Demystifying Digitization:
Planning a Digitization Project from Start to Finish

An Introduction to California Revealed

Thursday, June 29th, 2023
12pm - 1:30pm
Webinar Agenda

- Welcome + Introductions
- Brief History of *California Revealed*
- Why digitize?
- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Preservation + Access
- Q & A
California Revealed is an initiative of the California State Library that helps public libraries, archives, museums, historical societies, and other heritage groups **digitize, preserve, and provide online access** to materials documenting the state's history, art, and cultures.

Ahora Es Cuando, 1976/1977~
Center for the Study of Political Graphics
Our Team

Director
Pamela Vadakan

Outreach Coordinator
Guadalupe Martinez

Audiovisual Digitization + Preservation
Shahed Dowlatshahi – Manager
Willow Germs – Assistant
Alyse Hebert – Intern

Print Digitization + Preservation
Liz Seeley – Manager
StarRose Keyes-Lebergott – Assistant

Digital Preservation + Repository
Megan Lohnash – Manager
Natalie Martinez – Assistant
Opportunities to Participate

Digitization and Preservation Assistance

Description Assistance: Cataloging California

Memory Lab Training

Community Archiving Workshops

Barn owl raised in Topanga, California, 1960
LA County Library
Participating Organizations

400+ partners across the state!

African American Museum and Library
Autry Museum of the American West
The Bancroft Library, UC Berkeley
Bishop Chamber of Commerce
Butte County Library
California Academy of Sciences
California Historical Society
California State University, Sacramento
Center for Asian American Media
Center for Sacramento History
Daly City Public Library
Dixon Public Library
East Bakersfield High School
Eastern California Museum
Fresno County Public Library
Glendale Library
Graduate Theological Union
Headlands Center for the Arts
Inyo County Free Library
Kern County Library
Labor Archives, SF State University
Madera County Library, California History and Family Research Room
Manzanar National Historic Site
Napa County Historical Society
Oakland History Room
ONE National Gay & Lesbian Archives
Pacifica Radio Archives
San Diego Air and Space Museum
Shoshone Museum
UCLA, Ethnomusicology Archive
Upland Public Library
Whittier Museum
Yolo County Archives
YOU!
Why digitize?

- Access
  - Connect communities to local history
  - Provide researchers with online access
  - Publicize your organization’s collections
  - Identify and describe your collections

Hawksworth Home Movie, 1951
Madera County Library

https://californiarevealed.org/islandora/object/cavpp%3A161621
Why digitize?

- **Access**
  - Connect communities to local history
  - Provide researchers with online access
  - Publicize your organization’s collections
  - Identify and describe your collections

- **Preservation**
  - Opportunity to assess your collection
  - Preserve fragile or deteriorating materials
  - Address obsolescence (esp. magnetic media crisis)
  - Minimize risks of environmental threats and disasters
  - Provide long-term access

Hawksworth Home Movie, 1951
Madera County Library

https://californiarevealed.org/islandora/object/cavpp%3A161621
Planning for a digitization project . . .

- Assess the overall state of your collection
- Identify materials
- Set digitization standards
- Determine digitization workflow, equipment, software, roles
- Set quality control standards
- Plan for long-term storage + digital preservation
- Plan for how the digitized materials will be accessed
Digitization Steps

- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Digital Access + Preservation
Inventory + Assess

What do you have and where is it stored?

- Conduct a physical inventory of your materials to identify content for digitization
  - Go into the stacks, open boxes, and confirm what’s there!
- Use information gathered to set priorities
- Use your inventory to create initial descriptions (metadata) of the materials
  - Inventory can be used to create or update a finding aid or catalog records and for preservation planning, including digitization

University of Texas Library Storage Facility, Austin, TX
Levels of Description for Conducting an Inventory

Box Level

Folder Level

Item Level
Inventory + Assess: Format Identification
Why take the time to inventory?

- Inventorying is a first step in planning for current and future storage and preservation needs.

- Effective planning depends on knowing the extent and condition of your collection.
  - If you don't know what you have, and where it is, you won't be able to care for it, advocate for it, and fundraise.

- The description of your physical collection will be the basis of the description for your digital collection.

Jimmie Barroga and Gabriel Madriga, 1931
Center for Asian American Media
Digitization Steps

- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Digital Access + Preservation
Consider Preservation

Is the object at-risk?

- Archival: unique, original, master, rare, or “best available”?
- Obsolete format?
- Damaged or deteriorating?
- At-risk if use continues?
- At-risk if disaster strikes?
- For example, magnetic media, such as cassette tapes, are quickly deteriorating
Prioritize + Select

Consider Presentation

Would the object lend itself well to digitization?

- Able to be viewed by the public?
- Underutilized or underrepresented?
- Limited due to institutional factors?
- Limited due to format?
  - e.g. audiovisual, oversized, film negatives, slides

https://californiarevealed.org/islandora/object/cavpp%3A35307
Prioritize + Select

Consider Use
Will people use the object?
Is the object “important”?

- Historically significant?
- Requested by users?
- Part of a comprehensive collection?
- If damaged or lost, would you spend resources to replace it?
- Non-commercial or out of copyright?
- Your responsibility?

https://californiarevealed.org/islandora/object/cavpp%3A112324
Digitization Steps

- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Digital Access + Preservation
Metadata Records

- **Metadata** is structured, digital information that describes the nature of something.

- **Item-level description is ideal**
  - This improves accessibility, usability, discoverability, etc.

- A metadata record connects a physical object to its digital object (.xml record)

---

DATA ABOUT DATA

https://californiarevealed.org/islandora/object/cavpp%3A79030
Types of Metadata in Digitization

**Descriptive**
Enables *identification* and *discovery* of resources, such as title, creator, content, subjects, etc.

**Technical**
Includes *properties of a digital file*, such as format, size, and the *hardware* and *software involved* in the creation and/or maintenance of the file

**Administrative**
Relates to the *management of resources*, such as storage location, current workflow states, preservation information, etc.
REQUIRED METADATA FIELDS

- Institution
- Title
- Partner-supplied ID
- Created Date
- Creator(s)
- Significance/Priority
- Description
- Condition
- Media Type
- Format
- Extent
- Dimensions (Size)*
- Duration (Runtime)**
- Copyright Statement
RECOMMENDED METADATA FIELDS

- Generation
- Subject Topics
- Subject Entities
- Spatial Coverage
- Temporal Coverage
- Genre (AV only)
- Language
- Country of Creation
- Transcript
- Item Annotations
- Partner Administrative Notes

Location
RECOMMENDED METADATA FIELDS

- Generation
- Subject Topics
- Subject Entities
- Spatial Coverage
- Temporal Coverage
- Genre (AV only)
- Language
- Country of Creation
- Transcript
- Item Annotations
- Partner Administrative Notes
- Location
<table>
<thead>
<tr>
<th>California Revealed 2023/2024 Print Nominations</th>
<th><strong>Fields highlighted in red are not required. Use a selection for each:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Institution</strong> (1.1)</td>
<td><strong>Call Number</strong> (1.2)</td>
</tr>
<tr>
<td>Partner organization name</td>
<td>Field will be viewable to public once published.</td>
</tr>
<tr>
<td>Examples:</td>
<td>Examples:</td>
</tr>
<tr>
<td>Bishop Chamber of Commerce and Visitor Center</td>
<td>California Daily Herald 1930-05-03</td>
</tr>
<tr>
<td>California State Library</td>
<td>Damaged train at Donner Pass</td>
</tr>
<tr>
<td>Cal Fire Museum</td>
<td>USBF equipment building, California Division of Forestry, 1954</td>
</tr>
<tr>
<td>California Youth Authority work at Byers Ranch, No. 1104</td>
<td>California Youth Authority work at Byers Ranch, No. 1106</td>
</tr>
<tr>
<td>Cal Fire Museum</td>
<td>California Youth Authority camp at Whitemore, logging truck and equipment</td>
</tr>
</tbody>
</table>

**Metadata Records:** Spreadsheet to Publication

1. **Institution** (1.1) - The institution name will be viewable to the public once the publication is finalized.
2. **Call Number** (1.2) - A unique identifier, in the form of a Call Number or Temporary Identifier, is required to distinguish your items in our system. Do not use the same Call Number for multiple entries. If you provide a Call Number, you no longer need to provide a Temporary Identifier.
3. **Temporary Identifier** (1.3) - A unique identifier, in the form of a Call Number or Temporary Identifier, is required to distinguish your items in our system. Do not use the same Temporary Identifier for multiple entries. If you provide a Temporary Identifier, you do not need to provide a Call Number. (Field will not be viewable to the public.)
4. **Created Date** (1.4) - Date the resource was created.
5. **Creators** (1.5) - A person, family, or group responsible for, or involved in, the creation of the material, such as the author or photographer.
6. **Significance** (1.6) - A statement regarding the significance of the resource to California and/or local history. Justify why the object should be preserved and made accessible for future generations.
7. **Description** (1.7) - Summary about the intellectual content of the resource, including any contextual information necessary to understand the importance of the object from a research perspective.
8. **Condition** (1.8) - A short descriptive statement of the material’s physical condition. (Field will not be viewable to the public.)
9. **Condition Notes** (1.9) - A short descriptive statement of the material’s physical condition. (Field will not be viewable to the public.)
<table>
<thead>
<tr>
<th>Institution</th>
<th>Title* (1.2)</th>
<th>Call Number* (1.3)</th>
<th>Created Date* (1.4)</th>
<th>Creators* (1.5)</th>
<th>Significance* (1.6)</th>
<th>Description* (1.7)</th>
<th>Condition* (1.8)</th>
<th>Media Type* (1.9)</th>
<th>Format* (1.10)</th>
<th>Total Number of Parts* (1.11)</th>
<th>Type of Parts*</th>
<th>Length* (1.12)</th>
<th>Width* (1.12)</th>
<th>Unit of Measurement*</th>
<th>Copyright Statement* (1.13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cal Fire Museum</td>
<td>Schoolhouse Peak Lookout, repeater, power line and radio pole, 1950</td>
<td>BC-II-3149</td>
<td>1950-07</td>
<td>Siler, William</td>
<td>The &quot;Mark Cards&quot; collection of California Division of Forestry records housed at the Cal Fire Museum was created by Division of Forestry photographers who were tasked with documenting facilities and activities for the Division's internal records. The collection documents early fire control and lookout facilities, the control activities and wildlife prevention efforts by the California State Division of Forestry in the 1940s to 1960s. Some facilities still exist. Others have been built or remodeled or near the original site. The collection provides a snapshot in time of early state fire and forestry activities, and has been used in assessing the historic potential of some buildings and facilities still in use. The California Division of Forestry eventually became the Department of Forestry and Fire Protection, and has been known as CALFIRE since 2006.</td>
<td>Good</td>
<td>Still Image</td>
<td>Card</td>
<td>1</td>
<td>Page</td>
<td>9</td>
<td>5.25</td>
<td>in</td>
<td>Public Domain. No restrictions on use.</td>
<td></td>
</tr>
</tbody>
</table>
## Metadata Records: Spreadsheet to Publication

**Schoolhouse Peak Lookout, repeater, power line and radio pole, 1950**

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Schoolhouse Peak Lookout, repeater, power line and radio pole, 1950</td>
</tr>
<tr>
<td>Series Title</td>
<td>Black Card Photos</td>
</tr>
<tr>
<td>Date Created</td>
<td>July 1950</td>
</tr>
<tr>
<td>Description</td>
<td>This photograph of the Schoolhouse Peak Lookout in Humboldt County, California and the adjacent radio repeater, power line and radio pole was taken in July 1950 by William Siler. The photograph is from the &quot;Black Cards&quot; collection of California Division of Forestry records housed at the Cal Fire Museum. The collection was created by Division of Forestry photographers who were tasked with documenting facilities and activities for the Division's internal records. The collection documents early fire control and lookout facilities, fire control activities and wildfire prevention efforts by the California State Division of Forestry in the 1940s to 1960s. Some facilities still exist. Others have been rebuilt or remodeled at or near the original site. The collection provides a snapshot in time of early state fire and forestry activities, and has been used in assessing the historic potential of some buildings and facilities still in use. The California Division of Forestry eventually became the Department of Forestry and Fire Protection, and has been known as CalFIRE since 2006.</td>
</tr>
<tr>
<td>Creators and Contributors</td>
<td>Creator: Siler, William</td>
</tr>
<tr>
<td>Subject Topic</td>
<td>Fire lookouts</td>
</tr>
<tr>
<td>Subject Entity</td>
<td>California, Division of Forestry</td>
</tr>
<tr>
<td>Place(s)</td>
<td>Humboldt County, Calif.</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Media type</td>
<td>Text</td>
</tr>
<tr>
<td>Format</td>
<td>Card</td>
</tr>
<tr>
<td>Extent</td>
<td>1 Page</td>
</tr>
<tr>
<td>Dimensions</td>
<td>9 x 5.25 in</td>
</tr>
<tr>
<td>Generation</td>
<td>Original</td>
</tr>
<tr>
<td>Copyright Statement</td>
<td>Public Domain. No restrictions on use.</td>
</tr>
<tr>
<td>Country of Creation</td>
<td>United States</td>
</tr>
<tr>
<td>Source Institution</td>
<td>Cal Fire Museum</td>
</tr>
<tr>
<td>Link to Internet Archive</td>
<td><a href="https://archive.org/details/casbcfm_000142">https://archive.org/details/casbcfm_000142</a></td>
</tr>
</tbody>
</table>
As part of your application, CA-R requires metadata records for the items you wish to digitize:
- Must submit using CA-R’s Nominations Spreadsheet

Metadata records must follow CA-R’s Metadata Guidelines:
- Outlines specific formatting needs
- Offers guidance regarding the creation of titles, descriptions, and other free text fields
Digitization Steps

- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Digital Access + Preservation
Prepare for Digitization

- Pull selected items from their storage area
- Review the materials against their metadata records
- Identify and record specific digitization and/or special handling instructions
- Remove rubber bands, staples, clips, binders, fasteners, plastic bags, etc.
- Remove materials from mylar sleeves and folder them
- Tape down ends of open reels
- Arrange and clearly label materials to reflect the order you expect them digitized
When materials arrive at California Revealed . . .

- Confirm delivery, then conduct an inventory of the shipment to make sure all objects are accounted for and have arrived safely.
- Collection materials are processed for digitization separately through our Audiovisual, Print, and Digital “streams”
  - Audiovisual materials are shipped to an outside vendor for digitization
  - Print materials are either captured onsite or sent to a vendor
  - Digital materials are “captured” onsite
Digitization Steps

- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Digital Access + Preservation
Digitization Standards: Create a Project Plan

Capture workflow
Specify how the materials will be captured: equipment used, capture style

Final deliverables
Files created for each object: Preservation and access files, XML metadata files, checksums
Technical specifications for the files: File type, image/sound quality, color vs. greyscale, etc.

File naming conventions
Be consistent and use unique identifiers
Example: YYYYMMDD_seriestitle_#######

Metadata + metadata standard(s)
Information about the files: descriptive, technical, administrative, preservation
Map to a standard XML schema such as DublinCore or PBCore

Data storage
Estimate how much storage you’ll need
Specify the type of storage media you want the files delivered on

Quality Control
Who will check the files and to what extent

Cost
Timeline
Roles
Digitization Standards: In-house or Outsource?

**In-house Challenges:**
- Initial costs
- Requires staff

**In-house Advantages:**
- Total control of the process
  - Develop new skills

**Outsource Challenges:**
- Loss of control

**Outsource Advantages:**
- Less investment of resources (equipment and staff)
  - Potentially a lab can handle majority of formats
  - Bulk rate option
Digitization Standards: Capture Equipment

- Moving images
  - Playback equipment
  - AD converter (hardware + software)
  - TBC (time-based corrector)

- Sound recordings
  - Playback equipment
  - AD converter (hardware + software)

- Print (text and still image) materials
  - Digital camera + copy stand
  - Specialized film scanners for microfilm, photographic negatives, and slides
  - Photo editing software

- Born-digital materials
  - Digital converters (hardware + software)
**Preservation** = file containing the original, highest quality version of the data, *uncompressed*; used for long-term storage

**Access** = file containing the same “information”, but a smaller amount of data, *compressed*; used for streaming, circulating
Digitization Standards: Audiovisual

Moving image formats such as: VHS, Betacam, ¾” U-matic, 16mm film, MiniDV, DVD, etc.

Preservation = uncompressed; long-term storage
Digital File Type = .mov

Access = compressed; streaming, circulating
Digital File Type = .mp4
Digitization Standards: Audiovisual

Audio formats such as: audiocassette, ¼” open reel, DAT, CD, etc.

Preservation = uncompressed; long-term storage
Digital File Type = .wav

Access = compressed; streaming, circulating
Digital File Type = .mp3
Digitization Standards: Print

Print (still image and text) formats such as: books, pamphlets, slides, photographs, scrapbooks, etc.

Preservation = uncompressed; long-term storage
Digital File Type = .tif

Access = compressed; streaming, circulating
Digital File Types = .pdf, .jpg
Digitization Standards: **Digital**

**Born-digital formats** such as: files created and/or stored on a phone or computer, photos + videos taken using a digital camera or smartphone, emails, websites, etc.

**Preservation** = **uncompressed**; long-term storage

*Ideal* Digital File Types = `.mov, .wav, .tif`

**Access** = **compressed**; streaming, circulating

Digital File Types = `.mp4, .mp3, .pdf, .jpg`
Evaluating Digitization Standards: Quality Control (QC)

What are the goals of Quality Control (QC)?

- **Technical specifications are correct**
  - Digital file formats, # of copies, naming conventions

- **Digital copy is true to the original**: vendor didn’t introduce any errors
  - Refer to condition notes gathered during the inventory process

- Add, correct, and enhance **descriptive metadata**

- Add, correct, and enhance **technical metadata**

- Wrap up loose ends in **administrative notes**
Digitization Steps

- Inventory + Assess
- Prioritize + Select
- Describe + Generate Metadata
- Prepare for Digitization
- Digitization Standards: Capture + Evaluate
- Digital Access + Preservation
Your collections don't stop at California Revealed...
How do you ensure long-term access?

Digital preservation
Create a Digital Preservation Plan

- **LOCKSS** (Lots of copies keeps stuff safe)
- NDSA Levels of Preservation
- Digital Readiness Levels

Remember: 3–2–1!

Permalink: https://recollectionwisconsin.org/toolkit/storage
Your Digital Preservation Plan

- Order and store files on hard drives (avoid DVDs and CDs)
- Clearly label and organize hard drives
- Duplicate the files on a second drive and/or upload to the cloud
- Consult with IT dept (if available) about server storage
- Check files at least once a year to make sure hard drives are okay

Digital Readiness Toolkit:
recolletionwisconsin.org/digital-readiness-toolkit
CA-R’s Digital Preservation Plan

- Write to LTO data tape and store in 2 different geographic locations to mitigate risk of loss
  - Preservation copy
  - Access copy
  - MD5s
  - XML
- Internet Archive
  - Access copy + MD5s + XML
- Amazon S3 - for californiarevealed.org
  - Access copy + MD5s + XML
- Annual integrity checks on files on all files
Digitization & Preservation Assistance Program Timeline

Applications Due Friday, July 28, 2023

Digitization awards announced

Submit application + nominations

Prepare items + ship to CA-R

Files received + published online

Files approved

Files ordered for digital preservation

Items returned

Review files for Quality Control (QC)

Ongoing digital public access!

Digitize items + ship to CA-R

Files received + published online

Files approved

Files ordered for digital preservation

Application Form:
https://tinyurl.com/2023CRdigitize-preserve

Nominations Spreadsheet:
https://tinyurl.com/2023CRnoms-form
Guadalupe Martinez
Outreach Coordinator
gmartinez@californiarevealed.org

Pamela Vadakan
Director
pvadakan@californiarevealed.org

team@californiarevealed.org

916.603.6719
californiarevealed.org/partners

Keep in Touch!

https://californiarevealed.org/islandora/object/cavpp%3A28572
Q & A
What steps were you most interested in learning about or reviewing?

Do you feel more confident to start a digitization project?

Do you feel prepared to collaborate with California Revealed?

What challenges do you foresee in doing digitization, digital preservation, and access projects?

*Reflections on any part of the webinar are welcome at this time!*