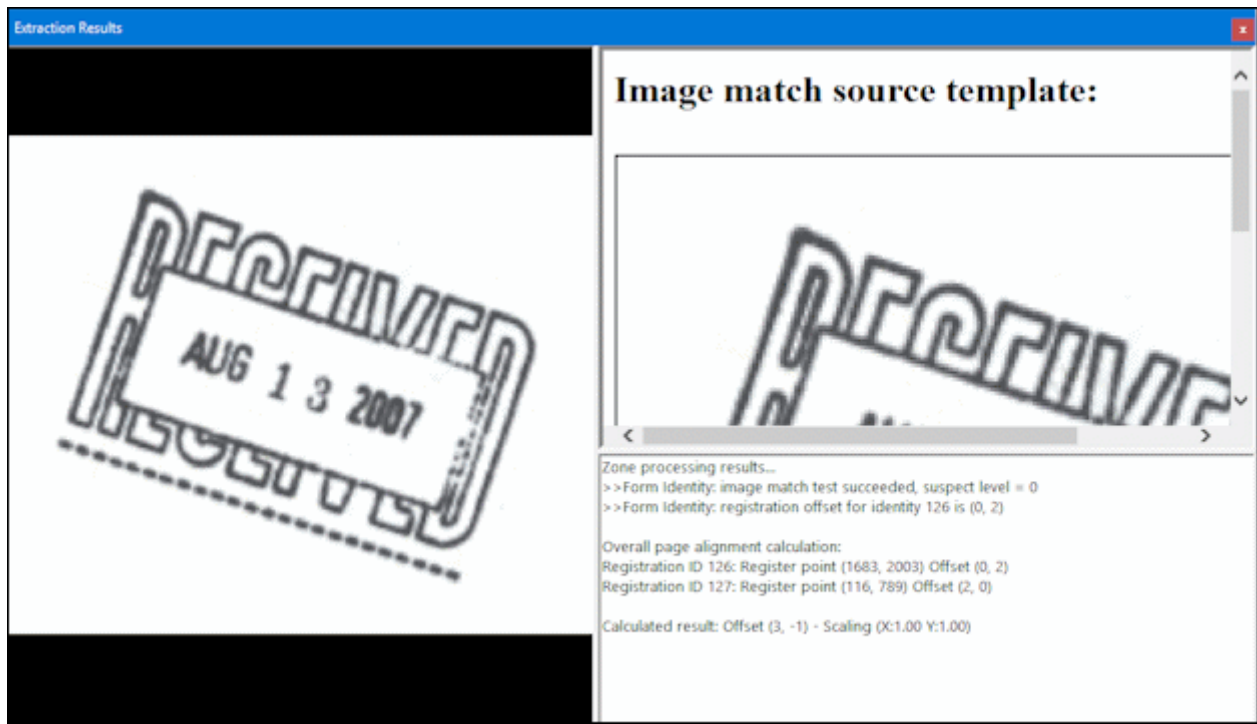


Information about the selected zone is displayed above the following options:

- **Modify Zone.** Click to display the **Page Registration Zone** dialog box to modify any of the Page Registration Zone's configuration settings.
- **Delete Zone.** Click to delete the selected Page Registration Zone.
- **Resize/Move Zone.** Click to make the Page Registration Zone editable on the Document Viewer. It can be resized and/or moved without changing any of its existing configuration settings using either the pointer or the following shortcut keys:

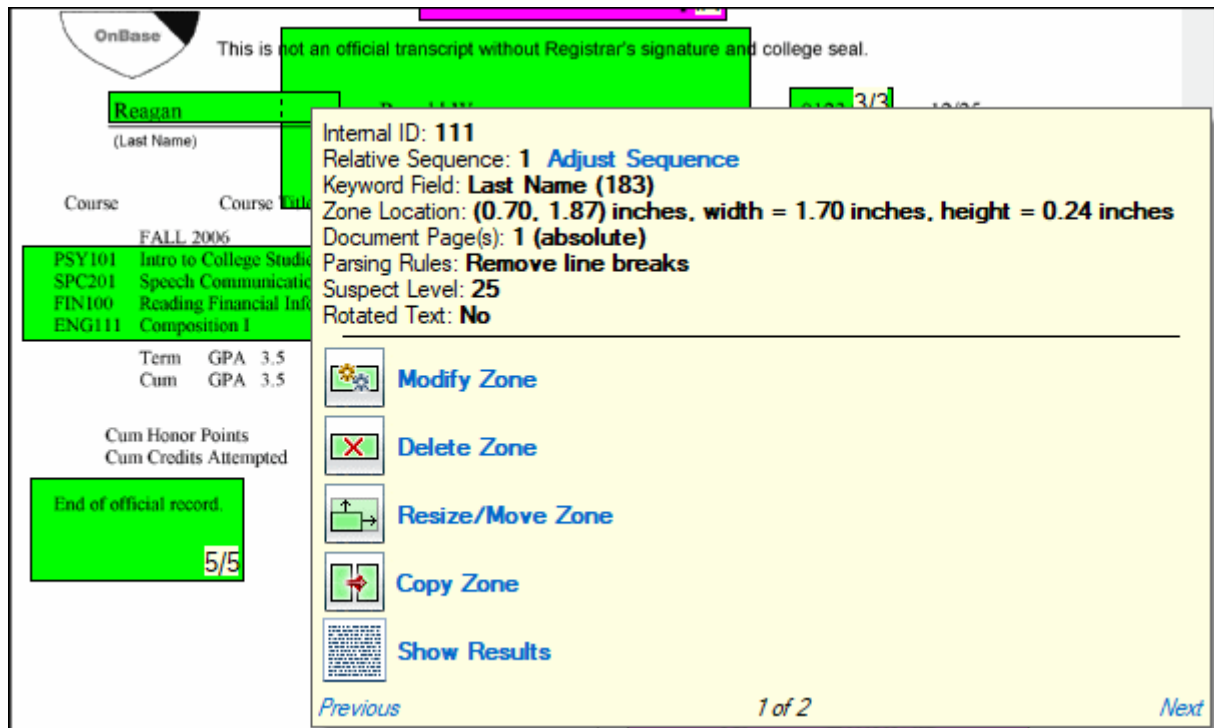
Shortcut Key	Action
Up / Down / Left / Right	Moves the zone one pixel in the specified direction.
Shift + Up / Down / Left / Right	Expands the zone by one pixel in the specified direction.
Ctrl + Up / Down / Left / Right	Shrinks the zone by one pixel in the specified direction.
Space / Enter	Saves the changes made to the size and position of the zone.
Esc	Cancels the changes made to the size and position of the zone.

- **Show Results.** Click to view the **Extraction Results** window, which displays the extraction area of the Page Registration Zone on the left, the text or image extracted from the zone in the upper right, and diagnostics information about the extraction process in the lower right.



Modifying or Analyzing Overlapping Zones

Depending on your document layout, zones may overlap. Overlapping zone borders are displayed with a dotted line.



When right-clicking on an area of the document with multiple zones, the name of the currently selected zone is displayed at the top of the dialog box. To switch between zones, click the **Previous** and **Next** links; once the desired zone is selected, it can be modified, deleted, resized/moved, copied, or analyzed.

Configuring Multiple Form Identification or Page Registration Zones for Scaling

You can configure multiple Form Identification Zones or multiple Page Registration Zones on a form to prompt the Advanced Capture engine to attempt to compensate for offset data. When a document's data appears larger than, smaller than, or unaligned with the data on the original form, the engine will attempt to calculate the magnitude of this difference. Using this calculation, the engine will then scale the form's configured Data Field Zones appropriately so that the zones' Keyword Values can still be captured accurately.

The calculated scaling values are listed as elements of the **Form Identity** in the **Result Verification** panel.

For example:

Form Identity: calculated scaling factor = **1.00, 1.00**, translation error = (1, 0) pixels.

- For scaling values close to **1.00**, no scaling error was detected.
- For scaling values less than **1.00**, the data on the current document appears smaller than the data on the original form.
- For scaling values greater than **1.00**, the data on the current document appears larger than the data on the original form.

Limitations and Requirements for Scaling

To improve the chances of successful scaling, consider the following:

- For the scaling value to be calculated, the Form Identification Zones must be configured as **AND** identification zones (i.e., the rules representing the Form Identification Zones in the **Combined rule expressions** field of the Form Identification Zone dialog box must belong to the same group).
- To ensure that the scaling value is properly calculated, set the Registration Point for each Form Identification Zone or Page Registration Zone using the same document.
- To ensure that identifiers appear within a configured zone, make larger Form Identification Zones or Page Registration Zones than you would normally make.
- If non-identifying data is detected within the Form Identification Zones or Page Registration Zone, the **Exact match** and **Fuzzy match** options might not be effective.
- If the difference in scale is too great, an Image Match for the Form Identification Zone or Page Registration Zone will not be able to match the form to the document.

The Regular Expression Library

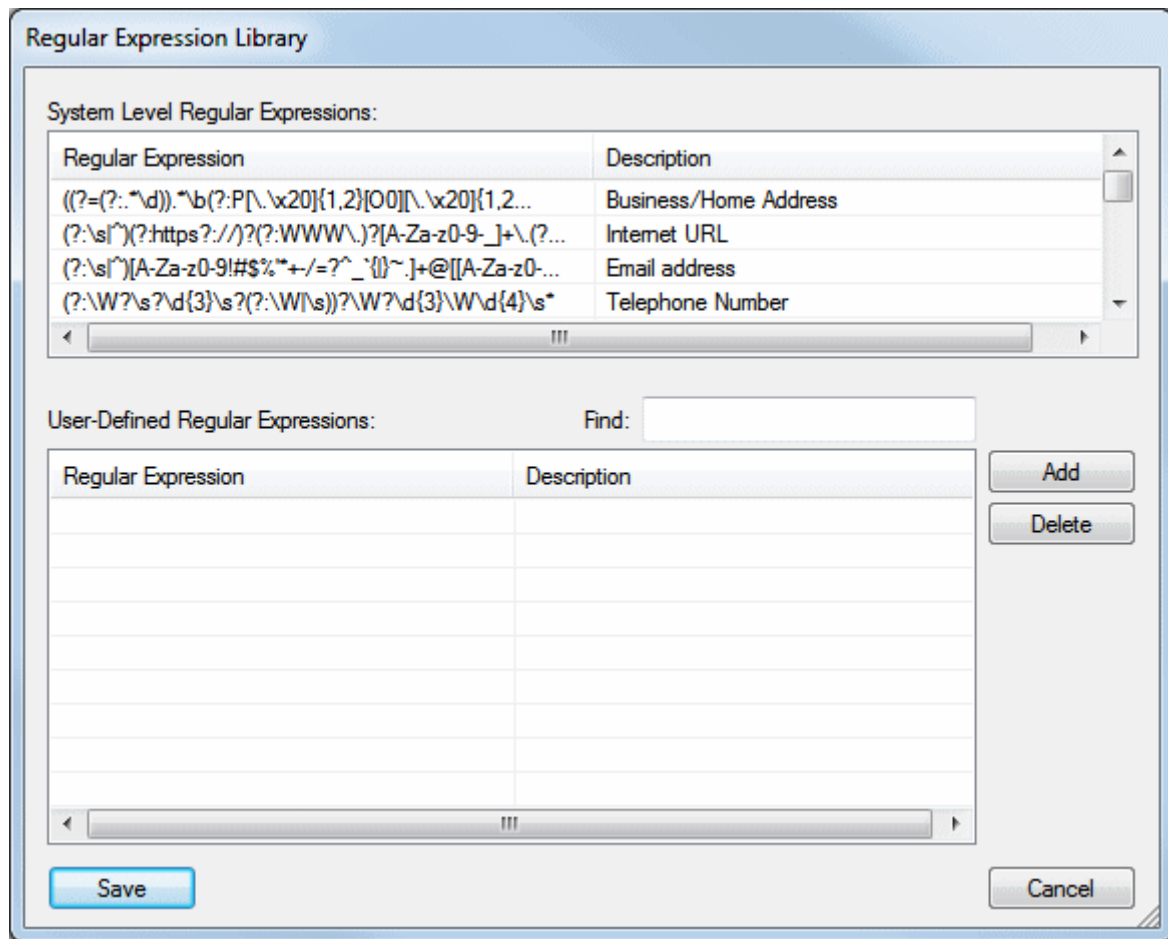
The Regular Expression Library contains all of the regular expressions that have been configured for your system. Within the library, you can edit your system's default, pre-populated expressions (i.e., for Address Extraction and Person Name Extraction) and add or delete your own custom expressions. By configuring the regular expressions in the library, you can make them readily available when configuring any form within your system and tailor your solution for specific Advanced Capture needs.

To configure regular expressions within the Regular Expression Library:

1. From the **Advanced Capture Configuration Window**, click the **Regular Expression Library** link in the **Tools** panel.

Note: You can also access the Regular Expression Library by clicking in a text field that can receive a regular expression (i.e., when configuring Data Field Zones or Keyword Lookup/Replace dictionary entries) and pressing **F2**. For more information on accessing the Regular Expression Library in this fashion, see [Configuring Regular Expressions From a Text Field on page 251](#).

2. The **Regular Expression Library** dialog box is displayed.



3. To quickly organize the list of system default expressions in the System Level Regular Expression section and/or the list of custom expressions in the User-Defined Regular Expressions section, click the column header of the desired column (e.g., **Regular Expression** or **Description**) to sort the list by the values in that column in ascending order. Click the same header again to sort the list in descending order.
4. To filter the list of custom expressions in the User-Defined Regular Expressions section, type the desired regular expression characters in the **Find** field.
5. Once you have organized the lists and/or located a specific regular expression, do one of the following:
 - Edit a system default expression (see [Editing Default Regular Expressions on page 249](#)).
 - Create a custom expression (see [Creating Custom Regular Expressions on page 250](#)).
6. To save any changes made to the Regular Expression Library, click **Save**.
7. To cancel any changes made to the Regular Expression Library, click **Cancel**.

Editing Default Regular Expressions

To edit a system default regular expression:

1. From the **Regular Expression Library** dialog, in the System Level Regular Expressions section, double-click on the regular expression's listing. The **Edit Regular Expression Entry** dialog box is displayed.

Edit Regular Expression Entry

Regular Expression: ((?=(?:.*\d)).*\b(?:P[\\.\x20]{1,2}[00][\\.\x20]{1,2}B|P[\\.\x20?]{00}[\\.\x20?B[00]X[\\.\x20?\d+B[IL1]N[\\.\x20?#?\\.\x20?\d+(?:INTERNATIONAL[\\.\x20?WA?Y|EXPRESS[\\.\x20?WA?Y(?:HIGH|FREE)WAY|INTERSTATE)[\\.\x20+(?:[NESW][\\.\x20+)?[\\.\x20?+\\.\x20+[NESW][\\.\x20+[\\.\x20+[NESW][\\.\x20+[\\.\x20+](?:BOULEVARD|BROADWAY|PARKWAY|AVENUE|CIRCLE|SQUARE|STREET|ROUTE|COURT|PLAZA|PLACE|POINT|ROAD))(?:BLVD|LANE|PARK|PIKE|PKWY|ROAD|AVE|BOX|HWY|HYW|LOTPKY|WAY|CR|DR|LN|RD|SQ|ST|CT|RT|WY))R[\\.\x20?R[\\.\x20?\d+U[\\.\x20?S

Set Default

☐ Case Sensitive

Sample Text:

Test

Description: Business/Home Address

Must contain exactly five capture groups:

1. Street address line 1
2. Street address line 2 (may be blank)
3. City name
4. State/province name
5. Postal code

Any letter case indicators in regular expression are processed as

Save Cancel

2. To edit the form of the regular expression, edit the text in the **Regular Expression:** field.

Note: To revert to the regular expression's default form, click **Set Default**.

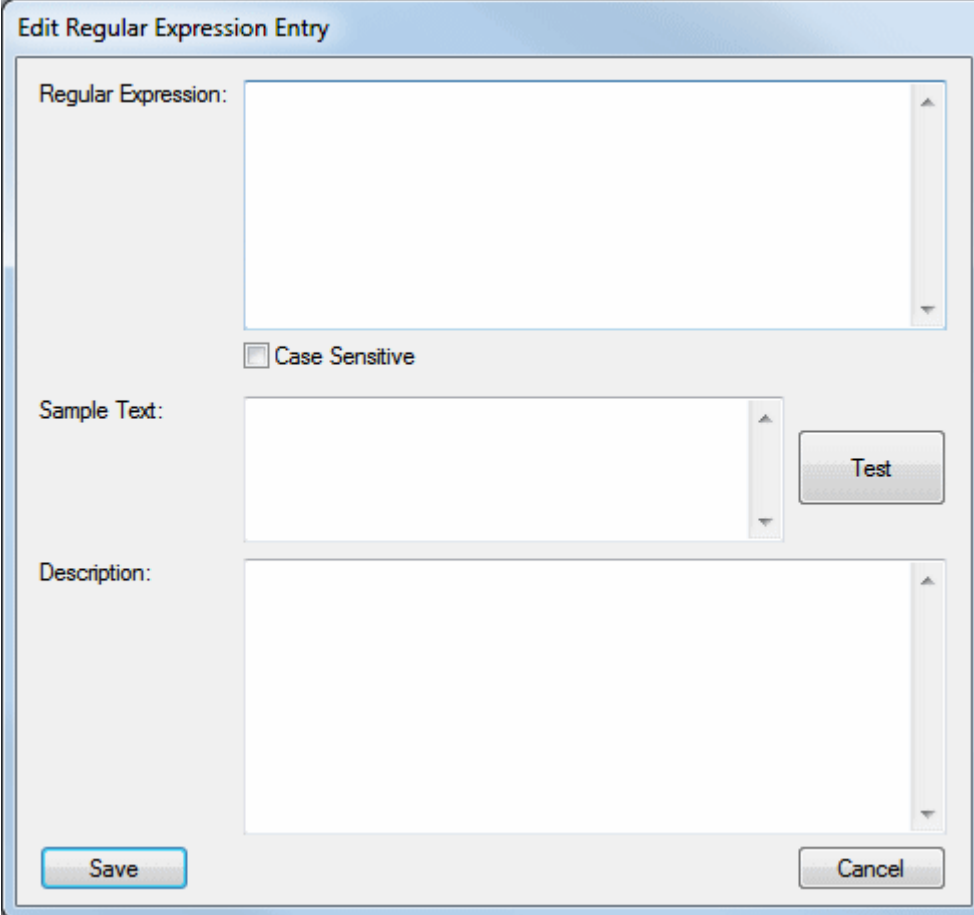
Note: For system default regular expressions, the **Case Sensitive** check box reflects the pre-defined case sensitivity of the expression; its status cannot be changed.

3. To test sample text against the form of the regular expression, enter the text in the **Sample Text:** field and click **Test**. A message is displayed, indicating whether the form of the regular expression was matched to the sample text.
4. To save your changes, click **Save**.
5. To cancel your changes, click **Cancel**.

Creating Custom Regular Expressions

To create a custom regular expression:

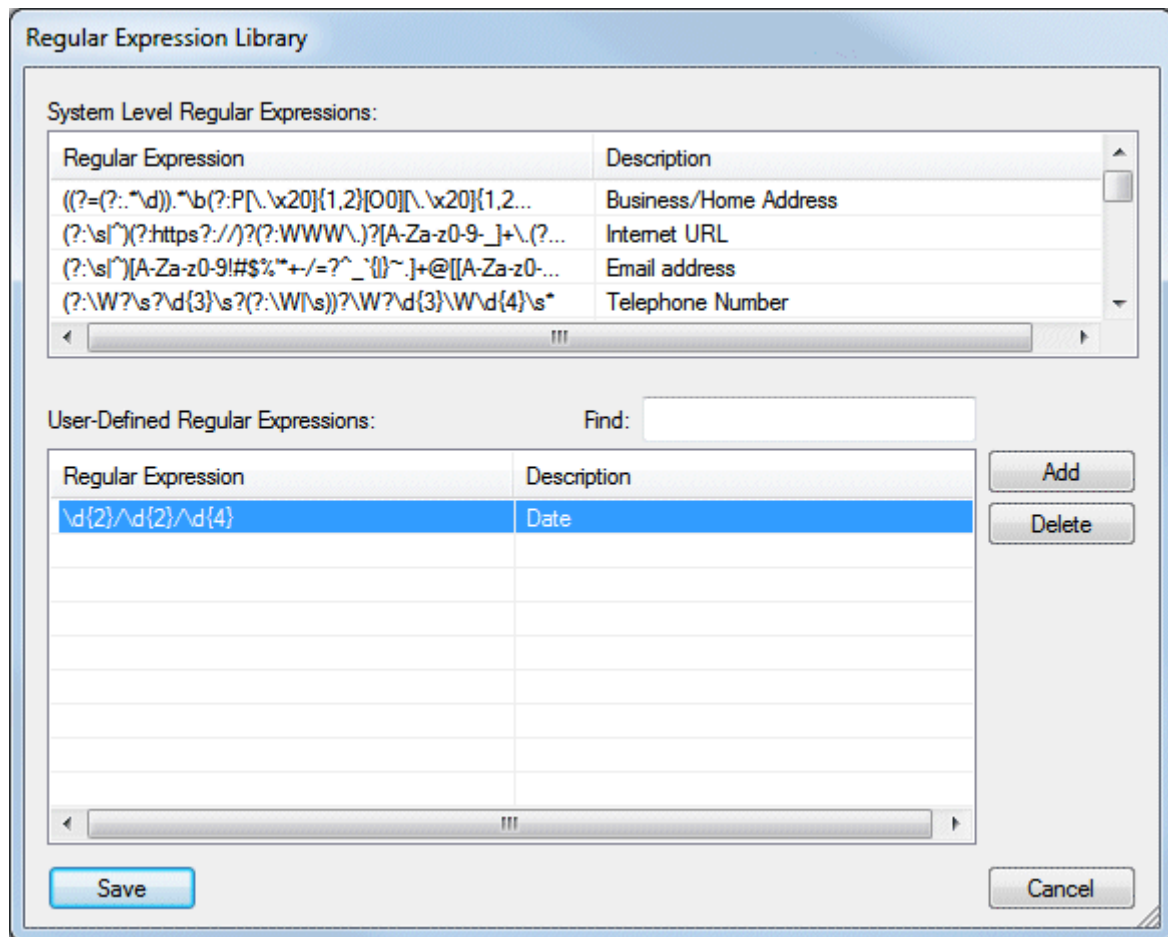
1. From the **Regular Expression Library** dialog, in the User-Defined Regular Expressions section, click **Add**. The **Edit Regular Expression Entry** dialog box is displayed.



Note: The **Edit Regular Expression Entry** dialog box displayed here is slightly different from the dialog box that is displayed when you are editing a system default regular expression.

2. To add a regular expression, enter the form of the expression in the **Regular Expression:** field.
3. To set case sensitivity as a consideration when matching extracted values to the regular expression, select **Case Sensitive**. If this option is not selected, case will not be considered when matching extracted values to the regular expression.
4. To test sample text against the form of the regular expression, enter the text in the **Sample Text:** field and click **Test**. A message is displayed, indicating whether the form of the regular expression was matched to the sample text.
5. Enter a description of the regular expression in the **Description:** field.
6. To save your changes, click **Save**.

7. To cancel your changes, click **Cancel**.
8. Once you have saved your changes to a custom expression, it is added to the User-Defined Regular Expressions section in the **Regular Expression Library** dialog.



9. To delete a custom expression, select it from the list of User-Defined Regular Expressions and click **Delete**.

Note: System default expressions, as well as custom expressions that have been assigned to a form's Data Field Zones or Keyword Lookup/Replace dictionary, cannot be deleted.

Configuring Regular Expressions From a Text Field

When configuring a regular expression for a Data Field Zone or a Keyword Lookup/Replace dictionary entry, you can access the Regular Expression Library from the text field that receives the regular expression. This gives you the ability to apply an existing expression from the library to the text field and to edit or create expressions in the library as you would when accessing the library from the **Tools** panel.

1. Click in a text field that will receive a regular expression and press **F2**. The **Match/Tag Value Configuration** dialog box is displayed.

Match/Tag Value Configuration

☐ Literal Text:

☐ Standalone Regular Expression:

☒ Regular Expression Library:

Regular Expression	Description
((?:\d+\x20(?:\x20[A-Z0-9/-]+\.)+)(?:\x...	Business/Home Address
\\$\?\-?(?=[1-9]{1}[0-9]{0,2}\,\d{3})~\,\d{0,2})?[1...	Currency
(?:\s ^)([A-Za-z0-9!#\$%&'*/+=?^_`{ }~]+)@[A-Za...	Email address
[+]?[0-9]*\.[0-9]+	Floating Point
(?:\s ^)(?=[https?://])(?=[WWW\.]?)[A-Za-z0-9-]...	Internet URL
([[:upper:]](?:[[:lower:]]*(?:[-]?[[:upper:]]...	Person Name - First Middle Last/Mixed Case
([[:upper:]](?:[[:upper:]]*(?:[-]?[[:upper:]]...	Person Name - First Middle Last/Upper Case
([[:upper:]] ([[:lower:]]*(?:[-]?[[:upper:]])))([[:lo...	Person Name - Last First Middle/Mixed Case
([[:upper:]] ([[:upper:]]*(?:[-]?[[:upper:]])))([[:un...	Person Name - Last First Middle/Inner Case

Create New Library Entry

Save Cancel

2. To enter a literal text string (i.e., specific text), select **Literal Text** and enter the text string, enclosed in quotation marks, in the corresponding field.
3. To enter a regular expression that will only be applied to this text field (i.e., an expression that will not be added to the Regular Expression Library), select **Standalone Regular Expression** and enter the expression in the corresponding field.

4. To apply a regular expression from the Regular Expression Library, select **Regular Expression Library**, and then select an expression from the list.
 - If you wish to edit an expression in the library, double-click on the expression in the list to open the **Edit Regular Expression Entry** dialog box. Then edit the entry as you would normally (see [Editing Default Regular Expressions on page 249](#)).
 - If you wish to add a new expression to the library, click **Create New Library Entry** to open the **Edit Regular Expression Entry** dialog box. Then create the new expression as you would normally (see [Creating Custom Regular Expressions on page 250](#)).

Note: Only one regular expression can be applied to each text field, except when configuring the **Begin table**, **End table**, and **Invalid row** fields for a Line Item Extraction Data Field Zone.

5. To apply the selected regular expression to the text field, click **Save**.
6. To cancel any selections or configuration changes, click **Cancel**.

Importing/Exporting Advanced Capture Forms

Advanced Capture forms can easily be imported into or exported from your Advanced Capture solution. This allows the forms to be shared between Advanced Capture solutions, reducing or eliminating the need to manually recreate the same forms across multiple solutions.

Advanced Capture forms are stored in the OnBase database and therefore do not need to be imported or exported between workstations that are part of the same Advanced Capture solution. Once created within your Advanced Capture solution, the Advanced Capture forms are available to be viewed and modified by users with the proper Advanced Capture configuration rights.

Exported Advanced Capture forms are saved as XML files for easy transmittal.

Importing Advanced Capture Forms

1. From the **Advanced Capture Configuration** window, click the **Import/Export Configuration** link in the **Tools** panel.
2. A confirmation dialog is displayed. Click **Import**.
3. The **Open** dialog box is displayed. Browse to the location of the XML file containing the Advanced Capture form data and click **Open**.
4. If the XML file contains data for a form that has the same global identification number as an existing form in your Advanced Capture solution, but a different name, a message is displayed prompting you to choose how to treat these form duplicates. See [Resolving Form Template Collision on page 255](#) for more information.

- Advanced Capture Configuration Import**

Created by Manager on workstation DOCm007163 (AUTO) at 2012-03-30 10:25:25
Source database 11.1 using schema version 3.5-324 and system install ID DEMO

Form Template Name	Needs Reconciliation	Already Exists in Database
<input checked="" type="checkbox"/> <Global Configuration Objects>	NO	N/A
<input checked="" type="checkbox"/> AI Invoice Form	NO	NO
<input checked="" type="checkbox"/> AI Transcript Form	NO	NO

Select All Deselect All

Import Cancel

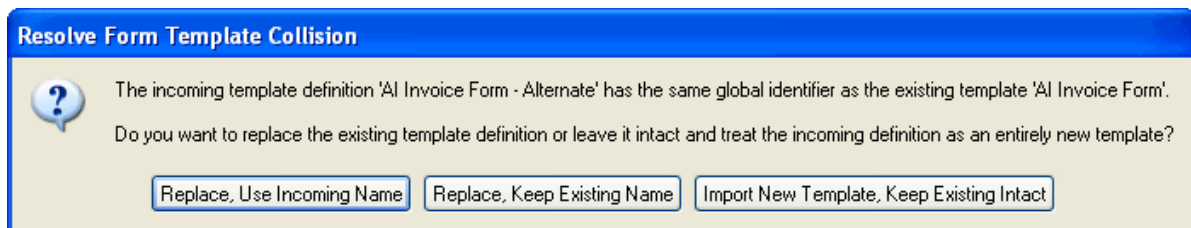
Information about each Advanced Capture form contained in the XML file is listed in the **Advanced Capture Configuration Import** dialog box (e.g., the forms' names, their reconciliation status, and if they already exist in your solution).

- Note:** If you selected **<Global Configuration Options>** to be imported in step 6, all custom, user-defined regular expressions are automatically imported.

9. Depending on the configuration of the solution that the Advanced Capture forms were created in, a message may be displayed warning you that the Advanced Capture forms contain unresolved configuration references. See [Reconciling Configuration References on page 256](#) for more information.
10. If the Advanced Capture forms you are importing have one or more field zones configured to use VB scripting and these VB script(s) have the same name as a script already existing in your solution, you can:
 - Re-name the incoming VB script before it is imported by clicking **Use Incoming VB Scripts** in the confirmation message.
 - Discard the incoming VB script and associate the incoming Advanced Capture form to the script with the same name already in your solution by clicking **Use Existing VB Scripts** in the confirmation message.All incoming VB scripts that do not have the same name as scripts already in your solution are always imported.
11. A message is displayed informing you that the import process was successful, and you are prompted to perform Advanced Capture form precedence order analysis.

Resolving Form Template Collision

When attempting to import Advanced Capture forms, you may be presented with a message notifying you that a form you are attempting to import has the same global identification number as an existing form in your Advanced Capture solution.



This message indicates that while the Advanced Capture form you are trying to import has the same identifier as an existing form, it does not have the same name. This might occur if you export a form from your solution into another solution, change the form's name either in the other solution or in the XML file, and then re-import the form back into your solution.

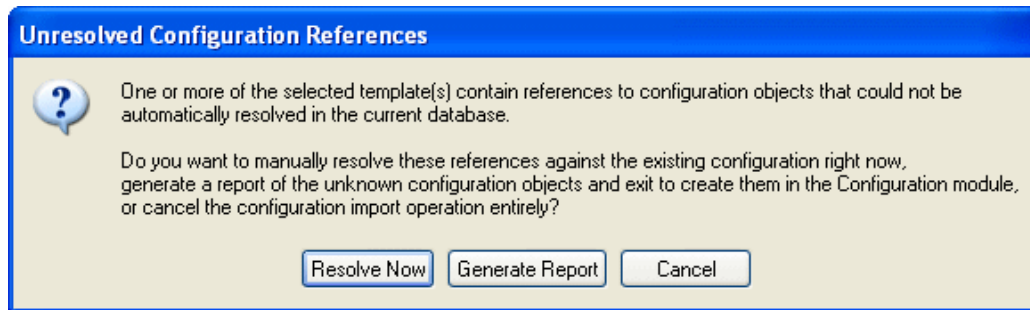
You have three options for resolving these collisions of "duplicate" forms on import:

- Click **Replace, Use Incoming Name** to replace the existing form with the form you are importing, using the name of the incoming form. The global identification number remains unchanged.
- Click **Replace, Keep Existing Name** to replace the existing form with the form you are importing, but to apply the name of the existing form to the incoming form. The global identification number remains unchanged.
- Click **Import New Template, Keep Existing Intact** to retain the existing form with its global identification number intact and to import the new form while assigning it a new, unique global identification number.

If you are attempting to import multiple form duplicates in the same XML file, you will be presented with this collision message for each separate form duplicate in succession. This allows you to choose different resolution options for different form duplicates, if desired. Once all collisions have been resolved, the import process can continue as normal.

Reconciling Configuration References

When attempting to import Advanced Capture forms, you may be presented with a message warning you that the forms you are attempting to import contain unresolved configuration references.



This message indicates that the Advanced Capture forms you are trying to import have been configured using elements (e.g., Document Types, Keyword Types, scan queues) that do not exist in your solution.

For example, a form you are trying to import may be configured to be matched to documents belonging to the **WSU - Transcripts** Document Type, and your solution does not contain a Document Type named **WSU - Transcripts**.

You have two options for resolving these references:

- Click **Generate Report** to generate a list of all of the unresolved configuration references.

A text document is displayed containing a list of all of the unresolved configuration references for the form(s) you are trying to import. You can use this report as a reference to create these elements in the Configuration module.

Note: While this report lists all of the unresolved configuration elements (e.g., Document Types, Keyword Types, scan queues, etc.), it does not give any information about their configuration settings.

- The **Configuration Object Mapping: <Configuration Element>** dialog box is displayed.

Select the appropriate Document Type for each unknown item listed
Select <None> for any items that do not have corresponding entries in this database

Set all references to:

OK Cancel

For example, an Advanced Capture form was associated with a Document Type named **HYL - Transcripts** in the solution it was exported from. However, this Document Type does not exist in the solution the form is being imported into, so you must use the **Corresponding Item** drop-down list (or the **Set all references to** drop-down list) to select a different Document Type to “substitute” for the **HYL - Transcripts** Document Type.

When you have selected a substitute for all of the unresolved elements listed in the dialog box, click **OK**. If you have unresolved configuration elements of a different

Once all unresolved configuration features have been resolved, the Advanced Capture forms are imported into your solution. You are given the option to save the reconciliation mapping information you specified to automatically resolve any future Advanced Capture forms imported from this source.

1. From the **Advanced Capture Configuration** window, click the **Import/Export Configuration** link in the **Tools** panel.
2. A confirmation dialog is displayed. Click **Export**. The **Advanced Capture Configuration Export** dialog box is displayed.

[illegible]

Deselect All

Browse

Cancel

The following information is displayed for each Advanced Capture form: the Form ID, the name of the Advanced Capture Form, the scan queue(s) that the Advanced Capture form is assigned to, and the Document Type that the documents matched to the Advanced Capture form are assigned to.

3. Select the check box next to each Advanced Capture form to be exported.
To select all Advanced Capture forms, click **Select All**. To deselect all Advanced Capture forms, click **Deselect All**.

Note: All entries in the Regular Expression Library will be exported automatically.

4. Click **Browse**. Browse to the location where you would like to save the XML file containing the exported Advanced Capture form data and enter a name for the Advanced Capture form data file.
5. Click **Save** to close the **Save As** dialog box.
6. Click **Export**. Once the Advanced Capture form data is successfully exported, a confirmation message is displayed.

Configuring Form Definition Groups

If you configure custom capture processes for Advanced Capture using the Capture Process Designer within Document Imaging, you can configure form definition groups within Advanced Capture to send the same documents through Advanced Capture processing multiple times, with only the specified types of information to be extracted each time the documents are processed.

For example:

You have configured an Advanced Capture form for a batch of documents, where you sometimes wish to only extract header data and other times wish to only extract detail data. With all extraction zones (for both header and detail data) configured for the form, you can create two form definition groups called **Header** and **Detail** and assign the form to both groups.

Once you have created the form definition groups, you can assign each one as an activation group on the form's individual extraction zones, as appropriate (i.e., assign the **[Header]** activation group to zones configured to extract header information, and assign the **[Detail]** activation group to zones configured to extract detail information).

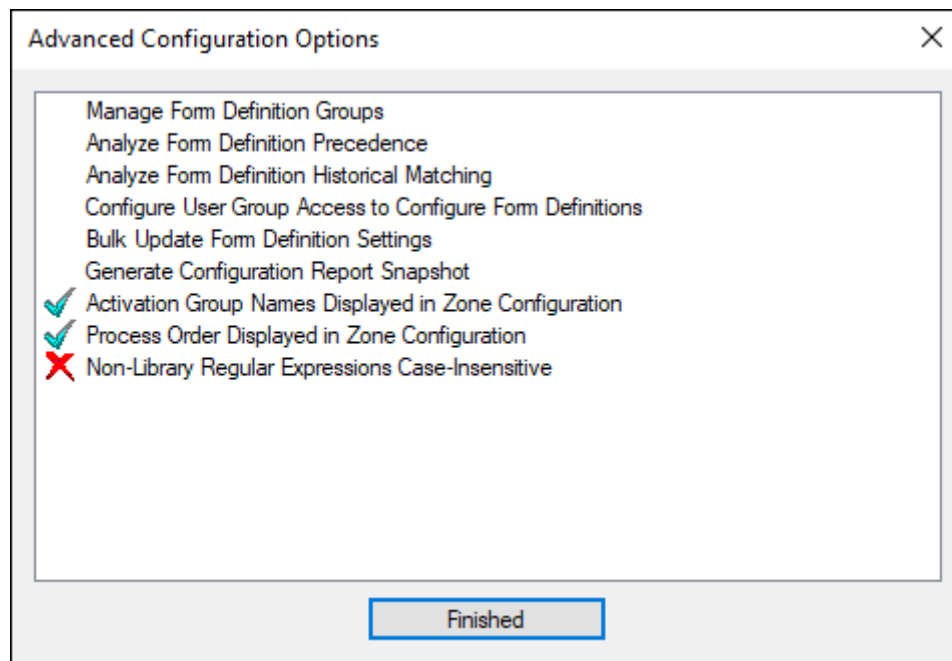
When configuring the related custom capture process using the Capture Process Designer in OnBase Studio, you can create two separate Advanced Capture status steps: one with **Header** as the active template group and one with **Detail** as the active template group. If you then configure this custom capture process to only route documents to one of these Advanced Capture status steps or the other, based on appropriate criteria, you can ensure that the process will only attempt to extract values for the active group.

Note: Other extraction zones that are not tied to activation groups would continue to be processed in either status step, as normal. Similarly, activation groups that are turned on or off by Form Identification or Page Registration Zones would continue to be processed normally in either status step.

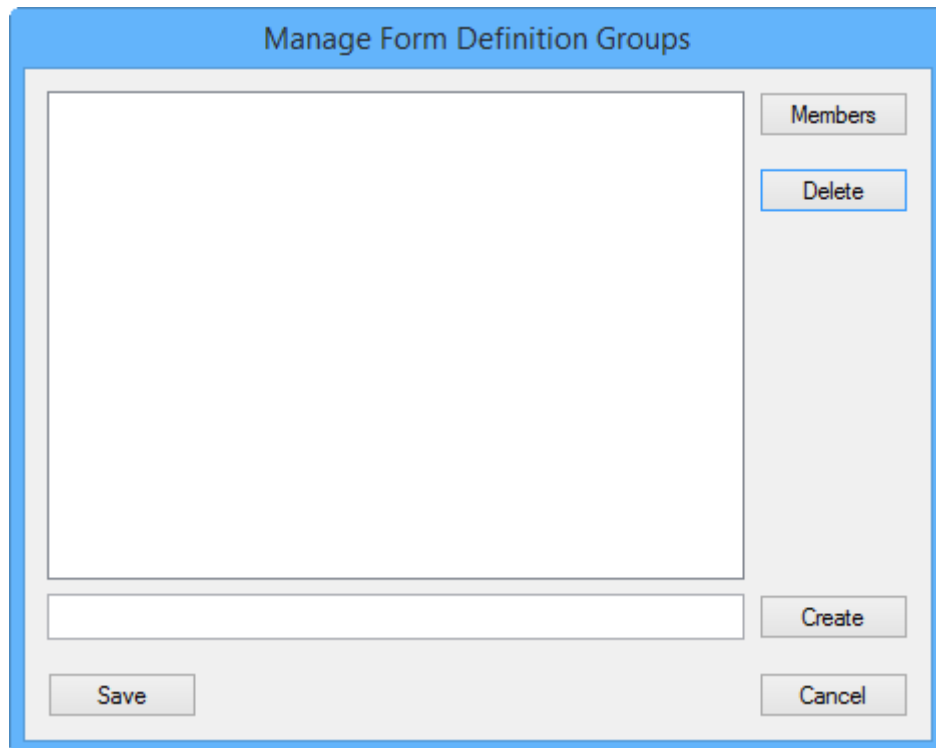
For more information on configuring custom capture process using the Capture Process Designer, see the **Document Imaging** module reference guide.

To configure a form definition group:

1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.



2. Click **Manage Form Definition Groups**. The **Manage Form Definition Groups** dialog box is displayed.



3. To create a new form definition group, type a descriptive name for the group in the text field and click **Create**. The **<Form Definition Group>** dialog box is displayed.

The screenshot shows a dialog box titled "Test Group". It features two list boxes. The left list box, labeled "Available", contains two items: "AC Transcript Form" and "HYL Transcript Form". The right list box, labeled "Assigned", is currently empty. Between these two list boxes are two buttons: "Add >>" and "<< Remove". At the bottom of the dialog, there are two buttons: "Save" on the left and "Cancel" on the right.

4. To assign a form to the form definition group, select the form in the Available list on the left and click **Add >>**.
5. To remove a form from the form definition group, select the form in the Assigned list on the right and click **<< Remove**.
6. Once you are finished adding forms to the form definition group, click **Save** to save your changes and return to the **Manage Form Definition Groups** dialog box.

Note: Any form can be assigned to any form definition group, no form definition group, or multiple form definition groups at the same time.

7. To edit a form definition group, select the group from the list and click **Members**. The **<Form Definition Group>** dialog box is displayed. From here, you can repeat steps 4 to 6 as appropriate to edit the group and save your changes.
8. To delete a form definition group, select the group from the list and click **Delete**. When the confirmation message is displayed, click **Yes** to confirm the deletion.

Note: When you delete a form definition group, only the group itself is deleted. The forms assigned to the group are not deleted, and they can continue to be used in your Advanced Capture solution.

9. Once you are finished configuring form definition groups, click **Save**.

Advanced Capture Form Order Analysis

Your Advanced Capture solution keeps track of how frequently each form is matched to documents that undergo Advanced Capture.

As an administrator, you can analyze these statistics, and, based on this analysis, your Advanced Capture solution can update the order of its forms and attempt to match the more frequently-used forms to documents first.

For example:

You scan a batch of **AP - Invoice** documents daily, and within these batches, there are three potential Advanced Capture forms available to these documents (**AP - Computers R Us**, **AP - The Phone Shack**, and **AP - Office Supply Warehouse**).

Based on the volume of documents that are scanned and indexed:

- The **AP - Office Supply Warehouse** form is matched to 75% of the documents
- The **AP - Computers R Us** form is matched to 15% of the documents
- The **AP - The Phone Shack** form is matched to 10% of the documents

Currently, Advanced Capture attempts to match the forms to incoming **AP - Invoice** documents in this order: **AP - The Phone Shack**, **AP - Computers R Us**, and then **AP - Office Supply Warehouse**.

In its current state, your **AP - Invoice** documents are not being indexed very efficiently because only 25% of all incoming documents are being matched to a form in less than three attempts.

After running the form order analysis, your Advanced Capture solution re-orders the forms so that it attempts to match all incoming **AP - Invoice** documents to the **AP - Office Supply Warehouse** form first; the **AP - Computers R Us** form second, and the **AP - The Phone Shack** form third.

Now, 90% of all incoming **AP - Invoice** documents are matched to a form in less than three attempts.

Note: If your solution is configured to use an Institutional Database, be aware that a super-user can view/re-order all forms across all institutions, while a non-super user can only view and re-order templates that are configured for his/her institution.

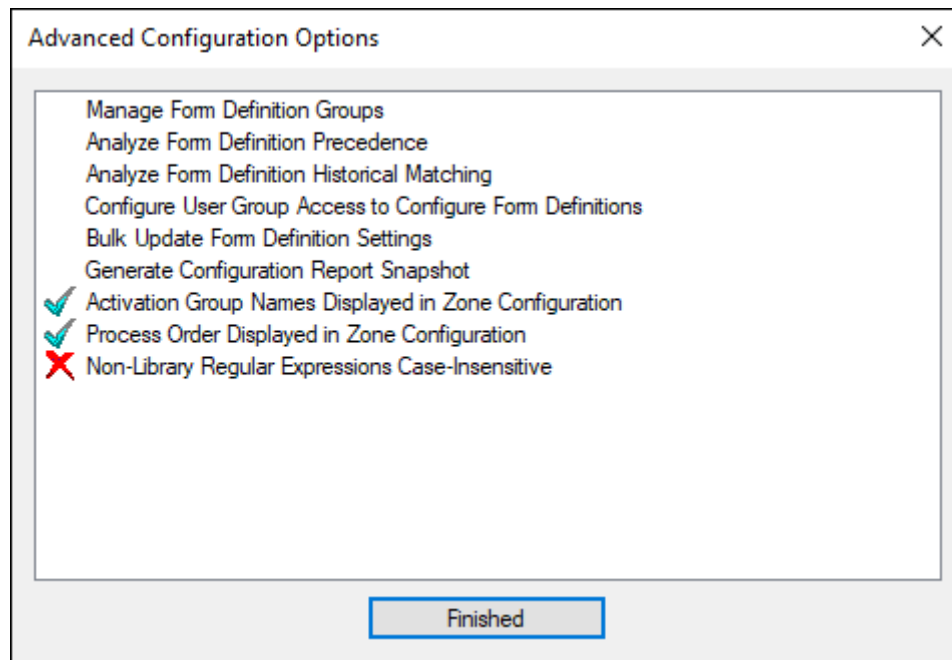
In order for the Advanced Capture form order analysis to provide accurate results, you must ensure that you have a sufficiently-large number of documents that have been successfully matched to a form currently in your system. By default, you can perform a form order analysis with a minimum of 100 documents and a maximum of 30,000 documents.

You can perform a form order analysis at any time from the **Advanced Capture Configuration** window. Additionally, you will always be prompted to perform a form order analysis after you have imported forms into your solution.

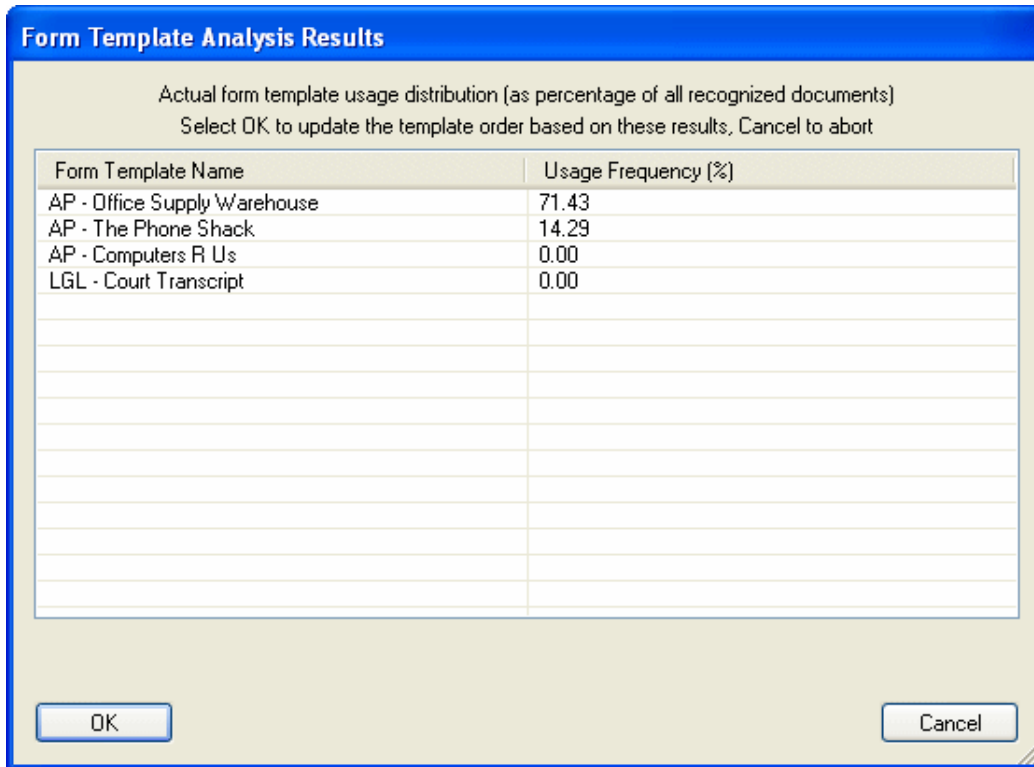
Tip: It is considered a best practice to periodically perform a form order analysis and re-order your forms. The frequency of this analysis should be based on the number of documents you are processing via Advanced Capture and the number of forms configured for your solution.

To perform a form order analysis:

1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.



2. Click **Analyze Form Definition Precedence** to trigger your Advanced Capture solution to analyze its form frequency statistics. Once completed, they are displayed in the **Form Template Analysis Results** window.



The screenshot shows a window titled "Form Template Analysis Results". Inside, there is a text box stating: "Actual form template usage distribution (as percentage of all recognized documents)" and "Select OK to update the template order based on these results, Cancel to abort". Below this is a table with two columns: "Form Template Name" and "Usage Frequency (%)". The table contains four rows of data, sorted in descending order of frequency.

Form Template Name	Usage Frequency (%)
AP - Office Supply Warehouse	71.43
AP - The Phone Shack	14.29
AP - Computers R Us	0.00
LGL - Court Transcript	0.00

At the bottom of the window are two buttons: "OK" and "Cancel".

The forms are displayed in descending order (i.e., the most-frequently used form is listed first and the least-frequently used form is listed last) along with their percentage of use.

3. Click **OK** to re-order the forms. Once the forms have been re-ordered, a confirmation message is displayed.

Now, your Advanced Capture solution will first attempt to match the most commonly used forms to incoming documents in order to more efficiently process documents.

Advanced Capture Form History Analysis

Your Advanced Capture solution keeps track of which form definitions are matched to documents that undergo Advanced Capture.

As an administrator, you can analyze these statistics. By default, you can perform a form definition historical matching analysis on a maximum of 30,000 documents.

You can perform a form definition historical matching analysis at any time from the **Advanced Capture Configuration** window.

To perform a form definition historical matching analysis:

1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.
2. Click **Analyze Form Definition Historical Matching** to trigger your Advanced Capture solution to analyze its form definition historical matching statistics. The **Form Definition Matching Analysis** window is displayed.
3. Using the Document Type Group and Document Type lists, select the Document Type that the documents you want to analyze belong to.
Once a Document Type is selected, the Keyword Type fields associated with the selected Document Type are displayed in the Keywords section.
4. Using the Date Range fields (i.e., **From** and **To**) and the Keyword Type fields, enter a date range and/or Keyword Values to retrieve the documents you want to analyze.
5. As an alternative to the above search options, you can use the **Scan Queue** and **Batch** drop-down lists in the **Find Active Scan Batch** section to search the documents in a specific batch. This allows you to search documents in uncommitted batches belonging to custom capture processes.
 - a. From the **Scan Queue** drop-down list, select the scan queue associated with the batch you wish to search. Note that all scan queues you have rights to, including those configured for custom capture processes, are available for selection.

Note: Once you select a scan queue, all of the above options are deselected and disabled. To re-enable these options for selection, select **<None>** from the **Scan Queue** drop-down list.

- b. From the **Batch** drop-down list, select the batch you wish to search.
-

Note: The **Batch** drop-down list is only enabled when a scan queue is selected in the **Scan Queue** drop-down list.

6. Click **Find** to perform the search.
7. The **Form Definition Match Results** window is displayed.
 - Documents that were retrieved by the historical matching analysis query, but have never undergone Advanced Capture, are listed on the **Not Processed** tab. If no documents meet this criteria, the tab is not displayed.
 - Documents that were retrieved by the historical matching analysis query and have undergone Advanced Capture, but were not matched to a form definition, are listed on the **Not Matched** tab. If no documents meet this criteria, the tab is not displayed.

- Documents that were retrieved by the historical matching analysis query and were matched to a form definition during Advanced Capture are listed on the **Matched** tab. The documents are listed under the form definitions to which they were matched. If no documents meet this criteria, the tab is not displayed.

Note: All documents listed on the **Not Processed**, **Not Matched**, and **Matched** tabs can be used to configure Advanced Capture forms in the **Advanced Capture Configuration** Window. To use a document to configure an Advanced Capture form in the **Advanced Capture Configuration** window, double-click on the document in the list. The document is displayed in the Document Viewer and added to the **Batch Documents** pane.

- Form definitions that were not applied to any of the documents retrieved by the historical matching analysis query are listed on the **Unused Form Definitions** tab. If no form definitions meet this criteria, the tab is not displayed.

Note: The form definitions listed on the **Unused Form Definition** tab may have been matched to other documents not retrieved by the historical matching analysis query.

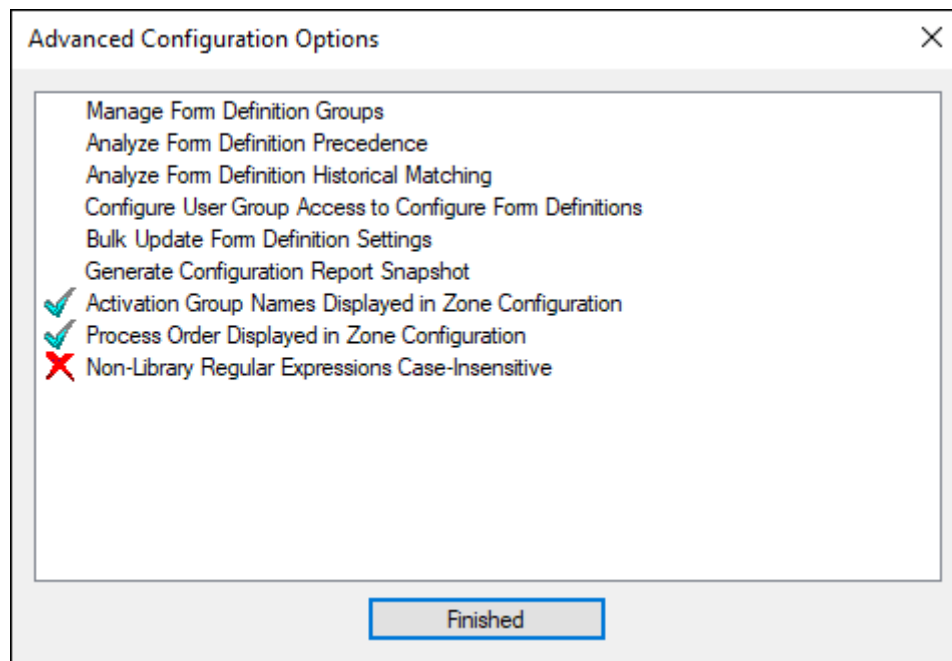
8. When your analysis is complete, and the **Form Definition Match Results** window is still displayed, click **Finished**.

Granting Configuration Rights to Additional User Groups

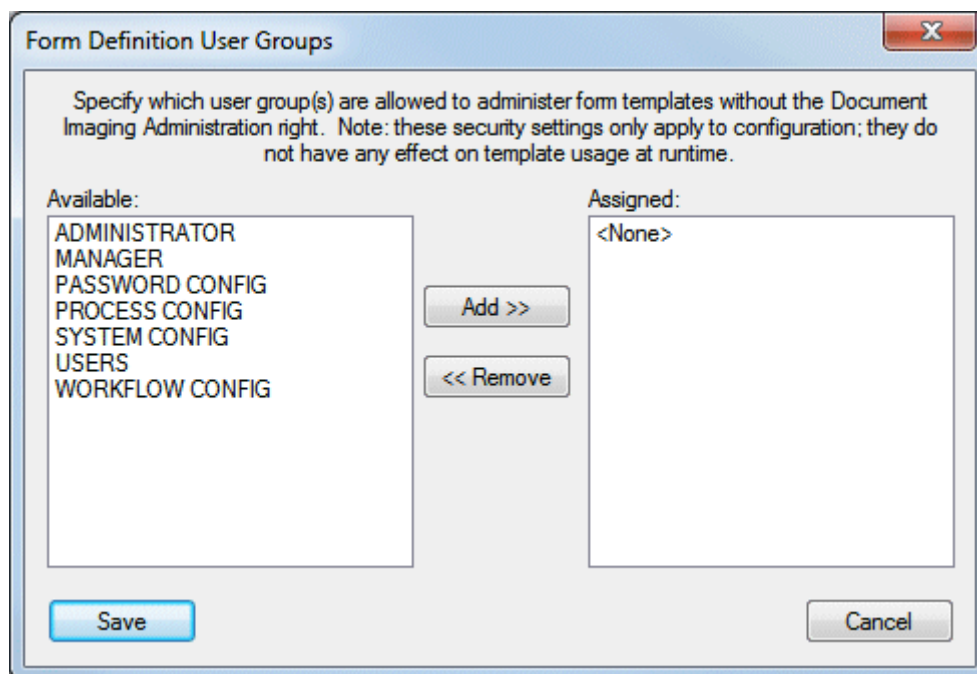
Users belonging to User Groups that have been given both the Advanced/Intelligent Capture product right and the Document Imaging administrative processing privilege have the ability to create, modify, or delete Advanced Capture forms and to associate Advanced Capture forms to existing documents. However, users can also perform these actions without being granted the Document Imaging privilege if they belong to User Groups that have been granted form definition rights.

To grant a User Group form definition rights:

1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.



2. Click **Configure User Group Access to Configure**. The **Form Definition User Groups** dialog box is displayed.



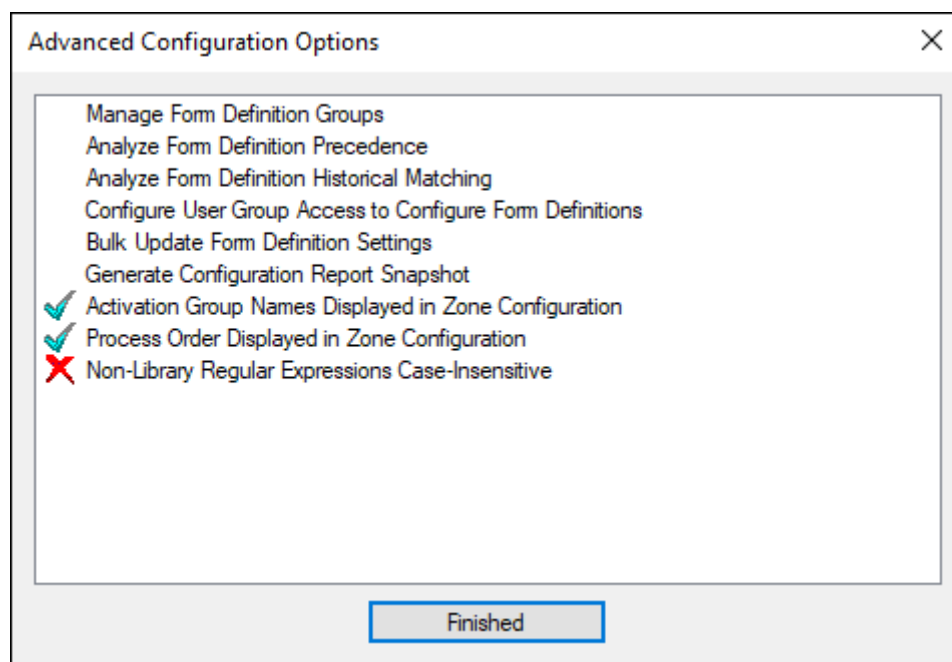
3. To assign a User Group form definition rights, select the User Group in the **Available** list and click **Add >>**.
4. To revoke a User Group's form definition rights, select the User Group in the **Assigned** list and click **<< Remove**.
5. To save your changes and exit the dialog, click **Save**.

Updating Multiple Form Definitions

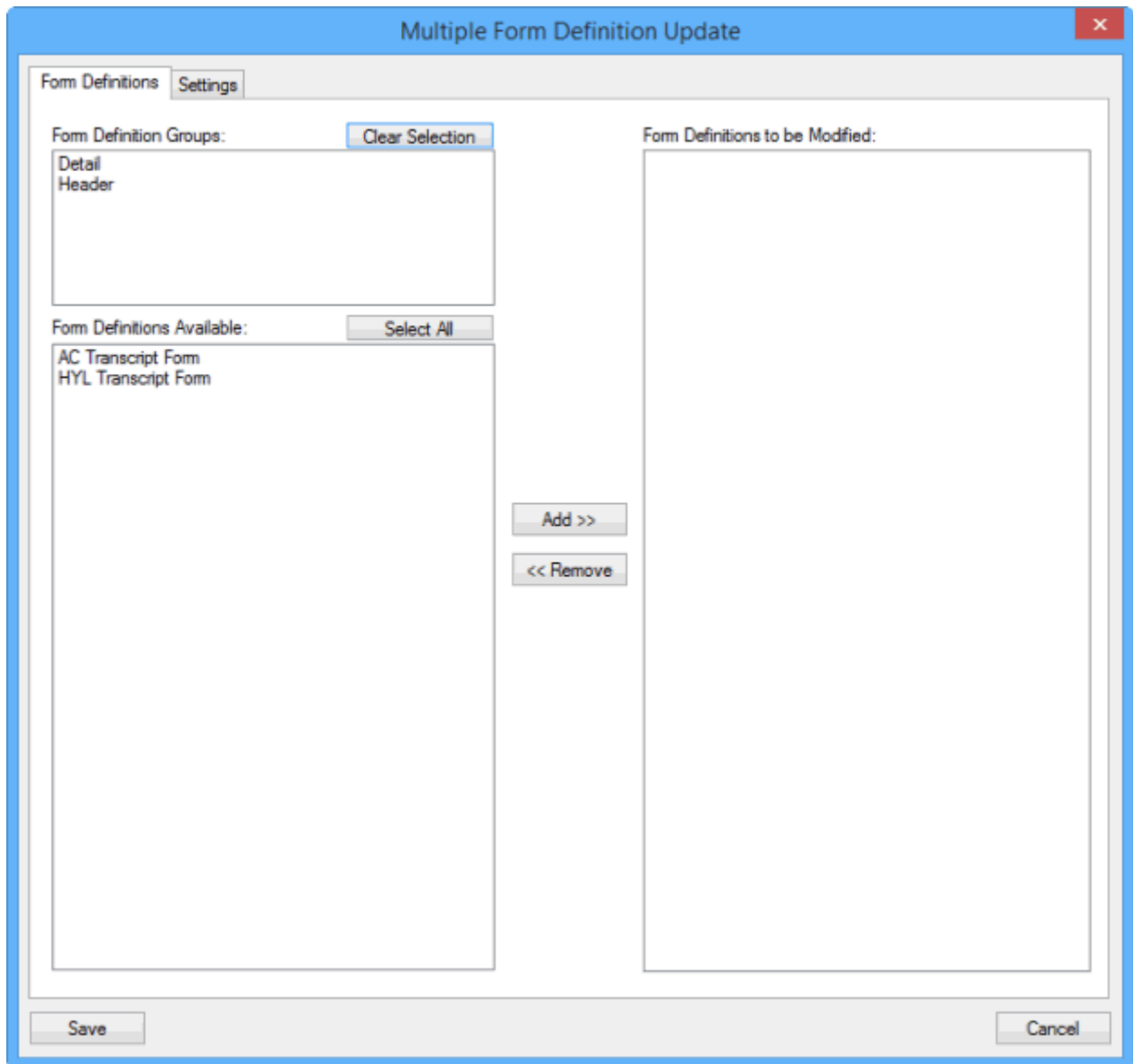
If you wish to apply the same changes to multiple forms in your system, you can save time by performing bulk updates to these multiple forms simultaneously.

To update the settings for multiple form definitions:

1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.



2. Click **Bulk Update Form Definition Settings**. The **Multiple Form Definition Update** dialog box is displayed.



3. From the **Form Definitions** tab, select all available form definition groups and form definitions you wish to modify, and click **Add >>**. The selected form definition groups and form definitions are moved from the available fields on the left to the **Form Definitions to be Modified** field on the right.
4. To remove any form definition groups or form definitions already set to be modified, select them and click **<< Remove**. They are moved from the **Form Definitions to be Modified** field on the right to the available fields on the left.

5. To access the form definition settings to be modified, click the **Settings** tab.

Multiple Form Definition Update

Form Definitions Settings

Assigned Document Type: <Unchanged>

☒ Documents require subsequent manual indexing <Unchanged>

☒ Carry keywords onto subsequent documents ☒ Auto-burn redactions

☒ Create XML data rendition ☒ Default for identification chain

☒ Create revision of single existing matching document

Suspect Level: (1 - 99)

Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: Maximum Page Count: 0 = no breaking

VB Script: <Unchanged>

Ancillary Form Settings

Scan Queue/Keyword Rules ☐ Update

Keyword Lookup and Replace ☐ Update

Form Definition Lookup Text ☐ Update

Image Processing Operations ☐ Update

Configuration User Groups ☐ Update

☐ Update

Document Types Allowed for Form:

Remove

Add

Configuration Note:

Indexing Note:

Leave items blank, indeterminate or <Unchanged> to leave unchanged on selected form(s). Otherwise, items set with a value or where "Update" is checked will be applied to all selected form definition(s).

Save Cancel

By default, the settings are blank (text fields), marked indeterminate (check boxes marked with black squares), or set to **<Unchanged>** (drop-down lists).

6. To keep settings unchanged for the forms you previously selected, leave them as blank, indeterminate, or **<Unchanged>**.

7. To modify settings for all of the forms you previously selected, change their values accordingly (e.g., enter values in the text fields, check or uncheck the check boxes, and/or select values other than **<Unchanged>** in the drop-down lists). To modify settings next to an **Update** check box (e.g., **Ancillary Form Settings** and **Document Types Allowed for Form**), you must also check **Update**.

Tip: For more information on these form definition settings, see [Defining Advanced Capture Forms](#) on page 44.

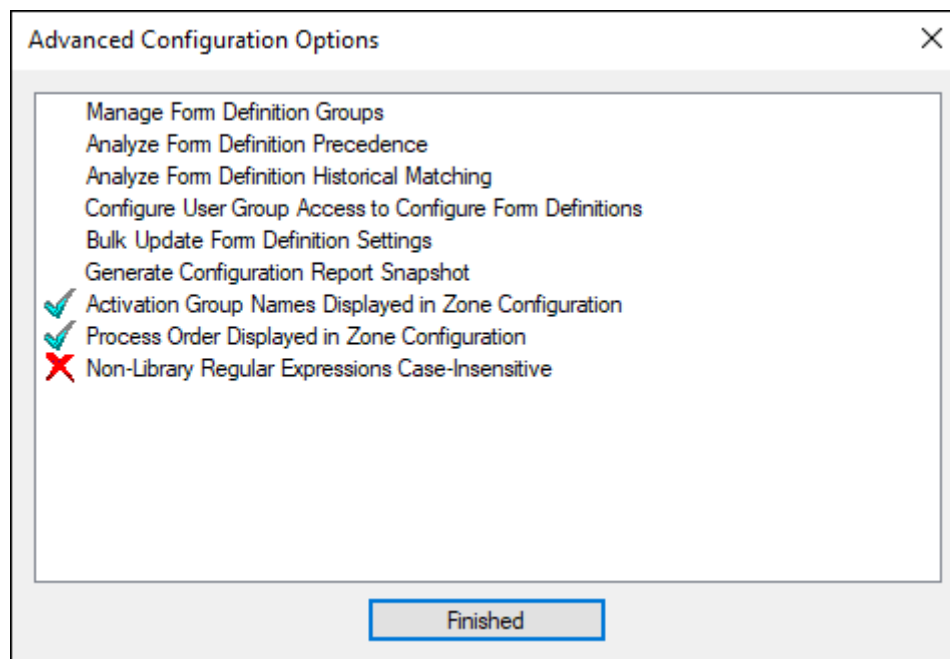
8. Once you have finished modifying the settings for the selected forms, click **Save**.
9. When prompted to confirm your changes, click **OK** to confirm or **Cancel** to cancel.

Generating Configuration Reports

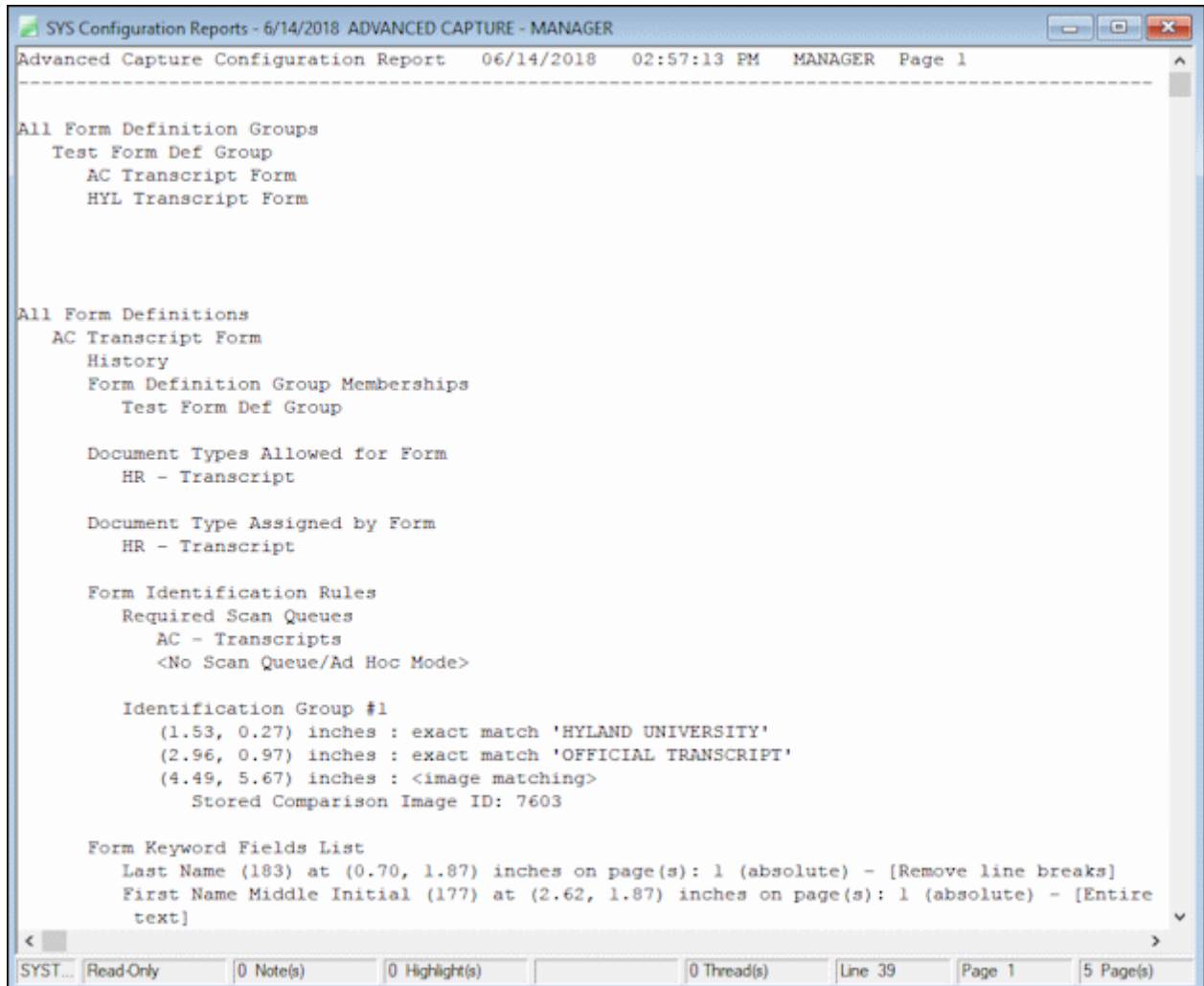
You can quickly generate a configuration report to show all form definitions and related information in an easy-to-read format.

To generate a configuration report:

1. From the **Advanced Capture Configuration** window, click the **Advanced Options** link in the **Tools** pane. The **Advanced Configuration Options** dialog box is displayed.



2. Click **Generate Configuration Report Snapshot**. A status bar is displayed to show the progress of the generation process. Once the process is complete, a configuration report is displayed.



The configuration report includes all form definitions and related information. Once run, the report is stored in OnBase as a document belonging to the **SYS Configuration Report** Document Type.

Configuring XML Renditions

The Advanced Capture module can be configured to generate an XML rendition of extracted Keyword Values when indexing documents. In order to configure XML rendition generation, you must perform the following configuration steps:

1. **Configure the Document Types being indexed to allow renditions.** For more information on enabling renditions, see the **System Administration** module reference guide or help file.
2. **Enable the creation of an XML data rendition on each form to be processed.** Advanced Capture Forms must have the **Create XML data rendition** option selected. To enable XML data rendition creation in new forms, see [Creating a New Advanced Capture Form on page 45](#). To enable XML data rendition creation in existing forms, see [Modifying a Form on page 52](#).
3. **Define XML nodes for all Data Field Zones.** Defining the XML node in a Data Field Zone generates the XML tags for each extracted Keyword Value. See [Configuring A Data Field Zone on page 121](#) for more information on defining XML Nodes in Data Field Zones.

Testing Advanced Capture Forms

Once you have matched a form to one or more sample document(s), you can run a test Advanced Capture process to see if the Form Identification, Registration, and Data Field Zones are processed correctly on the sample document(s).

Tip: It is considered a best practice to regularly test each Advanced Capture form to ensure that it does not require updating. All Advanced Capture forms should be tested after they are created and after any modifications have been made to them.

The test can be performed at the document level or at the batch level. The document level test is likely to be faster than the batch level test because only one document is being evaluated, but the batch level test can be useful because the OCR engine automatically attempts to match an existing form to all of the documents in the batch, and can help you determine which documents require a new or modified Advanced Capture form.

Note: All documents in the batch are listed in the **Batch Documents** panel, in the upper-right corner of the **Advanced Capture Configuration** window. A green check mark is displayed next to documents in the batch that have been matched to an Advanced Capture form, and a red X is displayed next to documents that are unclassified.

To test an Advanced Capture form:

1. From the **Advanced Capture Configuration** window, ensure that the document and its associated Advanced Capture form are displayed in the Document Viewer.
2. To test a single document, click **Process Current Document** in the **Tools** panel.
To test an entire batch, click **Process Entire Batch** in the **Tools** panel.
During processing, click **Cancel** at any time to cancel the test.

Note: If the **Process Entire Batch** test is canceled before it is complete, any documents in the batch that had not been processed remain listed as unclassified or they are matched to the Advanced Capture form that they were matched to during pre-processing or the last test process.

- If you are testing a single document, the **Test Result Verification** window is displayed once the test is complete.

Result Verification

- Form type: AI Transcript Form (103)
 - HR - Transcript
 - Calculated scaling factor (X:1.00 Y:1.00), offset (3,0) pixels
 - Field counts: 16 identified, 0 suspect
 - Line Item Counts/Page
 - Page 1 - Line Item Count: 4
 - Total Line Items: 4
 - Default suspect level: 75
 - Field results
 - Last Name (169) at (0.66, 1.86) inches on page(s): 1 (absolute) - [Entire text]
 - Text value: REAGAN
 - Calculated suspect level: 70, tolerance: 75
 - Document page: 1
 - First Name Middle Initial (170) at (2.63, 1.86) inches on page(s): 1 (absolute) - [Entire text]
 - Text value: RONALD W.
 - Calculated suspect level: 70, tolerance: 75
 - Document page: 1
 - ID # (171) at (5.59, 1.82) inches on page(s): 1 (absolute) - [Entire text]
 - Text value: 012345
 - Calculated suspect level: 70, tolerance: 75
 - Document page: 1
 - Issue Date (172) at (0.22, 4.68) inches on page(s): 1 (absolute) - [Find by regular expression]

```
>>Engine: OCR version 16.6 initialized
>>Engine: Barcode recognition support is licensed
>>Form Identity: begin new document operation [document type = HR - Transcript (137)]
>>Form Identity: begin identification test on form [AI Transcript Form (103)] at page 1
>>Form Identity: copy document page 1 to engine
>>Form Identity: beginning identification chain group 1
>>Form Identity: identity 103 zone = [142, 23] - [390, 72]
>>Form Identity: zone text on identity 103 [Hyland University]
>>Form Identity: exact text match test for [HYLAND UNIVERSITY]
>>Form Identity: match test succeeded at position 0
>>Form Identity: registration offset for identity 103 is [1, 0]
>>Form Identity: identity 125 zone = [298, 93] - [538, 142]
>>Form Identity: zone text on identity 125 [Official Transcript]
>>Form Identity: exact text match test for [OFFICIAL TRANSCRIPT]
>>Form Identity: match test succeeded at position 0
>>Form Identity: registration offset for identity 125 is [0, 1]
>>Form Identity: matched form ID = AI Transcript Form (103)
>>Form Identity: calculated scaling factor = [X:1.00 Y:1.00], translation error = [3, 0] pixels
>>Field recognition: copy document page 1 to engine
>>Field recognition: field ID (106) zone = [66, 186] - [234, 211]
>>Field recognition: field ID (106) scaling/translation applied, new position = [67, 186] - [235, 211]
>>Field recognition: field ID (106) raw result = 'Reagan', suspect level = 70
```

The **Test Result Verification** window is divided into two panes:

- The top pane contains a summary of information about the Advanced Capture form that was matched to the document. Information regarding Keyword Values that did not meet their respective fields' configured Suspect Level is displayed in red. Information regarding Keyword Values that did not meet their respective fields' configured formats (e.g., regular expressions' or tags' formats) is not displayed.
- The bottom pane contains information returned from the OCR engine during processing. Information regarding Keyword Values that did not meet their respective fields' configured formats (e.g., regular expressions' or tags' formats) is replaced by the following message: **Failed to extract meaningful data from raw zone result '<data extracted>.'**

Note: If you canceled the test, the **Test Result Verification** Window is displayed after the test is canceled. However, the top pane will display a message in red that indicates that the operation was canceled.

If you are testing an entire batch, the **Test Result Verification** window is not displayed. Instead, the results are only displayed in the **Result Verification** panel within the **Advanced Capture Configuration** window.

Tip: Double-clicking on the **Result Verification** panel exports the window's contents to a new text document. Double-clicking on the **Test Result Verification** window exports the bottom pane's contents to a new text document.

4. Review the information in the **Test Results Verification** window or the **Result Verification** panel in order to determine what, if any, changes need to be made to the Advanced Capture form(s).

Viewing Document History

To supplement the information provided in the **Result Verification** panel, you can review the document history. A document's history is a log of all actions performed on a selected document. This log is maintained by OnBase, and it contains information that is useful for reviewing batches.

The **Batch History** tab displays information about the batch in which a document was imported into OnBase. From an open document or the **Document Search Results** list, right-click and select **History**. The **Document History** dialog box displays all recorded batch actions in the **Batch History** tab.

Batch History

The following information is available on this tab:

- **Log Date** - the date the information was logged.
- **Log Time** - the time the information was logged.
- **User Name** - the name of the user who performed the interaction.
- **Batch Num** - the numeric label associating the batch with its column in the database.
- **Detail** - the type of interaction performed, such as the committal of the batch.

Generating a Document History Report

To generate a document history report, right-click in the **Document History** dialog box and select **Generate Report**. The new report is generated and displayed.

This report is stored in the **SYS - User Reports** Document Type and can be retrieved using this Document Type as a search criterion.

Configuring Advanced Capture to Use Intelligent Character Recognition

If your solution is licensed for Intelligent Character Recognition (ICR), you can configure Advanced Capture to read and capture Keyword Values from hand-written text.

ICR can only be configured to be performed in Text Field, Line Item Extraction, and Group Line Item Extraction Data Field Zones (not Optical Mark Field or Signature Detection Field Data Field Zones). In order to perform ICR, the Data Field Zone must be configured to identify hand-written text instead of machine-printed text.

To configure a Data Field Zone to read hand-written text (i.e., perform ICR processing):

1. From the **Data Field Zone** dialog box, ensure that either the Text Field tab, Line Item Extraction tab, or Group Line Item Extraction tab is displayed.

The screenshot shows the 'Data Field Zone' dialog box with the 'Text Field' tab selected. The 'Keyword type' is set to '<None>'. Under 'Data Application', 'Entire zone value' is selected. Under 'Page Location(s)', 'Absolute page:' is selected with the value '1'. The 'Advanced recognition' dropdown at the bottom is highlighted with a red oval and set to 'Machine print text/OCR'. Other options include 'Remove line breaks', 'First line only', 'Table style', 'Find by tag/expression', 'Person name extraction', 'Remove all whitespace from result', 'Redact zone from document after keyword assignment', 'Character search hints' (Numerals, Uppercase, Lowercase, Punctuation, Miscellaneous, MICR data), 'Advanced filtering' (VB script: <None>, Suspect level: 75), 'Specific allowed characters', 'Keyword association', and 'Activation groups'.

2. Using the **Advanced Recognition** drop-down select list, select one of the OCR/ICR recognition options:
 - **Machine print text/OCR.** The OCR engine will perform optical character recognition in order to read machine-printed text.
This option is selected by default.
 - **Handwriting/ICR - numerals/grouping punctuation (North American style).** The OCR engine will perform intelligent character recognition to read hand-written text.
This option will only attempt to recognize numeric characters written in the North American style (e.g., the 7 character is not crossed).
 - **Handwriting/ICR - numerals/grouping punctuation (European style).** The OCR engine will perform intelligent character recognition to read hand-written text.
This option will only attempt to recognize numeric characters written in the European style (e.g., the 7 character is crossed).
 - **Handwriting/ICR - alphanumeric/punctuation.** The OCR engine will perform intelligent character recognition in order to read hand-written text.
This option will attempt to recognize all alphanumeric characters.
 - **Auto-detect ICR/OCR - Default to ICR/Alphanumeric.** The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for alphanumeric characters will be used by default.
 - **Auto-detect ICR/OCR - Default to ICR/Numeric North American.** The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, North American-style characters (e.g., the 7 character is not crossed) will be used by default.
 - **Auto-detect ICR/OCR - Default to ICR/Numeric European.** The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, intelligent character recognition for numeric, European-style characters (e.g., the 7 character is crossed) will be used by default.
 - **Auto-detect ICR/OCR - Default to OCR.** The OCR engine will attempt to determine whether the value consists of hand-written or machine-printed text and then use the appropriate type of character recognition to read the text. When the type of text cannot be determined, optical character recognition will be used by default.

Note: The **Auto-detect ICR/OCR** options may not work properly if the Data Field Zone contains less than 25 characters.

Tip: Of the two **Auto-detect ICR/OCR** options, **Default to ICR** is more likely to produce the best results when the type of text cannot be determined. This is because the OCR engine's intelligent character recognition generally reads machine-printed text more accurately than the engine's optical character recognition reads hand-written text.

3. If one of the **Handwriting/ICR** or **Auto-detect ICR/OCR** options is selected for a Text Field Data Field Zone, the **Remove vertical/horizontal lines from zone before processing** check box is displayed below the Advanced Recognition drop-down select list.

Select this option if you would like the OCR engine to attempt to strip vertical or horizontal lines (i.e., constraint boxes) from the zone before it is processed.

4. Once you have selected the OCR/ICR recognition option for the Data Field Zone, click **Save** to close the **Data Field Zone** dialog box.

Limitations of Intelligent Character Recognition

Advanced Capture has the following limitations when attempting to perform ICR:

- ICR cannot be performed using Point and Shoot Indexing.
- Only characters belonging to single-byte character-based alphabets (e.g., English, French, Spanish, German, etc.) can be recognized via ICR processing. Characters belonging to double-byte character-based alphabets (e.g., Traditional Chinese, Japanese, Korean, etc.) cannot be recognized.
- ICR processing is available for only Text Field, Line Item Extraction, and Group Line Item Extraction Data Field Zones. It is not available for Optical Mark Field or Signature Detection Field Data Field Zones.

Improving the Accuracy of Intelligent Character Recognition

It is difficult to predict the accuracy of an ICR process due to the variability in the quality of the hand-written text. The following tips will help to improve the accuracy of ICR processing.

- The hand-written characters should be clearly separate from one another (i.e., they should not touch or intersect).
- The hand-written text should be unconstrained. If the characters are being entered into constraint boxes, do one of the following:
 - If the documents containing the constraint boxes can be modified before the Advanced Capture process is run, ensure that the constraint boxes are printed in red. Then, using Document Imaging's Image Processing options, configure a red Color Dropout process to remove the constraint boxes from the documents when scanning.
 - If the documents containing the constraint boxes cannot be modified before the Advanced Capture process is run, configure your OCR process format to automatically drop the constraint boxes by selecting **Remove vertical/horizontal lines from zone before processing** in the Data Field Zone dialog box.

Constrained Text:



Name

Unconstrained Text:

John Adams

Name

- The ICR numeric text processor (**Handwriting/ICR - numerals/grouping punctuation (North American style)**, **Handwriting/ICR - numerals/grouping punctuation (European style)**) is significantly more accurate than the ICR alphanumeric text processor (**Handwriting/ICR - alphanumeric/punctuation**) because the character set is much smaller.

If you are attempting to process numeric data only from a Data Field Zone, you will have more accurate results if you use the numeric text processor.

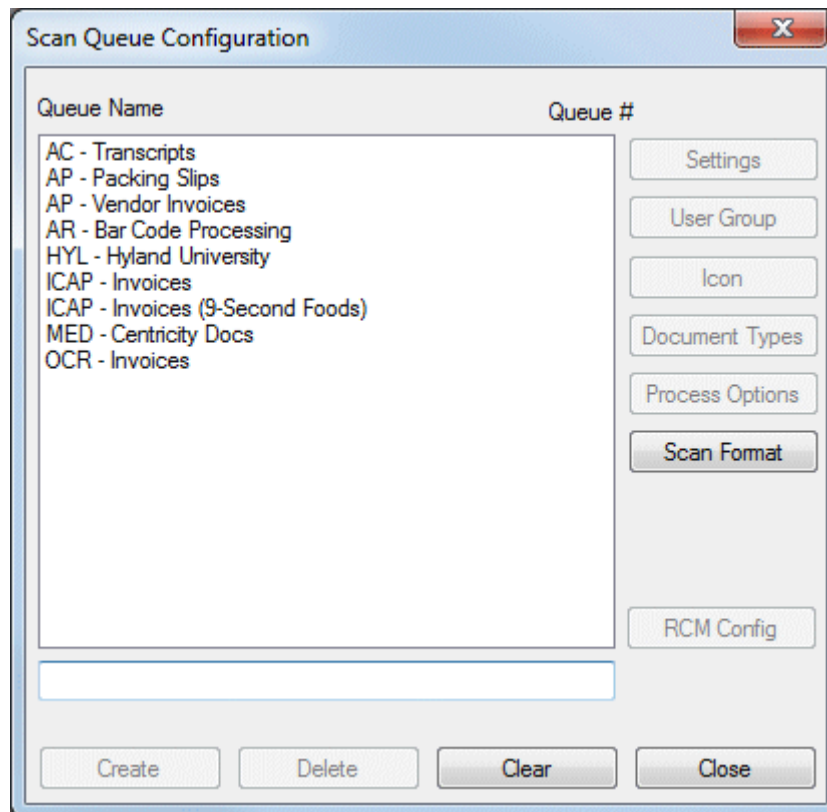
Configuring a Scan Queue for Advanced Capture

Scan queues associated with the batches that will undergo Advanced Capture must be configured to route those batches to the **Awaiting Advanced Capture** batch status queue.

Note: For information on creating and configuring scan queues, see the **Document Imaging** module reference guide or help files.

To configure a scan queue for Advanced Capture:

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.



2. Select the scan queue to be configured for Advanced Capture from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.

Assign Process Options for AC - Transcripts

Pre-Index Scan Queue Sorting		Post-Index Scan Queue Sorting		Processing Queue Warning Levels	
Batch Processing	Auto-Name	Capture	Indexing	Integrations	QA
<input type="checkbox"/> Separate Bar Code Processing Queue <input type="checkbox"/> Document Separation Queue <input type="checkbox"/> Show Existing Document Separations <input type="checkbox"/> Clear Keywords on Document Separation <input type="checkbox"/> Advanced Mode <input checked="" type="checkbox"/> Advanced Capture Processing Queue <input type="checkbox"/> Automatic Advanced Capture Processing <input type="checkbox"/> Skip Documents With Unidentified First Page <input type="checkbox"/> Interactive Data Capture Queue (Unity Client only) <input type="checkbox"/> OCR Indexed Documents <input type="checkbox"/> Disable OCR Text Popup On Image <input type="checkbox"/> Enable for Double-Blind Indexing Queue <input type="checkbox"/> Intelligent Capture for AP Processing Queue <input type="checkbox"/> Always Require Document Verification <input type="checkbox"/> Disable Automatic Document Separation <input type="checkbox"/> Double-Blind Indexing Queue <input type="checkbox"/> Document Re-Index Queue <input type="checkbox"/> Confirm Batch Exit from Re-Index / Keep Documents Available <input type="checkbox"/> Image Segment Archiver <input type="checkbox"/> Image Segment Archiver Queue <input type="checkbox"/> Single Document Mode - Only During Indexing Send to Queue: <None> Document Type: <None> <input type="checkbox"/> Continuous Page Slicing Mode Default Document Type: HR - Transcript <input checked="" type="checkbox"/> Default to Pre-Index Scan Mode	<input type="checkbox"/> Enable Capture Process Designer (Unity Client only) <input type="checkbox"/> Full-Page OCR Queue <input type="checkbox"/> Auto-OCR Batches <input type="checkbox"/> Respect Lines per Page <input type="checkbox"/> PDF Conversion Queue <input type="checkbox"/> Auto-PDF Conversion <input type="checkbox"/> Log Document Create Actions <input type="checkbox"/> Log Keyword Modifications <input type="checkbox"/> Auto-Perform Image Processing <input type="checkbox"/> Auto-Commit Batches <input type="checkbox"/> Auto-Name Batches <input type="checkbox"/> Auto-Folder Documents <input type="checkbox"/> Delay Disk Group Migration Until Batch Commit <input type="checkbox"/> Perform Custom Step <input type="checkbox"/> Auto-Run after Scan <input type="checkbox"/> Create Image Thumbnails On Commit <input type="checkbox"/> Do Not Allow Direct Scanning Into This Queue <input type="checkbox"/> Exclude Queue from Disconnected Scanning <input type="checkbox"/> Double-Click Initiates Queue Operation <input type="checkbox"/> Only Imaging Admin Modify Scan Formats <input type="checkbox"/> Prevent Scan Operation Without Required Keywords <div> Initiate Workflow <input checked="" type="radio"/> During Commit <input type="radio"/> During Index <input type="radio"/> Do Not Initiate Workflow </div> <div> Override Maximum Uncommitted Batches (This Scan Queue Only) <input type="text" value="0"/> 0 = Use Default and Check All Scan Queues </div>				

Save Cancel

3. Select the **Advanced Capture Processing Queue** check box. Batches imported via this scan queue will now be routed through the **Awaiting Advanced Capture** batch status queue.

4. If you would like batches associated with this scan queue to automatically undergo Advanced Capture as soon as the batch enters the **Awaiting Advanced Capture** batch status queue, select the **Automatic Advanced Capture Processing** check box.
If the **Automatic Advanced Capture Processing** check box is not selected, batches will reside in the **Awaiting Advanced Capture** batch status queue until they are processed manually; via a scheduled process, if your OnBase Client workstation is configured to do so; or until the workstation polls the batch status queue and processes any batches residing within it.
For more information on configuration options for automatic Advanced Capture processes, see [Configuring Automatic Advanced Capture Processes on page 286](#).
5. If you would like batches associated with this scan queue to skip the processing of documents having a first page that cannot be matched to an Advanced Capture form, select the **Skip Documents With Unidentified First Page** check box. Selecting this option can save processing time when the documents in the batch are already properly separated and only need to be indexed.
If the **Skip Documents With Unidentified First Page** check box is not selected, the Advanced Capture engine attempts to match subsequent pages of a document to an Advanced Capture form when the first page cannot be matched.
6. Click **Save** to save your scan queue configuration and close the **Assign Process Options for <Scan Queue Name>** dialog box.

Note: System documents cannot be processed by the Advanced Capture engine. If you assign a **Default Document Type** to the scan queue, ensure that this Document Type does not belong to the **System Documents** Document Type Group.

Disabling Images of Keyword Values from Displaying During Indexing

When a user is performing index verification on an image document and selects a Keyword Type field that has been populated with a Keyword Value via Advanced Capture, by default, an image of the Keyword Value identified by the engine is displayed above the Keyword Type field.

The screenshot shows a document form with the following fields and values:

- Document Date: 07/01/2010
- Keywords: 1001 (displayed as a large image above the field)
- Invoice #: 1001
- Invoice Date: 07/01/2010
- Vendor: COMPUTERS R US
- P.O. Number: 4279

Each field has a lock icon to its right, indicating it is locked for editing.

If desired, your scan queue can be configured to not display this image of the Keyword Value above the Keyword Type field.

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.
2. Select the scan queue to be modified from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.
3. Select the Indexing tab.
4. Select the **Disable Floating Image Snippets In Keyword Panel** check box.
5. Click **Save**.

Disabling Automatic Keyword Value Zoom

When a user is performing index verification on an image document and selects a Keyword Type field that has been populated with a Keyword Value via Advanced Capture, by default, the Working window will zoom to the area of the document from which the value was extracted.

If desired, your scan queue can be configured to not automatically zoom to this area.

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.
2. Select the scan queue to be modified from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.
3. Select the Indexing tab.
4. Select the **Disable Automatic Keyword Value Zoom to Image Zone** check box.
5. Click **Save**.

Note: If necessary, the Document Viewer automatically scrolls to show the area of the document from which the value for the selected Keyword Type field was extracted, without respect to the **Disable Automatic Keyword Value Zoom to Image Zone** option.

Navigating Only Between Suspect Keyword Values

By default, if a user presses the **Tab** key when performing index verification, the next Keyword Value, regardless of whether or not it is suspect, is selected in the **Indexing** dialog box.

If desired, you can configure the **Indexing** dialog box to only navigate between Keyword Values that have been identified as suspect when a user presses the **Tab** key (i.e., when a user presses the **Tab** key, the next suspect Keyword Value is selected; all non-suspect Keyword Values between the last-selected Keyword Value and the newly-selected Suspect Keyword Value are skipped). This function is enabled with the **TAB Cycles Through Suspect Keywords** indexing option.

Additionally, if a user is using Interactive Data Capture's Point and Click or Swiping indexing methods, enabling this option allows the user to navigate to the next suspect Keyword Value by clicking the highlighted text in the binding box on the image.

To enable this option for a scan queue using a standard capture process:

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.
2. Select the scan queue to be modified from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.
3. Select the **Indexing** tab.
4. Select the **TAB Cycles Through Suspect Keywords** check box.
5. Click **Save**.

To enable this option for a scan queue using a custom capture process:

1. From OnBase Studio, select the **Capture Process** tab in the **Repositories** pane.
2. Within the **Capture Process** tab, open the scan queue and capture process to be used for Advanced Capture processing.
3. In the design view pane, select the Index status step configured for performing Advanced Capture index verification. The indexing process options are displayed in the **General** tab of the **Properties** pane.
4. Select the **TAB Cycles Through Suspect Keywords** check box in the **Indexing Focus** section of the indexing process options.
5. Click **Save** on the **Home** ribbon.

When the **TAB Cycles Through Suspect Keywords** option is selected, a user may navigate between all Keyword Values in the **Indexing** dialog box (suspect or non-suspect) by pressing the **Up Arrow** or **Down Arrow** keys.

Note: If the **TAB Cycles Through Suspect Keywords** check box is checked, and if the **Tab to Next Field** radio button is selected in either the Main Enter Key or the Numeric Enter Key section, then pressing the **Enter** key will select the next suspect Keyword Value in the **Indexing** dialog box (i.e., non-suspect Keyword Values are skipped).

Configuring Automatic Advanced Capture Processes

Advanced Capture can be initiated on a batch manually, or it can be automated in several ways.

Note: For information on manually initiating Advanced Capture on a batch, see [Performing Batch Advanced Capture on page 372](#).

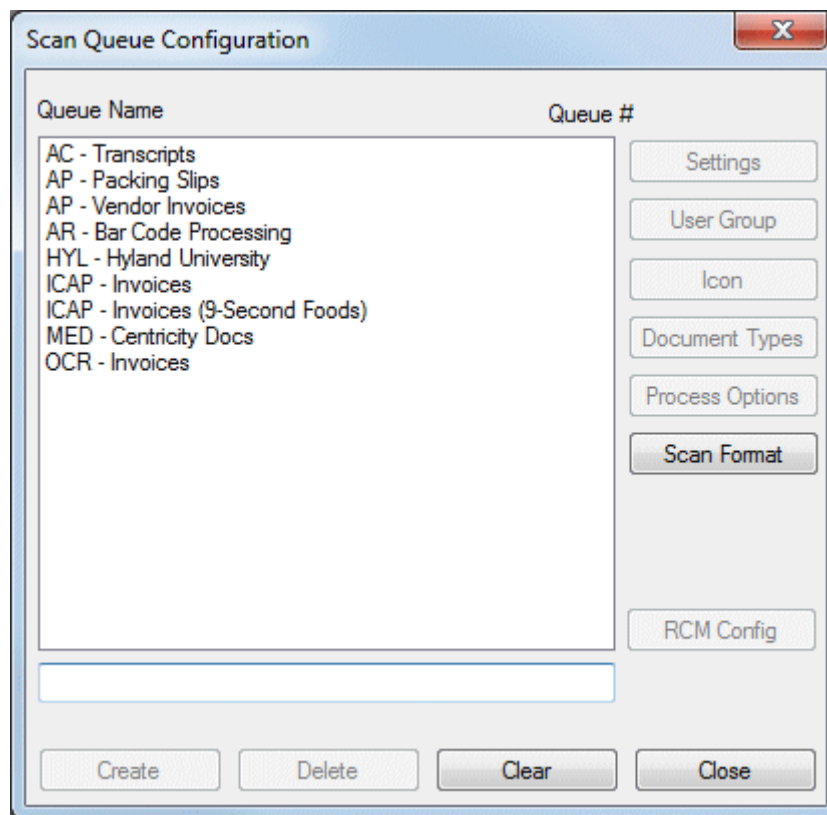
Automating Advanced Capture at the Scan Queue Level

Scan queues can be configured to automatically process batches once they enter the **Awaiting Advanced Capture** batch status queue.

Note: Only batches scanned through the OnBase Client can be processed automatically at the scan queue level. Batches scanned through other sources are sent to the **Awaiting Advanced Capture** batch status queue and remain there until they are processed manually, through scheduling, or through a command line switch.

To configure a scan queue to automatically process a batch once it enters the **Awaiting Advanced Capture** batch status queue:

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.



2. Select the scan queue to be configured to automatically process batches in the **Awaiting Advanced Capture** batch status queue from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.

- On the **Batch Processing** tab, select the **Automatic Advanced Capture Processing** check box.

Assign Process Options for AC - Transcripts

Pre-Index Scan Queue Sorting		Post-Index Scan Queue Sorting		Processing Queue Warning Levels	
Batch Processing	Auto-Name	Capture	Indexing	Integrations	QA
<input type="checkbox"/> Separate Bar Code Processing Queue <input type="checkbox"/> Document Separation Queue <input type="checkbox"/> Show Existing Document Separations <input type="checkbox"/> Clear Keywords on Document Separation <input type="checkbox"/> Advanced Mode <input checked="" type="checkbox"/> Advanced Capture Processing Queue <input checked="" type="checkbox"/> Automatic Advanced Capture Processing <input type="checkbox"/> Skip Documents With Unidentified First Page <input type="checkbox"/> Interactive Data Capture Queue (Unity Client only) <input type="checkbox"/> OCR Indexed Documents <input type="checkbox"/> Disable OCR Text Popup On Image <input type="checkbox"/> Enable for Double-Blind Indexing Queue <input type="checkbox"/> Intelligent Capture for AP Processing Queue <input type="checkbox"/> Always Require Document Verification <input type="checkbox"/> Disable Automatic Document Separation <input type="checkbox"/> Double-Blind Indexing Queue <input type="checkbox"/> Document Re-Index Queue <input type="checkbox"/> Confirm Batch Exit from Re-Index / Keep Documents Available <div> Image Segment Archiver <input type="checkbox"/> Image Segment Archiver Queue <input type="checkbox"/> Single Document Mode - Only During Indexing Send to Queue: <None> Document Type: <None> <input type="checkbox"/> Continuous Page Slicing Mode </div>	<input type="checkbox"/> Enable Capture Process Designer (Unity Client only) <input type="checkbox"/> Full-Page OCR Queue <input type="checkbox"/> Auto-OCR Batches <input type="checkbox"/> Respect Lines per Page <input type="checkbox"/> PDF Conversion Queue <input type="checkbox"/> Auto-PDF Conversion <input type="checkbox"/> Log Document Create Actions <input type="checkbox"/> Log Keyword Modifications <input type="checkbox"/> Auto-Perform Image Processing <input type="checkbox"/> Auto-Commit Batches <input type="checkbox"/> Auto-Name Batches <input type="checkbox"/> Auto-Folder Documents <input type="checkbox"/> Delay Disk Group Migration Until Batch Commit <input type="checkbox"/> Perform Custom Step <input type="checkbox"/> Auto-Run after Scan <input type="checkbox"/> Create Image Thumbnails On Commit <input type="checkbox"/> Do Not Allow Direct Scanning Into This Queue <input type="checkbox"/> Exclude Queue from Disconnected Scanning <input type="checkbox"/> Double-Click Initiates Queue Operation <input type="checkbox"/> Only Imaging Admin Modify Scan Formats <input type="checkbox"/> Prevent Scan Operation Without Required Keywords <div> Initiate Workflow <input checked="" type="radio"/> During Commit <input type="radio"/> During Index <input type="radio"/> Do Not Initiate Workflow </div>				
Default Document Type HR - Transcript <input checked="" type="checkbox"/> Default to Pre-Index Scan Mode	Override Maximum Uncommitted Batches (This Scan Queue Only) 0 0 = Use Default and Check All Scan Queues				

Save Cancel

Note: The **Advanced Capture Processing Queue** check box must be selected in order for the **Automatic Advanced Capture Processing** check box to be enabled.

- Click **Save** to save the configuration and exit the **Assign Process Options for <Scan Queue Name>** dialog box.

Automating Advanced Capture via a Scheduled Process

Advanced Capture can be configured as a scheduled process for an OnBase Client registered for Advanced Capture.

For more information, see [Scheduling on page 294](#).

Automating Advanced Capture via a Command Line Switch

If your workstation is registered for Advanced Capture, you can add a command line switch to your OnBase Client shortcut to automatically poll the **Awaiting Advanced Capture** batch status queue and automatically process the batches it finds there.

For more information, see your solution provider. Solution providers, see [Command Line Switches on page 8](#).

Exiting the Advanced Capture Form Configuration Interface

Once you have finished configuring your Advanced Capture form(s), exit the Advanced Capture configuration interface and return to the **Document Imaging** window. To exit:

- Close the **Advanced Capture Configuration** window by clicking the **X** in the upper-right corner.
- Click **Exit Configuration** in the **Tools** panel.

Tip: Prior to exiting the Advanced Capture configuration interface, be sure to save your changes by clicking **Save Configuration** in the **Tools** panel.

System Interaction

The following OnBase features and modules interact with Advanced Capture.

Automated Redaction

If your OnBase solution is also licensed for Automated Redaction, you can configure your Advanced Capture solution to mark Data Field Zones for Automated Redaction. Once the Data Field Zone has been processed and its Keyword Value captured and extracted, depending on your configuration, one of the following occurs:

- The zone is marked for redaction and the document is sent for Automated Redaction review and approval
- The zone is immediately redacted from the document without being sent for Automated Redaction review and approval.

Automated Redaction is only available for Text Field Data Field Zones.

Automated Redaction is configured at the Data Field Zone level. This allows you to redact some, but not all, Keyword Values from a document, or a Keyword Value (for example, a Social Security Number) from one Document Type but not another. Depending on the documents, you may be able to redact Keyword Values from some documents in a Document Type, but not others (depending on the configuration of the Advanced Capture forms matched to the documents).

For more information on Automated Redaction, see the **Automated Redaction** documentation.

Document Imaging

Orientation Detection & Automatic Page Rotation

If the scanning workstation is licensed and registered for batch Advanced Capture, it can be used to perform the Auto-Detect Orientation/Rotate Document Pages image process on batches of documents scanned via a scan queue configured to use image processing.

The Auto-Detect Orientation/Rotate Document Pages process uses the Advanced Capture engine to automatically detect if a page has been upside down or sideways and corrects them so that they are displayed at the correct orientation.

For more information on image processing and/or the Auto-Detect Orientation/Rotate Document Pages image process, see the **Document Imaging** module reference guide or help file.

Appending Documents to Existing Documents

Documents that are imported as part of a batch belonging to a scan queue configured to use the **Append to Existing Documents Matched on Keyword(s)** scan queue setting and are indexed via Advanced Capture (either batch or ad hoc Advanced Capture) are appended to existing documents with the same Document Type and Keyword Values.

See the **Document Imaging** module reference guide or help file for more information on the **Append to Existing Documents Matched on Keyword(s)** scan queue setting.

EDM Services

The EDM Services module allows you to configure Document Types to allow revisions. When documents are indexed via Advanced Capture (either batch or ad hoc Advanced Capture), they can be automatically checked to determine if they are revisions of an existing document.

Note: Documents must be imported as part of a batch belonging to a scan queue configured to check for revisions in order for Advanced Capture to identify a document as a revision of an existing document. Documents that were imported as a batch of a scan queue not configured for this feature, or documents imported via another import method, will not be compared against existing documents to determine if they are revisions.

If a scan queue that is configured to perform Advanced Capture is also configured to use the **Check for Revisions** option on the Indexing tab of the **Assign Process Options for <Scan Queue Name>** dialog box, OnBase will check the Document Type and Keyword Values of the indexed document against existing documents. If the newly-indexed document is a match, then it may be treated as a revision (or rendition, if the original document is a non-image document and the Document Type is renditionable) of the existing document (see the **Document Imaging** module reference guide or help file for more information on the behavior of the **Check for Revisions** setting).

However, because Advanced Capture is an automated process and no user interaction is needed to index documents, the behavior of documents identified as possible revisions of existing documents is different than if documents are manually indexed:

- When the scan queue's **Check for Revision** setting is used and more than one existing document with a matching Document Type and Keyword Values is found, the new document is routed to the **Awaiting Index** or **Index in Progress** queue, even though a Document Type has been selected for it and Keyword Values have been applied to it, where it can be reviewed and assigned as a revision of an existing document.
- When the scan queue's **Check for Revision** setting is used and a document indexed via Advanced Capture is determined to be a revision of an existing document and a revision comment is required for its Document Type, a revision comment is automatically generated indicating that the revision was identified via Advanced Capture.
- The scan queue's **Force Revision Prompt** setting is ignored when documents are indexed via Advanced Capture.

Intelligent Character Recognition (ICR)

Your Advanced Capture solution is able to identify and process machine-printed data via optical character recognition (OCR).

If your solution is also licensed for **ICR Support for Advanced Capture**, your solution is also able to identify and process hand-written text via intelligent character recognition (ICR).

In order to recognize hand-written text, the Data Field Zone must be configured to recognize hand-written text instead of machine-printed text.

Limitations of Intelligent Character Recognition

Advanced Capture has the following limitations when attempting to perform ICR:

- ICR cannot be performed using Point and Shoot Indexing.
- Only characters belonging to single-byte character-based alphabets (e.g., English, French, Spanish, German, etc.) can be recognized via ICR processing. Characters belonging to double-byte character-based alphabets (e.g., Traditional Chinese, Japanese, Korean, etc.) cannot be recognized.
- ICR processing is available for only Text Field, Line Item Extraction, and Group Line Item Extraction Data Field Zones. It is not available for Optical Mark Field or Signature Detection Field Data Field Zones.

Improving the Accuracy of Intelligent Character Recognition

It is difficult to predict the accuracy of an ICR process due to the variability in the quality of the hand-written text. The following tips will help to improve the accuracy of ICR processing.

- The hand-written characters should be clearly separate from one another (i.e., they should not touch or intersect).
- The hand-written text should be unconstrained. If the characters are being entered into constraint boxes, do one of the following:
 - If the documents containing the constraint boxes can be modified before the Advanced Capture process is run, ensure that the constraint boxes are printed in red. Then, using Document Imaging's Image Processing options, configure a red Color Dropout process to remove the constraint boxes from the documents when scanning.
 - If the documents containing the constraint boxes cannot be modified before the Advanced Capture process is run, configure your OCR process format to automatically drop the constraint boxes by selecting **Remove vertical/horizontal lines from zone before processing** in the Data Field Zone dialog box.

Constrained Text:



Name

Unconstrained Text:

Name

- The ICR numeric text processor (**Handwriting/ICR - numerals/grouping punctuation (North American style)**, **Handwriting/ICR - numerals/grouping punctuation (European style)**) is significantly more accurate than the ICR alphanumeric text processor (**Handwriting/ICR - alphanumeric/punctuation**) because the character set is much smaller.

If you are attempting to process numeric data only from a Data Field Zone, you will have more accurate results if you use the numeric text processor.

Unity Client

While batches that have undergone batch Advanced Capture can undergo additional indexing or verification in the OnBase Unity Client, be aware that Point and Shoot Indexing is not available in the OnBase Unity Client.

Additionally, ad hoc Advanced Capture cannot be initiated from the OnBase Unity Client.

Web Client

While batches that have undergone batch Advanced Capture can undergo additional indexing or verification in the OnBase Web Client, be aware that the index verification features available in the OnBase Client (e.g., Point and Shoot Indexing, suspect Keyword Values being highlighted in red, the Working window zooming in on the area of the document that the selected Keyword Value was extracted from, etc.) are not available in the OnBase Web Client.

Additionally, ad hoc Advanced Capture cannot be initiated from the OnBase Web Client.

Workflow

Ad hoc Advanced Capture can be performed as an action within a Workflow queue if your Workflow Life Cycle is properly configured and your OnBase solution is licensed for both Workflow and Advanced Capture.

For more information, see the **Workflow** module reference guide or help files.

Virtual Print Driver

Ad Hoc Advanced Capture can be performed when printing and importing an image document created via the Virtual Print Driver if your OnBase solution is licensed for the Virtual Print Driver and either Advanced Capture or Ad Hoc Advanced Capture, and your OnBase Client module is configured to use the Print Monitor.

For more information, see the **Virtual Print Driver** module reference guide.

Scheduling Overview

Scheduling processing for off-hours is an automated way to conserve system resources. Processing can be accelerated if the process is run from the database server.

Caution: Ensure that scheduled processes are not configured to run at the same time as a scheduled database backup. The database is locked while performing backups, preventing any processes from running.

Note: Purging documents from Document Maintenance can also be scheduled. For more information, see the **System Administration** module reference guide or help file.

Two types of processing activities may be scheduled with the Scheduler: a Process Format or a Process Job.

- A Process Format is used in processing modules and in scanning modules to specify how OnBase processes data being imported into OnBase. A Process Format is, basically, one individually-configured process.
- A Process Job is one or more Process Formats that have been configured to run sequentially. A Process Job does not have to consist exclusively of a single type of Process Format; it can contain multiple Process Formats from any module that allows scheduling.

Note: Process Formats created from Document Imaging sweep or scan from disk processes cannot be included in a Process Job.

Configuring & Using the Scheduler

Requirements for Configuring/Running a Scheduled Process

To configure a scheduled process, either a Process Format or a Process Job, a user must belong to a User Group with the **Client** and **Client Scheduler** product rights, and he/she must have rights to use the appropriate processing module. A scheduled process can be configured on any OnBase Client workstation, not just the processing workstation or a workstation running with the **-SCHED** command line switch.

To run a scheduled process, OnBase must be running with the **-SCHED** or **-SCHEDINST** command line switch on the processing workstation in order for the scheduled process to be executed at the configured time. The user account logged onto OnBase at this time needs only the **Client** product right in order for the process to be performed.

For more information on using command line switches with your OnBase solution, see the **Command Line Switches** module reference guide.

Using the -SCHED and -SCHEDINST Switches

This section explains the difference between the **-SCHED** and **-SCHEDINST** command line switches.

-SCHED

Some process formats or jobs can be scheduled to run automatically. The **-SCHED** switch causes the Client to queue these scheduled process formats and jobs for later processing; if the machine running the OnBase Client in Scheduler mode (i.e., running the OnBase Client with the **-SCHED** command line switch applied) is also the processing workstation, then the process formats or jobs will run at their scheduled times.

In order for the scheduled process format or job to be run, OnBase must be running in Scheduler mode on the processing workstation. If OnBase is not running, or if OnBase is not running in Scheduler mode, then the scheduled processes will not run.

A process format or job can be scheduled from any OnBase Client workstation by a user with the proper rights.

-SCHEDINST

The **-SCHEDINST** command line switch is very similar to the basic **-SCHED** switch. When you apply the **-SCHEDINST** switch to a Client shortcut, you can specify that the selected instance of the OnBase Client should only process jobs assigned to that Client instance's specific instance name.

The format of the switch is **-SCHEDINST="MyProcName"**, where **MyProcName** is the name of a specific processing instance. The OnBase Client that this switch is applied to will be unable to process any scheduled jobs that are not configured with a **Specific Processing Instance** of **MyProcName**.

A process format or job can be scheduled from any OnBase Client workstation by a user with the proper rights.

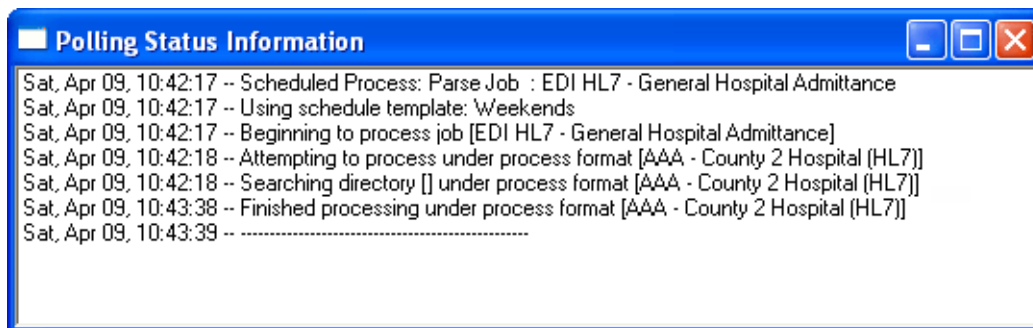
Note: If a scheduled process is assigned to a specific processing instance, it must be run from a client using the **-SCHEDINST** command line switch. If you try to run this process from a client using the **-SCHED** switch instead, the process will not be executed.

Verifying the Scheduler is Running

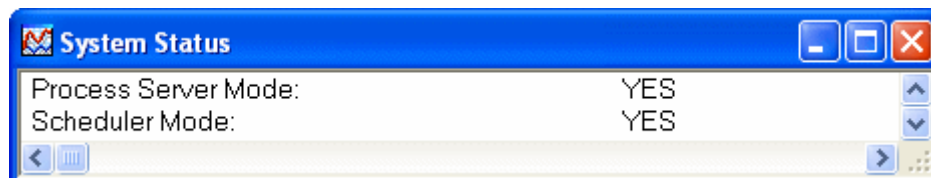
To verify that the Scheduler is running on the processing workstation, click **Window | Polling Status Information** in the OnBase Client.

Note: The **-SCHED** or **-SCHEDINST** command line switch must be applied to the Client shortcut to use this option.

The **Polling Status Information** window is displayed. Information about scheduled processes is displayed in it as the process is run. If this window exists, the Scheduler is running.



Another way to verify the Scheduler is running is to select **Window | System Status**. Both **Process Server Mode** and **Scheduler Mode** will be displayed as **YES**.



Running Multiple Scheduled Processes

Tip: Attempting to run more than one process job or format at once in the same session will result in a dramatic drop in all processing speeds. It is recommended to run a single automated process at a time.

If multiple jobs are configured, they can be performed sequentially in one OnBase Client session on the same workstation. Multiple sessions of the OnBase Client can be run simultaneously on one workstation to process these jobs in parallel; these sessions will coordinate processing tasks to ensure that each job is processed and that a job is not processed more than once.

In order to process jobs in parallel on multiple sessions of the OnBase Client, each session must be OnBase version 9.0 or later. If any one of the sessions is running an earlier version of OnBase, then none of the other sessions will perform any processing while it is processing.

Scheduled Process Configuration Reports

A user belonging to a User Group with the proper rights can run a Scheduled Processes Configuration Report.

This report provides information on all of the scheduled processes (process formats and process jobs) that have been scheduled to run. It is organized by processing workstation, and displays a weekly, monthly and end-of-month schedule, with jobs listed in order by starting time. Once run, this report is stored in OnBase as a document belonging to the **SYS Configuration Reports** Document Type.

Tip: It is considered a best practice to run a new Scheduled Process configuration report each time a new process (such as process format or process job) is scheduled. With the information stored in this report, troubleshooting and communications with Technical Support are greatly improved. Additionally, Configuration Reports are stored in OnBase, so there is a historical record of the structure of your OnBase solution.

For more information on Configuration Reports, including the Scheduled Processes Configuration Report, see the **System Administration** module reference guide or help file.

Working With Process Formats

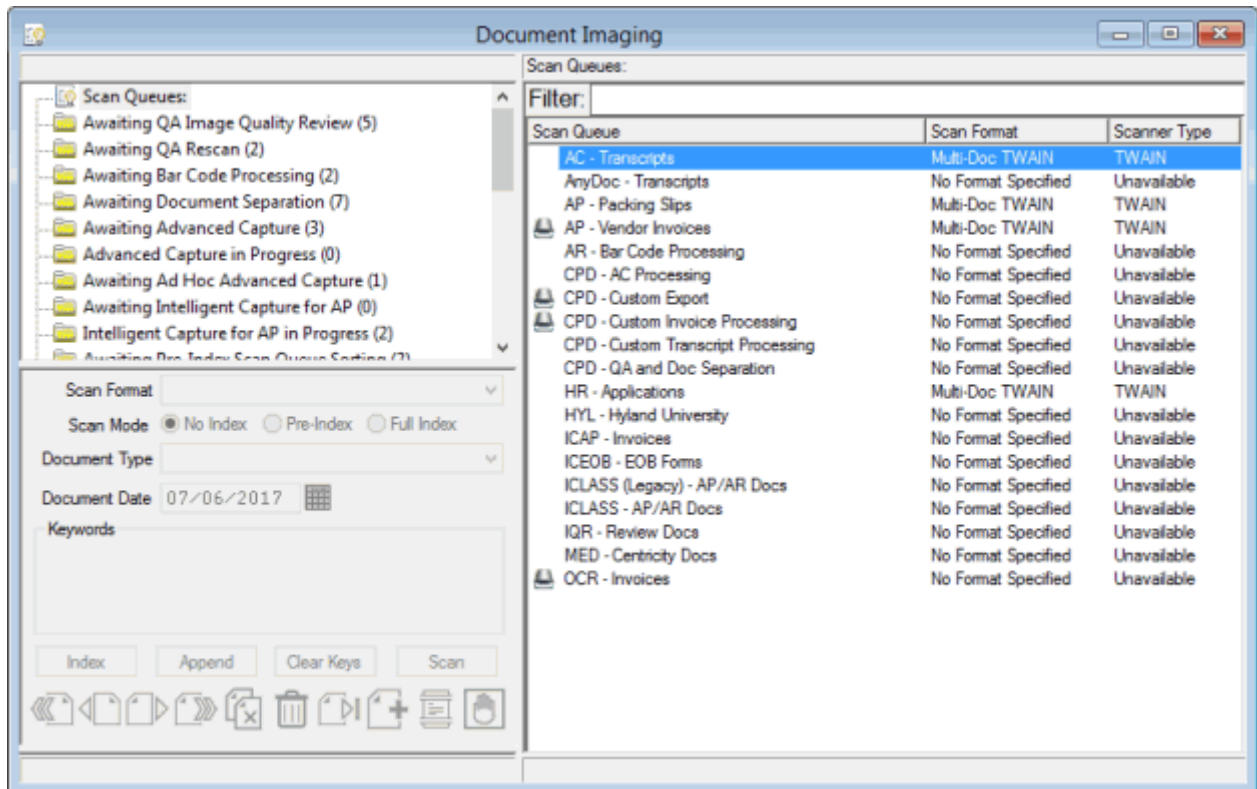
A Process Format is used in processing modules and in scanning modules to specify how OnBase processes data being imported into OnBase. A Process Format is, basically, one individually-configured process.

Creating a Scheduled Process Format

You can add a format to the Scheduler from its process queue by selecting the process format and selecting a scheduling option from the right-click menu.

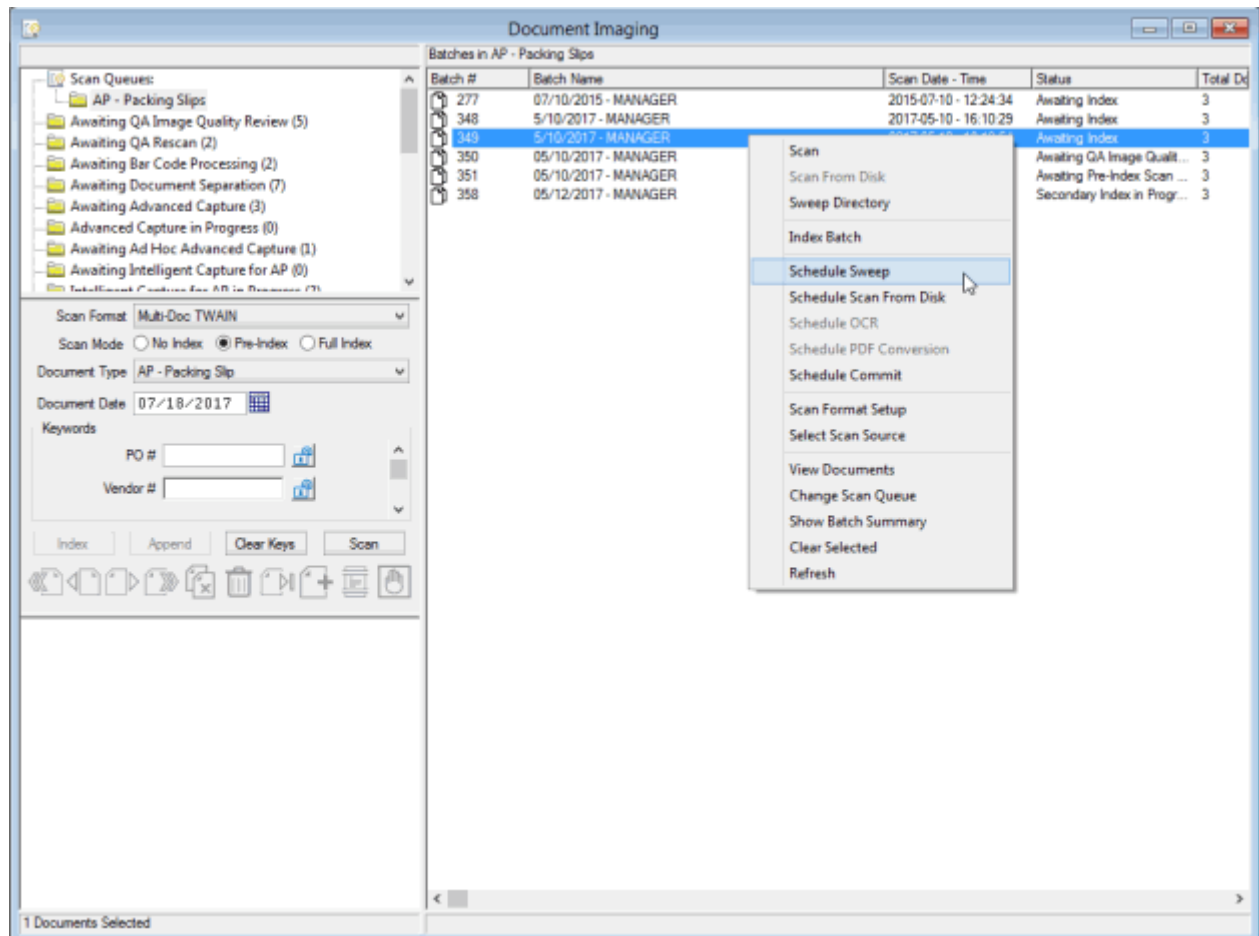
For example:

In the OnBase Client, click **Processing | Scan/Index**. The **Document Imaging** window is displayed.

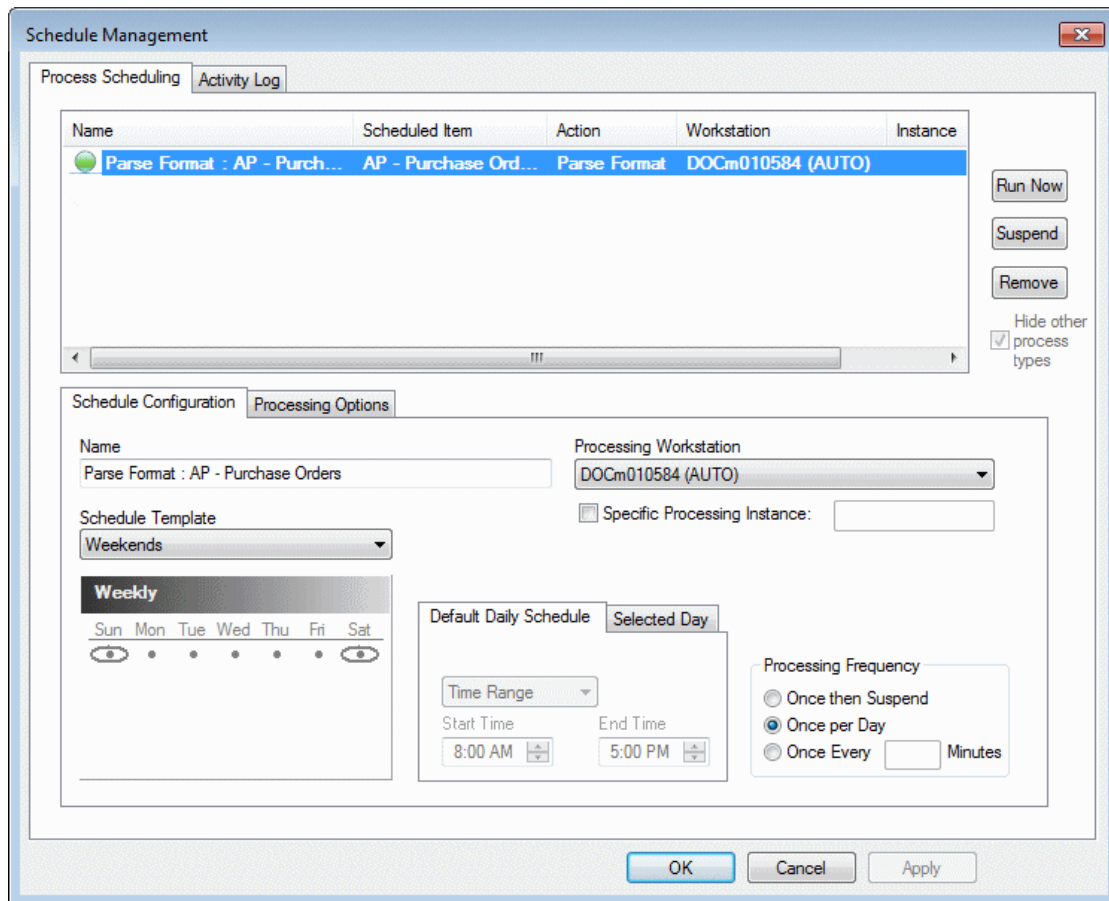


Double-click a process format you would like to add to the Scheduler. A list of available batches is displayed.

Select the batch you want to schedule, then right-click and select **Schedule Sweep**, **Schedule Scan From Disk**, **Schedule OCR**, or **Schedule Commit**.



The **Schedule Management** window is displayed.



A new Process Format is added to the **Scheduled Items** box. It is automatically selected.

By default, all scheduled Process Formats (e.g., COLD Process Formats, DIP Process Formats, etc.) are displayed in the **Scheduled Items** box when scheduling a new Process Format. For information on viewing only the Process Formats for the currently-selected process type, see [Viewing Scheduled Processes on page 309](#).

Schedule Configuration

The first options that must be configured for the scheduled process are the Schedule Configuration options on the **Schedule Configuration** tab. This tab is displayed by default.

1. In the **Name** field, enter a name for the scheduled process.
2. Using the **Processing Workstation** drop-down, select the workstation that will be used to run the scheduled process.

Note: This workstation will need to be running with the **-SCHED** or **-SCHEDINST** command line switch in order to run the scheduled process.

3. If you always want the scheduled process to be run from a specific instance of the OnBase Client, select the **Specific Processing Instance**, then enter the name of the instance in the **Specific Processing Instance** text field.

Note: If you select the **Specific Processing Instance** option but leave the **Specific Processing Instance** text field blank, the scheduled process can be run from any instance of the OnBase Client.

4. Using the **Schedule Template** drop-down, select one of the schedule templates for the process or select **<Custom Schedule>** to manually configure the schedule for this process.

Note: For information on creating a **Custom Schedule** or **Schedule Template**, see below.

5. Select how often you would like the scheduled process to run by selecting one of the Processing Frequency radio buttons.
 - **Once then Suspend.** The scheduled item will be processed once, then the scheduled process is suspended.
 - **Once per Day.** The scheduled item will be processed once per day.

Note: If the scheduled item is modified, the process may be run again on the same day.

- **Once every "" Minutes.** The scheduled item is processed in the interval (measured in minutes) entered in the field. The maximum number of minutes that can be entered is 99999.

Caution: This option is only supported when the **Default Daily Schedule** is set to **Time Range**. If your **Default Daily Schedule** is set to **Specific Time**, the scheduled item will only be processed at the specified time.

6. When you are finished setting the **Schedule Configuration** options, click **Apply**.

Calendar

The calendar is used to select the day(s) on which a scheduled process should be run.

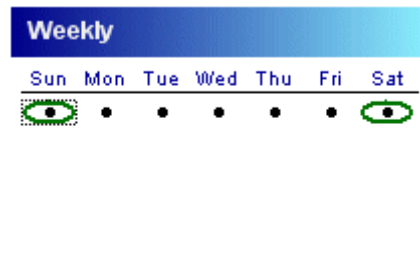
Note: The calendar is displayed based on your Workstation Regional Settings and the OnBase language DLL that you are using.

To change the view of the calendar, click the calendar heading (in the example above, **Weekly**) to display a menu. Select one of the following options to display a different calendar for configuration:

- **Weekly.** Allows you to configure a process to run on a certain day of the week (i.e., Thursday).
- **Monthly.** Allows you to configure a process to run monthly, on a particular date (i.e., the 1st and 15th of the month).
- **Monthly (Day-Relative).** Allows you to configure a process to run on a relative day of the month (i.e., the first Saturday of the month, the 2nd Wednesday of the month).

- **Annual.** Allows you to configure a process to run on a certain day of the year (i.e., June 30).
- **Full Calendar.** Allows you to configure a process to run on specified days of specified years (e.g., August 10, 2011 and/or July 17, 2012).

To select days that you would like to run a scheduled process, double-click the day on the calendar. The selected day is circled.

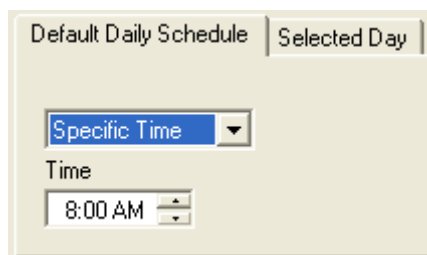
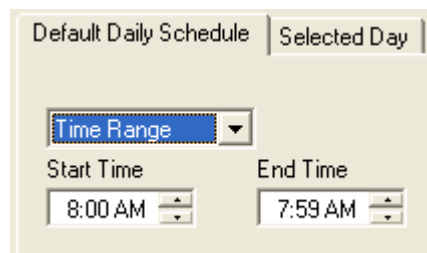


Note: In the example above, two days are selected but **Sunday** is the currently-selected day.

To deselect a day, double-click it.

Default Daily Schedule

The **Default Daily Schedule** tab allows you to configure the processing configuration for all days that do not have a **Selected Day** tab configuration.

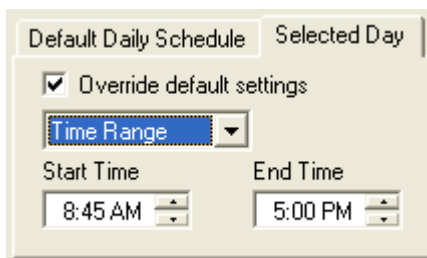


The drop-down list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time in which you want your job or format to begin processing. If you select **Specific Time**, a **Time** box is displayed. Select the time at which you want the job or format to begin processing.

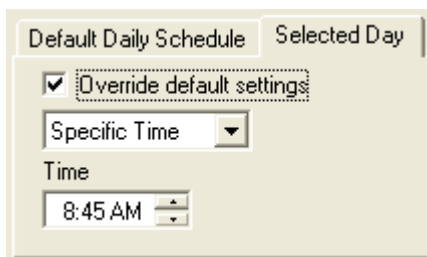
Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Selected Day

The **Selected Day** tab allows you to specify settings for the selected day that differ from the settings specified in the **Default Daily Schedule** tab. In order for the **Selected Day** tab to be enabled, you must click a day to select it and you must select the **Override default settings** check box.



The screenshot shows the 'Selected Day' tab in a software interface. At the top, there are two tabs: 'Default Daily Schedule' and 'Selected Day'. Below the tabs, there is a checked checkbox labeled 'Override default settings'. Underneath this, a drop-down menu is set to 'Time Range'. Below the drop-down, there are two time selection boxes: 'Start Time' with a value of '8:45 AM' and 'End Time' with a value of '5:00 PM'. Both time boxes have up and down arrows for adjustment.



The screenshot shows the 'Selected Day' tab in a software interface. At the top, there are two tabs: 'Default Daily Schedule' and 'Selected Day'. Below the tabs, there is a checked checkbox labeled 'Override default settings'. Underneath this, a drop-down menu is set to 'Specific Time'. Below the drop-down, there is a single time selection box labeled 'Time' with a value of '8:45 AM'. The time box has up and down arrows for adjustment.

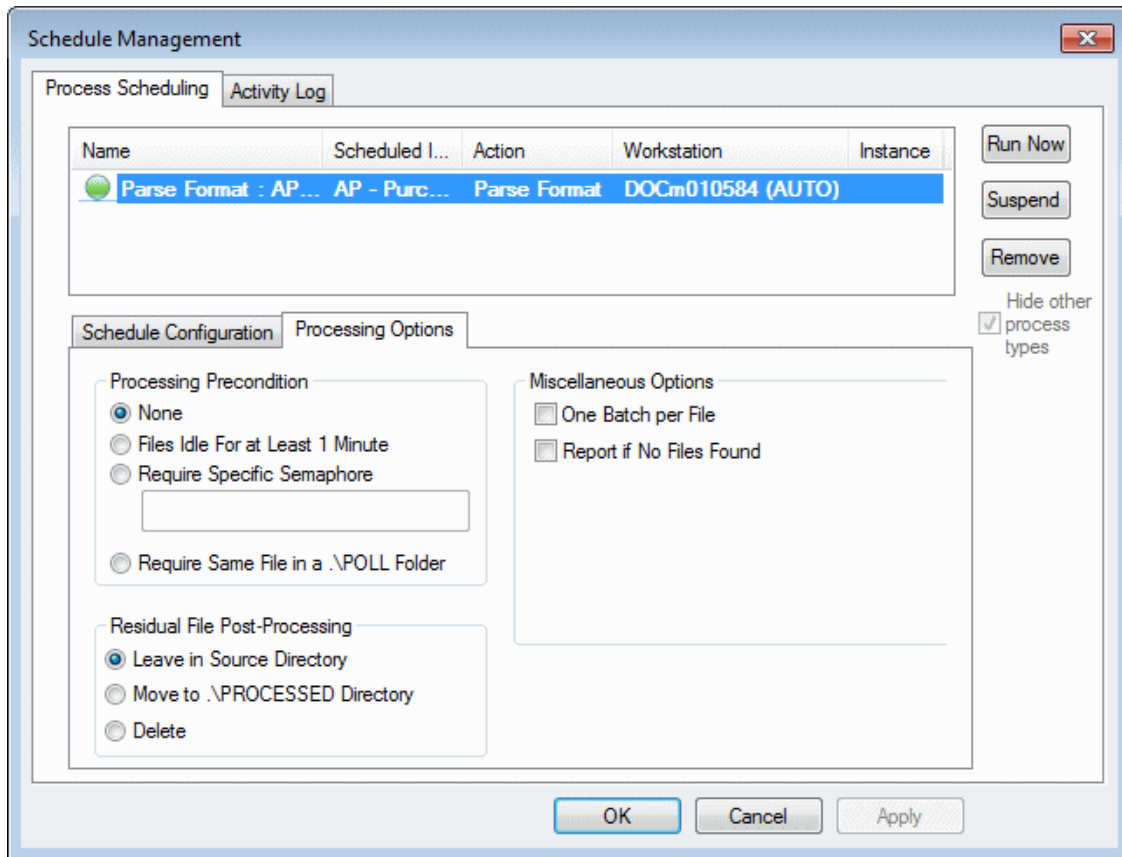
The drop-down list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time in which you want your job or format to begin processing. If you select **Specific Time**, a **Time** box is displayed. Select the time at which you want the job or format to begin processing.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

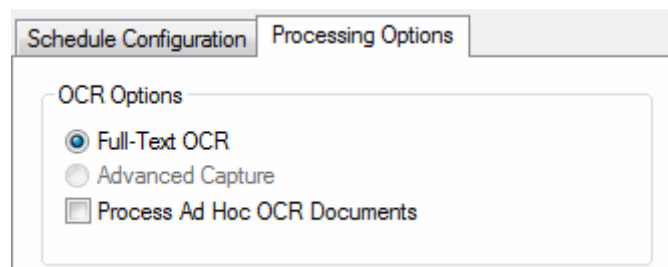
Processing Options

After the Schedule Options are configured on the Schedule Configuration tab, you must configure the Processing Options.

1. From the **Process Scheduling** tab of the **Schedule Management** window, click the **Processing Options** tab to display the Processing Options.



If you are scheduling a Full-Page OCR or an Advanced Capture process, the following options are displayed on the **Processing Options** tab.



2. Set the following Processing Options.

Option	Description
Processing Precondition	<p>The Processing Precondition options allow you to specify the conditions that must be met before processing can begin.</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Advanced Capture processes, Full-Text OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • None. If this option is selected, no processing precondition is necessary. • Files Idle For at Least 1 Minute. Select to indicate that processing must begin after the file indicated in the Default File Name of the processing format has been idle for at least one minute. • Require Specific Semaphore. Select to indicate that processing must begin after a trigger file is detected. The trigger file can be any file type/size/label and can be written to any location on the network. OnBase will only begin processing the processing file indicated in the Default File Name of the process format after the trigger file has been detected. How processing is triggered (definition of the file location and/or time variable) is defined by a semaphore. A semaphore is a technique for coordinating or synchronizing polling activity. A maximum of 255 characters can be entered in this field. The trigger file is deleted after processing. <hr/> <p>Note: If the trigger file is being accessed over FTP, it will not be deleted.</p> <hr/>

Option	Description
Processing Precondition (cont.)	<ul style="list-style-type: none"> • Require Same File in a .\POLL Folder. Select to indicate that processing must begin after a POLL file has been written to a specifically-configured POLL folder. The POLL file must appear in a folder labeled POLL, and the POLL folder must be created as a subfolder of the Default Directory of the process format. The name of the POLL file must be exactly identical to the name of the file to be processed. The value in the Default File Name field will be used to locate the POLL file. When OnBase locates the POLL file, the processor will attempt to process any file with that same name in the Default Directory. For example: The Default File Name is *.txt, and the Default Directory is C:\ProcessFiles. The file to be processed is stored in this directory. For this example, the file is named pf11x74.txt. The POLL file should be placed in C:\ProcessFiles\POLL, and named exactly the same as the process file (pf11x74.txt). OnBase will search C:\ProcessFiles\POLL for a file that matches the Default File Name of *.txt. Upon finding the pf11x74.txt file, the processor will return to the C:\ProcessFiles directory and search for the file named pf11x74.txt. This is the file that will be processed. <hr/> <p>Note: This option is not supported for use with the Directory Import Processor.</p> <hr/>

Option	Description
Residual File Post-Processing	<p>The Residual File Post-Processing options allow you to specify how residual files are processed (that is, files that have been processed but not deleted from the directory, such as read-only files).</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Advanced Capture processes, Full-Text OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • Leave in Source Directory. Select to leave any residual files in the folder they originated in. • Move to .\PROCESSED Directory. Select to move any residual files to the OnBase-generated PROCESSED folder located in the same folder the files were originally in. <hr/> <p>Caution: Depending on your system's configuration, processed files may be automatically deleted after an import process is run. In this situation, the processed files will not be moved to the PROCESSED folder because they have already been deleted from the folder they originated from.</p> <p>Depending on the processor you are using, you may be able to avoid this behavior by modifying the configuration of your import processor, or by marking the files to be processed as read-only.</p> <hr/> <ul style="list-style-type: none"> • Delete. Select to delete any residual files (that is, files that have been processed but not deleted from the directory) from the folder they originated in. <hr/> <p>Note: The Delete option is not available for Scheduled Sweeps or Scan from Disk processes.</p> <hr/>

Option	Description
Miscellaneous Options	<p>The Miscellaneous Options allow you to specify special scheduling options specific to the selected process. The availability of these options varies depending on the type of processor being scheduled. Many processing modules do not have some or all of these options.</p> <hr/> <p>Note: No Miscellaneous Options are available for scheduled PDF conversions, Advanced Capture processes, Full-Page OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • One Batch per File. Select to process each index file as one batch when multiple index files are being processed at once. This option is not supported for use with the Directory Import Processor. • Report if No Files Found. Select to create a Verification Report if no files are found when a scheduled format or job is run. <hr/> <p>Note: The Report if No Files Found option is only available when the None radio button is selected for the Processing Precondition. It is not available for scheduled Sweep or Scan from Disk processes.</p> <hr/> <ul style="list-style-type: none"> • Document Type. Available for certain scheduled Sweep processes. Use the drop-down to select the Document Type of processed documents. • Scan Format. Available for certain scheduled Scan from Disk processes. Use the drop-down to select the scan format to be used when processing documents. By default, the processor will use the last scan format that was assigned to the scan queue being processed. <hr/> <p>Note: Only Kofax scan formats can be selected from this drop-down.</p> <hr/>

Option	Description
OCR Options	<p>The OCR Options allow you to specify the configuration options for a scheduled Advanced Capture or Full-Text OCR process.</p> <hr/> <p>Note: These options are only available when scheduling an Advanced Capture or Full-Page OCR process (that is, the batch's scan queue has been configured for Advanced Capture or Full-Page OCR).</p> <hr/> <ul style="list-style-type: none"> • Full-Text OCR. Select this radio button if you are scheduling a Full-Text OCR process. • Advanced Capture. Select this radio button if you are scheduling an Advanced Capture process. • Process Ad Hoc OCR Documents. Select this radio button if you would like to perform Advanced Capture or Full-Text OCR on documents in the ad hoc batch status queues (Ad Hoc Advanced Capture or Awaiting Ad Hoc OCR).

3. When you are finished configuring the Process Options, click **Apply**.

Viewing Scheduled Processes

By default, only scheduled process formats and jobs of the currently-selected process type will be displayed in the **Schedule Management** window. To view scheduled process formats and jobs of all process types, deselect the **Hide other process types** check box.

To open the **Schedule Management** window, perform one of the following actions:

- Click **Processing | Scheduler | Schedule Management**.
- Open the **Scheduled Processes** queue and double-click on a scheduled process
- Right-click on a process format in its process queue and select **Schedule Format**.

Note: Additional Product Rights are required to view a scheduled purge process. For more information, see the **System Administration** module reference guide or help file.

Modifying a Scheduled Process Format

Once a scheduled process has been created, it can be modified as needed.

To modify an existing scheduled process:

1. Open the **Schedule Management** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select the process to be modified from the **Scheduled Items** box.

3. Modify the settings on the **Schedule Configuration** and **Process Options** tabs as needed.

For more information on the options on these tabs, see [Schedule Configuration on page 300](#) and [Processing Options on page 322](#).

Tip: You can modify the **Schedule Configuration** settings for multiple processes at the same time. To do so, use the **Shift** or **Ctrl** keyboard keys to select multiple processes before modifying the **Schedule Configuration** settings.

4. Once you have finished modifying the scheduled process, click **Apply**.

Deleting a Scheduled Process Format

Caution: If you delete a process format or process job that is scheduled, it will be deleted from the list of scheduled jobs.

Scheduled processes can be deleted from the **Schedule Management** window.


1. Open the **Schedule Management** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select the scheduled process(es) you would like to delete from the **Scheduled Items** box and click **Remove**.
3. Click **Apply**.




Running/Suspending a Scheduled Process Format

From the **Schedule Management** window, a scheduled process can be run immediately or it can be suspended.

1. Open the **Schedule Management** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select one or more scheduled processes from the **Scheduled Items** box.
 - To run the process(es) now, click **Run Now**. The processes are run the next time the processing workstation is polled.
 - To suspend the process(es), click **Suspend**. To resume one or more suspended processes, select those processes and click **Resume**.

An icon is displayed next to each scheduled process in the **Scheduled Items** box that indicates its status.

Icon	Description
	Run Now - Indicates that the user has clicked the Run Now button to cause the process to execute now instead of waiting for its scheduled time to run.

Icon	Description
	Suspend - Indicates a suspended process. The process will not run until a user selects it and clicks Resume .
	Active - Indicates an active scheduled process. An active process may be waiting to run or it may have already run at its scheduled time.
	Error - Indicates a process with a configuration error.

3. Click **Apply**.

Working With Process Jobs

A Process Job is one or more Process Formats that have been configured to run sequentially. A Process Job does not have to consist exclusively of a single type of Process Format; it can contain multiple Process Formats from any module that allows scheduling.

A few notes about Process Jobs:

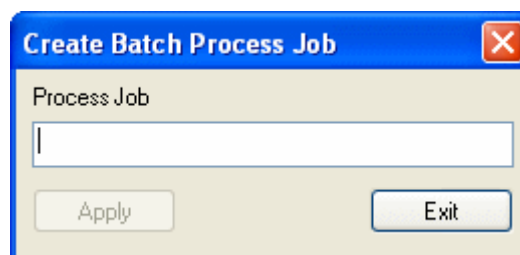
- Process formats must be created before they can be added to a job.
- AutoFill Keyword Import Processors can be scheduled from any **Process Job Queue**.
- Process Formats created from Document Imaging sweep or scan from disk processes cannot be included in a Process Job.

Creating a Job

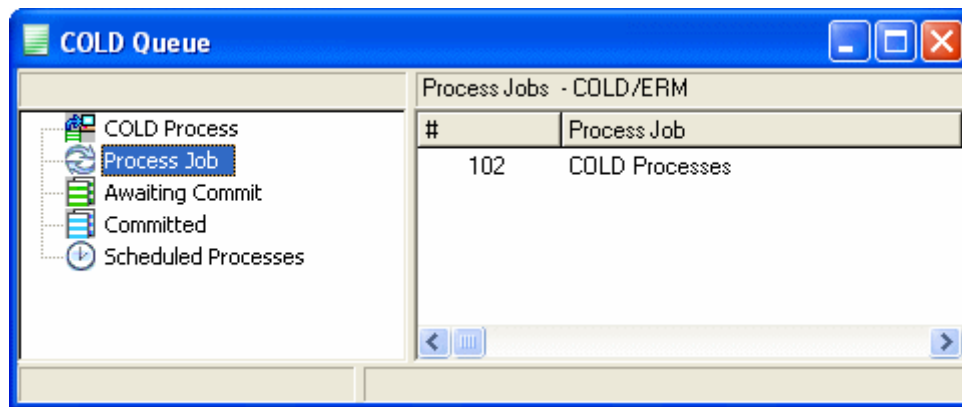
You can add a job to the Scheduler from a process queue (that is, the COLD Queue, the EDI Queue, and others).

To create a job, follow these steps:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on the window and select **Create New Job**.
Or, from the process queue, select **Process Job** and right-click in the **Process Jobs** window and select **Create New Job**. The **Create Batch Process Job** dialog box is displayed.



2. Enter a name for the job in the **Process Job** field and click **Apply**. The job is added to the process queue and is listed in the **Process Jobs** window.



Note: The process name must be 75 characters or fewer.

Note: If you are using the OnBase Client as a Windows Service, you must restart the OnBase Client after adding a new scheduled process.

Configuring a Job

To configure a job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Configure Job**.

Or, select the job to be configured from the **Process Jobs** window in the process queue, right-click and select **Configure Job**.

The **Process Job Configuration** window is displayed.

Process Job Configuration

Process Format	Date	File Name	Alternate Path	Alternate Disk Group

Up
Down

Process Format
▼

Alternate Path

Alternate File Name

Alternate Disk Group
▼

Language Conversion
< NO CONVERSION > ▼

Processing Options

☐ Store Document Indices

☐ Store Document Data Files

☐ Store Import File

☐ Create Auto-Folder

Default Keywords

Date

☒ Today

☐ Yesterday

☐ Last Week Day

☐ Last Business Day

☐ Specific Date / / [Calendar Icon]

☐ Disallow job date override

☐ Halt Processing After Error

Add Update Delete Clear Close

2. Configure a process format to add to the job:

Process Job Parameter	Description
Process Format	Select the process format to be incorporated in the process job. All available process formats are listed.
Alternate Path	Enter an alternate path to the data to be processed (i.e., the Default Directory) to use instead of the Default Directory configured for the selected process format. If an alternate path is not specified, the process format's Default Directory is used.
Alternate Filename	Enter an alternate file name for the data to be processed (i.e., the Default File Name) to use instead of the Default File Name configured for the selected process format. If an alternate file name is not specified, the process format's Default File Name is used.
Alternate Disk Group	Enter an alternate Disk Group to store the data being processed instead of the Disk Group configured for the selected process format. If an alternate Disk Group is not specified, the process format's default Disk Group is used.
Language Conversion	Select the language associated with the ASCII code page that created the import file. If a language conversion is not specified, the process format's Language Conversion setting is respected. <hr/> Note: This setting is only used for legacy language conversions. The option <NO CONVERSION> should be selected when configuring process settings. <hr/>
Store Document Indices	Select this option to store the processed documents in the database, along with their Keyword Values and document name. This option is enabled by default.
Store Document Data Files	Select this option to move the data file to the configured Disk Group after the process is complete. This option is enabled by default.
Store Import File	Select to store a copy of the index file used to import documents into OnBase for archive purposes. <hr/> Note: This option is not supported for use with modules that do not support the Store Import File processing option. See the configuration section of the appropriate module reference guide or help file to find out whether or not the Store Import File processing option is supported for a module. <hr/>

Process Job Parameter	Description
Create Auto Folder	<p>Select to provide the ability to Auto-Folder documents upon processing. See the Folders module reference guide or help files for additional information regarding Auto-Foldering.</p> <hr/> <p>Note: Not all processors offer the ability to Auto-Folder documents upon processing.</p> <hr/>
Default Keywords	<p>Click the Default Keywords button to select Keyword Types and Values that are displayed in the Batch Name for that Process Job when it is processed. These Keyword Types and Values are also displayed at the top of the Verification Report for that job.</p> <hr/> <p>Note: Only Keyword Types that have been configured for Document Types used in the Process Job are selectable.</p> <hr/> <p>Note: If a check process format is configured as part of the job, the Default Keywords button is disabled when the job is selected.</p> <hr/>
Disallow job date override	Select this option to prevent users from overriding the specified job date.
Halt Processing After Error	Select this option to halt processing for the process job if the configured process format generates an error. Any other process formats configured for the process job will not be processed.
Date	These settings allow a user-defined Document Date to be stored for the processed documents. This date is used as the %D parameter that appears in the document's Auto-Name string.

3. Click **Add**.

4. Repeat Step 2 for each process format that you would like to add to the job.

Process jobs are run in the order in which they display on the screen. Re-sequence a job by selecting it and clicking the **Up** or **Down** buttons.

Once you've added all process formats to the job, click **Close**.

Scheduling a Job

Once you have created and configured a job, you must schedule it in order for it to automatically run. A job is scheduled in almost the same way that a process format is scheduled.

To schedule a job, you must first open the **Schedule Management** window. To open it:

- From a process queue, select **Process Job** and then select the job to be scheduled in the **Process Jobs** window. Right-click and select **Schedule Job**.

- From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Schedule Job**.

Schedule Configuration

The first options that must be configured for the scheduled job are the Schedule Configuration options on the **Schedule Configuration** tab. This tab is displayed by default.

1. In the **Name** field, enter a name for the scheduled process.
2. Using the **Processing Workstation** drop-down, select the workstation that will be used to run the scheduled job.

Note: This workstation will need to be running with the **-SCHED** or **-SCHEDINST** command line switch in order to run the scheduled job.

3. Using the **Schedule Template** drop-down, select a schedule template for the process or select **<Custom Schedule>** to manually configure the schedule for this process.

Note: For information on creating a schedule template, see below.

To create a custom schedule, you will need to use the **Calendar** to select the day(s) you would like the scheduled job to run on and then you will need to specify the time the scheduled job will run using the **Default Daily Schedule** and/or **Selected Day** tabs. For more information, see those sections below.

4. Select how often you would like the scheduled job to run by selecting one of the **Processing Frequency** radio buttons.
 - **Once then Suspend**. The scheduled item will be processed once, then the scheduled process is suspended.
 - **Once per Day**. The scheduled item be processed once per day.

Note: If the scheduled item is modified, the process may be run again on the same day.

- **Once every "" Minutes**. The scheduled item is processed in the interval (measured in minutes) entered in the field. The maximum number of minutes that can be entered is 99999.
5. When you are finished setting the Schedule Configuration options, click **Apply**.

Calendar

To configure a job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Configure Job**.

Or, select the job to be configured from the **Process Jobs** window in the process queue, right-click and select **Configure Job**.

The **Process Job Configuration** window is displayed.

2. Configure a process format to add to the job:

Process Job Parameter	Description
Process Format	Select the process format to be incorporated in the process job. All available process formats are listed.
Alternate Path	Enter an alternate path to the data to be processed (i.e., the Default Directory) to use instead of the Default Directory configured for the selected process format. If an alternate path is not specified, the process format's Default Directory is used.
Alternate Filename	Enter an alternate file name for the data to be processed (i.e., the Default File Name) to use instead of the Default File Name configured for the selected process format. If an alternate file name is not specified, the process format's Default File Name is used.
Alternate Disk Group	Enter an alternate Disk Group to store the data being processed instead of the Disk Group configured for the selected process format. If an alternate Disk Group is not specified, the process format's default Disk Group is used.
Language Conversion	Select the language associated with the ASCII code page that created the import file. If a language conversion is not specified, the process format's Language Conversion setting is respected. <hr/> Note: This setting is only used for legacy language conversions. The option <NO CONVERSION> should be selected when configuring process settings. <hr/>
Store Document Indices	Select this option to store the processed documents in the database, along with their Keyword Values and document name. This option is enabled by default.
Store Document Data Files	Select this option to move the data file to the configured Disk Group after the process is complete. This option is enabled by default.
Store Import File	Select to store a copy of the index file used to import documents into OnBase for archive purposes. <hr/> Note: This option is not supported for use with modules that do not support the Store Import File processing option. See the configuration section of the appropriate module reference guide or help file to find out whether or not the Store Import File processing option is supported for a module. <hr/>

Process Job Parameter	Description
Create Auto Folder	<p>Select to provide the ability to Auto-Folder documents upon processing. See the Folders module reference guide or help files for additional information regarding Auto-Foldering.</p> <hr/> <p>Note: Not all processors offer the ability to Auto-Folder documents upon processing.</p> <hr/>
Default Keywords	<p>Click the Default Keywords button to select Keyword Types and Values that are displayed in the Batch Name for that Process Job when it is processed. These Keyword Types and Values are also displayed at the top of the Verification Report for that job.</p> <hr/> <p>Note: Only Keyword Types that have been configured for Document Types used in the Process Job are selectable.</p> <hr/> <p>Note: If a check process format is configured as part of the job, the Default Keywords button is disabled when the job is selected.</p> <hr/>
Disallow job date override	Select this option to prevent users from overriding the specified job date.
Halt Processing After Error	Select this option to halt processing for the process job if the configured process format generates an error. Any other process formats configured for the process job will not be processed.
Date	These settings allow a user-defined Document Date to be stored for the processed documents. This date is used as the %D parameter that appears in the document's Auto-Name string.

3. Click **Add**.

4. Repeat Step 2 for each process format that you would like to add to the job.

Process jobs are run in the order in which they display on the screen. Re-sequence a job by selecting it and clicking the **Up** or **Down** buttons.

Once you've added all process formats to the job, click **Close**.

Default Daily Schedule

The **Default Daily Schedule** tab allows you to configure the processing configuration for all days that do not have a **Selected Day** tab configuration.

The screenshot shows the 'Default Daily Schedule' tab selected. A drop-down menu is set to 'Time Range'. Below it, there are two time selection boxes: 'Start Time' set to 8:00 AM and 'End Time' set to 7:59 AM.

The screenshot shows the 'Default Daily Schedule' tab selected. A drop-down menu is set to 'Specific Time'. Below it, there is a single time selection box labeled 'Time' set to 8:00 AM.

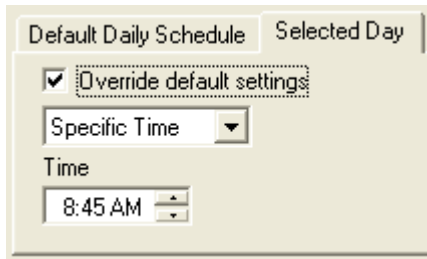
The drop-down list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time in which you want your job or format to begin processing. If you select **Specific Time**, a **Time** box is displayed. Select the time at which you want the job or format to begin processing.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Selected Day

The **Selected Day** tab allows you to specify settings for the selected day that differ from the settings specified in the **Default Daily Schedule** tab. In order for the **Selected Day** tab to be enabled, you must click a day to select it and you must select the **Override default settings** check box.

The screenshot shows the 'Selected Day' tab selected. The 'Override default settings' checkbox is checked. A drop-down menu is set to 'Time Range'. Below it, there are two time selection boxes: 'Start Time' set to 8:45 AM and 'End Time' set to 5:00 PM.



The screenshot shows a dialog box titled 'Default Daily Schedule' with two tabs: 'Default Daily Schedule' and 'Selected Day'. The 'Selected Day' tab is selected. Inside the dialog, there is a checkbox labeled 'Override default settings' which is checked. Below this is a dropdown menu labeled 'Specific Time' with a downward arrow. Underneath the dropdown is a 'Time' label and a time selection field showing '8:45 AM' with up and down arrows for adjustment.

The drop-down list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time in which you want your job or format to begin processing. If you select **Specific Time**, a **Time** box is displayed. Select the time at which you want the job or format to begin processing.

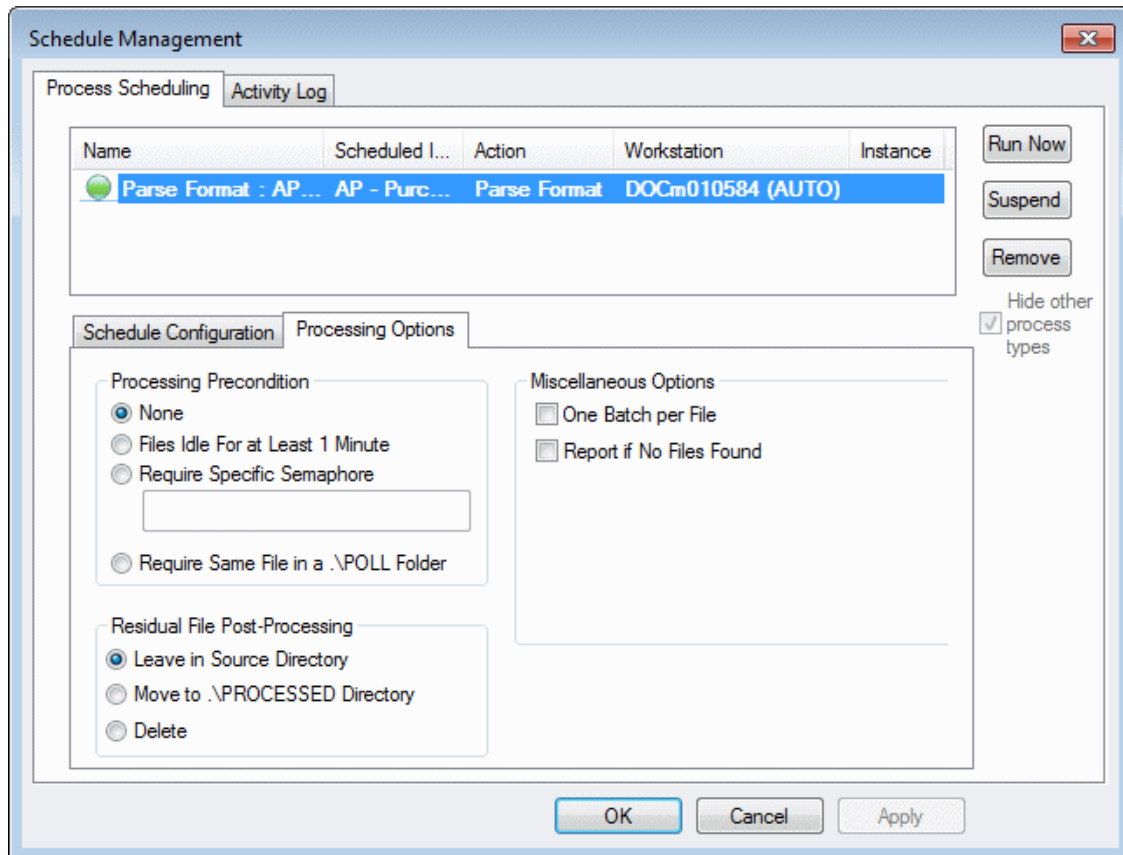
Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Processing Options

After the Schedule Options are configured on the **Schedule Configuration** tab, you must configure the Processing Options.

1. From the **Process Scheduling** tab of the **Schedule Management** window, click the **Processing Options** tab to display the Processing Options.

Note: This tab is only available if a single process is selected. If multiple processes are selected, the **Processing Options** tab is disabled.



2. Set the following Processing Options.

Option	Description
Processing Precondition	<p>The Processing Precondition options allow you to specify the conditions that must be met before processing can begin.</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Advanced Capture processes, Full-Text OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • None. If this option is selected, no processing precondition is necessary. • Files Idle For at Least 1 Minute. Select to indicate that processing must begin after the file indicated in the Default File Name of the processing format has been idle for at least one minute. • Require Specific Semaphore. Select to indicate that processing must begin after a trigger file is detected. The trigger file can be any file type/size/label and can be written to any location on the network. OnBase will only begin processing the processing file indicated in the Default File Name of the process format after the trigger file has been detected. <p>How processing is triggered (definition of the file location and/or time variable) is defined by a semaphore. A semaphore is a technique for coordinating or synchronizing polling activity. A maximum of 255 characters can be entered in this field.</p> <p>The trigger file is deleted after processing.</p> <hr/> <p>Note: If the trigger file is being accessed over FTP, it will not be deleted.</p> <hr/>

Option	Description
Processing Precondition (cont.)	<ul style="list-style-type: none"> • Require Same File in a .\POLL Folder. Select to indicate that processing must begin after a POLL file has been written to a specifically-configured POLL folder. The POLL file must appear in a folder labeled POLL, and the POLL folder must be created as a subfolder of the Default Directory of the process format. The name of the POLL file must be exactly identical to the name of the file to be processed. The value in the Default File Name field will be used to locate the POLL file. When OnBase locates the POLL file, the processor will attempt to process any file with that same name in the Default Directory. For example: The Default File Name is *.txt, and the Default Directory is C:\ProcessFiles. The file to be processed is stored in this directory. For this example, the file is named pf11x74.txt. The POLL file should be placed in C:\ProcessFiles\POLL, and named exactly the same as the process file (pf11x74.txt). OnBase will search C:\ProcessFiles\POLL for a file that matches the Default File Name of *.txt. Upon finding the pf11x74.txt file, the processor will return to the C:\ProcessFiles directory and search for the file named pf11x74.txt. This is the file that will be processed. The POLL file is deleted after processing. <hr/> <p>Note: This option is not supported for use with the Directory Import Processor.</p> <hr/>

Option	Description
Residual File Post-Processing	<p>The Residual File Post-Processing options allow you to specify how the processor will handle files that are left in the original folder after the import process has been run.</p> <ul style="list-style-type: none"> • Leave in Source Directory. Select to leave processed read-only files in the folder they originated in. • Move to ..\PROCESSED Directory. Select to move all processed files, regardless of read-only status, to the OnBase-generated PROCESSED folder located in the same folder the read-only files were originally in. <hr/> <p>Caution: Depending on your system's configuration, processed files may be automatically deleted after an import process is run. In this situation, the processed files will not be moved to the PROCESSED folder because they have already been deleted from the folder they originated from.</p> <p>This behavior can be avoided by modifying the configuration of your import processor, or by marking the files to be processed as read-only.</p> <hr/> <ul style="list-style-type: none"> • Delete. Select to delete the read-only files from the folder they originated in.
Miscellaneous Options	<p>The Miscellaneous Options options allow you to specify special scheduling options. Not all options are available for all processes.</p> <ul style="list-style-type: none"> • One Batch per File. Select to process each index file as one batch when multiple index files are being processed at once. <hr/> <p>Note: This option is not supported for use with the Directory Import Processor.</p> <hr/> <ul style="list-style-type: none"> • Report if No Files Found. Select to create a Verification Report if no files are found when a scheduled job is run.

- When you are finished configuring the Process Options, click **Apply**.

Viewing a Job

All scheduled process formats and jobs can be viewed in the **Schedule Management** window.

By default, the **Hide other process types** check box is enabled, so only the selected process type's process formats or process jobs are displayed.

To open the **Schedule Management** window:

- Click **Processing | Scheduler | Schedule Management** from the OnBase Client.
- From a process queue, select **Process Job** and then select a job in the **Process Jobs** window. Double-click on the job to display the process formats that compose it.

- From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed.

Modifying a Job

To modify an existing job:

From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Configure Job**.

Or, select the job to be modified from the **Process Jobs** window in the process queue, right-click and select **Configure Job**.

The **Process Job Configuration** dialog box is displayed.

Note: If you are using the OnBase Client as a Windows Service, you must restart the OnBase Client after modifying a scheduled process.

Note: For more information on configuring a process job, see [Configuring a Job on page 313](#) and [Scheduling a Job on page 315](#).

Renaming a Job

To rename an existing job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Rename Job**.
Or, select the job to be modified from the **Process Jobs** window in the process queue, right-click and select **Rename Job**.
The **Rename Process Job** dialog box is displayed.
2. Enter the new name for the job and click **OK**.

Deleting a Job

Caution: If you delete a process format or process job that is scheduled, it will be deleted from the list of scheduled jobs.

To delete an existing job:





1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Delete Job**.
Or, select the job to be modified from the **Process Jobs** window in the process queue, right-click and select **Delete Job**.
A confirmation message is displayed.
2. Click **OK**. The job is deleted.

Running/Suspending a Job

From the **Schedule Management** window, a job can be run immediately or it can be suspended.

1. Open the **Schedule Management** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select one or more jobs from the **Scheduled Items** box.
 - To run the jobs now, click **Run Now**. The selected jobs are run the next time the processing workstation is polled.
 - To suspend the jobs, click **Suspend**. To resume suspended jobs, click **Resume**.

An icon is displayed next to each scheduled job in the **Scheduled Items** box that indicates its status.

Icon	Description
	Run Now - Indicates that the user has clicked the Run Now button to cause the job to execute now instead of waiting for its scheduled time to run.
	Suspend - Indicates a suspended job. The job will not run until a user selects it and clicks Resume .
	Active - Indicates an active scheduled job. An active job may be waiting to run or it may have already run at its scheduled time.
	Error - Indicates a job with a configuration error.

3. Click **Apply**.

A job can also be run immediately from the process format queue or the **Process Jobs** window.

From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Process Job**.

Or, from a process queue, select **Process Job** and then select the job to be run in the **Process Jobs** window. Right-click in the **Process Jobs** window and select **Process Job**.

Viewing the Activity Log

The Activity Log provides visibility and control over the logging information generated during the execution of scheduled processes. To view the Activity Log, follow these steps:

1. From the OnBase Client, click **Processing | Scheduler | Schedule Management**. The **Schedule Management** dialog box is displayed.

The screenshot shows the 'Schedule Management' dialog box with the 'Activity Log' tab selected. The 'Process Scheduling' tab is also visible. The 'Activity Log' tab displays a table with the following data:

Name	Scheduled Item	Action	Workstation	Instance
Parse Format : AP - Purch...	AP - Purchase Ord...	Parse Format	DOCm010584 (AUTO)	

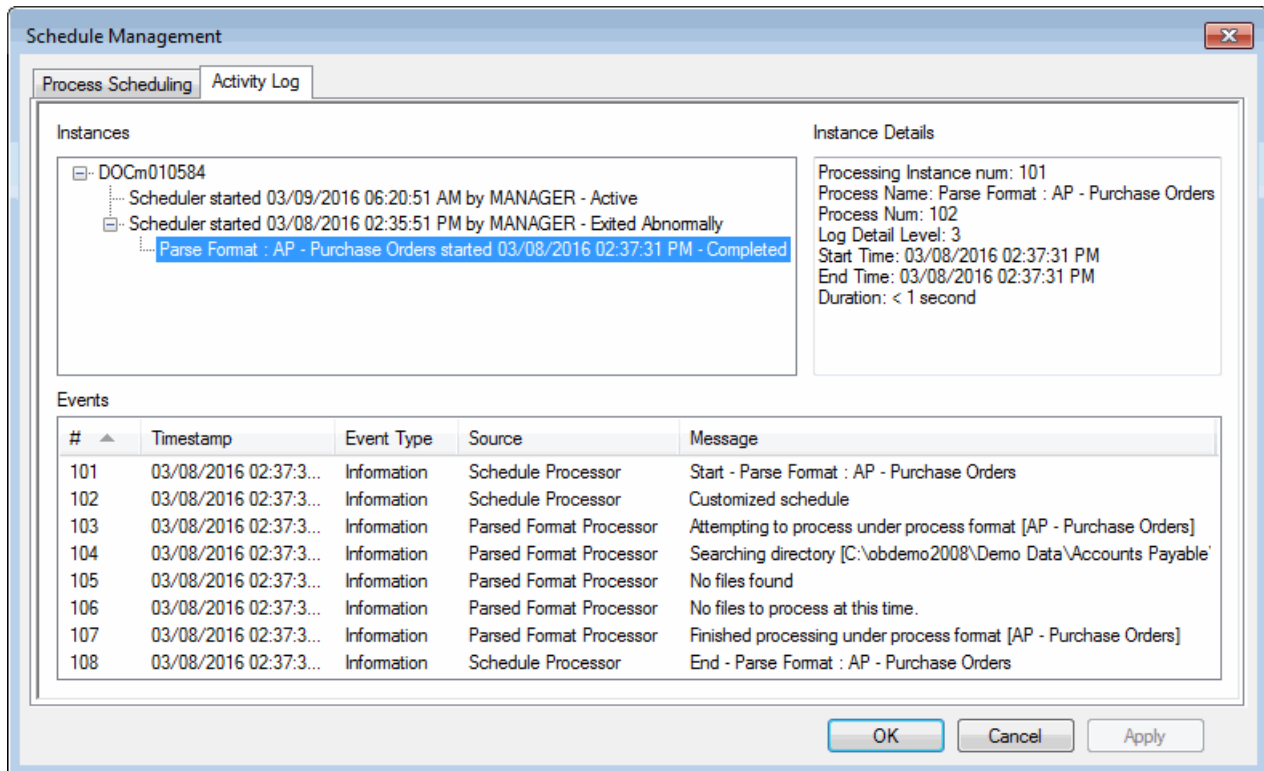
Buttons on the right side of the 'Activity Log' tab include 'Run Now', 'Suspend', 'Remove', and a checkbox for 'Hide other process types' (checked).

The 'Processing Options' tab is also visible, showing the following configuration:

- Name:** Parse Format : AP - Purchase Orders
- Processing Workstation:** DOCm010584 (AUTO)
- Schedule Template:** Weekends
- Specific Processing Instance:** (empty field)
- Weekly:** (radio buttons for Sun, Mon, Tue, Wed, Thu, Fri, Sat; Sun and Sat are selected)
- Default Daily Schedule:** (radio button selected)
- Time Range:** Start Time: 8:00 AM, End Time: 5:00 PM
- Processing Frequency:**
 - ☐ Once then Suspend
 - ☒ Once per Day
 - ☐ Once Every [] Minutes

Buttons at the bottom of the dialog box include 'OK', 'Cancel', and 'Apply'.

- Click the **Activity Log** tab. The **Activity Log** is displayed.



Note: The **Activity Log** tab is only available if logging is enabled and at least one log entry exists.

- Select a log entry to view more information about that processing instance. Details on the selected instance are displayed in the Instance Details section in the upper right corner of the dialog box, and details on each event within that instance are displayed in the Events section in the bottom of the screen.

Note: Depending on your assigned product rights, you may be able to delete unneeded entries from the Activity Log. See the User Group Configuration for Product Rights section of the **System Administration** documentation for information on product rights.

Creating Schedule Templates

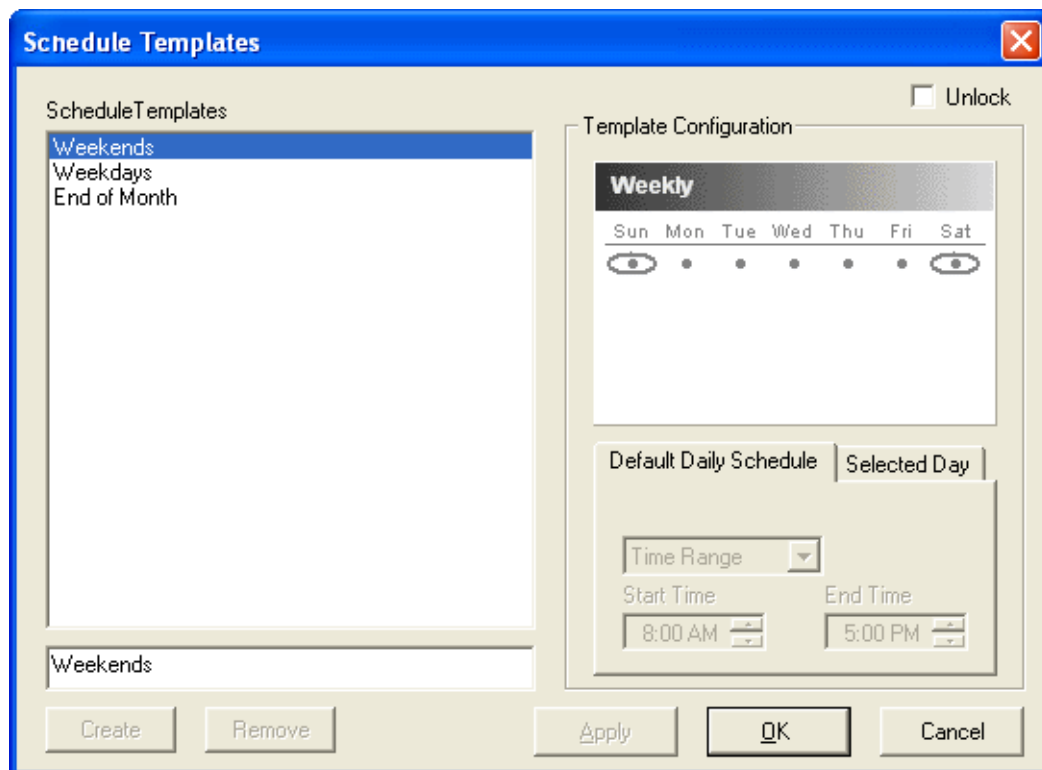
Creating Schedule Templates

A schedule template is used to create a processing schedule. These schedules can be used by multiple scheduled processes without having to be re-configured each time they are used.

Note: Any user with the Client and Client Scheduler product rights can create a schedule template. Once created, a schedule template is available to all users with Client and Client Scheduler product rights.

To create a schedule template:

1. From the OnBase Client, click **Processing | Scheduler | Schedule Templates**. The **Schedule Templates** window is displayed.



2. Enter a name for the new template and click **Create**.

Note: The maximum number of characters that can be used for a name is 80.

3. Configure the appropriate options. See the sub-sections below for more information on using the calendar, **Default Daily Schedule**, and **Selected Day** options under the **Template Configuration** area.
4. Once all Template Configuration options have been set, click **OK**.

To edit an existing template, select it from **Schedule Templates** list and select the **Unlock** check box. Once you have finished modifying it, click **OK**.

Calendar

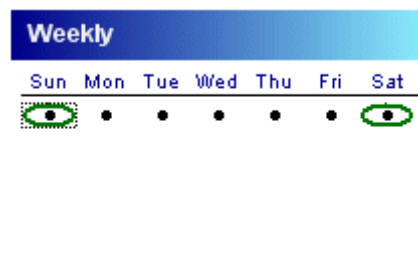
The calendar is used to select the day(s) on which a scheduled process should be run.

Note: The calendar is displayed based on your Workstation Regional Settings and the OnBase language DLL that you are using.

To change the view of the calendar, click the calendar heading (in the example above, **Weekly**) to display a menu. Select one of the following options to display a different calendar for configuration:

- **Weekly.** Allows you to configure a process to run on a certain day of the week (i.e., Thursday).
- **Monthly.** Allows you to configure a process to run monthly, on a particular date (i.e., the 1st and 15th of the month).
- **Monthly (Day-Relative).** Allows you to configure a process to run on a relative day of the month (i.e., the first Saturday of the month, the 2nd Wednesday of the month).
- **Annual.** Allows you to configure a process to run on a certain day of the year (i.e., June 30).
- **Full Calendar.** Allows you to configure a process to run on specified days of specified years (e.g., August 10, 2011 and/or July 17, 2012).

To select days that you would like to run a scheduled process, double-click the day on the calendar. The selected day is circled.



Note: In the example above, two days are selected but **Sunday** is the currently-selected day.

To deselect a day, double-click it.

Default Daily Schedule

The **Default Daily Schedule** tab allows you to configure the processing configuration for all days that do not have a **Selected Day** tab configuration.

The screenshot shows the 'Default Daily Schedule' tab selected. A dropdown menu is set to 'Time Range'. Below it, there are two time selection boxes: 'Start Time' set to 8:00 AM and 'End Time' set to 7:59 AM.

The screenshot shows the 'Default Daily Schedule' tab selected. A dropdown menu is set to 'Specific Time'. Below it, there is a single time selection box labeled 'Time' set to 8:00 AM.

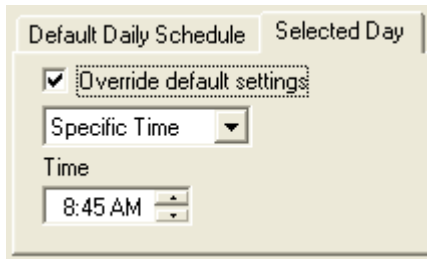
The drop-down list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time in which you want your job or format to begin processing. If you select **Specific Time**, a **Time** box is displayed. Select the time at which you want the job or format to begin processing.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Selected Day

The **Selected Day** tab allows you to specify settings for the selected day that differ from the settings specified in the **Default Daily Schedule** tab. In order for the **Selected Day** tab to be enabled, you must click a day to select it and you must select the **Override default settings** check box.

The screenshot shows the 'Selected Day' tab selected. The 'Override default settings' checkbox is checked. A dropdown menu is set to 'Time Range'. Below it, there are two time selection boxes: 'Start Time' set to 8:45 AM and 'End Time' set to 5:00 PM.

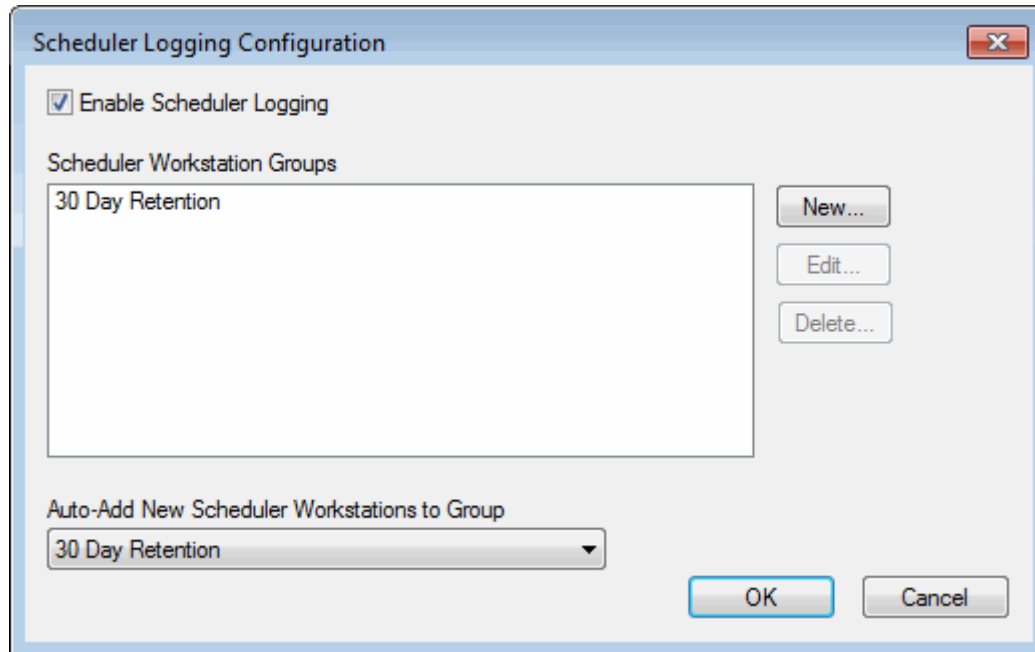


The drop-down list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time in which you want your job or format to begin processing. If you select **Specific Time**, a **Time** box is displayed. Select the time at which you want the job or format to begin processing.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Configuring Schedule Logging

Schedule logging is controlled at the workstation group level. Each workstation used to perform scheduled processing can only be a member of a single workstation group, and the settings defined for a workstation group are applied to all workstations within that group. Scheduler logging is configured from the **Scheduler Logging Configuration** dialog box, available from the OnBase Client under **Processing | Scheduler | Logging Configuration**.



Note: This dialog box is only available for selection if your user account has been assigned the required product right. See the User Group Configuration for Product Rights section of the **System Administration** documentation for information on product rights.

Select the **Enable Scheduler Logging** option to perform scheduler logging for all scheduler workstation group that have enabled the **Enable Logging for Group** option. If this option is not selected, no scheduler logging is performed for any scheduler workstation group.

By default, there is a single group named **30 Day Retention**. Other groups can be created as needed, depending on the logging requirements of different types of processing workstations. See the following topics for more information on creating, editing, and deleting scheduler workstation groups:

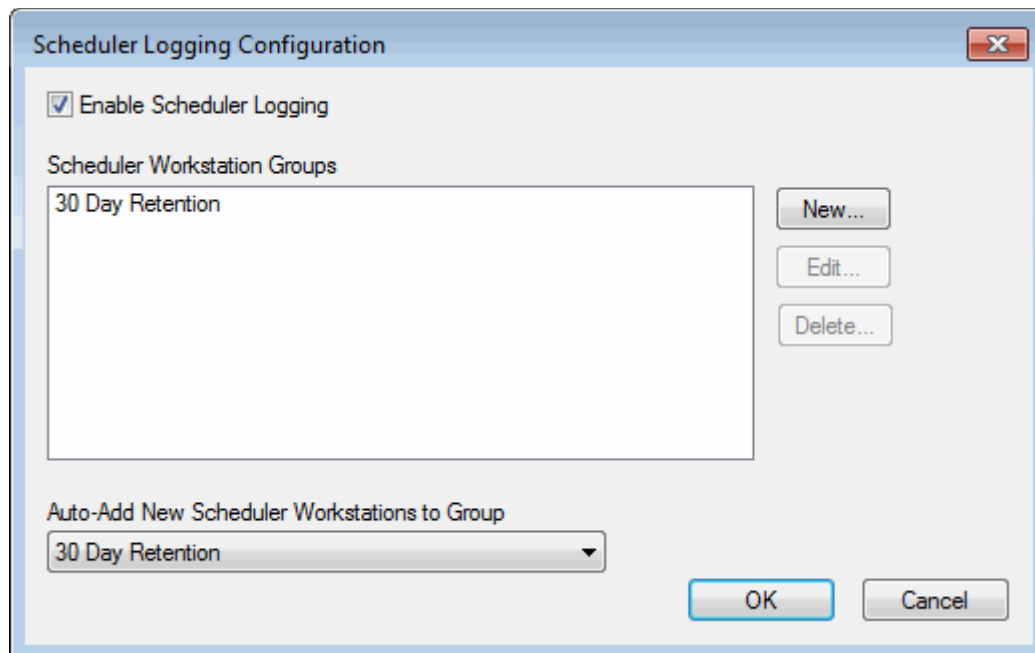
- See [Creating a Scheduler Workstation Group on page 335](#) for more information on creating a new scheduler workstation group.
- See [Editing a Scheduler Workstation Group on page 338](#) for more information on editing a scheduler workstation group.
- See [Deleting a Scheduler Workstation Group on page 341](#) for more information on deleting a scheduler workstation group.

The **Auto-Add New Scheduler Workstations to Group** setting controls whether or not new scheduler workstations will automatically add themselves to a scheduler workstation group. Select a scheduler workstation group from the drop-down list to automatically add new processing workstation to that group, or select <none> to disable automatic addition. By default, this is set to the **30 Day Retention** group.

Creating a Scheduler Workstation Group

Scheduler workstation groups control how schedule logging is performed by the assigned workstations. To create a new scheduler workstation group, follow these steps:

1. From the OnBase Client, click **Processing | Scheduler | Logging Configuration**. The **Scheduler Logging Configuration** dialog box is displayed.



2. Click **New**. The **Scheduler Logging Workstation Group** dialog box is displayed.

The screenshot shows the "Scheduler Logging Workstation Group" dialog box. It has a title bar with a close button. The main area contains several sections: "Workstation Group Name" with a text input field; "Detail Level" with a dropdown menu showing "<none>"; "Log Retention Settings" with four checkboxes: "Remove Log Entries After Retention Period" (checked, with a value of 0 days), "Retain Processing Summary After Retention Period", "Never Remove Logs Containing Processing Errors", and "Treat Warnings as Errors"; "Available Workstations" with a list box containing "SRVm006291 (AUTO)"; "Assigned Workstations" with an empty list box; and "Display Only Workstations with Scheduled Processes" with a checked checkbox. Between the two list boxes are ">>" and "<<" buttons. At the bottom right are "OK" and "Cancel" buttons.

3. Type a name for the scheduler workstation group in the **Workstation Group Name** field.
4. Select the **Enable Logging for Group** option so that logging is performed for workstations in the group. If this option is not selected, logging is not performed for this scheduler workstation group.
5. Select the desired amount of data to be logged from the **Detail Level** drop-down list. The higher levels of detail are most useful for new processes or processes that are experiencing issues.

6. If desired, you can configure a retention period for log entries. The following options are available:

Option	Description
Remove Log Entries After Retention Period: _ days	Select this option and enter a number in the available field to remove log entries from the scheduler log after the specified number of days.
Retain Processing Summary After Retention Period	Select this option to retain the processing instance record after the retention period has passed and all of the record's log entries have been removed.
Never Remove Logs Containing Processing Errors	Select this option to prevent the retention period from being applied to any processing logs that reported an error. This can provide an administrator more time to analyze any recorded issues.
Treat Warnings as Errors	<p>Select this option to treat warnings as errors for the purpose of log retention. When this option is selected, the retention period is not applied to any processing logs that reported a warning.</p> <hr/> <p>Note: This option is only available if the Never Remove Logs Containing Processing Errors option is selected.</p> <hr/>

7. Select all workstations you want to assign to this scheduler workstation group from the **Available Workstations** list, then click the >> button. The selected workstations are added to the **Assigned Workstations** list.

Because workstations can only be assigned to a single scheduler workstation group, the list of workstations in the **Available Workstations** list does not include any workstations that are already assigned to another scheduler workstation group.

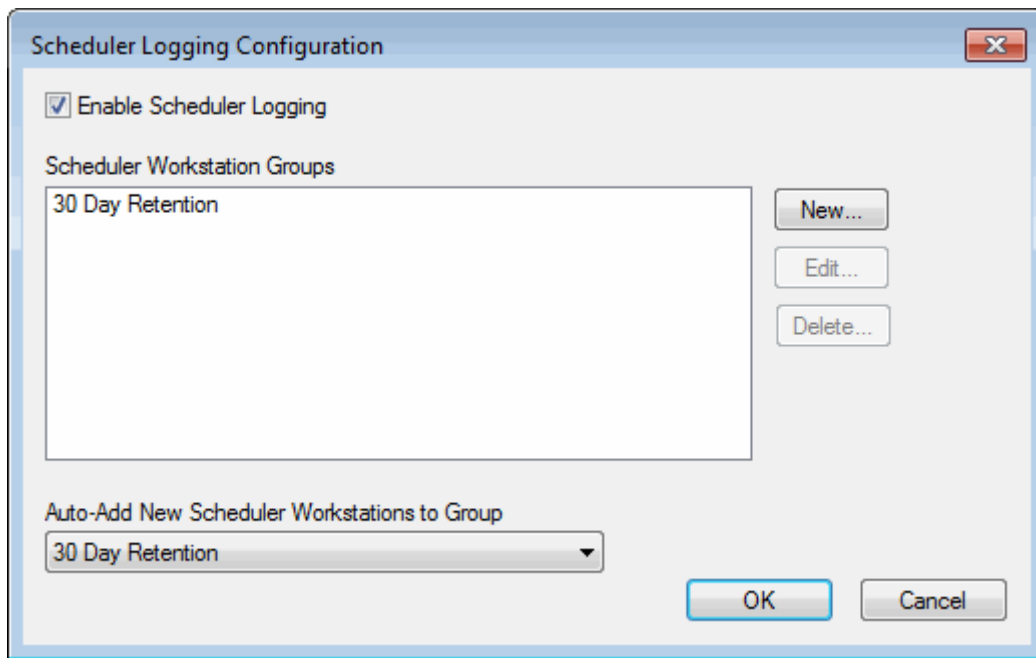
Tip: You can select the **Display Only Workstations with Scheduled Processes** option to limit the list of **Available Workstations** to those workstations that have scheduled processes assigned to them.

8. Click **OK**.

Editing a Scheduler Workstation Group

Scheduler workstation groups control how logging is performed by the assigned workstations. To edit an existing scheduler workstation group, follow these steps:

1. From the OnBase Client, click **Processing | Scheduler | Logging Configuration**. The **Scheduler Logging Configuration** dialog box is displayed.



2. Select a scheduler workstation group and click **Edit**, or double-click on a scheduler workstation group. The **Scheduler Logging Workstation Group** dialog box is displayed.

The screenshot shows the "Scheduler Logging Workstation Group" dialog box. It has a title bar with a close button. The main area contains several sections: "Workstation Group Name" with a text box containing "30 Day Retention"; "Detail Level" with a dropdown menu showing "Errors, Warnings, and Processing Information"; "Log Retention Settings" with checkboxes for "Remove Log Entries After Retention Period: 30 days", "Retain Processing Summary After Retention Period", "Never Remove Logs Containing Processing Errors", and "Treat Warnings as Errors"; "Available Workstations" and "Assigned Workstations" lists, both containing "DOCm010584 (AUTO)"; and a checkbox for "Display Only Workstations with Scheduled Processes". Navigation buttons ">>" and "<<" are between the lists. "OK" and "Cancel" buttons are at the bottom right.

Scheduler Logging Workstation Group

Workstation Group Name
30 Day Retention

☒ Enable Logging for Group

Detail Level
Errors, Warnings, and Processing Information

Log Retention Settings

☒ Remove Log Entries After Retention Period: 30 days

☐ Retain Processing Summary After Retention Period

☐ Never Remove Logs Containing Processing Errors

☐ Treat Warnings as Errors

Available Workstations

DOCm010584 (AUTO)
DOCm012604 (AUTO)
SRVm006291 (AUTO)

Assigned Workstations

DOCm010584 (AUTO)

>>

<<

☐ Display Only Workstations with Scheduled Processes

OK Cancel

3. Modify the scheduler workstation group's settings as desired. The following settings are available:

Option	Description
Workstation Group Name	The name of the scheduler workstation group.
Enable Logging for Group	The Enable Logging for Group option controls whether or not logging is performed for workstations in the group. Logging is only performed if this option is selected.
Detail Level	The Detail Level drop-down list controls the amount of data that is logged. Higher levels of detail are most useful for new processes or processes that are experiencing issues.
Remove Log Entries After Retention Period: _ days	When this option is selected, log entries are removed from the scheduler log after the specified number of days.
Retain Processing Summary After Retention Period	When this option is selected, the processing instance record is retained after the retention period has passed and all of the record's log entries have been removed.
Never Remove Logs Containing Processing Errors	When this option is selected, the retention period is not applied to any processing logs that have reported an error. This can provide an administrator more time to analyze any recorded issues.
Treat Warnings as Errors	<p>When this option is selected, warnings are treated as errors for the purpose of log retention. The retention period is not applied to any processing logs that have reported a warning.</p> <hr/> <p>Note: This option is only available if the Never Remove Logs Containing Processing Errors option is selected.</p> <hr/>
Available Workstations/Assigned Workstations	<p>The Available Workstations list contains all workstations that are available to be assigned to this scheduler workstation group. Because workstations can only be assigned to a single scheduler workstation group, the list of workstations in the Available Workstations list does not include any workstations that are already assigned to another scheduler workstation group.</p> <p>The Assigned Workstations list contains all workstations that have been assigned to this scheduler workstation group.</p>

Option	Description
Display Only Workstations with Scheduled Processes	When this option is selected, the list of Available Workstations is limited to those workstations that have scheduled processes assigned to them.

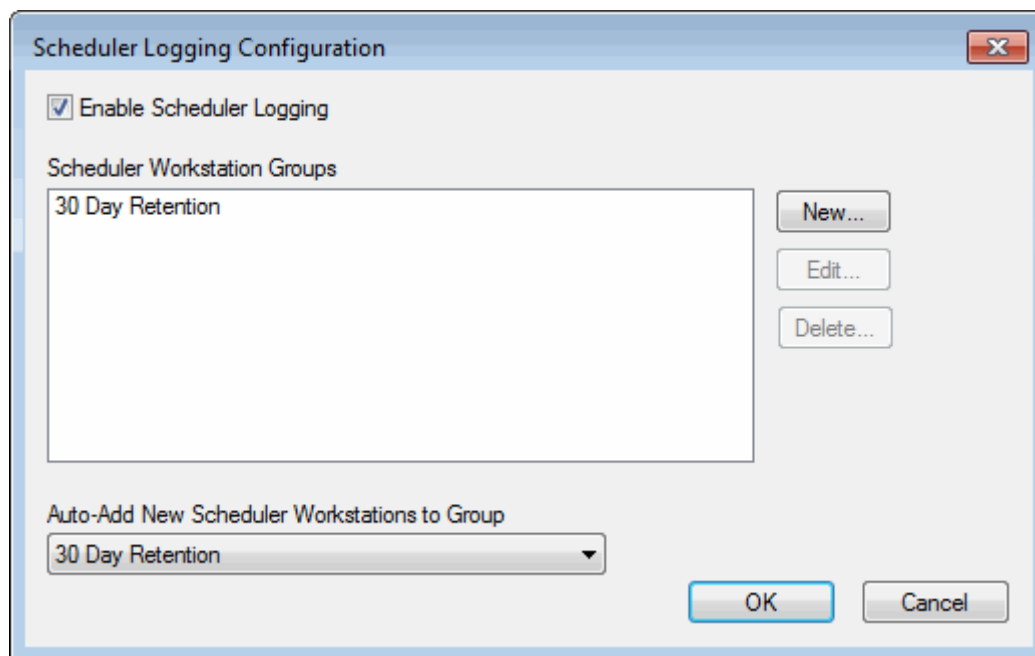
Note: After making a change to any of the options under **Log Retention Settings**, previously retained logs are rechecked to verify that they conform with the new settings. Logs which do not will be removed. For example, if you had previously configured the scheduler workstation group to **Retain Processing Summary After Retention Period** and then deselect that option, existing processing summaries older than the retention period will be removed.

4. Click **OK**.

Deleting a Scheduler Workstation Group

Scheduler workstation groups control how logging is performed by the assigned workstations. To delete a scheduler workstation group, follow these steps:

1. From the OnBase Client, click **Processing | Scheduler | Logging Configuration**. The **Scheduler Logging Configuration** dialog box is displayed.



2. Select a scheduler workstation group and click **Delete**. A confirmation dialog box is displayed.
3. Click **Yes**. The selected scheduler workstation group is deleted, and any workstations that were assigned to that group are available to be added to another scheduler workstation group.

The licenses listed in the following sections are applicable if you are upgrading from a version of OnBase prior to OnBase Foundation EP5.

Overview

Each Advanced Capture workstation requires a valid OnBase Client license, in addition to one of the following Advanced Capture licenses:

- **Ad-Hoc Advanced Capture.** An Ad-Hoc Advanced Capture license allows you to perform Point and Shoot Indexing and Advanced Capture on an open document from the right-click menu.
- **Advanced Capture.** An Advanced Capture license allows you to perform Point and Shoot Indexing and Advanced Capture on an open document as well as on batches residing in the **Awaiting Advanced Capture** or **Awaiting Ad Hoc Advanced Capture** batch status queues.

Note: If your solution is licensed for **Batch Automated Indexing** (a license available for OnBase solutions prior to the release of OnBase 10.0), contact your solution provider for information on obtaining a current Advanced Capture license.

Note: The Interactive Data Capture license is included with the Advanced Capture license. This means that users of Advanced Capture will automatically have access to full Interactive Data Capture functionality. For more information, see the **Interactive Data Capture** module reference guide.

If your solution is licensed for at least one Ad-Hoc Advanced Capture or Advanced Capture license, and if users have been granted the necessary product right, users of workstations not registered for Advanced Capture can route batches to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing by a centralized workstation that is licensed for Advanced Capture.

Additional Licensing

Additional licensing is required in order to use the following features in conjunction with Advanced Capture:

- A valid OnBase Workflow license is required to initiate Advanced Capture from within a Workflow Life Cycle. For more information, see the **Workflow** module reference guide or help files.
- In order to mark areas of a document that have been processed via Advanced Capture for redaction via Automated Redaction, your solution must also be licensed for Automated Redaction. For more information, see the **Automated Redaction** module reference guide or help files.
- To identify handwritten text in an Advanced Capture Data Field Zone using Intelligent Character Recognition (ICR), each workstation in your solution that is to perform ICR processing must be licensed and registered for ICR Support for Advanced Capture.
- To identify bar codes in an Advanced Capture Form Identification Zone or Data Field Zone, your solution must be licensed for the Bar Code Recognition Server. For more information, see the **Bar Code Process** module reference guide or help files.
- To perform Advanced Capture processing on Japanese, Korean, or Chinese languages, your solution must be licensed for Asian Language OCR. For more information, see the **Full-Page OCR** module reference guide or help files.

Check your current licensing status by selecting **Utils | Product Licenses** from the Configuration module.

The following best practice recommendations were assembled by a team of OnBase subject matter experts. They represent the accumulation of years of experience installing and configuring OnBase solutions.

The following recommendations are general in nature, and are applicable to most OnBase solutions and network environments. Depending on your solution design and your organization's needs, not all of the best practice recommendations listed below may apply to, or be recommended for, your OnBase solution.

Carefully consider the impact of making any changes, including those listed below, to your OnBase solution prior to implementing them in a production environment.

General Information

Recommended Standards for OCR Processing

Because OCR processing is an integral part of the Advanced Capture process, it is considered a best practice to ensure that all documents that undergo Advanced Capture meet the recommended criteria for OCR processing. This will ensure that the OCR engine returns the most accurate results possible.

- Scan documents at a minimum resolution of at least 240 dpi (300 dpi recommended) when preparing documents for OCR processing. Depending on the needs of your solution, a higher resolution may be required. The use of lower resolutions in OCR processing results in illegible text captured from the image.

The resolution should always be set to a squared value (such as 240x240 or 300x300 dpi). If poor or inaccurate OCR processing results are reported at 240 dpi, the documents should be discarded and re-scanned at incrementally higher resolutions until acceptable OCR results are achieved.

- Store bi-tonal (black and white) images using the TIFF-Group IV file format. Grayscale or color images should be saved using a lossless color-capable image file format.
- Scan documents as bi-tonal images. Bi-tonal images require far less disk space and load faster than grayscale or color images.

However, if documents are originally scanned as grayscale or color images, it is recommended to process the original grayscale or color document and then convert the image to a bi-tonal format. The OCR engine will produce better results with the original color or grayscale document than it will with a dithered document, especially if there are areas of the page that are similar in contrast.

- Always get the best scanner image possible before resorting to image cleanup.
- Use a dedicated workstation for OCR processing.

Recommended Standards for ICR Processing

If your solution is licensed for intelligent character recognition (ICR), you can use Advanced Capture to recognize and capture data from hand-written text.

It is difficult to predict the accuracy of an ICR process due to the variability in the quality of the hand-written text. The following tips will help to improve the accuracy of ICR processing.

- The hand-written characters should be clearly separate from one another (i.e., they should not touch or intersect).
- The hand-written text should be unconstrained. If the characters are being entered into constraint boxes, do one of the following:
 - If the documents containing the constraint boxes can be modified before the Advanced Capture process is run, ensure that the constraint boxes are printed in red. Then, using Document Imaging's Image Processing options, configure a red Color Dropout process to remove the constraint boxes from the documents when scanning.
 - If the documents containing the constraint boxes cannot be modified before the Advanced Capture process is run, configure your OCR process format to automatically drop the constraint boxes by selecting **Remove vertical/horizontal lines from zone before processing** in the Data Field Zone dialog box.

Constrained Text:



Name

Unconstrained Text:

A handwritten name 'John Adams' is shown without any constraint boxes. The text is written in a cursive style, with a horizontal line underneath the name.

Name

- The ICR numeric text processor (**Handwriting/ICR - numerals/grouping punctuation (North American style)**, **Handwriting/ICR - numerals/grouping punctuation (European style)**) is significantly more accurate than the ICR alphanumeric text processor (**Handwriting/ICR - alphanumeric/punctuation**) because the character set is much smaller.

If you are attempting to process numeric data only from a Data Field Zone, you will have more accurate results if you use the numeric text processor.

Date Strings

When using Advanced Capture to capture hand-written date strings that include slash (/) characters, results can vary greatly. To improve the consistency and accuracy of results, consider the following recommendations.

When Documents Can Be Modified

When the documents containing the date strings can be modified before the Advanced Capture process is run, the following modifications and corresponding configurations are recommended.

- Ensure that the slashes (/) in the date string are printed in red on the documents. Then, using Document Imaging's Image Processing options, configure a red Color Dropout process to remove the slashes from the documents when scanning.
- Replace the slashes (/) in the date string with dashes (-), then configure the date string's Data Field Zone to use the following **Specific allowed characters**: 1234567890. By only allowing the numbers in the date string, the dashes will be omitted from the result.

When Documents Cannot Be Modified

When the documents containing the date strings cannot be modified before the Advanced Capture process is run, the following configurations are recommended.

- If the Data Type of the Keyword Type configured for the date string's Data Field Zone is **Alphanumeric**, ensure that:
 - The Keyword Type is not configured to use Masking
 - The Data Field Zone does not define any **Specific allowed characters**
 - The Data Field Zone's **Advanced recognition** is configured for **Handwriting/ICR - alphanumeric/punctuation**
- If the Data Type of the Keyword Type configured for the date string's Data Field Zone is **Date**, ensure that:
 - The Data Field Zone defines the following **Specific allowed characters**:
1234567890/
 - The Data Field Zone's **Advanced recognition** is configured for **Handwriting/ICR - alphanumeric/punctuation**

OCR vs. AutoFill Keyword Sets

When configuring a solution that uses both OCR processes and AutoFill Keyword Sets, if the AutoFill Keyword Set has accurate data and the primary Keyword Value is able to be consistently and accurately recognized, it is considered a best practice to obtain the applicable values from the AutoFill Keyword Set instead of obtaining each of these values individually through OCR processing. Using reliable AutoFill Keyword Sets whenever possible reduces processing time and improves accuracy.

For more information, see the **AutoFill Keyword Sets** module reference guide or help files.

Usage

It is recommended that the following best practice information be considered before scanning and processing documents in the Advanced Capture client.

Advanced Capture

Ad Hoc Advanced Capture

Unlike batch Advanced Capture, ad hoc Advanced Capture offers no indication of values that were not confidently identified by the Advanced Capture engine (i.e., values that did not meet their configured Suspect Level). It is considered a best practice to always compare the indexing information assigned to the document by the ad hoc Advanced Capture process to the actual information displayed on the document to ensure that it is accurate.

Batch Advanced Capture

When scanning a batch of documents into a scan queue, it is considered a best practice to scan the batch as **Unindexed** or **Pre-Indexed** so that any documents that undergo batch Advanced Capture and are identified as having suspect Keyword Values are sent for user review in either the **Awaiting Index** or **Index in Progress** batch status queue. If the batch is scanned as **Fully Indexed**, batches containing documents that are associated with suspect Keyword Values are not routed for manual review in either batch status queue because the batch is already marked as indexed.

Performing Point and Shoot Indexing

It is considered a best practice to always compare the values returned from the OCR engine to the actual values displayed on the document when performing Point and Shoot Indexing.

Configuration

It is recommended that the following best practice information be considered before configuring your Advanced Capture solution.

Advanced Capture Form Configuration

The following best practice information should be considered when creating or modifying an Advanced Capture form.

- By default, the Suspect Level threshold is set to **75** and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.
- By default, the Assigned Document Type is set to **<Document Type Unchanged>**, which retains the existing Document Type of documents that will undergo Advanced Capture processing. When the Document Type is set at the scan queue level and does not need to be modified, it is considered a best practice to leave this setting as the default to avoid duplication of effort.
- When configuring a text Match Value for a Form Identification or Page Registration Zone, if the position of the text shifts from document to document to an extent that the exact text contained within the zone might change, it is considered a best practice to use the **Contains text** option to increase the chances of a desired match.
- When configuring an Advanced Capture form to use a **Regular expression match**, it is considered a best practice to test the expression with the OnBase Regex Validator, available from the Hyland Community or your solution provider. This helpful tool allows you to validate regular expressions and to view examples of the types of data typically captured by the Advanced Capture module.
- It is considered a best practice to save your Advanced Capture form after the form is created and after any configuration change is made (for example, any Form Identification, Registration, or Data Field Zones are added, deleted, modified, or moved).

- Because a limitless number of Advanced Capture forms can be configured for a single Document Type, let alone all Document Types configured for your Advanced Capture solution, it is considered a best practice to limit the number of Advanced Capture forms available to a Document Type as much as possible in order to save time and processing resources.
- It is considered a best practice to test each Advanced Capture form after it is created and after any modifications have been made to it. All Advanced Capture forms should also be regularly tested to ensure that they do not require updating.
- It is considered a best practice to periodically perform a form order analysis and re-order your forms. The frequency of this analysis should be based on the number of documents you are processing via Advanced Capture and the number of forms configured for your solution.
- If you would like your Advanced Capture solution to capture duplicate line item information (for example, when a student's school transcript displays duplicate line items reflecting values for a class taken multiple times within the same term), it is considered a best practice to use one of the following configurations:
 - For the MIKG configured for capturing the line items (for example, for course information on a school transcript), configure an additional Keyword Type representing a sequence number for the line items. Then, in the Sequencing Keyword Types for Multi-Instance Groups section of the **Form Definition Scan Queues/Keywords** dialog box, select this Keyword Type.
 - For the MIKG configured for capturing the line items (for example, for course information on a school transcript), configure two additional Keyword Types: **Line Item** and **Page**. Then, when configuring a Grouped Line Item Extraction Data Field Zone, configure the **Line Item** and **Page** Keyword Types as tags and assign them the **<Line Number>** and **<Page>** Meta Data Values, respectively.

Keyword Type Masking

When working with Keyword Types that have been configured for masking in your Advanced Capture solution, it is recommended that you configure regular expressions to compare and replace extracted values for these Keyword Types to ensure that these values are valid for the configured masks.

A regular expression is a rule that specifies a particular format for a value. Several OnBase modules (e.g., Automated Redaction, Advanced Capture, and Intelligent Capture for AP) are able to use regular expressions to identify data.

Tip: When working with regular expressions, it is considered a best practice to test the expression with the OnBase Regex Validator, available from the Hyland Community or your solution provider. This helpful tool allows you to validate regular expressions and to view examples of the types of data typically identified by the Advanced Capture module.

Commonly Used Characters

The following is a list of characters that are commonly used in regular expressions.

Character	Description
\d	Matches any numeric character
\D	Matches any non-numeric character
\s	Matches any white space character (including Tab and Alt characters, ASCII 32 and lower)
\S	Matches any non-white space character
\w	Matches any word character (i.e., A-Z , 0-9 , and _)
\W	Matches any non-word character
*	Denotes 0 or more instances of the preceding element
+	Denotes 1 or more instances of the preceding element
?	Denotes 0 or 1 instance of the preceding element
.	Matches any single character (i.e., wildcard)
^	Matches the starting position within the search string
\$	Matches the ending position within the search string
[]	Matches any single character included in the specified set of characters (e.g., [A-DF] or [ABCDF] would match A , B , C , D , or F)

Character	Description
[^]	Matches any single character not included in the specified set of characters (e.g., [^F] would match any character except F)
()	Denotes a logical grouping of part of an expression, as well as a SubMatch (e.g., if [A-Z]{3}\s(\d{3}) is matched against PSY 101 , then 101 is the SubMatch)
{ }	Denotes the minimum and maximum match counts (e.g., in the string ABC1990 , ABC\d{2,4} would match ABC19 , ABC199 , or ABC1990)
	Separates alternate possibilities (e.g., Bob Steve would match Bob or Steve)
?:	Denotes that the SubMatch it is contained within will not be stored (e.g., (abc)(?:defg)(123) would only have two SubMatches: abc and 123)

Commonly Used Regular Expressions

The following is a list of commonly used regular expressions.

Be aware that multiple expressions can be configured to specify each type of value; the following expressions are not the only expressions that can be used.

Also note that longer and more complex regular expressions might not be listed below. For a full list of your system's default, pre-populated regular expressions, access the Regular Expression Library.

Personal Information

Value Type	Example	Regular Expression
Name	Doe, John A	[A-Z]+,\s[A-Z]+\s[A-Z]+
Address	28500 Clemens Rd. Apt. 17b	For the full expression, see the default expression for Business/Home Address in the Regular Expression Library.
Social Security Number	123-45-6789	\d{3}-\d{2}-\d{4}

Value Type	Example	Regular Expression
Date of Birth (mm-dd-yy)	01-01-79 01/01/79	\d{2}(?:\W /)\d{2}(?:\W /)\d{2}
	01-01-1979 01/01/1979	\d{2}(?:\W /)\d{2}(?:\W /)\d{4}
	01-Jan-1979	\d{2}\W[A-Z]{3}\W\d{4}

Higher Education

Value Type	Example	Regular Expression
Term Name	Fall Semester 2000 Spring Semester 2001 Fall 2000 Spring 2001	(\S{4,6}\s)(?:SEMESTER)\s(\d{4})
Academic Year	Academic Year 1997 Academic Year 2001	(?:ACADEMIC YEAR)\s\d{4}
Course ID	PSY-101	\S{3}(?:\W _)\d{3}

General Examples

Value Type	Example	Regular Expression
Alphanumeric	ABC 123	[A-Z]{3}\s\d{3}
	AF77-12345678	[A-Z]{2}\d{2}-\d{8}
Decimal	3.00 or 12.0 or 8 (3.00) or [12.00] or (8)	\W?\d{1,2}(?:\W _)?\d{0,2}\W?
	\$156.55	\\$\d*\.\d{2}
	10.001257	\d+\.\d+
Date	June 06, 2011	[A-Z]+\s\d{2}[.,]\s\d{4}
	06 December, 2011	\d{2}\s[A-Z.,]+\s\d{4}

Value Type	Example	Regular Expression
Number	77-123456789	77-\d{9}
	0123456789	\d{10} (fixed length value)
		\d{5,20} (range)
		\d+ (variable, minimum length of one character to any maximum length value)

Use the following quick reference guides to help users or administrators perform their daily tasks.

Note: Quick reference guides are intended to serve as convenient guides for performing basic procedures involving specific features and functions of a module; they are not intended to serve as substitutes for the OnBase module reference guides. For comprehensive information on a module, see the appropriate module reference guide.

Quick Reference Guide	Description
Scanning and Index Verification (Document Imaging and Advanced Capture)	Describes basic procedures for preparing and scanning a batch (via Document Imaging), performing an Advanced Capture process, and performing index verification.
Scanning and Index Verification (Disconnected Scanning and Advanced Capture)	Describes basic procedures for preparing and scanning a batch (via Disconnected Scanning), performing an Advanced Capture process, and performing index verification.

Scanning a Batch

- From the Queue List window in the upper-left corner, select **Scan Queues**. All scan queues available to you are displayed in the Working window on the right.
- Double-click the appropriate scan queue.
- Load a batch of documents into the scanner, ensuring that the pages are positioned properly for scanning.
- In the Indexing window on the left (below the Queue List window), do the following:
 - From the **Scan Format** drop-down list, select an appropriate scan format for the batch.
 - In the **Scan Mode** section, select **Pre-Index**.
 - From the **Document Type** drop-down list, select an appropriate Document Type (or simply verify the default).
 - Click **Scan**.
- If prompted for a batch name, enter an appropriate name in the **Batch Name** dialog box and click **OK**.
- When scanning is complete, the **Scanning Complete** dialog box is displayed.
- Click **Done**. The batch is routed to the **Awaiting Advanced Capture** batch status queue.

Index Verification (cont.)

8. Repeat steps 3 to 7 for each document in the batch. To combine a displayed document with the previous document, click **Append**.
9. Once all documents in the batch have been indexed, the batch is automatically routed to the next configured batch status queue for the scan queue.

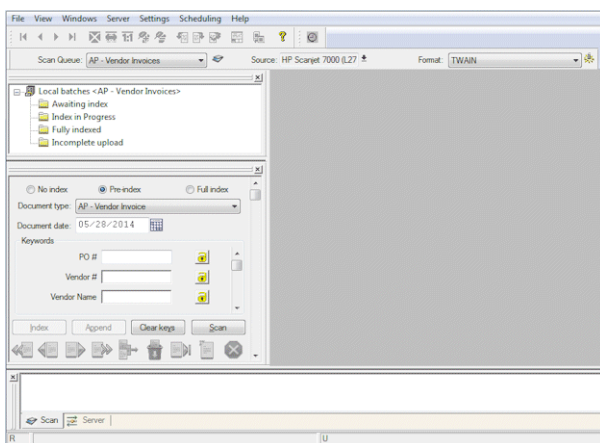
Document Preparation

Ensure that all documents are prepared and batched for scanning according to your Advanced Capture solution's requirements. For a complete list of recommended document preparation procedures, see the **Preparing Documents for Scanning** white paper.

Scanning a Batch

Note: The following procedures should be performed by scan machine operators. For the batch to be routed as described, the scan queue to which the batch belongs must be configured for Advanced Capture. For more information on Advanced Capture configuration and procedures, see the **Advanced Capture** module reference guide.

1. Log in to the Disconnected Scanning client. The **Disconnected Scanning** client window is displayed.



2. From the **Scan Queue** drop-down list in the upper-left corner (i.e., the list of all scan queues available to you), select the appropriate scan queue.
3. Click the **Select scanning method** button and select an appropriate scanning method (e.g., **Kofax**, **ISIS**, **TWAIN**, **Sweep**, or **PaperStream**) from the drop-down list.



4. Click the **Select source device** button to access a list of available scan sources.



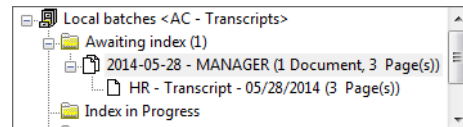
5. From the **Select Source** dialog box, select the appropriate scan source and click **Select**.
6. From the **Format** drop-down list, select the appropriate scan format.
7. From the **Indexing** window on the left, do the following:
 - Ensure that **Pre-index** is selected.

☐ No index ☒ Pre-index ☐ Full index

- From the **Document Type** drop-down list, select an appropriate Document Type (or simply verify the default).

Scanning a Batch (cont.)

8. Load a batch of documents into the scanner, ensuring that the pages are positioned properly for scanning.
9. From the **Indexing** window, click **Scan**.
10. If prompted for a batch name, enter an appropriate name in the **Batch Name** dialog box and click **OK**.
11. When scanning is complete, the **Scanning Complete** dialog box is displayed.
12. Click **Done**. The batch is routed to the **Awaiting index** batch status queue in the **Batches** window of the Disconnected Scanning client.



13. Select the batch, right-click, and select **Upload batch**. The batch is routed to the **Awaiting Advanced Capture** batch status queue within OnBase, where it can be processed.

Performing Advanced Capture

Note: The following procedures should be performed by indexing personnel with rights to perform Advanced Capture processing. If your solution is configured to perform Advanced Capture processing automatically, proceed to the **Index Verification** section below.

1. From the Queue List window, select the **Awaiting Advanced Capture** batch status queue. All batches residing in the queue are displayed in the Working window on the right.
2. Select the appropriate batch, right-click, and select **Perform Advanced Capture Processing**. A confirmation message is displayed.
3. Click **OK**. Advanced Capture processing is performed. If any processed values are marked as suspect, the batch is routed to either the **Awaiting Index** or **Index in Progress** batch status queue for index verification.

Index Verification

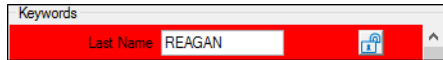
Note: The following procedures should be performed by index verification personnel.


1. From the Queue List window, select the **Awaiting Index** or **Index in Progress** batch status queue. All batches residing in the queue are displayed in the Working window on the right.
2. Select the appropriate batch, right-click, and select **Index Documents**. The first document in the batch is displayed in the Working window.

Index Verification (cont.)

3. In the Indexing window on the left (below the Queue List window), the Keyword Type fields are displayed. Verify that all indexing information captured via the Advanced Capture process is correct.

Note: If a captured Keyword Value is marked as suspect, its Keyword Type field is highlighted in red by default. Depending on your configuration, some Keyword Type fields may be highlighted in a different color.



4. Enter any remaining values for the Keyword Types that require manual entry as defined by your Advanced Capture solution.
 - To ensure that a Keyword Value is carried over to the next document in the batch, click the lock button to the right of the Keyword Type field.
 - To add an additional line for a Keyword Type field or a new line for a Keyword Type Group, select the field and press **F6**.
5. To delete any invalid Keyword Values, highlight the value and press **Delete** or **Backspace**.
6. To delete an invalid itemized line from a Multi-Instance Keyword Type Group, select the Keyword Type field of the itemized line and press **Ctrl+Delete**.
7. Once you have corrected and entered all appropriate Keyword Values, click **Index**. The next document in the batch is displayed in the Working window.
8. Repeat steps 3 to 7 for each document in the batch. To combine a displayed document with the previous document, click **Append**.
9. Once all documents in the batch have been indexed, the batch is automatically routed to the next configured batch status queue for the scan queue.



Advanced Capture

User Guide

Advanced Capture allows you to automatically extract indexing values, such as Document Type identifiers, the Document Date, and Keyword Values, directly from an image or PDF document.

Point and Shoot Indexing allows you to extract the Document Date and Keyword Values directly from any image document displayed in the Document Imaging Working window.

Advanced Capture

Both ad hoc and batch Advanced Capture allow you to index a document by automatically matching it to Advanced Capture forms. Advanced Capture forms are pre-configured templates that have been created with certain zones, or fields, that are analyzed for a specified type of indexing information (that is, a Document Type identifier, the Document Date, or Keyword Values).

The Advanced Capture engine compares all available Advanced Capture forms to a document until it is able to match a form to the document's physical layout. Once a match is determined, the document is processed and the indexing information is automatically applied to the document.

You can perform ad hoc Advanced Capture from a number of places within the Client module, and, if you have been granted the necessary product right but your workstation is not registered for Advanced Capture, you can queue documents for processing at a later time or on another workstation. Depending on your licensing, you may also be able to perform batch Advanced Capture on a batch of documents from within the Document Imaging interface.

If your workstation is not registered for the Bar Code Recognition Server, bar code processing cannot be performed. For more information, see the **Bar Code Process** documentation.

Performing Ad Hoc Advanced Capture

Ad hoc Advanced Capture allows you to perform Advanced Capture on a document while working with the document. Depending on how you are interacting with the document, ad hoc Advanced Capture can be initiated via right-click option or by pressing a button on a dialog box.

Ad hoc Advanced Capture can be performed from the following areas of the OnBase Client:

- **From the Document Search Results list.** See From the Document Search Results List on page 362 for more information.
- **From an open document.** See From an Open Document on page 363 for more information.
- **From the Document Imaging window.** See From the Document Imaging Window on page 364 for more information.
- **From the Import Document dialog box.** See From the Import Document Dialog Box on page 367 for more information.

- **From a Folder.** See From a Folder on page 368 for more information.
- **From a Workflow Queue.** See the Workflow entry in the System Interaction section on page 293 for more information.
- **From a document being created and automatically imported by the Virtual Print Driver.** See the Virtual Print Driver entry in the System Interaction section on page 293 for more information.

If you are at a workstation that has either an Advanced Capture or Ad Hoc Advanced Capture license registered to it, the document will immediately undergo Advanced Capture.

If your workstation is not registered for Advanced Capture or Ad Hoc Advanced Capture, but your OnBase solution is licensed for one or both of these products, the document will be routed to the **Awaiting Advanced Capture** batch status queue and will be processed later, when a license becomes available, or by another workstation licensed for Advanced Capture.

Tip: Unlike batch Advanced Capture, ad hoc Advanced Capture offers no indication of values that were not confidently identified by the Advanced Capture engine (i.e., values that did not meet their configured Suspect Level). It is considered a best practice to always compare the indexing information assigned to the document by the ad hoc Advanced Capture process to the actual information displayed on the document to ensure that it is accurate.

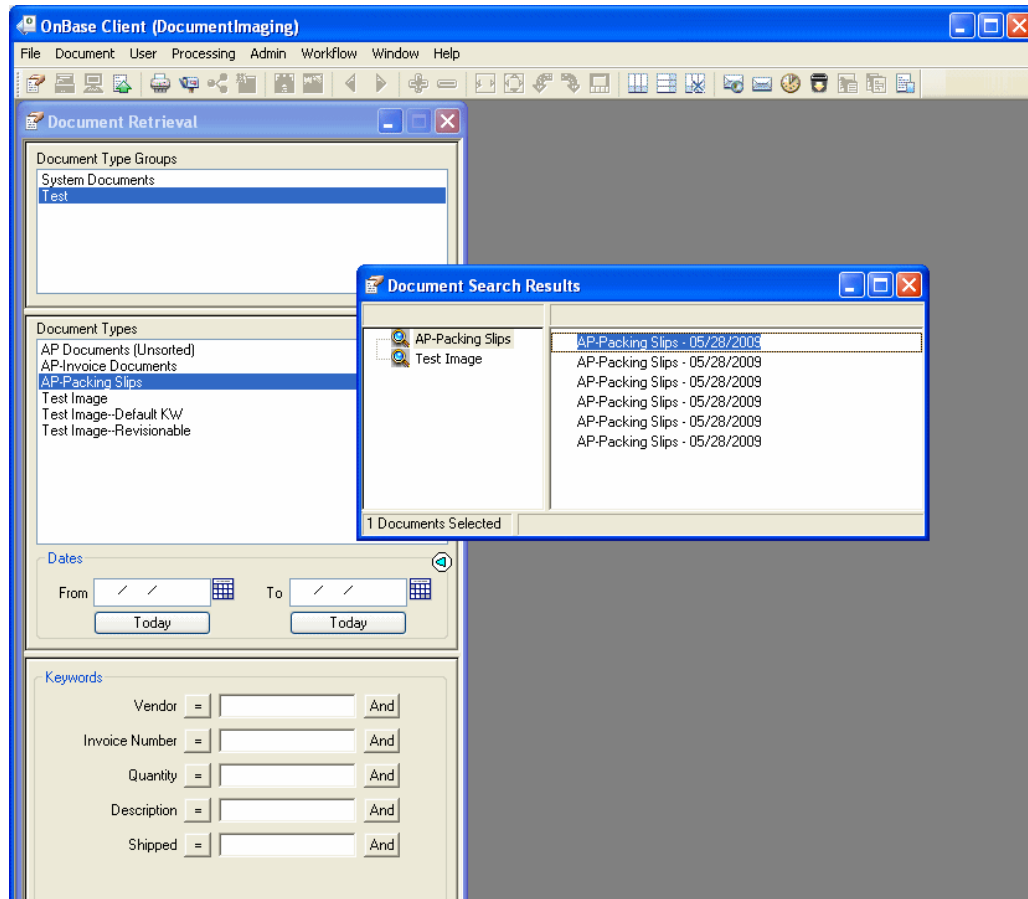
When performing ad hoc Advanced Capture processing, also note the following:

- If you are using ad hoc Advanced Capture to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document.
- If a document causes an ad hoc Advanced Capture process to fail, the document is automatically removed from the batch so that the batch can be reprocessed without repeating the same error.
- When you view the list of documents in a batch, any document containing restricted Keywords or belonging to a Document Type you do not have rights to access is given a generic **RESTRICTED: <Document Type name>** label. You cannot view these restricted documents in the Document Viewer.

From the Document Search Results List

To perform ad hoc Advanced Capture from the **Document Search Results** list:

1. From the OnBase Client, search for a document. The **Document Search Results** list is displayed.



2. Right-click on the document that you would like to perform ad hoc Advanced Capture on and select **Perform Document Advanced Capture....**

The Advanced Capture process is performed.

Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

From an Open Document

To perform ad hoc Advanced Capture on an open document:

1. From the OnBase Client, retrieve and open the document you wish to perform ad hoc Advanced Capture on. The document is displayed in the Document Viewer.

AP-Packing Slips - 05/28/2009

Packing Slip

Invoice No: 3330
Date: March 3, 2008

RECEIVED
MAR 5 4 08

Phone Shack
5546 South Street
Johnstown, PA 15901

To: 9-Second Foods, Inc.
19000 Quikky Circle
Columbia, MD 65201

Ship To: Same

Salesperson	P.O. Number	Date Shipped	Shipped Via	F.O.B. Point	Terms
Joe Smith	558	3/3/08	UPS		2% 15 Net 30

Quantity	Description	Shipped	Received
40	722569 Siemens Black Phones	40	

Quantity Shipped	40
Quantity Received	
Accepted By	
Date Received	

Make all checks payable to: Phone Shack
If you have any questions concerning this Packing Slip, call:
Joe S., 814-555-2234

Thank you for your business!

Test... **Page 1** **1 Page(s)**

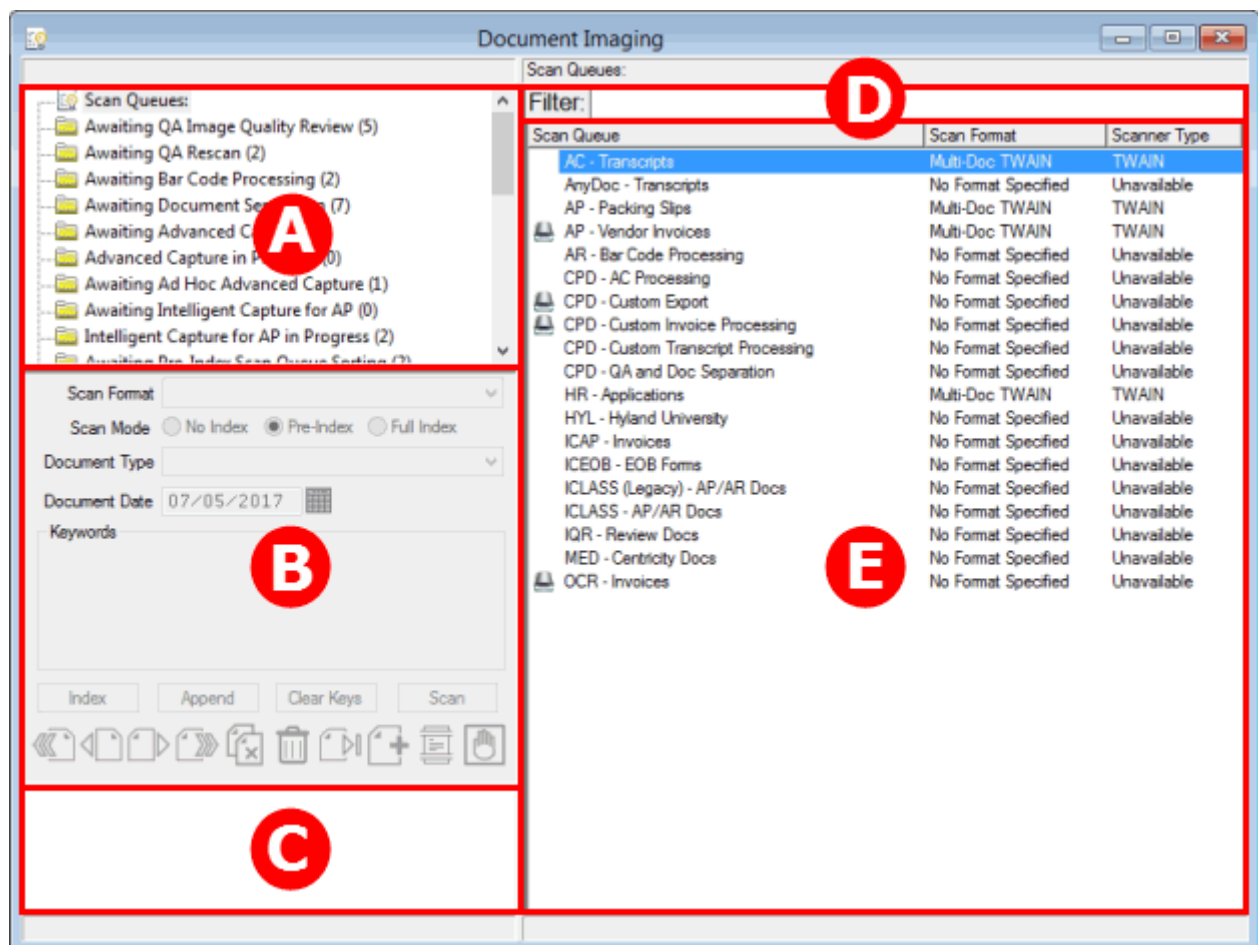
- Right-click on the document and select **Perform Document Advanced Capture....**
The Advanced Capture process is performed.

Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

From the Document Imaging Window

To perform ad hoc Advanced Capture on a document from within the Document Imaging interface:

- From the OnBase Client, click **Processing | Scan/Index**. The **Document Imaging** window is displayed.

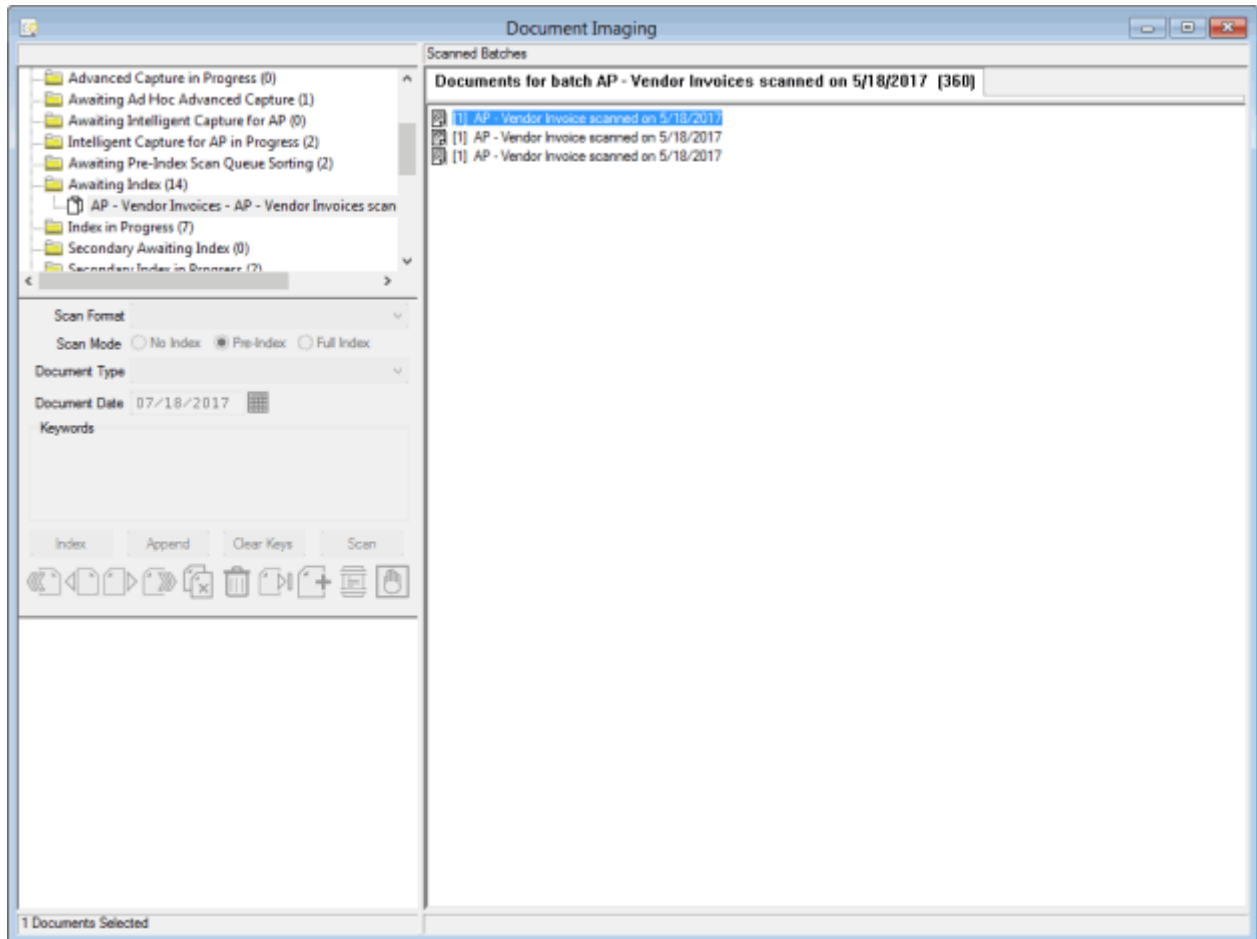


The **Document Imaging** window consists of the following unique sections.

Diagram Section	Section Name	Description
A	Queue List Window	The Queue List window displays entries for Scan Queues , Scheduled Processes , and the various batch status queues to which you have access. Select an entry to view in the Working window the individual scan queues, scheduled processes, or batches belonging to that entry. For more details on the entries in the Queue List window, see the table below.
B	Indexing Dialog Box	The Indexing dialog box contains settings for the document being indexed. These settings relate to scan formats, Document Types, and Keyword Values.
C	Indexing Status Window	The Indexing Status window displays messages regarding scanning and indexing activity. Depending on your system configuration, this information may also be written to a log file. Contact your solution provider for more information.
D	Filter Field	The Filter field allows you to type characters by which to filter the individual scan queues, scheduled processes, or batches that are displayed in the Working window for the entry selected in the Queue List window.
E	Working Window	The Working window displays the individual scan queues, scheduled processes, or batches for the entry selected in the Queue List window.

2. In the Queue List window, select the batch status queue in which the document you wish to index resides. All batches residing in that batch status queue are listed in the working window.

3. In the Working window, double-click the batch containing the document you wish to perform ad hoc Advanced Capture on. All documents residing in that batch are listed in the Working window.



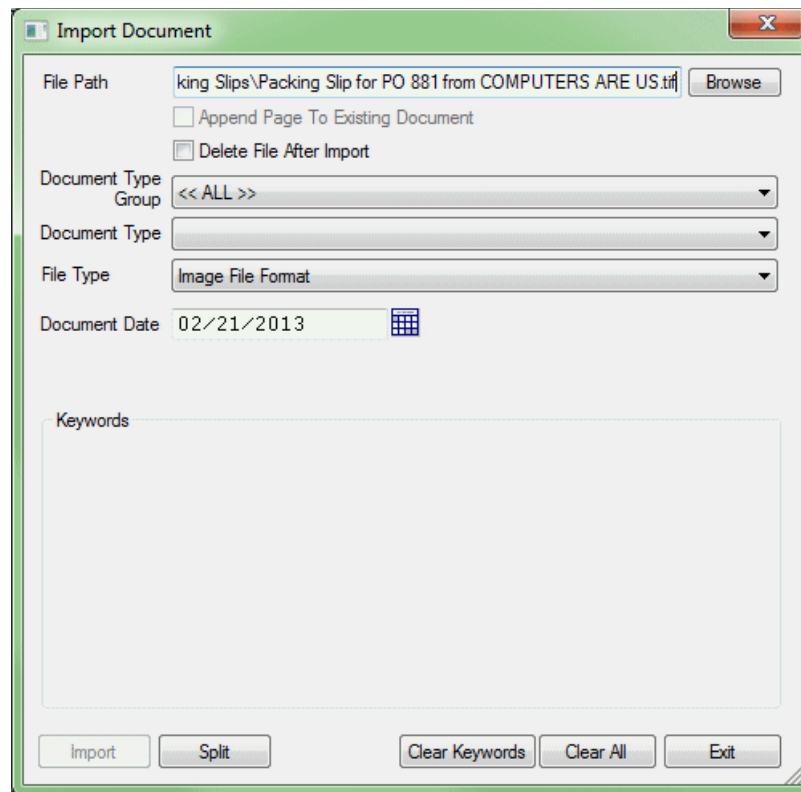
4. In the Working window, select the document you wish to perform ad hoc Advanced Capture on. Right-click and select **Perform Document Advanced Capture....**
The Advanced Capture process is performed. The document and its associated batch remain in the **Awaiting Index** batch status queue.

Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

From the Import Document Dialog Box

To perform ad hoc Advanced Capture from the **Import Document** dialog box:

1. Import a document into OnBase. The **Import Document** dialog box is displayed.



Note: For information on the various methods to ad hoc import documents, see the **Client** module reference guide or help file.

- Using the Document Type drop-down, select the Document Type for the document being imported. Once the Document Type is selected, the **Queue for Advanced Capture** check box is displayed.

- Select the **Queue for Advanced Capture** check box and, if desired, enter any other indexing information for the document.
- Click **Import**.

The Advanced Capture process is performed.

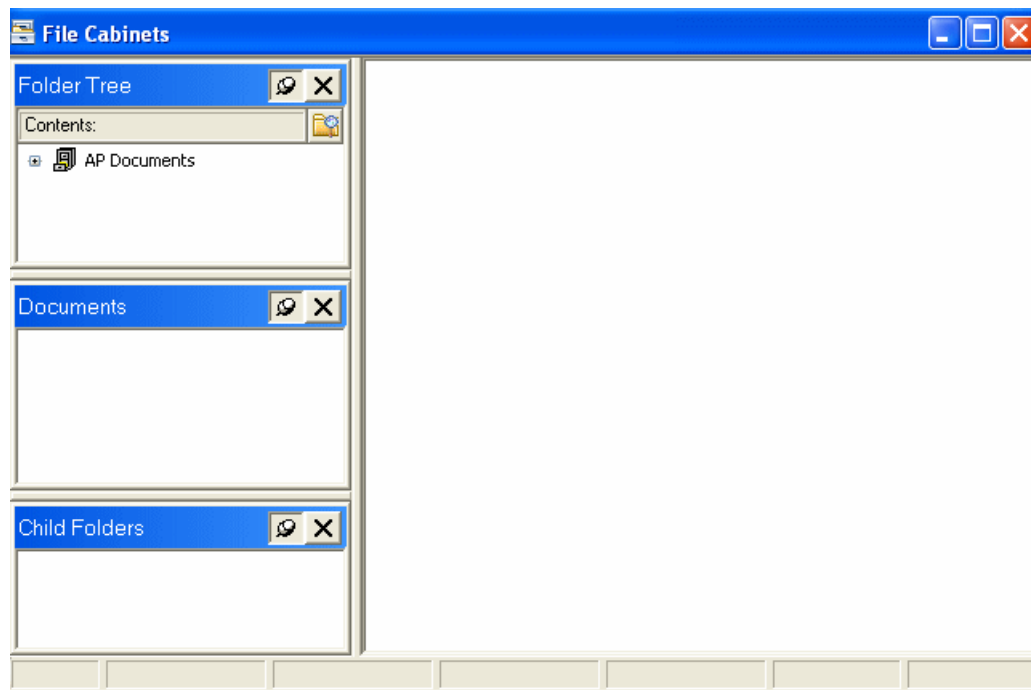
Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

From a Folder

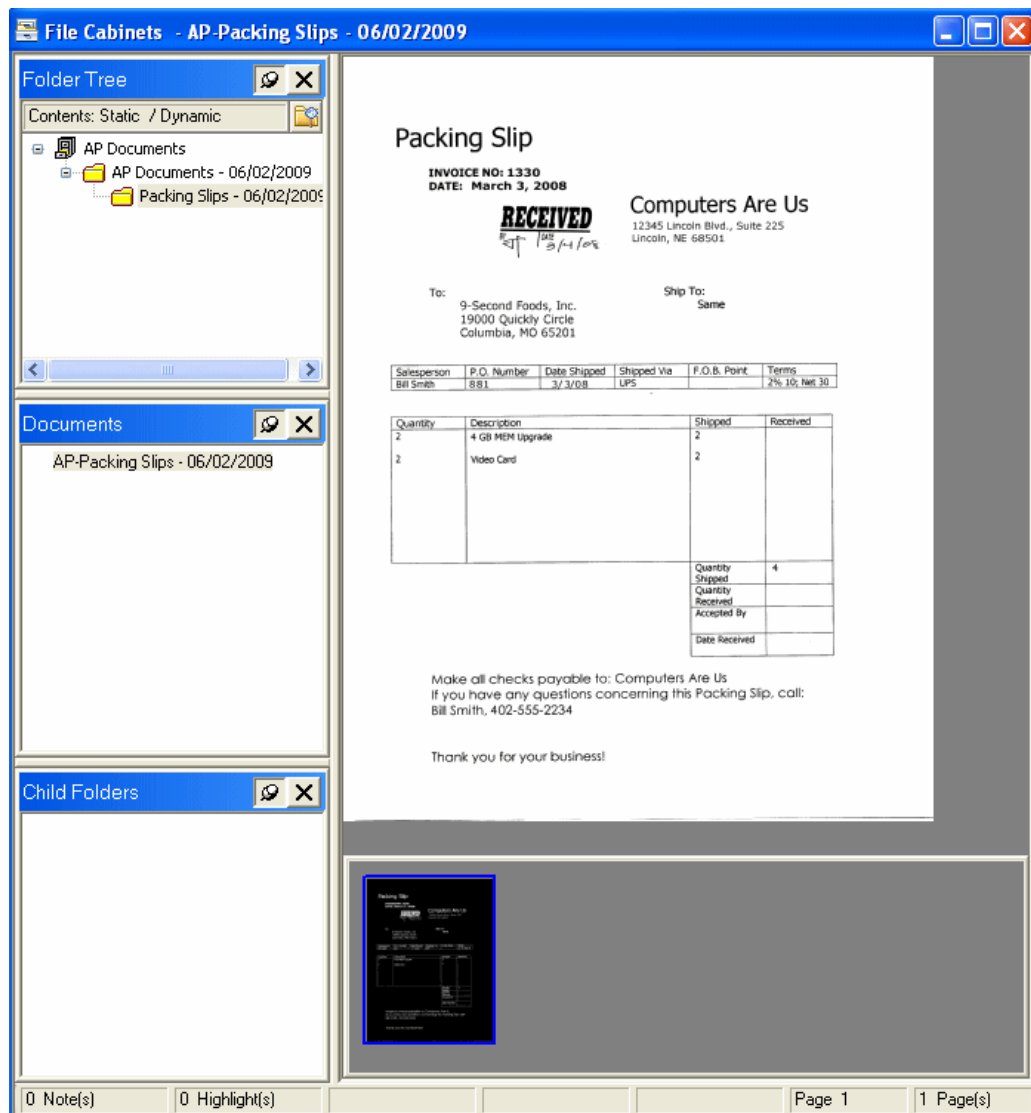
Note: Ensure that Folders have been properly configured for your OnBase solution. For more information, see the **Folders** module reference guide or help files.

To perform ad hoc Advanced Capture from a Folder:

1. From the OnBase Client, open the File Cabinets window by clicking **File | File Cabinets** or by clicking the **Open File Cabinets** toolbar button.



- Expand the folder until you reach the document that you would like to perform ad hoc Advanced Capture on. The document is displayed in the Document Viewer.



3. To perform Advanced Capture on the displayed document, perform one of the following:
 - With the document selected in the Documents pane, right-click within the Documents pane and selected **Perform Document Advanced Capture....**
 - Right-click on the document in the Document Viewer and select **Perform Document Advanced Capture....**

The Advanced Capture process is performed.

Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

From a Workflow Queue

Note: Ensure that your OnBase solution is licensed for Workflow and that a Workflow Life Cycle has been properly configured. For more information, see the **Workflow** module reference guide or help files.

To perform ad hoc Advanced Capture from a Workflow queue, right-click on the document in the Document Viewer and select **Perform Document Advanced Capture....** The Advanced Capture process is performed, with its progress displayed in the status bar.

Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

From the Virtual Print Driver

Note: Ensure that your OnBase solution is licensed for the Virtual Printer Driver and that the OnBase Client has been configured to use the Print Monitor. For more information, see the **Virtual Print Driver** module reference guide.

When the Print Monitor detects that an image document has been printed by the Virtual Print Driver, the **Import from Virtual Printer** dialog box is displayed. Once the Document Type of the document has been selected via the **Document Type** drop-down, the **Queue Document for Advanced Capture** check box is displayed.

To perform ad hoc Advanced Capture, select the **Queue Document for Advanced Capture** check box, enter any additional Keyword Values, and click **Import**. The Advanced Capture process is performed, with its progress displayed in the status bar.

Note: If your workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Advanced Capture** batch status queue for processing. In order to see the **Awaiting Ad Hoc Advanced Capture** batch status queue, as well as process batches residing within it, the workstation must be registered for Advanced Capture.

Performing Batch Advanced Capture

Advanced Capture can be performed on batches of documents within the Document Imaging interface.

Batch Advanced Capture allows you to immediately and automatically perform Advanced Capture on a large number of documents imported into a scan queue via a batch import module, such as Document Imaging, Disconnected Scanning, Document Import Processor, etc., after import.

Tip: When scanning a batch of documents into a scan queue, it is considered a best practice to scan the batch as **Unindexed** or **Pre-Indexed** so that any documents that undergo batch Advanced Capture and are identified as having suspect Keyword Values are sent for user review in either the **Awaiting Index** or **Index in Progress** batch status queue. If the batch is scanned as **Fully Indexed**, batches containing documents that are associated with suspect Keyword Values are not routed for manual review because the batch is already marked as indexed.

One of the benefits of performing batch Advanced Capture is the identification of values that are not confidently identified by the Advanced Capture engine. Each indexing value identified by the Advanced Capture engine is given a score, and if this score does not meet the specified Suspect Level, the value is highlighted and the document is brought to the user's attention.

Note: When a Keyword Type is configured to use a Data Set, a blank Keyword Value is not automatically marked as suspect unless the Keyword Type is required for the Document Type.

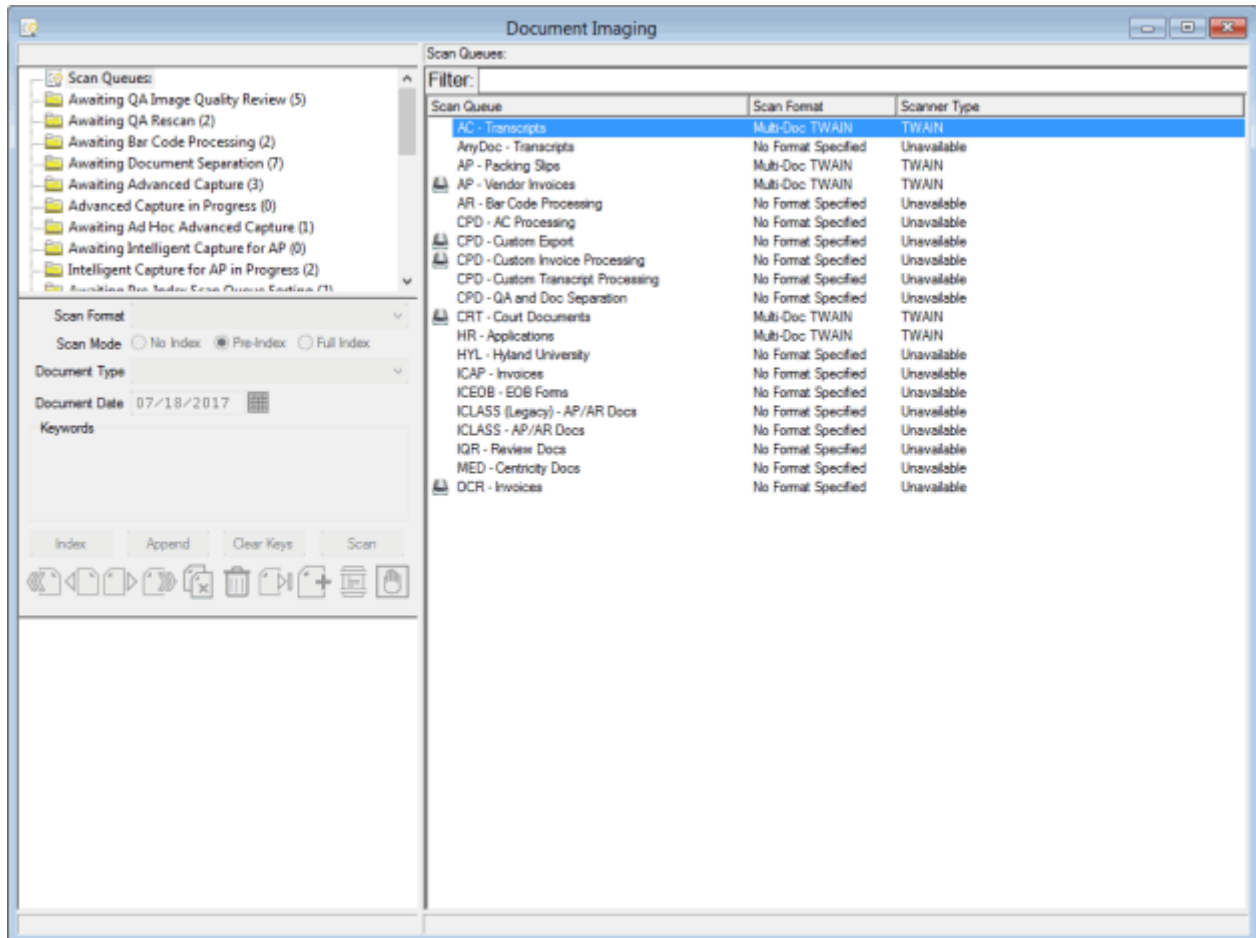
Batch Advanced Capture also allows you to perform Advanced Capture on documents that have been queued for ad hoc Advanced Capture by a workstation not registered for Advanced Capture.

Note: If you are using batch Advanced Capture to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set and multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, either of two actions may occur. Depending on the configuration of the AutoFill Keyword Set, either all instances of the AutoFill Keyword Set are assigned to the document, or the batch is routed to either the **Awaiting Index** or **Index in Progress** batch status queue so that you can select the instance(s) of the AutoFill Keyword Set to assign to the document.

On a Newly Imported Batch

To perform batch Advanced Capture on a newly imported batch of documents:

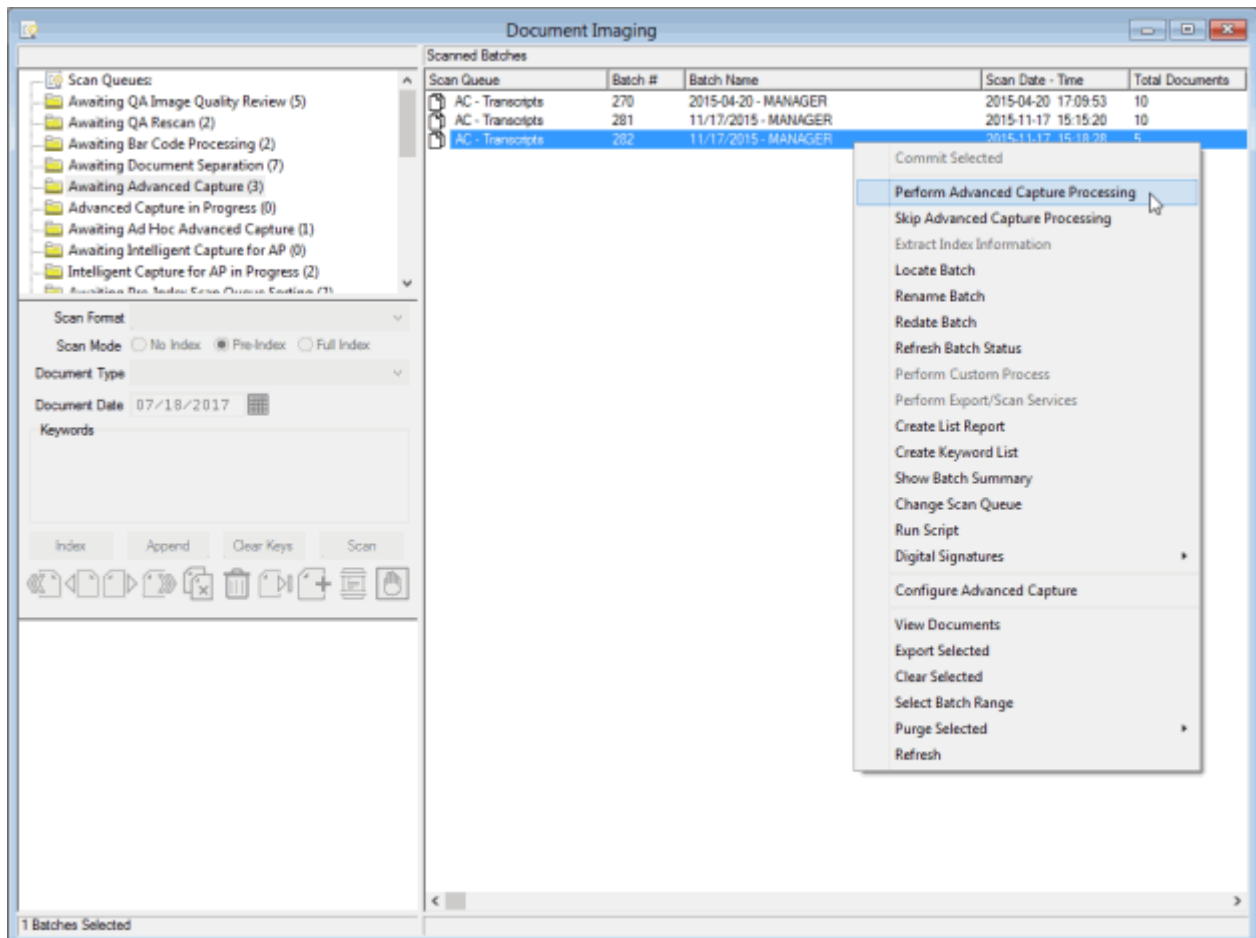
1. Import a batch of documents into OnBase via a scan queue configured to use Advanced Capture.
2. From the OnBase Client, click **Processing | Scan/Index**. The **Document Imaging** window is displayed.



Note: Depending on your scan queue configuration, the batch you imported in Step 1 may have been routed to a batch status queue prior to the **Awaiting Advanced Capture** queue. If necessary, perform any necessary activities to route the batch to the **Awaiting Advanced Capture** batch status queue prior to continuing to Step 3.

3. In the Queue List window, select the **Awaiting Advanced Capture** batch status queue. All batches residing in the queue, including the batch you imported in Step 1, are displayed in the Working window.

4. In the Working window, select the batch on which you would like to perform batch Advanced Capture. Right-click and select **Perform Advanced Capture Processing**.



5. A message is displayed asking you to confirm that you would like to perform batch Advanced Capture on the selected batch. Click **OK**.
The batch Advanced Capture process is performed.

Note: Depending on your configuration, some batches in the **Awaiting Advanced Capture** batch status queue may have been marked for OnBase AnyDoc processing with Infiniworx. If you attempt to perform Advanced Capture on any of these batches, an error message will be displayed. These batches will be picked up by the Data Capture Server and automatically routed to the **Advanced Capture in Progress** batch status queue, where they will eventually be picked up by Infiniworx for processing. For more information on the **Advanced Capture in Progress** batch status queue, see [Resolving In-Progress Batches on page 375](#). For more information on Infiniworx processing, see the **Infiniworx User Guide**.

6. Once the Advanced Capture process is complete, the batch is transitioned out of the **Awaiting Advanced Capture** batch status queue.
 - If the Advanced Capture process was successful, and all values processed by the Advanced Capture engine meet the acceptable Suspect Level, then the batch is transitioned to the next batch status queue configured for the scan queue.
 - If any values processed by the Advanced Capture engine are marked as suspect, or if any forms matched to documents in the batch are configured to also require documents to undergo manual indexing, the batch is transitioned to either the **Awaiting Index** or **Index in Progress** batch status queue for additional indexing or index verification.

Resolving In-Progress Batches

Batches are routed to the **Advanced Capture in Progress** batch status queue in the following scenarios:

- Advanced Capture processing on a batch is interrupted before the processing can finish.
- A batch in the **Awaiting Advanced Capture** batch status queue has been configured for OnBase AnyDoc processing with Infiniworx.

If a batch has been marked for OnBase AnyDoc processing with Infiniworx, it will eventually be picked up by Infiniworx for processing. While residing in the **Advanced Capture in Progress** batch status queue, such a batch is locked; it cannot be manually routed out of this queue. For more information on Infiniworx processing, see the **Infiniworx User Guide**.

If a batch has partially undergone Advanced Capture processing before being interrupted, however, you have the following options to manually route the batch out of the **Advanced Capture in Progress** batch status queue:

- To forgo the originally intended Advanced Capture processing, right-click on the batch and select **Skip Advanced Capture Processing**. The batch is routed to the next batch status queue configured for the scan queue.
- To redo the originally intended Advanced Capture processing, right-click on the batch and select **Move to Awaiting Advanced Capture**. The batch is routed back to the **Awaiting Advanced Capture** batch status queue, where it can then undergo Advanced Capture from the beginning of the process.

Index Verification

Index verification must be performed if indexing values for documents that undergo batch Advanced Capture do not meet their configured Suspect Levels during processing.

If at least one value on one document in a batch is determined to be suspect, then the entire batch is moved to either the **Awaiting Index** or **Index in Progress** batch status queue for index verification, according to the following criteria:

- If at least one value on all documents in a batch is determined to be suspect, all documents remain unindexed and the batch is routed to the **Awaiting Index** batch status queue.
- If some documents in a batch are fully indexed through Advanced Capture processing (that is, no values are determined to be suspect on at least one document in the batch) and other documents in this batch remain unindexed after Advanced Capture processing (that is, at least one value is determined to be suspect on at least one document in the batch), then the entire batch is routed to the **Index in Progress** batch status queue.

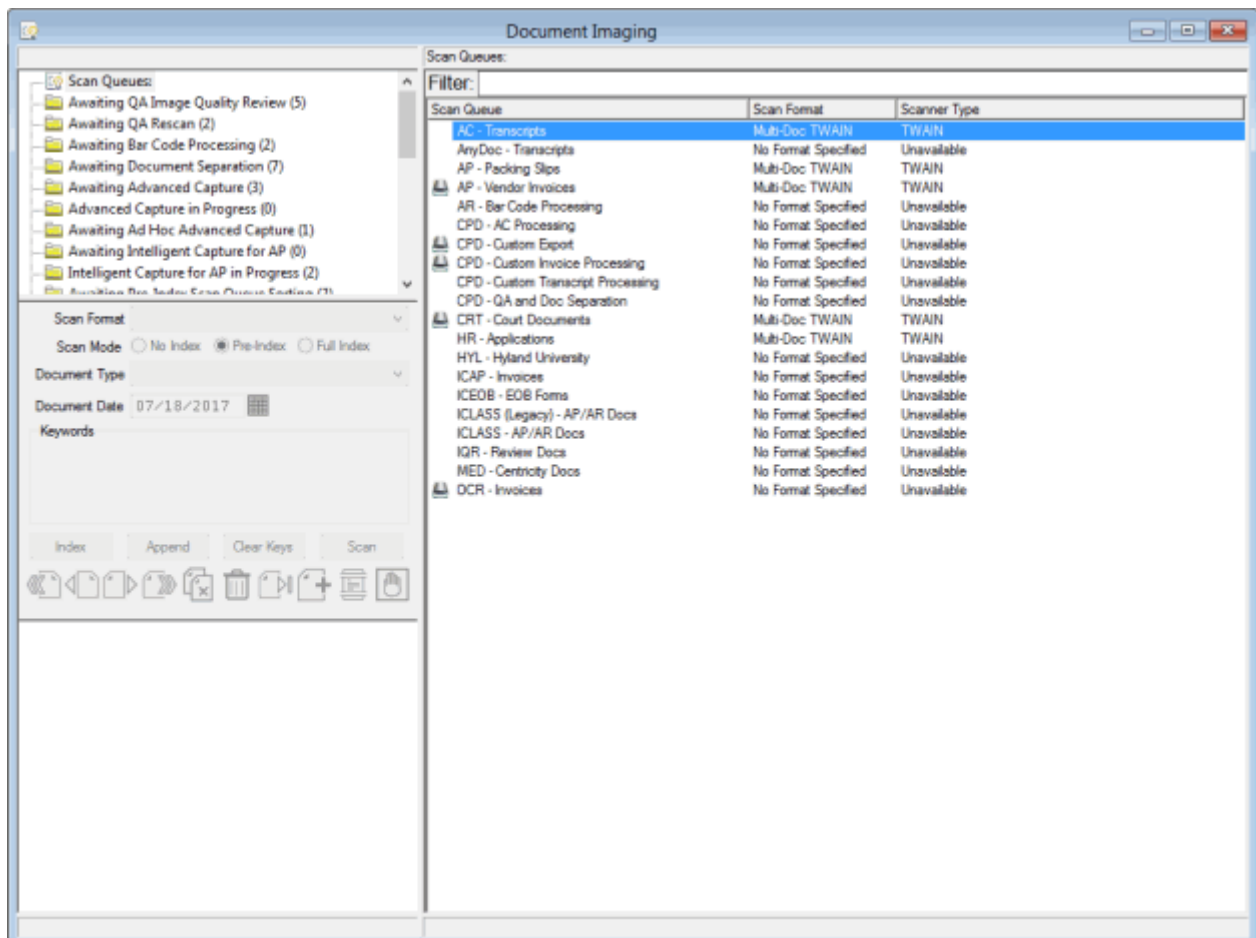
When performing index verification, please note the following:

- Index verification is only available for batch Advanced Capture. When performing ad hoc Advanced Capture on a document, no indication is given if any detected values are suspect.
- While a batch that has undergone batch Advanced Capture can be reviewed, corrected, or indexed further in the OnBase Web Client, the index verification features available in the OnBase Client described below (for example, suspect Keyword Values being highlighted in red) are not available in the OnBase Web Client.

- The index verification features available in the OnBase Client described below (e.g., suspect Keyword Values being highlighted in red) are also available in the OnBase Unity Client. When using the Unity Client to perform index verification, you must also consider the following:
 - Whenever new Advanced Capture forms are created or zones on existing forms are changed, the Unity Client must be restarted to reflect the changes.
 - Appropriate color contrasts should be used so that clear distinctions can be made between Keyword Type names, Keyword Value text, and color highlights in both the standard Keyword panel and any Unity Form panel used for indexing. If color contrasts are not clearly distinguishable, contact your system administrator.
 - Unity Forms cannot display multiple Keyword Values for the same Keyword Type if the Keyword Type does not belong to a Multi-Instance Keyword Type Group.

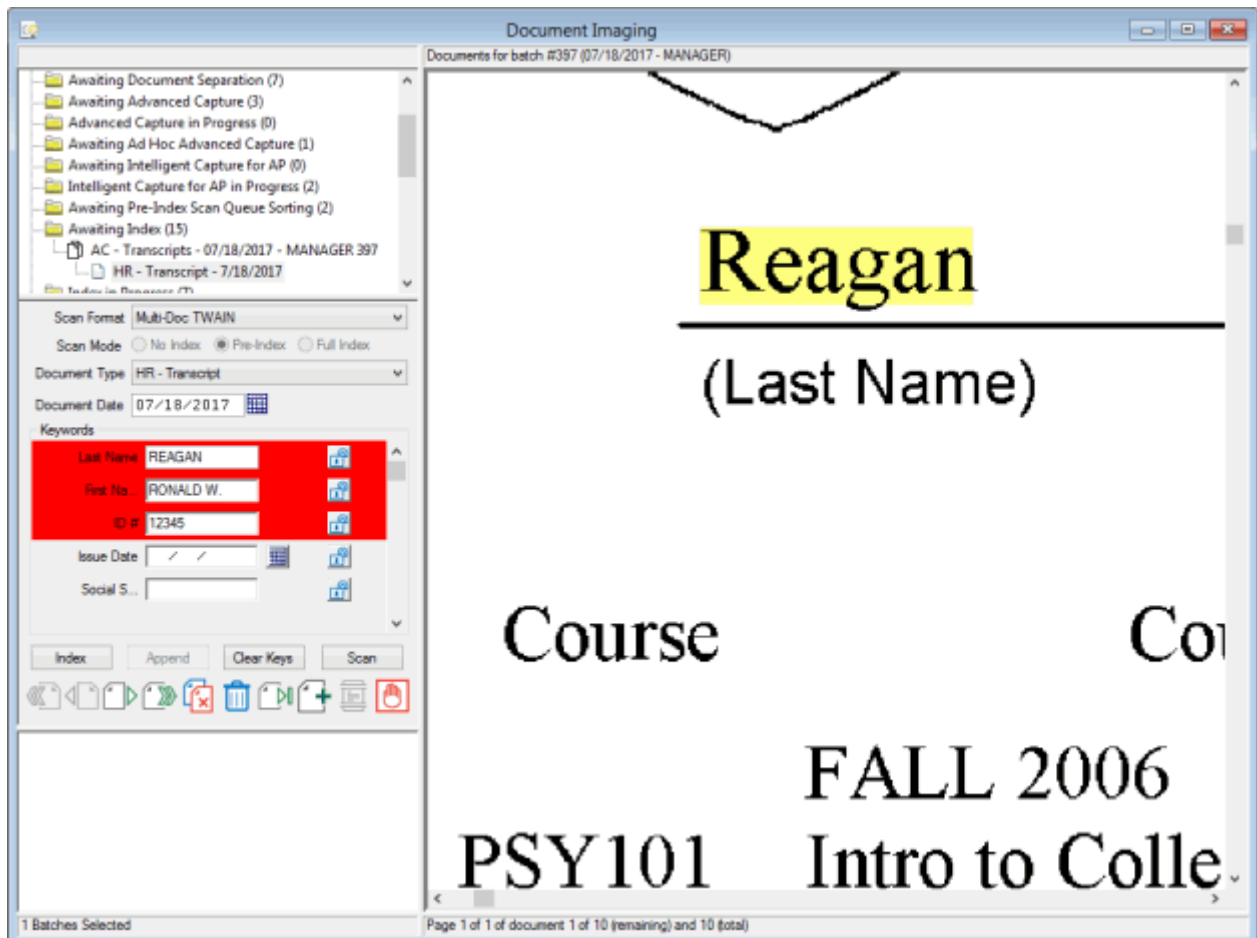
To perform index verification:

- From the OnBase Client, click **Processing | Scan/Index**. The **Document Imaging** window is displayed.



- In the Queue List window, select the **Awaiting Index** or **Index in Progress** batch status queue. All batches residing in the queue are displayed in the Working window.

3. In the Queue List window, select the batch that you would like to perform index verification on. Right-click and select **Index Documents**.



Note: When performing index verification, only documents in the batch that have a suspect value or that were left unindexed will be displayed.

The first document in the batch is displayed in the Working window and the **Indexing** dialog box is populated with its indexing information. Indexing values that were marked as suspect are highlighted in red by default in the **Indexing** dialog box. Depending on your configuration, some indexing values may be highlighted in different colors.

Note: If a Keyword Type name is displayed in white, it is a required Keyword. While required Keywords are normally displayed in red, they are displayed in white when their values have been marked as suspect so that the Keyword Type name does not blend in to the red highlight.

Note: If a Keyword Value is displayed for review, but is not highlighted in red, the Keyword Value is the primary Keyword Value of more than one instance of an AutoFill Keyword Set. Once you review and accept the Keyword Value, you are asked to select the correct instance, or

instances, of the AutoFill Keyword Set to assign to the document.

-
4. Click in the first suspect Document Date or Keyword Value field in the **Indexing** dialog box. Depending on your configuration, the area of the document that the value was extracted from is displayed in detail in the Working window and is highlighted in yellow, and the value recognized by the Advanced Capture engine is displayed in detail above the Keyword Type field.
-

Note: The detailed views of the extracted value in the Working window and above the corresponding Keyword Type field are only supported on image documents; they are not supported on PDF documents.

5. Compare the Document Date value or Keyword Value to the value displayed on the document.
If the value is correct, navigate to the next suspect value on the document, and repeat Step 5. If no suspect values remain on the current document, click **Index** to display the next document in the batch that has at least one suspect value.
If the Document Date or Keyword Value is incorrect, manually enter the correct value in the field or extract it from the document using Point and Shoot Indexing.
Once the value has been corrected, navigate to the next suspect value on the document and repeat Step 5. If no suspect values remain on the current document, click **Index** to display the next document in the batch that has at least one suspect value.
6. Once all documents in the batch requiring index verification have been verified, the batch is automatically routed to the next configured batch status queue for the scan queue.

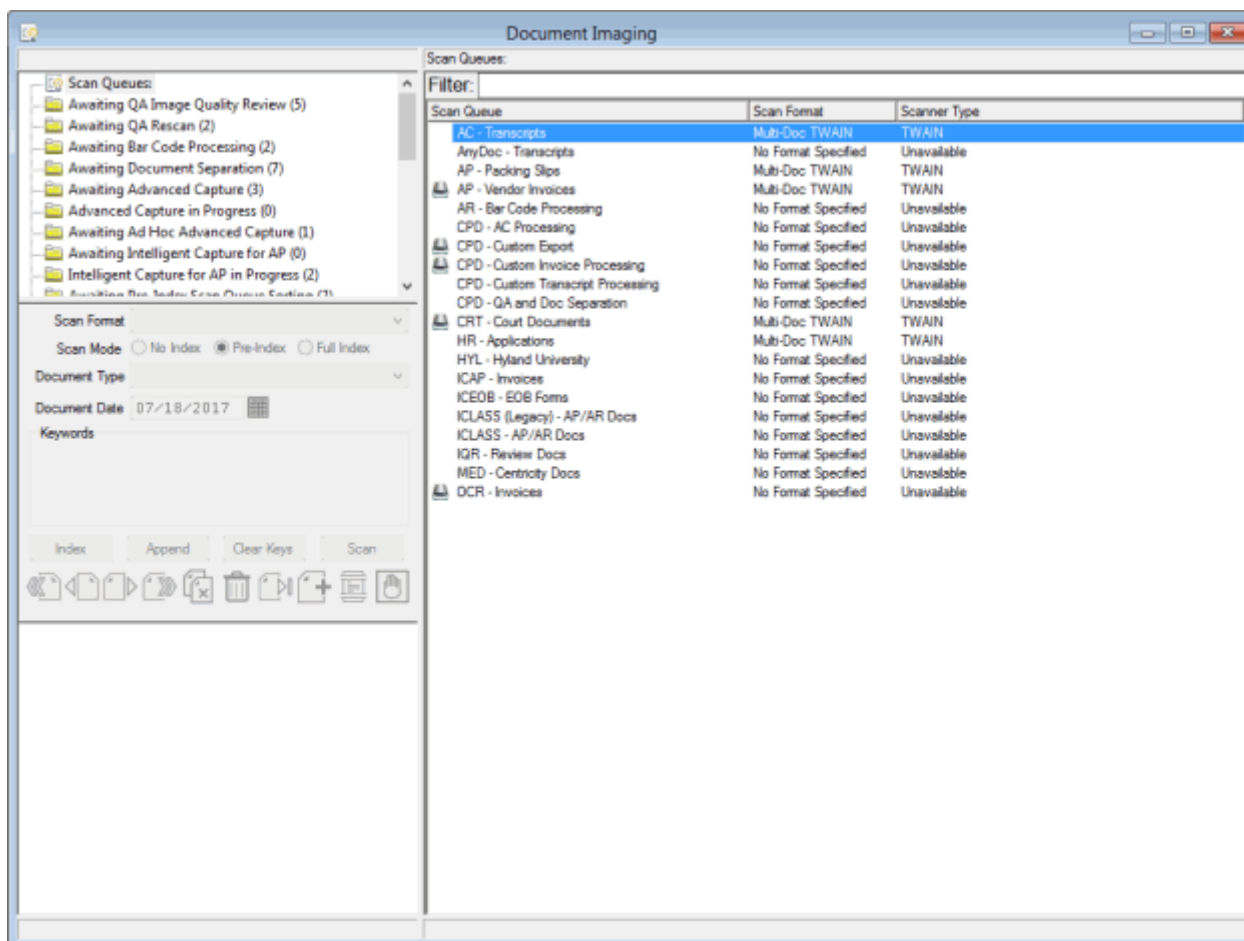
On Documents Queued for Ad Hoc Advanced Capture

A workstation connected to an OnBase solution licensed for Advanced Capture or Ad Hoc Advanced Capture has the ability to queue individual documents for Advanced Capture if the workstation is not registered for Ad Hoc Advanced Capture or Advanced Capture and if the user has been granted the necessary product right.

These queued documents reside in the **Awaiting Ad Hoc Advanced Capture** batch status queue. They can be processed by a workstation registered for Advanced Capture.

To perform Advanced Capture on documents that have been queued for ad hoc Advanced Capture:

1. From the OnBase Client, ensure that the workstation is registered for Advanced Capture and then click **Processing | Scan/Index**. The **Document Imaging** window is displayed.

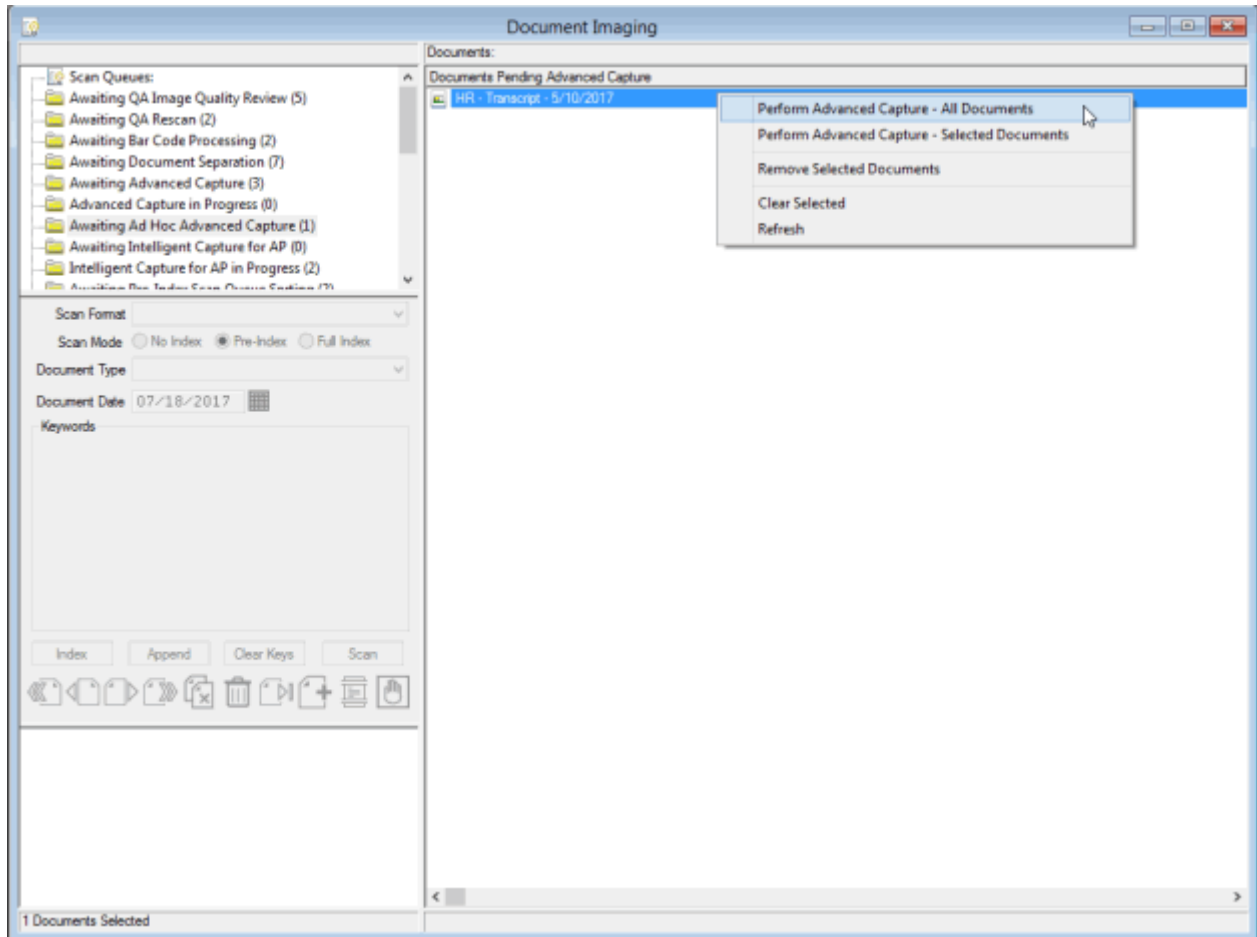


2. In the Queue List window, select the **Awaiting Ad Hoc Advanced Capture** batch status queue. All documents residing in the queue are displayed in the Working window.

Note: Despite the fact that they reside in a Document Imaging batch status queue, be aware that these are individual documents that are not affiliated with a scan queue. No scan queue processing options will apply to the documents that undergo ad hoc Advanced Capture.

3. In the Working window, select the document that you would like to perform Advanced Capture on. Right-click and select **Perform Advanced Capture Processing - Selected Documents**.

To perform Advanced Capture on all documents residing in the batch status queue, right-click and select **Perform Advanced Capture Processing - All Documents**.



4. A message is displayed asking you to confirm that you would like to perform Advanced Capture on the selected document(s). Click **Yes**.
The Advanced Capture process is performed.
5. Once the process is complete, the documents are removed from the **Awaiting Ad Hoc Advanced Capture** batch status queue.

Tip: Processing documents in the **Awaiting Ad Hoc Advanced Capture** batch status queue is considered ad hoc Advanced Capture, despite the fact that an Advanced Capture license is required. Ad hoc Advanced Capture offers no indication for values that did not meet the configured Suspect Level. It is considered a best practice to always compare the indexing information assigned to the document by the ad hoc Advanced Capture process to the actual information displayed on the document to ensure that it is accurate.

Keyword Considerations

Consider the following when performing Advanced Capture processing and index verification.

Keyword Type Masking

Advanced Capture attempts to store an extracted value for a Keyword Type even if the Keyword Type is configured for masking and the value is not valid for the mask. Depending on which client you are using for indexing, this invalid value may not be displayed or may be displayed as a truncated value.

If you are using Advanced Capture to extract values for a masked Keyword Type, ensure that the extracted value is valid for the configured mask.

Default Keyword Values

If a Default Keyword Value is configured for a Keyword Type on a document, and no other Keyword Value is assigned to that Keyword Type in the course of Advanced Capture processing (that is, a value is applied through Data Field Zone processing, an Advanced Capture form default, or manual pre-scan indexing), then the Default Keyword Value is automatically applied to the document. If Advanced Capture processing does assign a value to that Keyword Type, then the Default Keyword Value is not applied.

For more information on configuring Default Keyword Values, see the **System Administration** module reference guide.

Point and Shoot Indexing

Point and Shoot Indexing allows you to extract Keyword Values from any image document during indexing or re-indexing from within the OnBase Client's Document Imaging interface.

Point and Shoot Indexing is not specific to Advanced Capture solutions. It can be performed by any user with rights to index documents on a workstation licensed for any of the following:

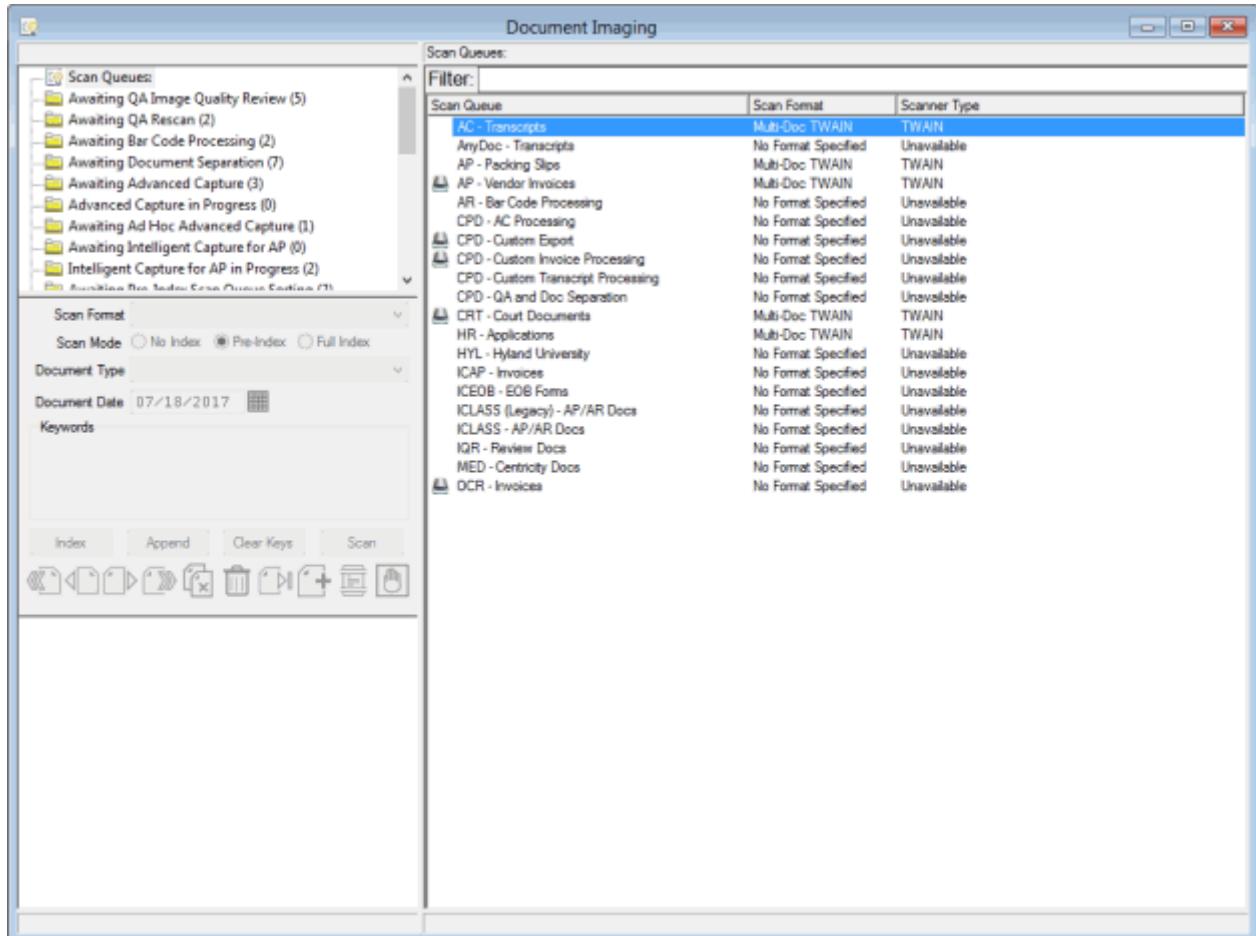
- Advanced Capture
- Ad-Hoc Advanced Capture
- Intelligent Capture for AP
- Interactive Data Capture

Note: While a batch that has undergone batch Advanced Capture can be reviewed, corrected, or indexed further in the Web Client or the Unity Client, Point and Shoot Indexing can only be performed in the OnBase Client. However, additional indexing features in the Unity Client are available through Interactive Data Capture. For more information on these additional indexing features, see the **Interactive Data Capture** module reference guide.

Performing Point and Shoot Indexing

To perform Point and Shoot Indexing:

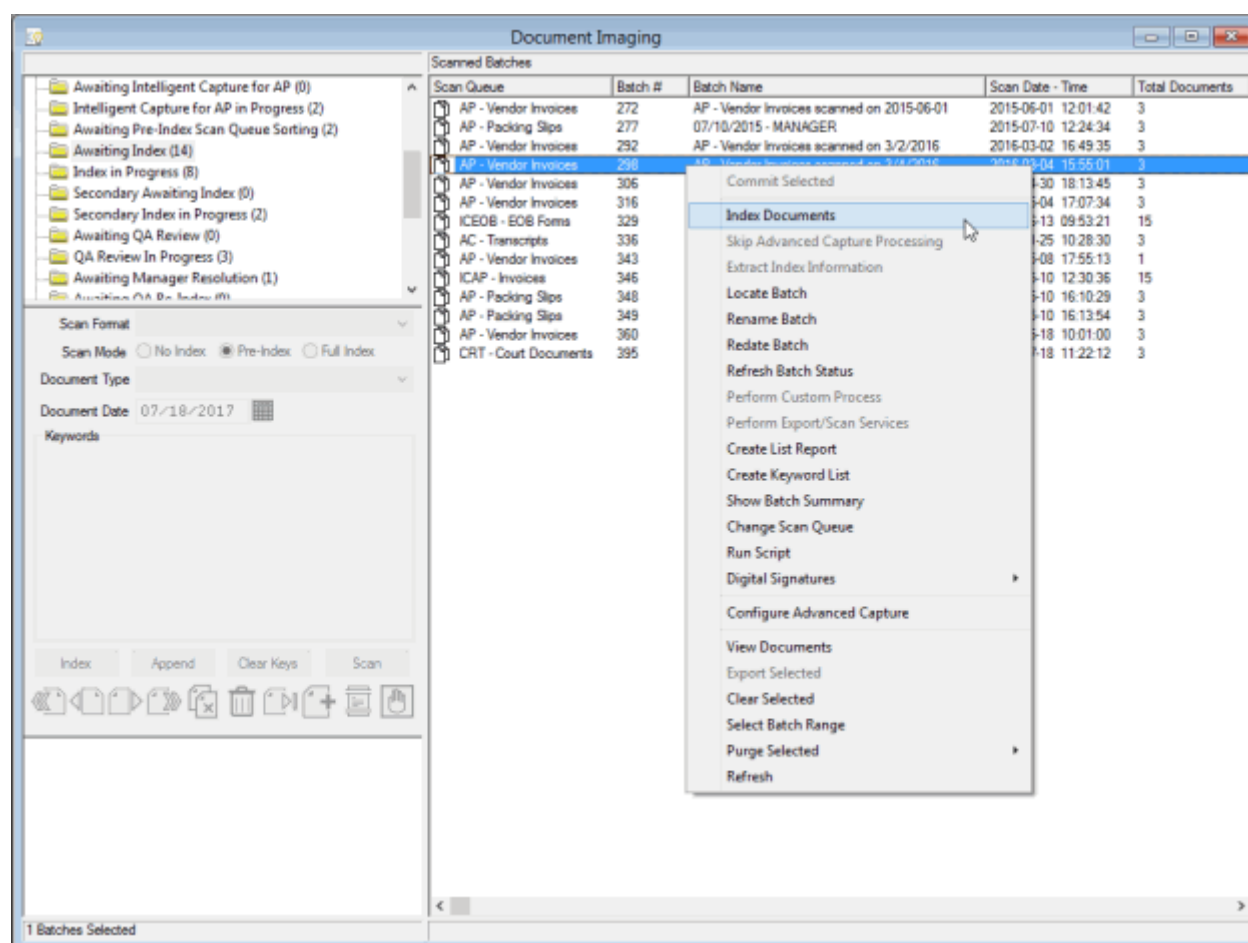
1. From the OnBase Client module, click **Processing | Scan/Index**. The **Document Imaging** window is displayed.



2. In the Queue List window, select an indexing or re-indexing batch status queue (i.e., **Awaiting Index, Index in Progress, Awaiting Re-Index, Re-Index in Progress, Awaiting QA Re-Index**, etc.). All batches residing in the selected batch status queue are displayed in the Working window.

Note: The following example demonstrates how to perform Point and Shoot Indexing on an invoice document residing in the **Awaiting Indexing** batch status queue. Depending on the batch status queue you are using, some menu options, button names, and types of documents may be different. Consult the **Document Imaging** module reference guide or help files for more information.

3. Right-click on a batch in the Working window and select **Index Documents**.

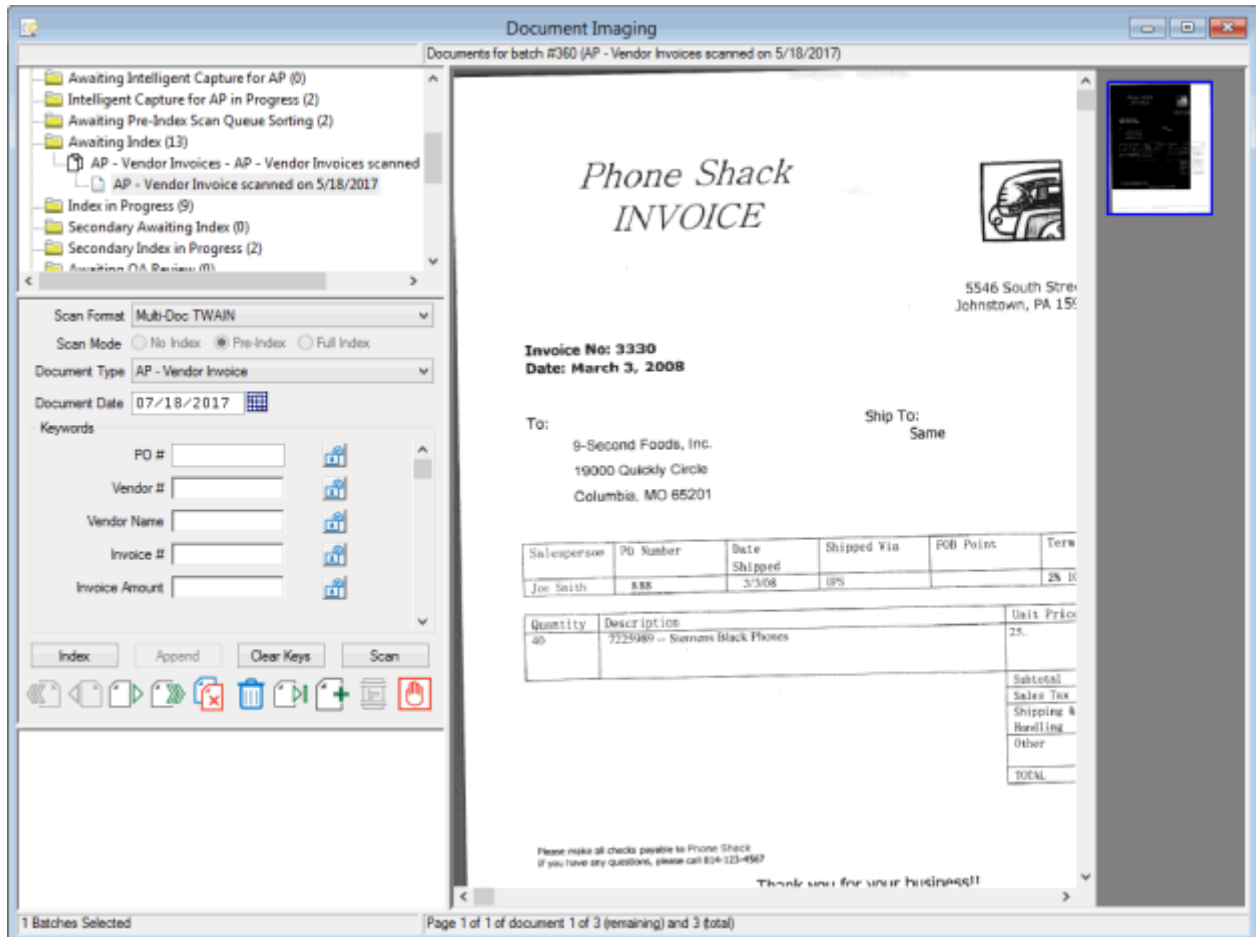


The first document in the batch is displayed in the Working window.

- If a Document Type has not already been selected for the displayed document, select one from the Document Type drop-down for the displayed document.

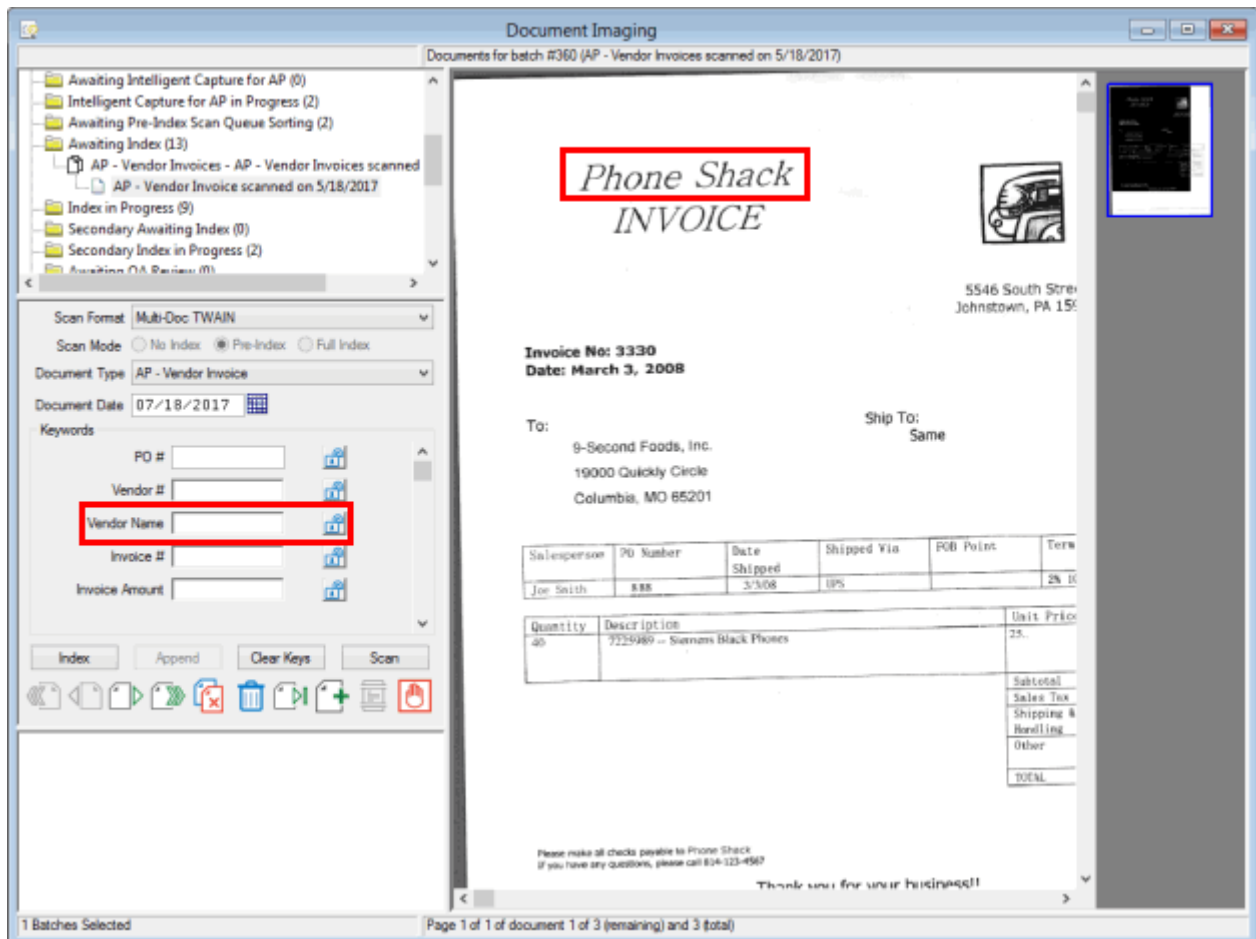
Once a Document Type is selected, the **Indexing** dialog box is populated with the Keyword Types associated with the Document Type.

For illustration purposes, the following example demonstrates the selection of a Document Type for an invoice, but your solution may involve other types of documents.



- In the **Indexing** dialog box, set the focus to the field, either the Document Date field or one of the Keyword Type fields, that you would like to extract a value for.

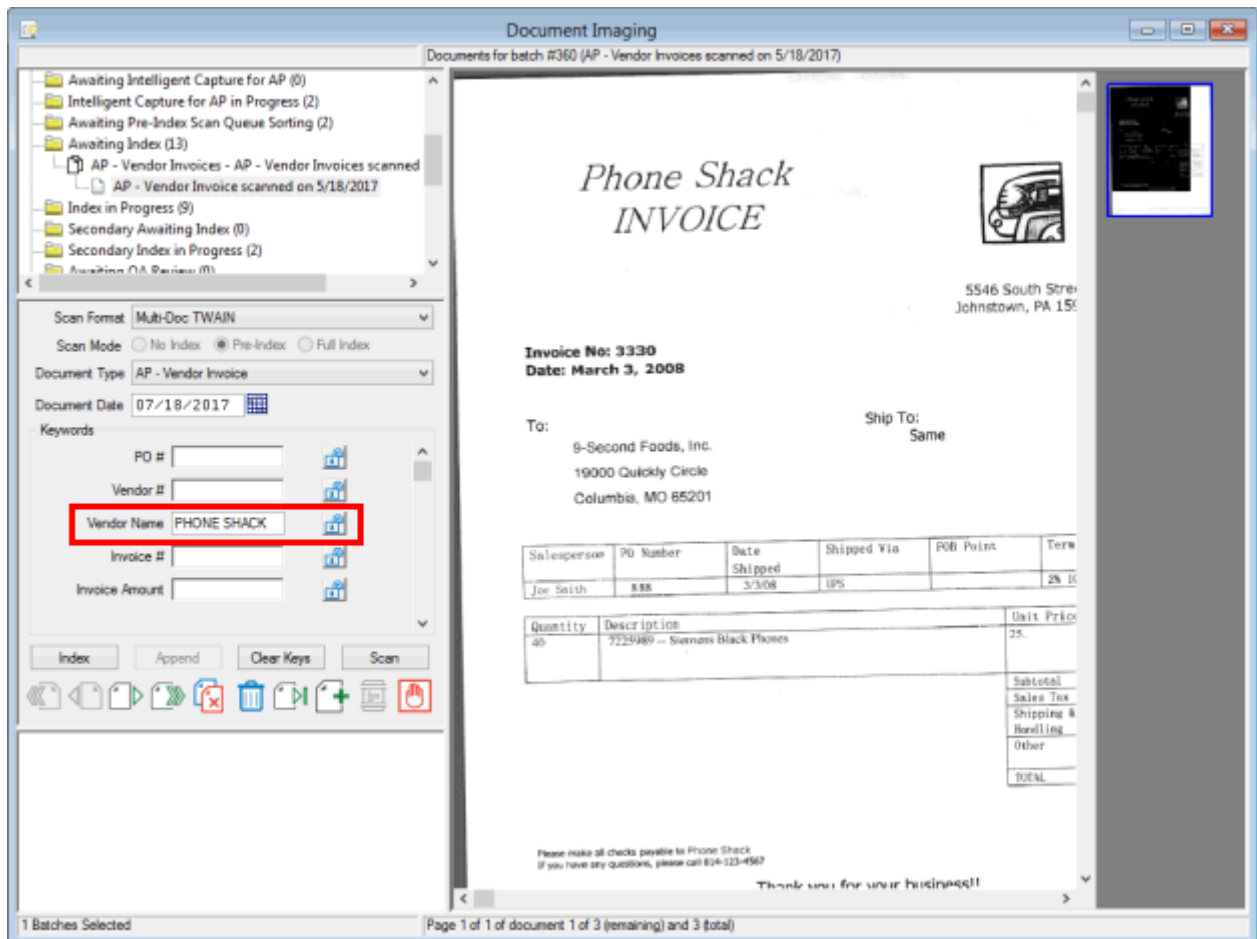
- In the Working window, press and hold the **Shift** key and click and hold the left mouse button while using the pointer to draw a box around the value that is to be extracted. For illustration purposes, the following example identifies the Vendor Name on an invoice, but your solution may involve other kinds of documents or data.



- Release the **Shift** key and the left mouse button. There may be a brief pause while the OCR engine processes the area you selected.

8. The value returned from the selected area is automatically populated in the selected field. Note that the focus is still set to the field after the value is populated.

For illustration purposes, the following example identifies the Vendor Name on an invoice, but your solution may involve other kinds of documents or data.



Note: If you are using Point and Shoot Indexing to extract a Keyword Value that is the primary value for an AutoFill Keyword Set, be aware that the AutoFill Keyword Set is not triggered until you set the focus to another Keyword Type field.

9. Repeat Steps 5 through 8 for each value that you would like to extract for the currently displayed document.

Tip: It is considered a best practice to always compare the values returned from the OCR engine to the actual values displayed on the document when performing Point and Shoot Indexing.

10. Once you are finished indexing the current document, click **Index**. The next document in the batch is displayed.

Repeat Steps 4 through 9 for the remaining documents in the batch. Once all documents in the batch have been indexed, the batch is automatically routed to the next batch status queue configured for the scan queue.

A Note About Single and Double-Byte Characters

If your Advanced Capture solution is configured to identify Latin characters and/or Arabic numerals (i.e., single-byte characters such as those found in English, French, German, etc.) and you wish to use Point and Shoot Indexing to identify Asian characters on an image, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the point to draw a box around the value to be extracted.

Likewise, if your Advanced Capture solution is configured to identify Asian characters (i.e., double-byte characters such as those found in Chinese and Japanese) and you wish to use Point and Shoot Indexing to identify Latin characters and/or Arabic numerals, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the point to draw a box around the value to be extracted.

The language, or languages, that are identified by your Advanced Capture solution are set when configuring an OCR format. Multiple OCR Formats can be created, but only one can be assigned per Document Type.