

**California Revealed
Statement of Work
for Print Materials
January 2020**

Project description

California Revealed (CA-R) is a California State Library initiative based in Sacramento which helps California's public libraries and other non-profit local heritage groups reformat, preserve, and make accessible historic materials that are significant to California history. Participating organizations select items from their collections, create discovery metadata, and send materials and metadata to CA-R, then CA-R oversees the digitization and reformatting processes (outsourced to vendors), as well as provides online access and long-term preservation.

Our goal is to digitize approximately 9,805 primary source materials (e.g., newspapers, scrapbooks, documents, photographs), totaling over 235,000 pages and drawn from 60 public libraries, historical societies, archives, and other organizations. CA-R staff will work with these partner organizations, vendors, and digital repositories to accomplish this goal. To see a list of CA-R's partner organizations from past years and to browse the California Revealed collection, please visit: californiarevealed.org.

Collections will start arriving from partners to our Sacramento office in January 2020. We do not yet know the exact item count, dimension, or condition of the physical materials but expect to receive and outsource digitization for:

Format	# of Objects	# of Pages
Newspapers (Bound, Loose)	4153	81071
Periodicals, Newsletters	379	12551
Microfilm (Reels)	150	Unknown
Transparencies: Negatives; Slides; Glass	600; 1691; 171	1391; 6056; 171
Photographs	495	990
Scrapbooks and Albums	22	1589
Bound: Books, Volumes, Yearbooks	275	88066
Loose: Correspondence, Manuscripts, Documents, Transcripts, Clippings	310	20398
Printed Materials: Booklets, Catalogs, Programs, Pamphlets, Brochures	451	7465
Posters, Maps, Drawings, Postcards	687	709
Ephemera (Flyers, Invitations, Cards)	126	220

We anticipate starting shipment of these physical materials to the vendor in January 2020. The shipments will be made in batches and organized by partner organization; each shipment might include a variety of formats. Shipments will continue through the end of June 2020. The vendor must complete digitization and delivery of initial files within 8-16 weeks of receiving the original physical materials from CA-R. For a more detailed timeline, please refer to Appendix A.

Please direct all correspondence related to this project to:

Pamela Vadakan, Director, pvadakan@californiarevealed.org

Theresa Berger, Print Preservation Manager, tberger@californiarevealed.org

California Revealed
California State Library
900 N Street
Sacramento, CA 95814
916.653.5074

PROJECT WORK PLAN

Technical specifications

This section outlines the technical specifications for digital files created from print materials, beginning with general requests for handling and digitization, followed by specific target format specifications for each digital file instantiation. Technical specifications assume a 100-megapixel array camera.

General requests

- Target specifications assume materials are being captured with a 100-megapixel camera. The goal is to capture at the highest resolution possible based on the size of the object without stitching. If stitching is needed, vendor must contact CA-R to approve additional costs.
- Capture both color and black-and-white originals as color. No color/tone enhancement.
- Excess copy board will be cropped to leave a narrow border around the entire document. For bound material, this border will extend from the visible gutter.
- For bound, multi-page items with foldouts and inserts, leave pieces in position to maintain context.
 - The two-page spread will be captured as many times as necessary to replicate the experience of physically viewing the page. Each capture of a fold-out will include the various layers of the fold-out as well as its opposite page in order to maintain object pagination. Before unfolding, a 2-up capture will also be included of the object as-is, in order to demonstrate the untouched view of the page.
 - For example, see [Program: Autumn Festival, October 16-17, 1908](#) or [Green Letts's No. 53 Diary](#).
 - The opposing page to the foldout or insert should have similar digital dimensions to maintain optimal viewing in the Internet Archive viewer when the user views two pages at a time.
 - However, when a multi-page, unbound object is delivered, pages will be captured as-is and dimensions will not be altered. Item will be presented and viewed 1-up to replicate experience of looking through individual documents.
 - For example, see 1-up presentation of: [Film Series Three Correspondence, Mar-Jul 1947: Art in Cinema collection](#)
- California Revealed scans cover to cover, but prefers not to scan more than five consecutive blank pages. If there are more than five blank pages, vendor will skip pages and add a "Pages Skipped" target-page supplied by CA-R.
 - When a target is inserted, a second blank target page will be inserted to maintain optimal viewing on the Internet Archive viewer when the user views two pages at a time.
 - CA-R will note in the "Notes for the Vendor" field of the XML spreadsheet if pages are to be skipped within an object and if a "Pages Skipped" target page is required.
 - The front and back of the front cover of the item should be the first two captures, in order to maintain pagination on the CA-R book reader.
- California Revealed prefers to flag potentially sensitive content with a "Sensitive Materials" target-page supplied by CA-R. If sensitive materials are discovered within an object, vendor will place the Sensitive Materials target before the content in question.
 - When a target is inserted, a second blank target page will be inserted to maintain optimal viewing on the Internet Archive viewer when viewing two pages at a time.
 - CA-R will note in the "Notes for the Vendor" field of the XML spreadsheet if and where a Sensitive Materials target is required within an object. If vendor notices sensitive content that is not already flagged, vendor will contact CA-R for appropriate instructions.
 - The front and back of the front cover of the item should be the first two captures, in order to maintain pagination on the Internet Archive viewer.

Print materials

e.g. books, directories, photograph albums and scrapbooks, newspapers, letters, brochures, etc.

- Preservation master: TIFF, 400 ppi (up to 28 x 21 in.) / 300 ppi (up to 38 x 28 in.), 24-bit RGB color, uncompressed
- Access copy: PDF with uncorrected raw OCR.
- Checksums: .md5 for each preservation and access file.
- Capture front and back of postcards (unless otherwise noted).
- After capturing two-up, split images to single page TIFFs at the gutter, minimizing the 1/8" margin without eliminating original content.
- Insert target pages when and where directed.

Microfilm

- Preservation master: TIFF, 400 ppi (for original titles less than 18 in.) / 300 ppi (for original titles larger than 18 in.), 24-bit RGB, uncompressed.
- Access copy: PDF with uncorrected raw OCR.
 - Multi-page, searchable PDF files, bound by issue or segment.
- Checksums: .md5 for each preservation and access file.
- Reels filmed as two-page spreads will be cropped to one page per image.
- Vendor will supply issue-level titles, dates, and page counts as soon as the information is available. CA-R will update object-level metadata and provide a new metadata spreadsheet accordingly.

Photographic materials

e.g. photographic prints, maps, art prints, and posters

- ⊘ Preservation master: TIFF, 400 ppi (up to 28 x 21 in. / 300 ppi (up to 38 x 28 in.), 24-bit RGB color, uncompressed
- ⊘ Access copy: JPG for single images, and a PDF for multi-image objects.
- ⊘ Checksums: .md5 for each preservation and access file.
- ⊘ Capture front and back of photographic prints (unless otherwise noted).
 - Confirm single vs back-page capture with CA-R for all other objects.
- ⊘ Borders, mattes, and mounts for photographic or pictorial materials will be counted towards the 1/8" border required for all captures, unless otherwise specified.

Photographic negatives and transparencies

- Preservation master: TIFF, 3,000 ppi (up to 3.5 x 2.5 in.) / 1,500 ppi (up to 5.5 x 7.5 in.), grayscale or color as appropriate, uncompressed
- Access copy: JPG, and a PDF for multi-page objects
- Checksums: .md5 for each preservation and access file.
- Use an overhead array camera and cold light transilluminator light box with glass.

Treatment

For print materials, treatments such as page splitting of newspapers or disbinding of bound objects may be necessary for optimal capture. If additional preparation to stabilize, repair, or apply other irreversible treatments is needed, vendor must contact CA-R to request permission from the partner organization in order to proceed, and approve additional costs if needed. Ideally all such actions should be recorded in the technical metadata by the vendor.

Technical requirements

CA-R asks partners to prepare their materials so they objects are secure in their folders and boxes and "scanner ready" before they package them for shipment (see [Shipping Guidelines](#)). CA-R will create an electronic inventory for each shipment for the vendor. Upon arrival at their facility, the vendor will carefully unpack and confirm contents and report back any discrepancies within 30 days of receiving, prior to scheduling digitization. During this 30-day period, the vendor

will perform a comprehensive review and inspection of the materials, evaluate overall condition, and inform CA-R if items will incur fees beyond the original estimate due to any reason.

Materials shipped to the vendor will be handled with clean hands (gloves when necessary) and in a clean work environment with clean digitization equipment. The vendor will inspect, prepare, and digitize the object exactly according to CA-R specifications (outlined above) and will include a condition report/technical evaluation about the original source that includes any damage, special handling considerations, or errors discovered upon check-in or transfer. Examples include, but are not limited to, the following: excessive dust/dirt, stains, tears or faint text/image. If the vendor determines that an object cannot be digitized without damage to the material, CA-R should be notified and the material set aside, pending the partner organization's decision.

Special handling needs will be noted by the partner organization and CA-R in the "Notes to Vendor" field and shipping inventory letter. This includes specific, clear, digitization instructions regarding mounts, inserts, foldouts, and rectos/versos. A physical note will also be included with the item when appropriate. **The vendor will contact CA-R if the general condition of an object may impact the quality of the capture results, and if any special handling - such as repair or treatment - is necessary, or if handling costs exceed the initial estimate per item. CA-R can ask the partner organization for permission to proceed and possibly cover the additional handling cost if needed. In particular, please be attentive to fragile materials.** The vendor will record any condition or treatment notes as Vendor Notes.

Once the materials are prepared for digitization, the vendor will make every effort to handle them carefully. The vendor will not introduce any image processing or enhancement at any point in the creation of the preservation master files. The vendor will ensure that the content is captured faithfully and accurately with no information loss or degradation. The vendor will take responsibility for the safe care and handling of materials at all times. When not being prepared or captured, materials will be stored in a locked, climate-controlled, room. After capture, the vendor will return materials to their original arrangement, as they were packed and shipped to the vendor.

Special handling requirements for print materials

The vendor will make every effort to maintain original condition, handling the materials carefully to avoid or minimize damage. The vendor will limit any factors that may pose a risk to the objects being captured. The vendor will not use any sheet-fed scanners or any devices with automatic page turning.

The vendor will clean materials as needed using canned air or a soft Chinese Hake brush. Severe dirt or debris that cannot be cleared with canned air will be noted in the Vendor Quality Control/Transfer Notes.

Materials should be illuminated using low ultraviolet-emitting light.

If glass is used to flatten documents, the glass will be cleaned with deionized water between captures to minimize transfer of dust and fibers between objects.

Review before digitization

The partner organization is responsible for identifying duplicate content; CA-R will also attempt to highlight potential duplicates and choose the best source to digitize when possible. Upon arrival at the facility and within 30 days of receiving, the vendor will carefully unpack, confirm contents, and report back any duplicates or discrepancies between the title/metadata supplied and the content.

Review after digitization

The vendor is expected to review 100% of the digital files and all associated metadata (in particular title, date, and page counts), for quality, completeness, and accuracy. Still image and text objects will be viewed at 100% using imaging software. The vendor will check the digital file against the original source to ensure a faithful reproduction of the original and confirm that no artifacts were introduced in the digitization process. The vendor will note inherent abnormalities in the "Vendor Quality Control/Transfer Notes" field in the metadata - these notes will be particularly helpful for California Revealed staff when they inspect the files.

CA-R, with partner participation, will review deliverables within 60 days of receipt of the files to ensure that all project specifications are met. Original materials will be returned directly to the partner after all parties have confirmed that the files are acceptable and there is no necessary rework. Please see Appendix A below for vendor workflow/timeline.

Subcontracting

The vendor will perform all digitization on its premises. No subcontracting of this work will be permitted without prior communication and approval from California Revealed and partner.

File naming and directory specifications

File names are based on the Object Identifier (e.g., casmim_000003), which includes the partner's Marc organization code followed by a unique, sequential number. The Object Identifier serves as the prefix for all file instantiations associated with the digital object. Please see the metadata spreadsheet supplied by California Revealed with each shipment for Object Identifiers.

The directory structure for the digital files and the supporting metadata and documents must be "flat" (without subdirectories) for each object.

Create a folder for each partner within the shipment by job number, labeled with partner's Marc organization code, followed by a subfolder for each object that is labeled by the Object Identifier (e.g. CA-R2082/cwh/cwh_000003). The following items should be within each folder per object:

- preservation file(s)
- preservation file .md5(s)
- access file(s)
- access file .md5(s)
- xml metadata record
- technical evaluation form or notes about the transfer (if not already supplied in the xml record)

Label preservation masters [ObjectIdentifier]_prsv.[fileExtension] (e.g., casmim_000003_prsv.tif)

Label access files [ObjectIdentifier]_access.[fileExtension] (e.g., casmim_000003_access.jpg)

Label .md5 checksums in the following manner:

- Media file name: casmim_000003_prsv.tif
- Checksum file name: casmim_000003_prsv.tif.md5

A single page still image object would consist of the following:

- ∄ casmim_000003_prsv.tif
- ∄ casmim_000003_prsv.tif.md5
- ∄ casmim_000003_access.jpg
- ∄ casmim_000003_access.jpg.md5
- ∄ casmim_000003_metadata.xml

A single page text object would consist of the following:

- ∄ csfpal_000155_prsv.tif
- ∄ csfpal_000155_prsv.tif.md5
- ∄ csfpal_000155_access.pdf
- ∄ csfpal_000155_access.pdf.md5
- ∄ csfpal_000155_metadata.xml

If a still image or text object consists of multiple parts or pages, add "_p####" to the file name. For example, a multipage text object would have the following files:

- cwh_000003_access.pdf
- cwh_000003_access.pdf.md5
- cwh_000003_p0001_prsv.tif
- cwh_000003_p0001_prsv.tif.md5
- cwh_000003_p0002_prsv.tif
- cwh_000003_p0002_prsv.tif.md5

- cwh_000003_metadata.xml

For complex pages, such as fold-outs and inserts captured multiple times, files will be named as “_p#####” and “_###” for each sub-page to the file name. For example:

- cgl_000002_p0001_001_psv.tif
- cgl_000002_p0001_001_psv.tif.md5
- cgl_000002_p0001_002_prsv.tif
- cgl_000002_p0001_002_prsv.tif.md5
- cgl_000002_access.pdf
- cgl_000002_access.pdf.md5
- cgl_000002_metadata.xml

XML metadata

The vendor will deliver two XML metadata records per object. One record will use the [Dublin Core](#) schema and the other will be a raw or standard XML record generated by the vendor as materials are prepared and digitized. These metadata records will combine descriptive, rights, and administrative metadata supplied by the partner and California Revealed, along with technical and preservation metadata supplied by the vendor. They should document information about the equipment used for digitization as well as transfer and quality control notes. Please see California Revealed’s [Metadata Guidelines](#) for a complete list of required fields per object.

For newspaper microfilm records, California Revealed will provide a metadata template with record placeholders for the vendor to fill in issue dates, masthead titles, and page counts. The vendor will deliver the spreadsheet with this metadata information so that California Revealed can ingest the information in our content management system, prior to digitization. CA-R will then return a new spreadsheet for the vendor to use for generating XML and embedding files.

Sample Dublin Core records

Image object (one page)

https://archive.org/download/c_000167/c_000167_metadata.xml

Text document (two pages)

https://archive.org/download/cscrm_000380/cscrm_000380_metadata.xml

Bound document

https://archive.org/download/cgl_000007/cgl_000007_metadata.xml

Embedded metadata

For still image and text objects, the vendor will ensure embedded technical metadata is accurate for all file instantiations, including file size and type, width, height, color channels, and resolution.

The vendor will embed the following descriptive metadata supplied by California Revealed for each access and preservation file:

- **Embedded Title** (e.g. Wedding photo of Shinichi and Sawano Yamashita)
- **Embedded Institution** (e.g. Tulare County Library)
- **Embedded Comment** California Revealed
- **Embedded Copyright** Public Domain or Copyrighted. First two sentences of the copyright statement. Please note if copyright status is unknown, California Revealed will leave the Embedded Copyright field blank and no copyright information should be embedded. (e.g. Copyrighted. Rights are owned by Tulare County Library.)

Deliverables

Please see Appendix A below for vendor workflow/timeline. California Revealed’s files will be retained on the vendor’s storage system until they are checked and accepted by California Revealed and the partner organization; the deadline for this approval is six months after CA-R receives the files. The vendor will confirm the deletion of the files via email within

six months or once California Revealed has given permission to delete the files, after quality control has been completed and all parties approve.

The vendor will batch-deliver digital files to CA-R for quality control on borrowed exFAT-formatted hard disk drives (HDDs), preferably with Firewire or USB 3.0 ports. The vendor will produce an .md5 checksum per file state (preservation and access) immediately after digitization to verify that files were not modified in transfer.

CA-R will perform quality control on the deliverables and may ask the vendor to review the original physical materials to determine whether or not artifacts are inherent to the source. CA-R will report quality control feedback from the partner and will ask the vendor for replacement files for any deliverables that do not meet specifications. Replacements should be received from vendor within 30 days of request.

After quality control and upon approval of the files by CA-R and the partner organization, the vendor will deliver a set of authoritative files (including preservation files, access files, .md5 checksums, xml metadata, and tech sheets, if included) on a mirror set of two LTFS-formatted LTO8 tapes. Include only the final approved files for each object on the LTO. Please ensure that files from the same partner are not divided across more than one LTO tape, and please ensure all the associated files that represent an object are on the same LTO tape, unless they do not fit on one tape. The vendor will check the checksums on the LTO after writing the tapes to confirm that files were not altered during writing. California Revealed will return the borrowed hard drives upon receipt and approval of the final LTO deliverable.

The vendor will create a manifest of the LTO tapes in an Excel spreadsheet, including the following per file, in order:

- LTO_MediaLabel/LT_[Unique Identifier] (e.g., CA0010L8 and/or barcode)
- MARC Organization Code: (e.g., csat)
- Object Identifier (e.g., csat_000012)
- Folder Name (e.g., csat_000012)
- General File Name (e.g., csat_000012_p001_prsv.tif)
- Path and File Name (e.g., CA-R2019\csat\csat_000012\ csat_000012_p001_prsv.tif)

If vendor is unable to provide final deliverables on LTO tapes, vendor will deliver a set of authoritative files (including preservation files, access files, .md5 checksums, xml metadata, and tech sheets, if included) on exFAT-formatted hard disk drives, preferably with Firewire or USB 3.0 ports. Include only the final approved files for each object. Please ensure that files from the same partner are not divided across more than one hard drive, and please ensure all the associated files that represent an object are on the same hard drive, unless they do not fit on one. Vendor will also include copies of all final metadata spreadsheets for objects represented on the drive. The vendor will check the checksums on the hard drive after writing to confirm that files were not altered during transfer. California Revealed will return any borrowed hard drives upon receipt and approval of the final deliverables.

Partner organizations will contact the vendor directly for copies of their approved files and associated metadata – either on a mirror hard drive purchased from the vendor or on one supplied by the partner – within 60 days upon the vendor's file delivery to California Revealed, after California Revealed's approval of the files. The partner organization is expected to pay only for storage media and shipping for copies of their files. California Revealed will inform the vendor which partners expect copies at the start of digitization work. When confirming shipment of the backup copies to the partner, the vendor will cc: team@californiarevealed.org.

Shipping

All original physical materials will be shipped from CA-R offices at the California State Library via the vendor's FedEx account. The vendor will invoice CA-R for shipping.

Vendor will send deliverables to:

California Revealed
California State Library
900 N Street
Sacramento, CA 95814
916.653.5074
team@californiarevealed.org

After final approval of the files and metadata by CA-R and the partner organization, the vendor will return the materials, as they were originally shipped, directly to the partner via the vendor's FedEx account. The vendor will ship the originals back

early in the week so that they are not in transit more than 3 days or over a weekend/holiday. The vendor will invoice CA-R for shipping. CA-R will confirm and provide shipping addresses and contact information for partners at the time of return request. The vendor should send shipping and tracking information to both the partner and CA-R when shipping has been scheduled.

Please send all shipping notices to: tberger@californiarevealed.org and cc pvadakan@californiarevealed.org.

Billing

Califa, California Revealed's fiscal agent, handles all payments for California Revealed. Mail or email invoice(s) to Pamela Vadakan; approved invoices will be forwarded to Califa for payment. Ensure invoices are labeled with the following Bill to/Send to information:

Bill to:

Califa Group
330 Townsend Street, Suite 133
San Francisco, CA 94107
415.796.3901
billing@califa.org

Send to:

Pamela Vadakan
California State Library
900 N Street
Sacramento, CA 95814
916.653.5074
pvadakan@californiarevealed.org

California Revealed

Appendix A: Vendor Workflow and Timeline

Process	Time Frame
California Revealed (CA-R) calls for nominations	60 days
CA-R and State Library review nominations	30 days
CA-R sends award letters to partner organizations	30 days
Partner ships originals to CA-R	30 days
CA-R checks inventory against award letter	Upon receipt of originals
CA-R processes and ships batches of originals (minimum 3 partners per batch) to vendor in receipt order	30-90 days
Vendor digitizes physical materials upon receipt (for most materials)	8-10 weeks
Vendor QAs files (for most materials)	4-6 weeks
CA-R QC's new files and metadata and sends feedback to vendor per batch/partner. CA-R uploads access files and metadata online per partner and sends first email notification to partner to check files.	60 days upon receipt of files from vendor
Partner QC's, updates metadata, checks image quality of access files, and sends feedback to CA-R. CA-R emails follow-up to partner to order backup files from vendor.	30 days upon delivery of email to partner
CA-R responds and relays technical questions from partner, if any, to vendor	30 days upon delivery of email to partner
Vendor responds and provides replacement files if requested by CA-R	Within 30 days of receiving CA-R feedback
Vendor sends CA-R invoice for digitization, including additional costs for rework if needed	30 days upon receipt of CA-R feedback
CA-R QC's replacement files	15-30 days upon receipt of replacement files
CA-R signs off on files and approves LTO creation	Within 90 days of CA-R's receipt of files, including rework
CA-R pays invoice for digitization	30 days upon approval of replacement files
Vendor returns originals directly back to partner and provides backup files, if requested by partner	1-2 weeks following final approval
Vendor ships final LTO deliverables to CA-R	Within 30 days following final approval
Vendor sends CA-R invoice for LTO	30 days upon receipt of LTO
CA-R QC's LTO and approves LTO deliverables	30 days upon receipt of LTO