# Cal**SAWS** | Notes from Imaging Committee

| Date: | July 15, 2021 | Notes Location: | CalSAWS Web Portal > Meetings > Committees > Imaging > 2021 > 2021 July 15 |
|---|---|---|---|
| Time: | 9:00 am – 12:00 pm | Meeting Materials: | CalSAWS Imaging OCR Stats Update.ppt OCR Stats 7.8.21.xlsx |
| Meeting Called by: | Rhiannon Chin | | |

Attendees:

| | | NAME | | | NAME | | | NAME |
|---|---|---|---|---|---|---|---|---|
| ☒ | R1 | Jack Seng | ☒ | R4 | Chris Gomez | ☒ | RM | Matthew VanderEyck |
| ☐ | R1 | Terri Rose | ☒ | R4 | Martha Esparza | ☐ | RM | Ayana Alvarez |
| ☒ | R1 | Christine Alvarez | ☒ | R4 | Louis Cuellar | ☒ | Proj. | Project Staff |
| ☒ | R1 | Brent Wong | ☒ | R4 | Cheryl Armstrong | | | |
| ☒ | R1 | Todd Estabrooks | ☐ | R4 | Aaron Gomes | | | |
| ☒ | R2 | Beth Andrews | ☒ | R5 | Phi Phi Thai | | | |
| ☒ | R2 | Shawna Reed | ☒ | R5 | Tony Baker | | | |
| ☐ | R2 | Michelle Fell | ☒ | R5 | Laura Alba | | | |
| ☐ | R2 | Hortencia Hernandez | ☒ | R5 | Felix Sanchez | | | |
| ☒ | R2 | Nataliya Kurrina > Don | ☒ | R5 | Eric England | | | |
| ☐ | R3 | Heather Brantley | ☒ | R6 | Arin Shahgholi | | | |
| ☐ | R3 | Crystal Kehle | ☒ | R6 | Juan Herrera | | | |
| ☒ | R3 | Dayna Boggs | ☒ | R6 | Mario Palacios | | | |
| ☒ | R3 | Julie Evinger | ☒ | R6 | Dianna Crowley | | | |
| ☒ | R3 | Michelle Smith | ☒ | R6 | Mohsin Khan | | | |

| Topic | Important Points |
|---|---|
| **Welcome & Introductions**<br>• Roll Call | |
| **OCR Levels and Statistics**<br>• OCR Training Process<br>• Overlapping Terms | • Terms – OCR, Brainware, Confidence Level, Exception Rate, Splitting, and Index<br>• OCR – Optical Character Recognition<br>  o OCR will change a document type if it has high enough to VERY confidence of it being in a different classification<br>• Brainware reviews each document and records words and phrases found on the document along with location, page number and frequency.<br>• Our Hyland partners review the results to determine which or phrases should be used in the final configuration.<br>• OCR is unable to translate phrases or other languages<br>• Initial training and Tuning have been completed |

| Topic | Important Points |
|---|---|
| | • Initial training consists of Person level documents provided by C-IV Customers |
| | • Overlapping Forms will continue to need monitoring until the confidence level increases. It's important to know this is currently happening, and efforts will continue for improvement. |
| |     o Were physical documents used to train the system? Images used are from CIV data base, they were physical originally |
| |     o Housing Receipts from hotels/motels will fall under shelter |
| |     o If we image packets that include customer notices & earnings verification. Will OCR re-index or split our documents? Assuming that you use something like single or multi case capture, yes OCR will split the documents right now |
| |     o An example of additional samples needed – High school ID card vs Employer ID. HS ID was too vague but the employee ID had more information |
| |     o Do we have the ability to put a number in the corner to have the document go to a certain location? Yes you have the option. OCR will attempt to read the bottom left |
| |     o The recommendation for county forms is to print a form number on the bottom left and OCR can classify off of that number based on our current system mappings |
| **OCR Levels and Statistics-Cont'd**<br><br>• OCR Index Rates-<br>    o Form Name Statistics<br>• Next steps | • Excel Column -<br>    o **Indexed** -The percent of sample training documents were accurately indexed and archived with the correct form name<br>    o The threshold confidence for person level documents will increase as we work through the new imaging system<br>    o Case level documents are classified by form number and Barcode, they have extremely high success rates.<br>    o If the Barcode is cut off will OCR look at Form Name? There is a hierarchy. It will go a step above and look for the QR Barcode first and then the other bar code number and then form number and name. OCR will go with the one closest to the bottom of the document itself.<br>    o How will new county forms be addressed in the future? As The County creates them and start incorporating the imaging form numbers |

| Topic | Important Points |
|---|---|
| | in to your documents. CalSAWS reports new form numbers daily to OCR.<br> o Can a county add barcodes to existing county forms to improve classification? No. the bar codes are from CalSAWS and are unique. They cannot be at the county level<br> o OCR will not be used during document migration efforts? No not at this time<br> o The only way to add classification for county documents is to place the CalSAWS form number in the footer? That is correct |
| **Imaging Auditing Live Review** | Definition of "managed" county again? Is that where the county has a specified agreement to have their CalSAWS systems managed by the Project? Managed counties have their hardware controlled/updated/etc. by the Project<br><br>Will these reports allow us to know how many documents a certain person imaged? Documents captured is already a report in CalSAWS<br><br>Will we be able to view this during UAT or any time before Go-Live? It is currently available in UAT<br><br>Will Modifications made by a user only show when searching actions taken in a particular queue? Depends on what you are searching for. These reports are at a high level.<br><br>Is there a way to keep track of documents routing to other counties? That would be captured in the Index Change Audit if a user updated a document to a different county code and sent it forward to that "Other County Document" queue |
| **Wrap – up Items** | CalWIN and LA county documents will be used once they are live in the imaging solution<br><br>Next meeting – August 19, 2021 9:00 am – 12:00 pm Region 3 is responsible to take notes. |