

**Suggestions for Metadata Cleaning for Sequoia and Kings Canyon
Roberts and Historic Images Collections**

Bailey Berry, MLIS

UCLA Graduates School of Education and Information Science

Collections Background:

Sequoia and Kings Canyon National Park (referred to here as SEKI) has several digitized and digital collections of images, which it has stored and made available through the NP Gallery portal. NP Gallery is a digital asset management system that was designed to aggregate digital and digitized collections from the National Parks. SEKI has four digitized collections which it has made available or will make available through NP Gallery: The Roberts Collection, Historic Images, Resource Management Slides, and Civilian Conservation Corps. For the purposes of this project, I will be focusing on the Roberts Collection and Historic Images.

The Roberts Collection consists of materials belonging to Henry E. Roberts, who was photo concessioner for General Grant National Park between 1900-1930. Included in the collection are glass plates, prints, stereographs, slides, and correspondence relating to the history of SEKI ranging from 1888-1960. The materials digitized and available on NP Gallery are the Slides and Prints Series in the collection.

Historic Images represent a collection of images from the early history of the park taken by a variety of photographers, some of whom are unknown. They represent subjects related to park operations such as buildings, events and museum collections.

Project Overview:

Initial metadata creation for these collections on NP Gallery was done by a small staff with limited time, thus metadata for these collections succumbs to what describe as “Lag”- “The dissemination of metadata is not necessarily synchronized with the dissemination of the object to

which it applies”.¹ NP Gallery uses a local schema for metadata terms and arrangement. SEKI’s use of these terms is not consistent and should be clarified and standardized across collections. Description of objects within these elements is often incomplete, with vital information about the contents of photographs or their date of creation absent. SEKI would like to increase the discoverability of its collections by allowing California Digital Library to harvest it and include it in Calisphere. Before this process can be completed, metadata must be cleaned and enhanced both to help with the harvesting process and to increase access and discoverability both in NP Gallery and on Calisphere.

Recommendations for enhancing this metadata have been made based on access to the front-end user experience of NP Gallery and available metadata. They will be amended once more information is available.

Intended Audience/ Use Case:

The Roberts and Historical Images Collections have three principle uses.

- NP Gallery is useful for potential or past visitors to the park who want to learn more about it to plan their visit or who have a connection with the park and would like to understand more about its history. For example, someone who had hiked in a part of SEKI may want to look through the Historic Images or Roberts Collection files to understand how related landmarks have changed over time.

¹ Bruce, Thomas R. and Diane I. Hillman. “The Continuum of Metadata Quality: Defining, Expressing, Exploiting”. *Metadata in Practice* American Library Association (Chicago, 2004), 238-256.

- The National Parks collect material and have a large following from enthusiasts for nature and nature photography. Both collections offer beautiful examples of historical park photography.
- Both collections also document elements of California's history, especially the early history of logging in California.

For these three use cases, it is important that the associated metadata include relevant information about where the photographs were taken, any contextual information about events, landmarks and people featured in photographs, the dates of creation and metadata about the original format of the digitized file.

Existing Schema:

The metadata schema in NP Gallery has its own set of terms. The intended use for these terms is open-ended and both collections have decided to use some of them in different ways. For example, "Description" is used by the Roberts Collection for comments about processing the images, which I believe are in Henry E. Roberts' own words and relate to the margins and exposure time. Historical Images has used this section to define the collection, populating "Description" with the phrase "SEKI Historical Images". In another example, collections have chosen a different term to describe the image creator: The Roberts Collection has chosen the term "PhotoCredit" while Historical Images uses the term "Photographer". Other uses of terms seem inappropriate or like a poor fit for their content. For example, the Roberts Collection uses "AtlText" to provide information about the Collection and Series. Finally, some metadata is repeated across several terms. This is especially true in historical collections. In some instances, the same information included in the "title" is included again in "Comment", "Image Details" and "Locations". As Bruce and Hillman describe, this means the metadata lacks logical

consistency both within each collection and across the two collections. “Users expect to be able to search collections of similar objects using similar criteria, and increasingly they expect search results and indicative indexes to have similar structures and appearance.”²

Figure 1 is a chart with the current terms used and the ways each collection is defining and populating them. In order to provide the user with a coherent and logically consistent set of metadata, Figure 2 is a proposal for how SEKI should define terms for its metadata across collections, how they would map to the Calisphere schema and whether terms are relevant for the Calisphere catalog. The proposed changes should allow for a more accurate set of metadata elements which more logically describe the collections. For example, the creator of the photograph is uniformly referred to as “PhotoCredit”. “Comment” is used to retain the original language of the cataloger and of Henry E. Roberts in describing the images and “Description” contains a more accessible and informative piece of information about the contents of the image. “Source Information” is added to include information about the original type of material digitized (Negative, Print, Glass plate, etc.). The misleading “Camera Information” element is replaced with the more accurate “Scanner Information”. “Alt Text” has been removed- series and subseries information may be helpful to include under an element such as “source location” if available. This information is only useful to a user if properly contextualized by its schema element. Series and subseries are also used to populate keywords, which helps users to locate materials related to images they are interested in.

² Bruce, Thomas R. and Diane Hillman

Metadata Content/ Resource Description Issues:

The Roberts and Historic Images collections were digitized and added to NP Gallery for discovery using a limited staff and with incomplete source metadata. Thus, metadata for images in both the Roberts Collection and the Historic Images Collection is often incomplete and inaccurate. While metadata for these collections will continue to be limited by a lack of source information, there are ways that the content of metadata elements for both collections can be enhanced to increase their usefulness and relevance.

There are some simple inconsistencies that can be easily corrected in the content metadata. The format of item titles is not consistent across collections. For example, in the Robert's Collection, specific waterfalls are given their full name ("Tornado Creek Falls"), while unknown waterfalls are labeled "falls", "Falls", or "Waterfall". Some animals are labeled "deer", "Deer 2" or qualified with an action "Feeding Deer" or "Woman feeding deer". Consistency in capitalizations and use of numbers could and should be corrected. Titles in Historic Images are grouped into general series ("Vehicles and Equipment", "Aerial Views", "Buildings and Utilities") followed by more specific information ("Vehicles and Equipment, Trucks, Ash Mountain Headquarters"). These titles have many inconsistencies in the way they are formatted – which words are capitalized, the use of commas vs. periods, whether the series information goes before or after the location of the photograph. These should also be corrected and unified. Assuming that these titles were devised by catalogers adding metadata to digitized images, titles should be edited to be more informative and uniform. Cataloging of Cultural Objects (CCO) offers some guidance on how to construct titles for objects that have no existing one. Many of these titles are long and descriptive, CCO emphasizes the need for more concise titles and

recommends that descriptive information be included in descriptions or subjects, instead.³ Titles from the Roberts Collection could be more descriptive. For example, “Falls” could be qualified with a type of fall, its location or name.

There should also be better conformance with metadata value standards within metadata records. Create Date in the Roberts Collection should conform to the standard date format used by the Historic Images collection (mm/dd/yyyy) instead of the form it is currently using (yyyymmdd). Time Submitted, Last Edited, Embedded Timestamp and Metadata date all use different date formats, as well, and should be unified. Geographic Locations should also be unified and clarified. The Roberts Collection gives the latitude and longitude for both Kings Canyon National Park and Sequoia and Kings Canyon National Park. Historic Collection does not give any coordinates, but does include what “locations” are present in the photograph along with the county information and state for Sequoia and Kings Canyon. Both collections must determine what location information to include and which order. In their Geographical and Temporal Guidelines for mapping, Digital Public Library of America recommends Country, State, City, County and Coordinates.⁴

Much of the material described in both collections contains names and terms specific to Sequoia and Kings Canyon National Park and is, therefore, not described by controlled vocabularies, authority files or thesauri. However, there needs to be control for how certain subjects or terms are used across these collections. For example, both “Sequoia & Kings Canyon” and “Sequoia and Kings Canyon” are used. The collections are titled the abbreviated

³ “Rules for Title” *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*. Edited by Murtha Baca, et al. American Library Association (Chicago, 2006): 58.

⁴ “DPLA Geographic and Temporal Guidelines for MAP 3.1”. *Digital Public Library of America*. (December, 2017). http://bit.ly/dpla-geo-styleguide-3_1

form of the park name: “SEKI Historical Images” and “SEKI Roberts Collection” instead of using the full name, which can make browsing for these collections difficult: users may not know that SEKI is synonymous with Sequoia and Kings Canyon. I propose listing the names of these collections as “Sequoia and Kings Canyon Roberts Collection” and “Sequoia and Kings Canyon Historic Images”. Some cultural terms used are outdated and should be improved with a current preferred term. For example, the word “Indian” is used instead of “Native American” for the title “Indian Baskets”. In this example, the use of the preferred term used in both the Library of Congress Subject Headings and in the Getty Art and Architecture Thesaurus would be “Native American Baskets”.

Finally, inaccuracies and lack of detail should be corrected in the metadata. “CreateDate” in the Roberts Collection is not only in a different format, but is an incorrect date from 2018. While the exact year of the creation of each negative may be unknown, it would be more helpful to include the date range of the collection (1888-1950) or simply, “Unknown”. I would argue that in this case some information, even a wide date range, is better than no information. The descriptions used in each collection should be edited to include more useful information. The descriptions in the Roberts Collection include notes from Henry E. Roberts about how negatives should be processed or about the contents of an image. Descriptions in the Historic Images simply state “SEKI Historic Images”. In order to be more clear and useful to a user, descriptions should be rewritten to include any known details about the image in clear language. Information can simply be taken from the Title, Comment, Original Description or from the Finding Aid associated with the collection. The information currently populating the “Description” element should then be moved to more appropriate elements in the metadata schema.

Discoverability:

The use of “Keywords” should help users browse for certain subjects within NP Gallery and within the collection. The current keywords used, however, are not necessarily useful. For example, the keywords “Water” and “Building” lack specificity. The keyword “H. E. Roberts” is used instead of “Henry E. Roberts” to describe the Roberts Collection, which is inconsistent with how his name is written throughout the rest of the catalog. “Box 1” may help link objects in a particular location in the archive, but are not helpful as a keyword for searching. Keywords are essential in NP Gallery. A user logging in is prompted to use a keyword search to find any item, which they can refine by “state” and “Park”. It is important that keywords include the controlled versions of the object title, the creator, the collection, the series and any names or landmarks included in the description. They should contain determined preferred terms and common equivalents (“Henry E. Roberts”, “H.E. Roberts”, and “Roberts” or “National Park Service Rustic” and “Parkitecture”).

A user logging in to NP Gallery who wants to find material only related to Sequoia and Kings Canyon can search by “Park”. However, this brings up any image from the park in no particular order- not organized by collection, subject or time period. If a User wants to filter results, they can only do so by “Categories” (Scenic, Historic, Events, etc.), “NPS Units and Sites”, which includes all National Parks referenced in the collection, or “File Type” (Image, Document or Album). Within NP Gallery, better use of Keywords and the Albums feature can help organize items from the same collection and series to help with discovery. If a user wants to search for a “Historic” image or document, an entry they find useful should have keywords that point to related images and should be part of an album that reflects its collection and series.

Aggregation with Calisphere is an exciting way for Sequoia and Kings Canyon to increase discovery of its collection, and this is why quality metadata is important. By clarifying subject, description, locations, create date and by creating more concise and descriptive titles, SEKI can maximize access to their digital and digitized contents both in Calisphere and for users directed back to NP Gallery.

Rights Metadata:

Calisphere and its associated organizations (The California Digital Library and the Digital Public Library of America) have helped develop Rightsstatements.org in order to unify the rights information being submitted by its partner institutions. According to their use of Rightsstatements.org, partners are encouraged to submit a URI indicating whether the work is under copyright restrictions or not and a Rights Statement or narrative that clarifies any specifics of this status.

Currently, most work in both collections are part of the Public Domain. This is currently listed under “Constraints Information”. Each collection has its own Rights Statement under “Copyright”, which – using different language – clarifies that “Restrictions apply on use and/or reproduction” or that “Permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this website. Digital assets without any copyright restrictions are public domain.” This language is confusing, especially when constraints information for these works is listed as “Public Domain”, which typically means that there are no copyright restrictions on the works listed. If there are restrictions on the use of these items even though they are in the Public Domain, these restrictions should be clear in the Copyright Statement. If these restrictions only apply to certain items in the collection which are copyrighted, this should also be clear to the user. Finally, users should be directed to where to

find more information about restrictions, either through a link to a page with more details – as recommended by rightsstatements.org⁵ or by directing the user to the contact information included in the metadata under “Contacts”. I have included a sample statement in Figure 2 under “Constraints Information”.

Workflow/ Project Management:

I believe that improving the quality of this metadata is important, especially as Sequoia and Kings Canyon move to make their collections more available through Calisphere. I also recognize that it requires a lot of work. There are some ways that can help streamline the process. Metadata can be taken from the front end and converted into a .csv format through Simple Text and updated to OpenRefine for bulk editing. Alternatively, metadata already compiled in a .csv file can easily be updated and edited and will streamline the process even further. Once metadata has been cleaned through OpenRefine, my hope is that it can be updated in bulk. If not, OpenRefine will still make individual changes to records easier to manage and aggregating metadata will help identify more large-scale inconsistencies. Minor changes to formatting, such as unifying the date format, are easy to make and can help foster trust from the user in the accuracy of the metadata. More complicated changes- such as rewriting the titles and descriptions- can be done simply using existing metadata included in titles, comments, image details and keywords and expanded later through research, if necessary.

I understand that initial NP Gallery metadata was limited by the amount of information available in the physical archive. To continue improving metadata for these collections in the future, crowdsourcing could be used to provide more detailed descriptions of many of the

⁵ “Contractual Restrictions”. *Rightsstatements*. [Rightsstatements.org](https://rightsstatements.org/page/1.0/?language=en#collection-nc).
<https://rightsstatements.org/page/1.0/?language=en#collection-nc>

landmarks included in the Roberts Collections. For example, visitors to the park may be able to help identify waterfalls or other landmarks featured in photographs, which would be time-consuming for an individual cataloger to research. Crowdsourcing could also help to improve coordinates of landmarks included in object records, which would be valuable information for many users who would like to locate these areas of the park during their visit.

Figure 1

Label	Description	Example	Appears in Both?
Asset ID	The unique digital hash of the photograph	e773efdbb4424875a2e1300dc90daaec	Yes
Title	Title of digital object, describes contents of photograph	Falls (Roberts Collection); Buildings and Utilities, tent frame seasonal housing (Historical Images)	Yes
AltText	Information about archive collection and series related to image	SEKI Roberts Collection Series: 2 Negatives; 2 Falls	No, only Roberts Collection
Description	Not standardized. Varies from information about type of resource ([print]) to technical information about photograph (Roberts Collection); labels the collection (Historic Images)	North Grove, 13x17 - contrast PMC, Expose full neg. to 11/2 min, shade big tree 1/2 min longer Then shade center of g - ?, 0+ trees 1 1/2 min -, Exposure 3 1/4 min altogether, 2 - 6 1/2 x 8 1/2 will borders, 2 - 5x7 without " " (Roberts Collection); SEKI Historical Image (Historical Images)	Yes
Comment	Notes from cataloger	Unknown Photographer, Lodgepole, SNP, Buildings and Utilities, tent frame seasonal housing. 490100	No, only Historical Images
Image Details	Information about archive series related to image (Roberts Collection); Alternative version of image title (Historic Images)	Series: 2 Negatives; 2 Falls (Roberts Collection); Buildings and Utilities, tent frame seasonal housing (Historical Images).	Yes
Publisher	Claims ownership of National Park Service	U. S. National Park Service	No, only Roberts Collection
Copyright	Rights Statement	Restrictions apply on use and/or reproduction (Roberts Collection); Permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this website. Digital assets without any copyright restrictions are public domain (Historical Images)	Yes

PhotoCredit	Image creator	Henry E. Roberts	No, only Roberts Collection
Photographer	Image creator	Unknown Photographer	No, only Historical Images
Constraints Information	Rights Statement	Public Domain	Yes
Keywords	Hyperlinks to image collections that contain related keywords, describe archival location (Box 1), contents of photograph and types of resource (negative)	Falls , H. E. Roberts , Kings Canyon National Park , Negatives , Roberts Collection , Sequoia National Park , Water	Yes
Resource Type	Describes the category of object the resource falls into	Image	Yes
NPS Units	Park location of collection and abbreviation	Sequoia and Kings Canyon National Parks, Code: SEKI Kings Canyon National Park, Code: KICA	No, only Roberts Collection
Locations	The geographic Coordinates of the image (Roberts Collection); More specific information about landmarks or locations in the image (Historic image)	Kings Canyon National Park, California Latitude: 36.8744888305664, Longitude: -118.593803405762 Sequoia and Kings Canyon National Parks, California Latitude: 36.7172584533691, Longitude: -118.537902832031 (Roberts Collection); Lodgepole ; tent frame seasonal housing, Sequoia National Park, Tulare County, California (Historical Images)	Yes
Related Collections	Name of Collection as hyperlink to collection in NP Gallery	SEKI Roberts Collection; SEKI Historic Images	Yes
Create Date	Date original (not digital) image was created- Roberts Collection has non-standardized number format	20180221 (Roberts Collection); 01/01/1949 (Historical Images)	Yes
Contacts	Contact information for the curator of the collection	Person: Curator Organization: US National Park Service Position: SEKI Resource Management Address: Sequoia and Kings Canyon National Park, Three Rivers CA Email: seki/superintendent@nps.gov	No, only Roberts Collection



Intended Audience	Refers to a cataloging category that is unclear	Option 3	No, only Historical Images
Time Submitted	Date digital copy was added to repository	Wednesday, January 11, 2017 11:22:50 AM	Yes
Last Edited	Date metadata of digital copy was last changed	Thursday, June 6, 2019 10:37:17 AM	Yes
Original File Name	Digital image file name	00075.tiff	Yes
Camera Information	Information about camera used to create image	EPSON EPSON, Perfection V700/V750	No, only Historical Images
Embedded Timestamp	Date digital image was created? (all timestamp dates have the same date and format)	Monday, January 1, 0001 12:00:00 AM	No, only Historical Images
Resource Format	Type of File	Tiff or Jpeg	Yes
File Size	Expressed in bytes	35.9 MB	Yes
Metadata Date	Date original metadata was created?	2015-04-16T11:33:51-07:00	No, only Historical Images
Rating	? Based on a five star system		Yes
Categories	Describes the types of content in the image	Scenic, Historic	Yes
Related Portals	Links to Sequoia and Kings Canyon NP Gallery portal	Sequoia & Kings Canyon National Parks	Yes

Figure 2

Label	Calisphere Term	Description	Example	Appears in Calisphere?
Asset ID	Identifier	The unique digital hash of the photograph	e773efdbb4424875a2e1300dc90daaec	Yes
Title	Title	Title of digital object, describes contents of photograph	Falls (Roberts Collection); Buildings and Utilities, tent frame seasonal housing (Historical Images)	Yes
Description	Description	Additional information about contents of the image, stated concisely and clearly	Image of waterfall from the early development of Sequoia and Kings Canyon (Roberts Collection) Lodgepole Campground, tent frame seasonal housing (Historic Images)	Yes

Comment	N/A	Notes from original catalog entry	North Grove, 13x17 – contrast PMC, Expose full neg. to 11/2 min, shade big tree ½ min longer Then shade center of g - ?, 0+ trees 1 ½ min -, Exposure 3 ¼ min altogether, 2 – 6 ½ x 8 1/2 will borders, 2 – 5x7 withought “ “ (Roberts Collection); Unknown Photographer, Lodgepole, SNP, Buildings and Utilities, tent frame seasonal housing. 490100 (Historic Images)	No, Description has enough information
Publisher	Publisher	Claims ownership of National Park Service	U. S. National Park Service	Yes
Copyright	Rights	Whether the item is under copyright restrictions	Public Domain	Yes
PhotoCredit	Creator	Image creator	Henry E. Roberts	Yes
Constraints Information	Rights	Rights Statement	Digital assets without any copyright restrictions are public domain and can be used freely. Permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this website. Please see “Copyright” and “Contacts” for more information.	Yes
Keywords	Subject	Hyperlinks to image collections that contain related keywords created by crowdsourcing or from original metadata	Negatives, Falls, Henry E. Roberts, Roberts, Waterfall (Roberts Collection); Buildings and Utilities, Tent Frame Seasonal Housing, Buildings, Housing, Tent Frame, National Park Service Rustic, Parkitecture (Historic Images)	Yes, for browsing/ discovery
Resource Type	Type	Describes the category of object the resource falls into	Image	Yes
NPS Units	N/A	Park location of collection and abbreviation	Sequoia and Kings Canyon National Parks, Code: SEKI Kings Canyon National Park, Code: KICA	No, only relevant for internal Park knowledge
Locations	Place	More specific information about landmarks,	Kings Canyon National Park, California Latitude: 36.8744888305664, Longitude:	Yes

		Vegetation Zones, Trails or buildings in the image	-118.593803405762 Sequoia and Kings Canyon National Parks, California Latitude: 36.7172584533691, Longitude: -118.537902832031 (Roberts Collection); Lodgepole ; Tent frame seasonal housing, Sequoia National Park, Tulare County, California (Historical Images)	
Related Collections	Collection and Series Information	Name of Collection as hyperlink to keyword search in NP Gallery	Sequoia and Kings Canyon Roberts Collection (Roberts Collection); Sequoia and Kings Canyon Historical Image Series (Historical Images)	Yes
Create Date	Date Created	Date original (not digital) image was created- Roberts Collection dates are unknown, but date range is helpful	1888-1950 (Roberts Collection); 01/01/1949 (Historical Images)	Yes
Source Information	Source	Information about original type of resource digitized: composition and creation of image	Negative (Roberts Collection); Negative (Historic Images)	Yes
Contacts	N/A	Contact information for the curator of the collection	Person: Curator Organization: US National Park Service Position: SEKI Resource Management Address: Sequoia and Kings Canyon National Park, Three Rivers CA Email: seki/superintendent@nps.gov	No, Publisher information is enough (with option to see original resource in NP Gallery)
Intended Audience	N/A	Refers to a cataloging category that is unclear	Option 3	No, only relevant for internal park knowledge
Time Submitted	N/A	Date digital copy was added to repository	Wednesday, January 11, 2017 11:22:50 AM	No, only relevant for internal park knowledge
Last Edited	N/A	Date metadata of digital copy was last changed	Thursday, June 6, 2019 10:37:17 AM	No, only relevant for internal park knowledge
Original File Name	N/A	Digital image file name	00075.tiff	No, only relevant for internal park knowledge

Scanner Information	N/A	Information about scanner used to create image	EPSON EPSON, Perfection V700/V750	No, only relevant for internal park knowledge
Embedded Timestamp	N/A	Date digital image was created? (all timestamp dates have the same date and format)	Monday, January 1, 0001 12:00:00 AM	No, only relevant for internal park knowledge
Resource Format	Format	Type of File	Tiff or Jpeg	Yes
File Size	N/A	Expressed in bytes	35.9 MB	No, only necessary for image download- which is accessed through NP Gallery
Metadata Date	N/A	Date original metadata was created?	2015-04-16T11:33:51-07:00	No, only relevant for internal park knowledge
Rating	N/A	? Based on a five star system		No, only relevant on NP Gallery
Categories	N/A	Describes the types of content in the image	Scenic, Historic	No, only relevant for NP Gallery
Related Portals	N/A	Links to Sequoia and Kings Canyon NP Gallery portal	Sequoia & Kings Canyon National Parks	No, only relevant for NP Gallery
Albums in Which This Asset Appears	N/A	Links to related albums	Negatives; Negatives, Falls (Roberts Collection); Buildings and Utilities; Buildings and Utilities, Tent Frame Seasonal Housing (Historic Images)	No, only relevant for NP Gallery

References

Bruce, Thomas R. and Diane I. Hillman. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting". *Metadata in Practice*. American Library Association (Chicago, 2004), 238-256.

Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images. Edited by Murtha Baca, et al. American Library Association (Chicago, 2006).

"Contractual Restrictions". *Rightsstatements*. Rightsstatements.org.
<https://rightsstatements.org/page/1.0/?language=en#collection-nc>

"DPLA Geographic and Temporal Guidelines for MAP 3.1". *Digital Public Library of America*. (December, 2017). http://bit.ly/dpla-geo-styleguide-3_1