CalSAWS

California Statewide Automated Welfare System

Plan Document

DDID 2162 - Document Migration

DRAFT

CalSAWS	DOCUMENT APPROVAL HISTORY	
	Prepared By	CalSAWS Imaging Team
	Reviewed By	

DATE	DOCUMENT VERSION	REVISION DESCRIPTION	AUTHOR

Table of Contents

1 OVERVIEW

Document Migration of the C-IV, Los Angeles, and CalWIN Imaging systems into the new CalSAWS Imaging Solution. The goal being to centralize all documents into one location in order to unify the counties into a cloud-based statewide Imaging Solution.

1.1 Requests

This migration will populate the CalSAWS Imaging Solution with images and metadata provided by the counties. This metadata will need to contain required indexing information for the migrated images. Once metadata has been acquired, the data will be processed.

1.2 Overview of Recommendations

The goal of Document Migration is to consolidate all documents from the 58 CalSAWS counties into one centralized Imaging repository.

Counties will provide converted imaging data to CalSAWS in multiple batches. This will consist of test batches, and multiple waves of county imaging migration data. With the use of the test batches, CalSAWS will verify accuracy of imported data. The multiple migration waves will ensure all county data from legacy imaging systems is available upon go-live.

1.3 Assumptions

1.3.1 Overview

All C-IV data and images will be converted by the CalSAWS Project. Los Angeles, and CalWIN counties will need to prepare their data to meet the CalSAWS Imaging Solution requirements. Additionally outlined in this section are export deadlines, and data transfer methods.

1.3.2 Data Purge

Counties will want to purge any data that falls outside of the data retention policies decided upon by CalSAWS.

- C-IV counties will have a data purge effort made to remove documents that do not meet the data retention policy specifications.
- Los Angeles county will receive a list of retained cases from CalSAWS Project, based on the CalSAWS data retention policy.
- CalWIN Project will generate a list based on the CalSAWS data retention policy.

1.3.3 Kick-Off Meeting

The CalSAWS Project will work directly with counties to establish key contacts and decide upon detailed timelines for major county milestones. During this meeting, a detailed RACI will be presented outlining the tasks and responsibilities required for migration to progress smoothly. This meeting will be an opportunity to ensure essential parties are accounted for and that counties have all the vital information needed to prepare properly for the export of data.

1.3.4 Discovery Session

Prior to document and metadata export, discovery sessions will be held with appropriate county contacts established during the kick-off meetings. During these sessions, CalSAWS will detail required metadata for the destination system, while counties will be responsible for transforming their data appropriately to fit the agreed upon values.

1.3.5 Data Storage Configuration

Counties will need to export all data in a File System Storage(FSS) configuration. Any counties using differing configurations (Centera) will need to convert prior to export.

1.3.6 Data Transfer Methods

The imaging data from counties will transferred to CalSAWS via one of the following methods:

- SFTP
- Amazon Snowball
- AWS Datasync

1.3.7 File Formats

The CalSAWS Imaging Solution will support all incoming file formats. If counties have the compacity to convert their existing data, below is a list of the recommended file formats.

Recommended File Formats	
TIF (Hyland Native Format)	PNG
TIFF	PDF (Unlocked/Unsecured)
JPG	WAV (Audio)
JPEG	MP3 (Audio)

1.3.8 Metadata

Metadata will be necessary for context to properly index documents imported into the system. Those metadata fields can be broken into two categories, Required, and Optional.

Required Metadata	Optional Metadata
Filename/Unique Document Identifier	CalSAWS Document Type
Case Number (Case Level Documents)	Form Number
CIN (Person Level Documents)	Legacy Creation Date
Form Name	Received Date
Applicable Date	State Hearing Number
Case/Person Flag (If Applicable)	E-Application Number
County Code	Case Name
	Notes (Varchar(2048))

1.3.8.1 Date Format

As part of metadata formatting, date formats need to match the following template.

[Year-Month-Day](2020-12-30)

1.3.8.2 Metadata Page Association

Documents will require metadata to be configured appropriately to identify multi-file documents.

If a document is made up of multiple files (i.e. one file per page) each file will require metadata indicating the relation with a unique document identifier and order. Below are 3 examples of how this association would have to be addressed:

1. Document A is comprised of 5 files, one jpg for each page of the document. Metadata for each file will need to include a unique document identification number, and page number to associate to each file.
2. Document B is comprised of 1 multipage tiff, this file will only require metadata for the one file with a unique document identification number, as the pages are inherently organized by the multipage tiff.
3. Document C is comprised of a multipage tiff file, and a jpg file, metadata will be needed for the multipage tiff indicating that its pages precede or follow the jpg file, and a corresponding unique document identifier.

Additionally when submitting multiple metadata files it is important that all pages of a document are with the same metadata file. Document page files should not be split amongst different metadata files.

1.3.9 County Milestones

To ensure go-live dates are met, the following milestones will occur. A CRFI will go out to obtain a primary point of contact for each county. Additionally this CRFI will establish specific date for the Kick-off Meeting and Discovery Session.

County Milestones
Kick-Off Meeting
Discovery Session
Test Batch Export
Test Batch 2 Export
Initial Data Export
Delta Data Export(s)
Final Data Export

2 RECOMMENDATIONS

2.1 Testing Batches

Test batches will be a sample of each counties imaging data, and metadata tying to the images. CalSAWS will use this sample data to validate the import method and anticipate any issues with metadata configuration, prior to the primary data import.

Test Batches should be submitted meeting criteria outlined in the Assumptions

2.1.1 Test Batch 1

Test batch 1 will be a data migration and configuration test. This will allow both the counties and the CalSAWS Project to verify the import and export methods.

- Full sample of the scope of data formats (tif, tiff, jpg, jpeg, gif, etc..)
- Sufficient batch sample size (250 - 1000 documents)
- A representation of the largest file size
- Batch samples should contain a wide range of different cases
- Subsequent imports may be requested for proper error mitigation

2.1.2 Test Batch 2

Test batch 2 will use information gathered from test batch 1. This will call for additional batch requests above and beyond this list.

- Full sample of the scope of data formats (tif, tiff, jpg, jpeg, gif, etc..)
- Sufficient batch sample size of new data (250 - 1000 documents)
- A representation of the largest file size
- Batch samples should contain a wide range of different cases
- Based on data gathered from Test Batch 1 CalSAWS will make recommendations for improving the import results
- Subsequent imports may be requested for proper error mitigation

2.2 C-IV Migration

2.2.1 Discovery Session

CalSAWS alongside Hyland will work with the C-IV counties to establish finalized conversion process for importing and exporting data. During this session, document type mappings and other aspects of data conversion will be evaluated. The findings during these sessions will finalize the configurations of the conversion tools used upon metadata export and import.

2.2.2 Export/Import Test Batches

CalSAWS will import test batches of C-IV data to identify any unforeseen issues with data migration. Any issues detected within the sample size batches will then be mitigated and retested prior to initial full data import.

2.2.3 Image Data Export

The existing document images data will be converted from CAS into the AWS S3 cloud storage. The C-IV imaging system configurations will be modified to direct all new incoming image to be stored in AWS S3 by default, eliminating the need to perform follow-up conversions from the CAS.

2.2.4 Metadata Export

The C-IV metadata will be exported to Hyland and processed by conversion scripts into the CalSAWS imaging database. During this process detailed logging will be generated for data verification.

2.2.5 Delta 1/2 Import

Each Delta phase of the import process will be a checkpoint to import newly ingested metadata and update previously imported metadata from the C-IV System.

2.2.6 Final Import

The final import will be any remaining new metadata and changes, bringing the CalSAWS Imaging Solution up to date as go live is approached.

2.2.7 Go Live

Counties should expect downtime the weekend (Friday 9:00 PM – Monday 6:00 AM) prior to go live.

2.3 Los Angeles Migration

2.3.1 Discovery Session

CalSAWS alongside Hyland will sample Los Angeles metadata to establish a mapping process for exported data. During this session, values such as form names, date formatting, index field length, and other aspects of data, will be evaluated. The findings during these sessions will finalize the configurations of the conversion tools used upon data export and import.

2.3.2 Export/Import Test Batches

Los Angeles will submit test batches to CalSAWS to identify any unforeseen issues with data migration. These batches will be done over the network to test Los Angeles network stability with the solution. Any issues detected within the sample size batches will then be mitigated and retested prior to initial full data import.

- See Test Batches for details on the sample requirements.

2.3.3 Metadata/Images Migration

Los Angeles will prepare their data in a format appropriate for the CalSAWS Imaging Solution to ingest.

- See Assumptions for appropriate format requirements.
- Additional details about the metadata preparation will be outlined during the Discovery Sessions

2.3.4 Data Export

Los Angeles will export the data onto an Amazon Snowball for physical migration of data, or use a network transfer with the aid of AWS Datasync. The exact means of data transfer will be determined during the discovery session and with the aid of data collected during test batches.

2.3.5 Initial Import

Once test batches are complete, the initial import of imaging document data will begin. This will include all existing data up to a point in time.

2.3.6 Delta 1 / 2 Import

Each Delta phase of the export process will be a checkpoint to import newly ingested data and update previous metadata from the Los Angeles system.

2.3.7 Final Import

The final export will be any remaining new data and changes, bringing the CalSAWS Imaging Solution up to date as go live is approached.

2.3.8 Go Live

Counties should expect downtime the weekend (Friday 9:00 PM – Monday 6:00 AM) prior to go live.

2.4 CalWIN Waves 1 – 6 Migration

2.4.1 Discovery Session

CalSAWS alongside Hyland will sample each CalWIN counties metadata to establish a mapping process for exported data. During this session, values such as form names, date formatting, index field length, and other aspects of data, will be evaluated. The findings during these sessions will finalize the configurations of the conversion tools used upon data export and import.

2.4.2 Export/Import Test Batches

CalWIN counties will submit test batches to CalSAWS to identify any unforeseen issues with data migration. These batches will be done over the network to test each counties network stability with the solution. Any issues detected within the sample size batches will then be mitigated and retested prior to initial full data import.

- See Test Batches for details on the sample requirements.

2.4.3 Metadata/Images Migration

CalWIN counties will prepare their data in a format appropriate for the CalSAWS Imaging Solution to ingest.

- See Assumptions for appropriate format requirements.
- Additional details about the metadata preparation will be outlined during the Discovery Session

2.4.4 Data Export

CalWIN counties will export the data onto an Amazon Snowball for physical migration of data, or use a network transfer with the aid of AWS Datasync. The exact means of data transfer will be determined during the discover sessions and with the aid of data collected during test batches.

2.4.5 Initial Import

Once test batches are complete, the initial import of imaging document data will begin. This will include all existing data up to a point in time.

2.4.6 Delta 1 / 2 Import

Each Delta phase of the export process will be a checkpoint to import newly ingested data and update previous metadata from the CalWIN counties. In addition any modified document(s)/metadata will be updated.

2.4.7 Final Import

The final import will be any remaining new data, bringing the CalSAWS Imaging Solution up to date as go live is approached.

2.4.8 Go Live

CalWIN counties should expect downtime the weekend (Friday 9:00 PM – Monday 6:00 AM) prior to go live.

2.5 Data Validation

2.5.1 Overview

The purpose of data validation is to ensure accuracy of imported data during the migration process. Using sample sizes of original data, CalSAWS will verify that metadata mappings match design specifications of the CalSAWS Imaging Solution.

Data validation tests will be performed on an ongoing basis throughout the migration.

2.5.2 Validation Tests

Data will be validated using the log files generated during the conversion of the data into the CalSAWS Imaging Solution.

2.5.3 Converted Data Tests

Converted data will be verified by counties during the UAT timeframe. Data that has been fully converted will be available in the UAT environments for verification.

3 SUPPORTING DOCUMENTS

Number	Functional Area	Description	Attachment

4 OUTREACH

This section will list out outreach performed to the counties, such as CRFIs and CITs related to the document migration process.

- CRFI 20-003 Imaging – County-Specific Documents/Forms
- CRFI 20-0XX Imaging – County Kick-Off/Discovery Sessions

5 APPENDIX

Project Deliverable or Activity	CalSAWS	Counties		
	Accenture Consortium	C-IV County	Los Angeles County	CalWIN Counties
Kick-Off Meeting				
Respond to CRFI regarding PPOC		**R**	**R**	**R**
Present county specific RACI	**R A**	**I**	**I**	**I**
Discovery Session(s)				
Discovery Session for metadata configuration	**R C**	**A**	**A**	**A**
Verify Data Transport Requirements	**R A**	**C**	**C**	**C**
Verify Network Infrastructure	**R A**	**C**	**C**	**C**
Evaluate C-IV Metadata Configuration for Export	**R A**	**I**		
Evaluate LRS Metadata Configuration for Export	**C I**		**R A**	
Evaluate CalWIN Metadata Configuration for Export	**C I**			**R A**
Establish Test Batch Details	**R A**	**C I**	**C I**	**C I**
Develop process to track newly ingested images and metadata after initial export	**C I**		**R A**	**R A**
Test Batch(es)				
Export test batch images and metadata	**R A**		**R A**	**R A**
Import/Process/Map exported test batch images and metadata	**R A**	**I**	**I**	**I**
Verify test batch of imported images and metadata	**A C I**	**R**	**R**	**R**
Reconfigure export metadata/images per test batch results	**R A**	**I**	**R A**	**R A**
Data Export/Import				
Export Initial batch of images and metadata	**R A**		**R A**	**R A**

Import/Process/Map exported initial batch images and metadata	**R A**	**I**	**I**	**I**
Verify initial batch of imported images and metadata	**A C I**	**R**	**R**	**R**
Export Delta(s) images and metadata	**R A**		**R A**	**R A**
Import/Process/Map exported Delta(s) images and metadata	**R A**		**I**	**I**
Verify imported Delta(s) imported images and metadata	**A C I**	**R**	**R**	**R**
Export Final batch of images and metadata	**R A**		**R A**	**R A**
Import/Process/Map exported Final batch images and metadata	**R A**		**I**	**I**
Verify imported Final batch of imported images and metadata	**A C I**	**R**	**R**	**R**
Data Conversion				
Convert C-IV UID To Migrate into CalSAWS	**R A**			
Append UIDs to CalWIN documents	**R A**			
UAT Image/Metadata Validation	**R**	**R**	**R**	**R**

R=Responsible (owner - Individual with primary responsibility for completing the deliverable)
A=Accountable (Signs off on work - individual who is accountable for ensuring the deliverable is complete)
C=Consulted (has information/capability - individuals who have knowledge or content needed to complete the deliverable)
I=Informed (notified of results -individuals who should be notified of deliverable content)