

Stereograph Card Images Data Package README

Version information

Version 1.1 | Last updated 2024-04-17

- *1.1 (2024-04-17)* LC Labs consolidated coversheet and README content with minor formatting updates
- *1.0 (2022-09-27)* First version

CONTENT ADVISORY

Please note that terminology in historical materials and in Library descriptions does not always match the language preferred by members of the communities depicted, and may include negative stereotypes or words that offend.

For questions or more information about this material, please contact Prints and Photographs Division staff through the [Ask a Librarian service](#).

About the source data or collection

Brief description & background of collection

Stereographs consist of two nearly identical photographs or photomechanical prints, paired to produce the illusion of a single three-dimensional image, usually when viewed through a stereoscope. Typically, the images are on card mounts, but they may take the form of daguerreotypes, glass negatives, or other processes. The Prints & Photographs Division's holdings include images produced from the 1850s to the 1940s, with the bulk of the collection dating between 1870 and 1920, when they were most popular. The online images feature cities and towns around the world, expeditions and expositions, industries, disasters, and portraits of Native Americans, presidents, and celebrities.

Approximately 52,000 stereographs are organized and available in the Stereograph Cards collection for researchers onsite at the Library; a growing proportion are also available online. The Library acquired many of the stereographs through copyright deposit, and has since focused particularly on acquiring views made before 1870—the year that the Library became home to the Copyright Office. Many of these new acquisitions were unpublished or published in small editions by local photographers.

The stereographs are arranged as follows:

- A portion of the organized stereos are filed in the Reading Room by subject, geographic location, or portrait sitter. (Sample call number: STEREO US GEOG FILE - Georgia--Atlanta)
- Another portion of the stereos are organized into groups (LOTs) by subject, photographer, publisher, or donor. (Sample call number: LOT 4164)

More information can be found at: * <https://www.loc.gov/pictures/collection/stereo/> *
<https://www.loc.gov/collections/stereograph-cards/about-this-collection/>

Original format

Mostly photographic prints mounted on cards. Some items are images of original glass plates, some produced anaglyphs (red-blue) acquired by LC appear in collection.

Library of Congress reading room

Prints and Photographs Division, <https://www.loc.gov/rr/print/>

Contact

For more information please contact the Prints and Photographs reference specialists at <https://ask.loc.gov/prints-photographs>.

Metadata type

MARC

Scale of description

A growing proportion of the stereographs have been digitized and described individually (at the item level). Usually known is publisher, date, subject, location. Important stereos may have added subject, authorized name or location entries.

Rights information

The Library's stereograph collection includes thousands of photographs submitted for copyright protection. Those that were copyrighted or published in the U.S. more than 95 years ago are in the public domain because the copyright has expired. The collection also includes images not submitted for copyright that may not have been published. The term of copyright for unpublished images is the life of the creator plus 70 years. For unpublished anonymous works and works where the death date of the creator is not known, the copyright term is 120 years from the date of creation. For more information, see the [Rights and Restrictions page](#).

Digitization information

The Library began putting scans of copy negatives online in the 1990s. Direct digitization in small chunks happened in the 2000s and steady digitization work began around 2013. As part of routine digitization workflows, the stereograph cards were scanned and put online. The loc.gov website makes the digital surrogates available in multiple formats including multiple sizes of JPEG, GIF, and TIFF. As described above, this dataset contains the medium-sized JPEG files.

About this exploratory data package

Stereographs consist of two nearly identical photographs or photomechanical prints, paired to produce the illusion of a single three-dimensional image, usually when viewed through a stereoscope. The Prints & Photographs Division's [holdings](#) include images produced from the 1850s to the 1940s, with the bulk of the collection dating between 1870 and 1920. The online images feature cities and towns around the world, expeditions and expositions, industries, disasters, and portraits of Native Americans, presidents, and celebrities.

They include views from nearly all of the fifty states and the District of Columbia, as well as more than twenty foreign countries.

Please note: the Stereograph Cards digital collection continues to grow as more stereographs are digitized and put online. Therefore, the images in this dataset do not constitute the entirety of what may be available on loc.gov.

What's included?

The data package contains:

- A folder containing 39,597 stereograph item records with their corresponding image (.JPG) files pulled from P&P Stereograph Cards collection (There are some known duplicate images--the same hash but different LOC item ID's)
- `metadata.json`: a JSON file containing the metadata for all 39,597 images in the data set
- `metadata.csv`: a CSV transformation of the original JSON metadata
- `manifest.txt`: a text file listing the id (combination of the image ID and image file name), MD5 hash, and location of the images in the data set
- `README`: An overview of the source data or collection provenance, the contents of the data package, and how the data package was created. Available as `.md`, `.html`, and `.pdf`.
- `dpp`: The data processing plan. Available as `.md`, `.html`, and `.pdf`.
- sample data: 1,000 randomly selected items from the 39,597 set with their corresponding image files have been provided as sample data. Included with this are a

metadata.csv metadata.json, and manifest.txt.

Computational readiness and possible uses

The JPGs in this dataset are flat image files and require datafication for further computational analysis. Possible research uses of this dataset include image clustering, visualization, producing visual 3D renderings and possibly 3D printables, geolocating and mapping interfaces, calculating position/direction of images, or sequencing into timelines.

How was it created?

This dataset was created using the [LOC JSON/YAML API](#) and comprises a scoped portion of the stereographs and not every item in the collection. Subject matter experts were consulted in the creation of a JSON API query

(<https://www.loc.gov/collections/stereograph-cards?dates=1800/1924&fa=access-restricted:false&q=no%20text>) to produce rights free stereographs from the 1850s through 1924, as a subset of what was available online in the collection on loc.gov in August 2022. This original query returned 44,694 results.

The JSON results were then placed in a new JSON structure of key/value pairs where the key is the LCCN and the value its corresponding entry in the `content.results` section of the original LOC API query.

The images chosen for the dataset were those in the *medium* range. First, if applicable, the `item.service_medium` field was used. If this field was not available, an image with a width of 640px was selected from the `image_url` field. Finally, if there were no images with a width of 640px, then one with a height of 640px was used.

The final dataset, after filtering out duplicates and items that had no images available, is comprised of 39,597 items.

Pre-processing

When the stereograph cards first arrived at the Library of Congress, they were sorted and arranged into subject groupings with many available for direct research in open files in the Prints & Photographs Reading Room by subject, geographic location, or portrait sitter. (Sample call number: STEREO US GEOG FILE - Georgia--Atlanta).

Another portion of the stereos are organized into groups (LOTs) by subject, photographer, publisher, or donor. (Sample call number: LOT 4164).

A large portion of the collection has been scanned in recent years, with most of the geographic images and many of the subject groupings available online. Item-level records are produced for

individual stereographs that have been scanned. The loc.gov website makes the digital surrogates available in multiple formats including multiple sizes of JPEG, GIF, and TIFF. Most of the original scans of the stereographs were produced at 700 or 800ppi. As described above, this dataset contains the medium-sized JPEG files.

Dataset field descriptions

The data fields that follow are directly translated from the `metadata.json` file. The JSON file is highly nested in nature, and that nested structure is not strictly carried over into the CSV. The CSV data fields contain the top level keys and, where applicable, one nested level below. In these cases, the field names are signified by the top level key . secondary key; for example: `item.call_number`.

All values in each column are stored as they would be found in the JSON metadata. Meaning, that if the column's value is a list or array, it is stored as a string representation of that value. For example: the `aka` field's value is in list format:

```
[ 'http://www.loc.gov/item/2015651359/' ,  
'http://www.loc.gov/pictures/item/2015651359/' ,  
'http://www.loc.gov/pictures/collection/stereo/item/2015651359/' ,  
'http://hdl.loc.gov/loc.pnp/stereo.1s04563' ,  
'http://hdl.loc.gov/loc.pnp/stereo.2s04563' ,  
'http://www.loc.gov/resource/stereo.1s04563/' ,  
'http://www.loc.gov/resource/stereo.2s04563/' ,  
'http://lccn.loc.gov/2015651359' ]
```

Each of the fields described below appears for a result under the `content.results` section of the API response for this query. Please note that not all elements appear for each result. Elements appearing in only some results have been marked with an asterisk.

- `manifest_id`: added to the metadata to represent the ID of the object in the `manifest.txt` file. It is a combination of the LCCN and filename of the image separated by an `_`.
- `access-restricted`: used for limiting access to some items from offsite
- `aka`: alternative identifiers for documents (e.g., shortcut urls)
- `campaigns`: added at display time if a search result needs campaign codes added to the published url
- `contributor`: generally used for name/related names such as photographer, author, etc. but is distinct from names in subject field. For these images, this fields usually cites the photographer or copyright claimant
- `coordinates`: a text version of geographical coordinates; typically drawn from a MARC field

- **date:** a date chosen from all available dates to be the sortable date. Can be the creation date, publication date, or a date referenced in the item depending on website target and cataloging; entries in this field could be formatted as a year or YYYY-MM-DD. Items are sortable by this date.
- **dates*:** List of dates related to the item. In ISO 8601 format, UTC. Items are facetable by these dates.
- **description:** often includes a short, summary description of the original physical item written to accompany the item in a list of search results; for these images, descriptions can be created from MARC records content and/or pulled from the [Prints and Photographs Online Catalog](#)
- **digitized:** true/false; indicates whether a digital surrogate exists
- **extract_timestamp:** timestamp of most recent ETL process
- **group:** ETL group
- **hassegments:** true/false for having segmented data (pages, bounding boxes of images, audio segmentation, etc.) in the index
- **id:** uniform resource identifier, aka uri, for an item; resolves as a url beginning in http. Note: for historical reasons, the ID follows the pattern of an HTTP URL, not an HTTPS URL, even though loc.gov now supports only HTTPS.
- **image-url:** URLs for images in various sizes, if available. If the item is not something that has an image (e.g. it's a book that's not digitized or an exhibit), the URL for the image might be for an icon image file. The url is meant to convey enough information to be used for various result displays (i.e. list, grid, gallery, slideshow) described in more detail in the 'resources' element.
- **index:** the index number of the results among all results. This starts with 1 and continues through all of the results in the whole set (not just this page).
- **language:** languages associated with the item
- **latlong:** latitude and longitude from a lookup table based on the location cited in the catalog record
- **location:** place(s) related to the item. These are extracted from subject headings and other metadata, so there may be duplicates.
- **location_city*:** field for cities related to the item
- **location_country*:** field for countries related to the item
- **location_state*:** field for state related to the item
- **location_county*:** field for counties related to the item
- **location_str:**
- **locations:** locations expressed as searches in facets
- **lonlat:** latitude and longitude from a lookup table based on the location cited in the catalog record
- **mime-type:** the formats available for a digitized item
- **number*:** Swiss-Army knife for "numbers" - e.g., OCLC number, shelf number, etc.
- **numbercarriertype***

- `numberformerid*`
- `number_lccn*`
- `number_oclc*`:
- `numbersourcemodified*`
- `online-format`: text name of the online format, usually derived from mime type; superset of mime-type(s)
- `original-format`: the kind of object being described (not the digitized version). If the record is for an entire collection, that is included here.
- `other-title`: alternative language titles, other alternative titles
- `partof`: collections, divisions, units in the Library of Congress, or any of a number of less formal groupings and subgroupings used for organizing content.
- `related`: helps the website make connections between related content on the website.
- `reproductions`: advisory note from the Prints and Photographs division on downloading images offsite or purchasing copies through Library of Congress Duplication Services.
- `resources`: details of digital content and structure
- `shelf-id`: primary sorting field of item records; variously derived based on website target; for these images, shelf id contains the unique call number citation for the item in the Prints & Photographs Division
- `site`: originally showed the source system, used in transform for display methods, tracks ETL target closely
- `subject`: list of subjects. These are separated elements of the Library of Congress Subject Headings. Geography is not shown here, see the location element.
- `timestamp`: time record inserted in the web index
- `title`: title of the item
- `type*`: medium of original item, i.e. "still image"
- `unrestricted`: true or false
- `url`: URL on the loc.gov website. If the items is only in the library catalog, the URL will start with `lccn.loc.gov`.

The following `item` subfields of the `content.results` section are mainly for display of the item on the loc.gov website. These subfields may pull information from the Prints and Photographs Online Catalog API and/or website target-specific interpretations of MARC records.

- `item.accessadvisory*`: *information provided about whether the item described can be accessed directly* - `item.callnumber` - `item.contributors` - `item.controlnumber` - `item.created` - `item.createdpublished` - `item.createdpublisheddate` - `item.creator`: - `item.creators` - `item.contributor_names`: - `item.date` - `item.digitalid` - `item.displayoffsite`: similar to `access_restricted`; used for limiting access to some items from offsite - `item.format` - `item.formats` - `item.genre` - `item.id` - `item.language` - `item.link` - `item.location` - `item.marc` - `item.medium` - `item.mediumbrief` - `item.mediums` - `item.modified` - `item.numberformerid*`: - `item.notes` - `item.othercontrolnumbers*`: *ex. "22762560"* - `item.othertitle`: - `item.place` - `item.part_of`: - `item.rawcollections*`: *ex: diof, pp, stereo* - `item.repository` -

item.reproductionnumber - *item.resource links* - *item.restriction**: - *item.rights advisory* - *item.rights information* - *item.servicelow* - *item.servicemedium* - *item.stmtofresponsibility**: for stereographs, this is usually information transcribed from the item stating the name of the photographer or publisher - *item.sortdate* - *item.sourcecreated* - *item.sourcecollection**: - *item.sourcemodified* - *item.subjectheadings* - *item.subjects* - *item.summary* - *item.thumb_gallery* - *item.title*

The following are *related* subfields of the *content.results* section: - *related.grouprecord* - *related.lotlink**: - *related.neighbors*

Rights Statement

The Library's stereograph collection includes thousands of photographs submitted for copyright protection. Those that were copyrighted or published in the U.S. more than 95 years ago are in the public domain because the copyright has expired. The collection also includes images not submitted for copyright that may not have been published. The term of copyright for unpublished images is the life of the creator plus 70 years. For unpublished anonymous works and works where the death date of the creator is not known, the copyright term is 120 years from the date of creation.

Creator and contributor information

Creator: Chase Dooley

Contributors: Eileen J. Manchester, Phil Michel, Dave Woodward, Patrick Rourke, Meghan Ferriter, Alice Goldfarb

Contact information

Please contact LC-Labs@loc.gov with any questions or suggestions!