

☒ CalSAWS M&E

<b>Distribution Date:</b>	July 3, 2024
<b>To:</b>	PPOC.All; Committee.Imaging.All; Consortium.RegionalManagers.All
<b>CIT Name:</b>	<b>CalSAWS Imaging Optical Character Recognition (OCR) Recommendations</b>
<b>From:</b>	CalSAWS Project

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

- |  |   |
|--|---|
| <input type="checkbox"/> General                 | <input type="checkbox"/> Reports              |
| <input type="checkbox"/> Policy                  | <input type="checkbox"/> Fiscal               |
| <input type="checkbox"/> CW                      | <input type="checkbox"/> Caseload Movement    |
| <input type="checkbox"/> CF                      | <input type="checkbox"/> Management           |
| <input type="checkbox"/> MC                      | <input type="checkbox"/> Fiscal               |
| <input type="checkbox"/> CMSP                    | <input type="checkbox"/> Security             |
| <input type="checkbox"/> FC/KG/AAP               | <input type="checkbox"/> Batch and Interfaces |
| <input type="checkbox"/> Child Care              | <input checked="" type="checkbox"/> Imaging   |
| <input type="checkbox"/> WtW                     | <input type="checkbox"/> Migration            |
| <input type="checkbox"/> Other Program(s) _____  | <input type="checkbox"/> Conversion           |
| <input type="checkbox"/> BenefitsCal             | <input type="checkbox"/> Technical            |
| <input type="checkbox"/> Customer Correspondence | <input type="checkbox"/> Training             |
| <input type="checkbox"/> OCAT                    | <input type="checkbox"/> Help Desk            |
| <input type="checkbox"/> Other _____             |   |

Description:	<p><b>Purpose</b></p> <p>The purpose of this CIT is to refresh the CalSAWS counties with information on the Optical Character Recognition (OCR) override flags and recommended OCR best practices.</p> <p><b>Background</b></p> <p>The CalSAWS Imaging Solution has Optical Character Recognition (OCR) built into the solution. OCR is an automatic process that attempts to select form information after a document has been submitted into the Imaging Solution. Selecting the form information is commonly called indexing.</p> <p>OCR looks for specific items on a page to index a document to the correct CalSAWS or Imaging form information for the document. The form information may be a CalSAWS generated form or an approved Imaging form name. The list of form information currently available are listed in the Imaging Form Name Matrix.</p> <p>OCR looks for the following information to index a document, in this order:</p> <ul style="list-style-type: none"><li>• A CalSAWS barcode.</li><li>• A CalSAWS or Imaging form number on the bottom left of the document.</li></ul>
--------------	--

- Certain keywords or phrases on the page to identify a person level document.

If OCR can confidently use this information to index a document, it is routed forward.

If OCR cannot confidently index a document, the form information 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. If the form name is considered time sensitive (see the "Routes to Barcode Verification" column in the Imaging Form Name Matrix) but a barcode is not found, the document will route to the Barcode Verification queue.

Note, most BenefitsCal document uploads will not route to a queue after OCR processing. This information and process is outlined in CIT 0027-24 "CalSAWS Imaging - Re-Enablement of Imaging OCR for BenefitsCal Document Uploads."

### OCR Override Flags

SCR CA-228869 implemented two flags to give CalSAWS users more control over how OCR processes documents. County policy guides how and when these flags should be changed.

- **OCR Bypass flag:** This flag may 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:

- Single Case
- Multi Case
- Returned Mail
- 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. However, time sensitive documents will still route to Barcode Verification if the Ignore Barcode capture mode is not used.

- **OCR Split Override flag:** This flag may 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

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 screenshot shows the 'Capture and Indexing' window in CalSAWS. At the top, it displays 'Batch 0002189842'. Below this, there are fields for 'QUEUE' (County (QA - San Bernardino)), 'ORIGIN' (36 - San Bernardino - County), 'CONTENT' (1 document(s), 1 page(s)), and 'BUNDLE ID'. The main area shows a document titled 'Print Receipt' with fields for 'Date', 'To', 'From', 'Amount', 'Pay by', and 'Check'. The right-hand sidebar contains various settings, including 'TASK OVERRIDE' (false), 'NO CHANGE SAR7/QR7' (false), 'PERSON OVERRIDE' (false), 'PROGRAM LIST' (false), 'PROGRAM UID', 'BARCODE', 'OCRSPLITTOVERRIDE' (True), 'OCRBYPASS' (False), 'COLOR MODE' (Color), 'WF PROCESSING QUEUE', and 'NOTES'. The 'OCRSPLITTOVERRIDE' setting is highlighted with a red box.

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

### Recommendations for OCR Success

The following are best practices recommended for OCR success.

#### 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 onto their forms not generated from CalSAWS, so OCR can leverage the printed numbers to assist with indexing.

**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 select a form name or number at the point of scan. Selecting a form name or number adds weight to the selection. If OCR is trying to decide between two names, it will lean toward the name the user selected. Even if

a form is selected, it is possible that OCR will override the user-selected value if it is confident in its selection.

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.

### **Correct and/or set the barcode at point of scan**

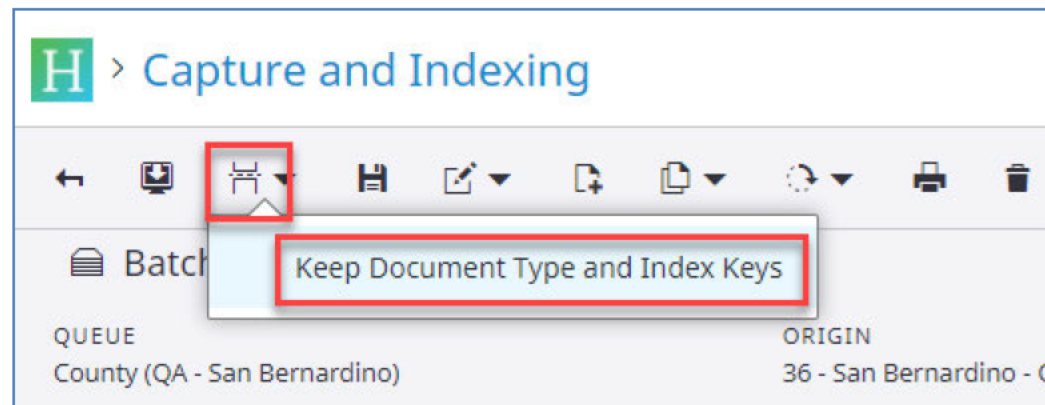
In QA & Indexing, users may enter the barcode number at the point of scan. The Imaging Solution will index the document based on entered barcode after clicking on "Submit." These documents automatically skip OCR.

### **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.



### **Organize pages in the correct order**

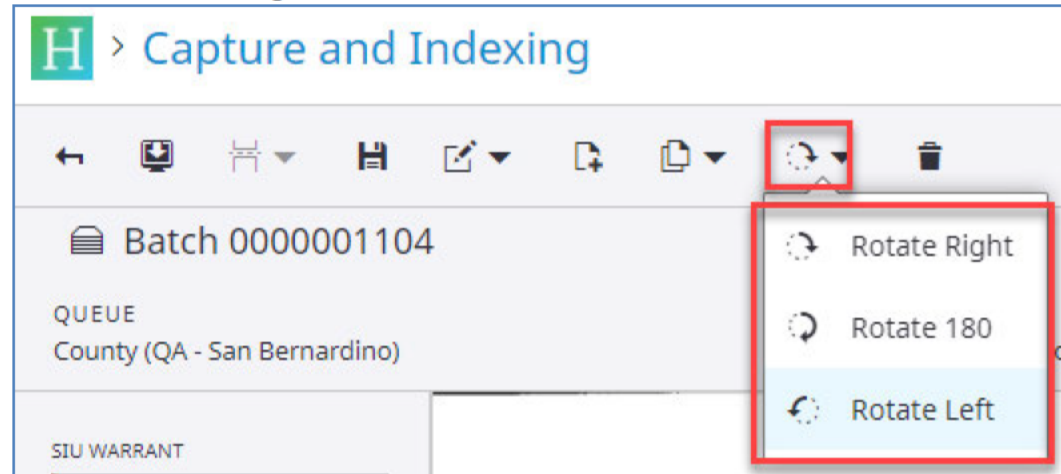
Physical page organization should be done before a document is first scanned. If something is out of order after scanning, users should drag and drop the



thumbnails of a document into the correct order while the document is in QA & Indexing.

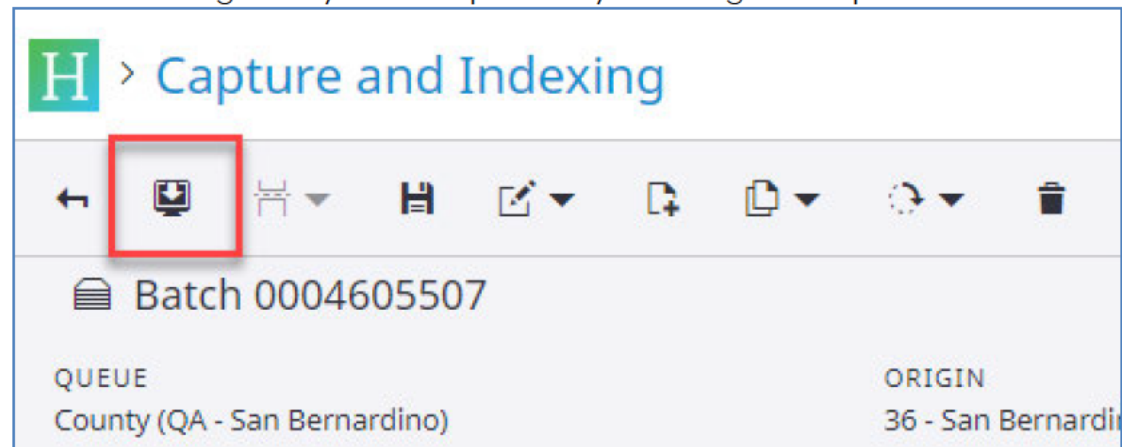
**Submit images in the correct orientation**

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



**Replace blurry images**

During QA, users should review the batch to ensure all pages are legible. If a page is blurry, OCR will not be able to read the information on the page to index the document. Pages may be re-captured by selecting the Capture button.



**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 improve the quality of the image, allowing OCR to better read the document.

**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

	<p>not be valid. When the invalid barcode is read by Imaging, the form and case information will not be returned.</p> <p><b>Ensure the barcodes on CalSAWS documents printed in-office is not cutoff</b> When printing a document in-office (Locally), users should ensure the barcode is not cutoff. CalSAWS documents should be printed using the "Fit to Page" or "Best Fit" option.</p> <p><b>Scanner settings</b> Scanners should be set to 300 dpi for optimal OCR extraction.</p> <p><b>Scanner maintenance</b> Scanners should be serviced and cleaned per the manufacture instructions. It is also recommended to check for and install updated Twain or Pixel drivers once a year.</p> <p><b>County Action</b> The CalSAWS Project recommends that counties share this communication with all Imaging users.</p> <p>If you have questions on this CIT, please reach out to the Primary Contacts and cc your Regional Manager(s).</p>
Primary Project Contact:	Rhiannon Chin <a href="mailto:ChinR@CalSAWS.org">ChinR@CalSAWS.org</a>
Backup Project Contact:	Darcy Alexander <a href="mailto:AlexanderD@CalSAWS.org">AlexanderD@CalSAWS.org</a>
Attachments:	CIT 0107-24 Imaging Form Name Matrix- Updated 07.02.24.xlsx
Web Portal Link:	<div style="background-color: black; width: 100px; height: 20px; margin-bottom: 10px;"></div> <p>OR</p> <p>You may also retrieve the CIT document and attachments by following these steps:</p> <ol style="list-style-type: none"> <li>1. Click on the CRFIs &amp; CITs link at the top of the page.</li> <li>2. Click on the "CalSAWS Information Transmittal (CIT)" folder.</li> <li>3. Click on the "2024" folder.</li> <li>4. Click on the appropriate CIT # folder.</li> </ol>