CalSAWS | Information Transmittal (CIT)　　　0063-22

☒ **CalSAWS M&E**　　　☐ **CalWIN Migration**

Distribution Date:	March 8, 2022
To:	PPOC.All; Committee.Imaging.All; Consortium.RegionalManagers.All
CIT Name:	**CalSAWS Imaging Optical Character Recognition (OCR) Updates**
From:	CalSAWS Project

PPOCs, please forward to the appropriate impacted staff in your county:

☐ General　　　　　　　　　　　　　　　☐ Reports
☐ Policy　　　　　　　　　　　　　　　　　☐ Fiscal
　☐ CW　　　　　　　　　　　　　　　　　☐ Caseload Movement
　☐ CF　　　　　　　　　　　　　　　　　☐ Management
　☐ MC　　　　　　　　　　　　　　　☐ Fiscal
　☐ CMSP　　　　　　　　　　　　　　☐ Security
　☐ FC/KG/AAP　　　　　　　　　　　☐ Batch and Interfaces
　☐ Child Care　　　　　　　　　　　☒ Imaging
　☐ WtW　　　　　　　　　　　　　　☐ Migration
　☐ Other Program(s)_____　☐ Conversion
☐ BenefitsCal　　　☐ Your Benefits Now!　☐ Technical
☐ Customer Correspondence　　　　　☐ Training
☐ Other_____　　　　　　☐ Help Desk

Description: (Including any step-by-step instructions)	**Purpose** The purpose of this CIT is to provide the CalSAWS counties with information on the new functionality that will be implemented with CA-228869 Create OCR Override and OCR Split Override. **Background** New functionality for the CalSAWS Imaging Solution will be implemented with CA-228869, which is currently targeted for release on **March 16, 2022**. This SCR includes three key changes to give users more control over how Optical Character Recognition (OCR) processes documents. • **OCR Bypass flag:** This is a new flag that can be set at the point of capture in QA & Indexing or before submitting a document from the No Case workflow queue. When it is set to "True," the document will not be processed by OCR. When documents are scanned, this flag will default to "False." The flag applies to the document and does not apply to a batch. It will display for the following capture modes: 　o　Single Case 　o　Multi Case 　o　Returned Mail 　o　Ignore Barcode

If the document is time-sensitive, the barcode is not read at the point of scan, and the OCR Bypass flag is set to "True," users will need to manually enter the barcode before Submitting a document to ensure the CalSAWS Customer Reporting page updates. The Imaging Solution will not attempt to read or insert a barcode number after a document is submitted from QA and Indexing.

- **OCR Split Override flag**: This is a new flag that can be set at the point of capture in QA & Indexing or before submitting a document from the No Case workflow queue. When it is set to "True," it prevents OCR from splitting a document.

    When documents are scanned, this flag will default to "True." The flag applies to the document and does not apply to a batch. It will display for the following capture modes:
    - Single Case
    - Multi Case
    - Returned Mail
    - Ignore Barcode

- **New logic if the confidence threshold is not met**: If OCR cannot confidently index a document, the form name will default to the value a user selected in QA & Indexing. These documents will not go to the Exception queue. If the form name is <Pre-OCR> or Unknown, the document will go to the Exception queue.

The OCR Bypass and OCR Split Override flags are available to users with the Imaging Capture security right.

Users can set the OCR Bypass and OCR Split Override flags for an entire batch of documents before starting a scan or file upload using Single Case capture mode.

When the flags are set before clicking on "Start," the flag will apply to all documents scanned.

Capture and Indexing ⚙

* **Capture Profile**

Single Case Scan ▾

* **Applicable Date**

02/25/2022 📅

* **Received Date**

02/25/2022 📅

Bundle ID

* **Origin**

36 - San Bernardino - County ▾

OCRBypass

False ▾

OCRSplitOverride

True ▾

Cancel   **Start**

Users can manually change the flags per document after opening the batch in QA & Indexing.

The flags can manually be set per document for all other capture modes. Note, Single Case, Multi Case, Returned Mail, and Ignore Barcode remain the only capture modes that use OCR.

Additional Information

The following are the practices previously recommended for OCR success that still apply. A new recommendation to correct/enter the barcode in QA & Indexing has been added to this list.

Add Imaging form numbers to county forms

OCR will attempt to read the form number in the bottom left of a document. Counties may incorporate the Form Numbers on the Imaging Form Name Matrix into their forms so OCR can leverage the printed numbers to assist with classification.

Location: The last line on the bottom left, with no text to the immediate right of the form number

Font: Arial, Times New Roman, or Courier

Font size: 8,10, or 12

Text direction: Horizontal

Margins at minimum: Left .3 in, bottom .3 in (make sure to consider your county's scanner margins)

Any combination of the above requirements can be used. In general, a blank line between the imaging form number and other information on the page is recommended.

Handwritten form numbers may be used but may not be reliably read by OCR.

Set form name at point of scan

In QA & Indexing, users may enter a form name at the point of scan. Selecting a form name adds weight to the selection, meaning if OCR is trying to decide between two names, it will lean toward the name the user selected. Even if a form

name is selected, it is possible that OCR will override the user-selected value if system confidence threshold is met.

All forms (excluding flyers or brochures) generated from CalSAWS have an exact match in Imaging. The exact match should be used in most situations, unless dictated otherwise by County policy.

The Imaging Form Names Matrix lists all the Imaging Form Names and their associated document properties. The matrix is posted here: ██████████
██
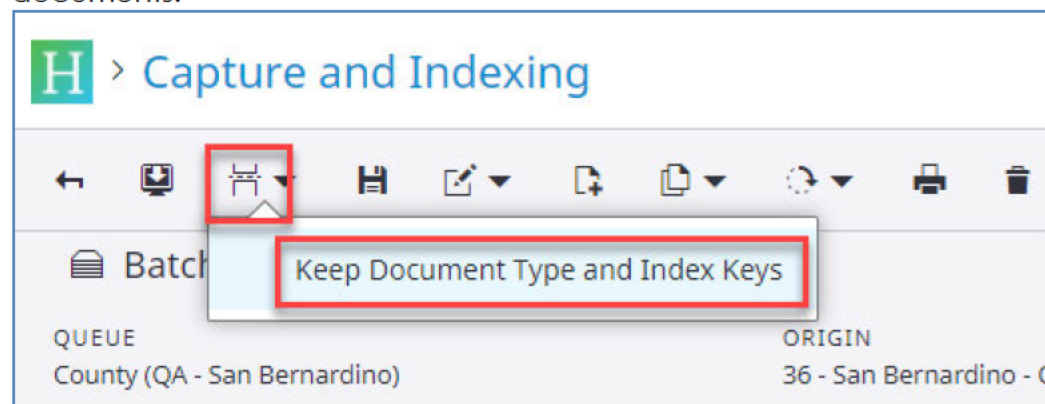
Correct and/or set the barcode at point of scan
In QA & Indexing, users may enter the barcode number at the point of scan. The Imaging Solution will index the document based on entered barcode after clicking on "Submit." These documents will automatically skip OCR.

Split documents at point of scan
When scanning multiple documents at the same time and setting OCR Split Override to "False," use of **separator sheets**, and/or **manually splitting** documents when performing quality assurance before submitting a scanned batch, will **greatly reduce** the likelihood of them being routed to exception.
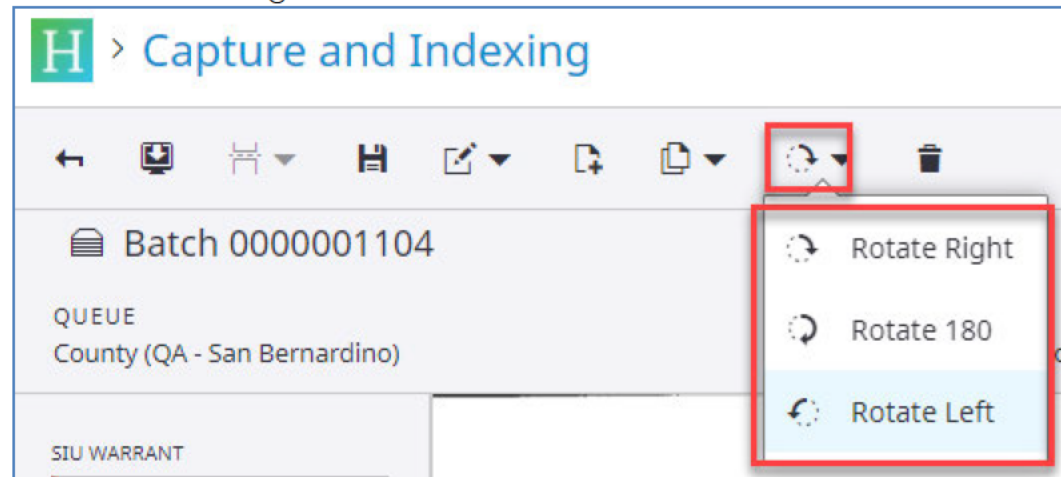
When scanning multiple documents at the same time **without separator sheets** (e.g., a shelter expense, a miscellaneous document, and a utility expense), if the System can correctly identify the form name for pages 1, 2, 4, and 5 but **cannot confidently categorize** page 3, the **entire document** will be sent to exception. Since the System does not know whether page 3 is part of the shelter expense, its own document, or part of the utility expense, it **must err on the side of caution** and route the entire 5-page document to the exception queue for staff to review and make sure it is split correctly.

You can separate documents with **barcoded case coversheets** or **generic separator sheets**; or take a moment to manually split them during quality assurance, to increase OCR's chance of classifying and indexing them correctly. The "Keep Document Type and Index Keys" action will create a document split above the selected page and retain the current case and form context for both documents.

Submit images in the correct orientation

After scanning a document, users can use the rotate buttons to ensure a document is facing the correct direction.

Organize pages in the correct order

Physical page organization should be done before a document is first scanned. If something is out of order after scanning, users should drag and drop the thumbnails of a document into the correct order while the document is in QA & Indexing.

Ensure documents generated from CalSAWS are for programs already linked to a case

When generating a document, users are prompted to select a program. Always select a program that is on the case for which the document is being generated. If the document is generated for a program that is not on the case, the barcode will not be valid. When the invalid barcode is read by Imaging, information for the form and case will not be returned.

Use Photo scan for items with a background (ID, passport, birth certificate, Green card, etc.)

Documents with a color background, such as IDs, passports, etc., can be imaged using Photo scan. This is a color capture mode that will allow OCR to better read the document.

Scanner settings

Scanners should be set to 300 dpi for optimal OCR extraction.

County Action

The CalSAWS Project recommends that counties distribute this communication to **all Imaging users** with county business process instructions on when the OCR Bypass and OCR Split Override flags should be used.

For additional Imaging documentation, click the following URL or navigate to this location for Imaging Fact Sheets: ████████████████████

	████████████████████████ If you have questions on this CIT, please reach out to the Primary Contacts and cc your Regional Managers.
Primary Project Contact: (Name, email address)	Rhiannon Chin ChinR@CalSAWS.org
Backup Project Contact: (Name, email address)	Inez Finnigan Finnigani@CalSAWS.org
Attachments:	None
Web Portal Link:	████████ OR You may also retrieve the CIT document and attachments by following these steps: 1. Click on the CRFIs & CITs link at the top of the page. 2. Click on the "CalSAWS Information Transmittal (CIT)" folder. 3. Click on the "2022" folder. 4. Click on the appropriate CIT # folder.