Good morning and thank you panelists and moderator.
• Provide an overview of the Carol M. Highsmith Archive
• Outline the Library of Congress strategy for quickly bringing tens of thousands of images online
• Discuss some of the implications of processing digital images in bulk

This morning I will provide an overview of the Carol M. Highsmith Archive maintained at the Library of Congress; outline our strategy at the Library of Congress for quickly bringing tens of thousands of images online; and discuss some of the implications of processing these digital images in bulk.
This large archive of primarily born-digital photographs expresses the vision of Carol Highsmith. Carol is a highly energetic photographer who has been photographing the United States for decades. Her work covers all aspects of the American landscape including many images of interest to ARLIS members and their patrons. Highsmith has dedicated all her rights in these images to the public so they are particularly useful to researchers looking for images that they can publish without a rights hassle. However, there are some caveats I will talk about later.
http://www.loc.gov/pictures/item/2011634808/
The collection is especially rich in architecture. Hartford, Connecticut.
http://www.loc.gov/pictures/item/2012630744/
The Rosenbaum House, Florence, Alabama.
http://www.loc.gov/pictures/item/2010640717/
Jasmine Hill Gardens, near Wetumpka, Alabama.
http://www.loc.gov/pictures/item/2010639995/
"Grey Towers," the 1892 turreted home of sugar refinery owner William Welsh Garrison in Glenside, outside Philadelphia, Pennsylvania
http://www.loc.gov/pictures/item/2011633556/
Gaineswood, a plantation house in Demopolis, Alabama
http://www.loc.gov/pictures/item/2010641094/
Peery’s Egyptian Theater, Ogden, Utah. http://www.loc.gov/pictures/item/2011631166/
Roofless church by Philip Johnson, New Harmony, Indiana.
http://www.loc.gov/pictures/item/2010630152/
http://www.loc.gov/pictures/item/2010630429/
The mosaic above the doorway of St. Anthony's (1952) Roman Catholic Church in Long Beach, California, depicts Pope Pius XII witnessing the assumption of the Virgin.
http://memory.loc.gov/phpdata/pageturner.php?type=contactminor&cmIMG1=/pnp/pplot/13700/13725/01405t.gif&agg=pplot&item=13725&caption=103
San Jose de Gracia Mission, Las Trampas, New Mexico.
http://www.loc.gov/pictures/item/2011631082/
Weisman Art Museum, Minneapolis, Minnesota.
http://www.loc.gov/pictures/item/2010630457/
Missoula County Courthouse, Missoula, Montana.
http://www.loc.gov/pictures/item/2011635388/
http://www.loc.gov/pictures/item/2011631631/
Modern office building in northern Virginia.
http://www.loc.gov/pictures/item/2011634307/
Much of Carol’s architectural work includes details that show the decorative arts. Arctic building seal architectural details, Seattle, Washington. http://www.loc.gov/pictures/item/2010630411/
Exterior detail, Border Patrol Station, Naco, Arizona.
http://www.loc.gov/pictures/item/2010719676/
Interior of the Grand Canyon watchtower, Arizona.
http://www.loc.gov/pictures/item/2011636465/
The Cadet Chapel in Colorado Springs is the most popular man-made attraction in Colorado, with more than a half million visitors every year.
http://www.loc.gov/pictures/item/2011632993/
Stairs, U.S. Court House, Austin, Texas. http://www.loc.gov/pictures/item/2013634326/
Entrance to Johnson Wax Headquarters, Racine, Wisconsin.
http://www.loc.gov/pictures/item/2011631297/
Railing detail, U.S. Custom House, Charleston, South Carolina.
http://www.loc.gov/pictures/item/2010719458/
Bas relief sculpture, Apex Building, Washington, DC.
http://memory.loc.gov/phpdata/pageturner.php?type=contactminor&cmIMG1=/pnp/pplot/13700/13734/01531t.gif&agg=pplot&item=13734&caption=6
Interior door detail, U.S. Custom House, Charleston, South Carolina.
http://www.loc.gov/pictures/item/2010719459/
Art at the Baltimore Basilica, Baltimore, Maryland.
http://www.loc.gov/pictures/item/2011634915/
Newell post carved by WPA workers during the Great Depression at Timberline Lodge, midway up Oregon's soaring Mount Hood.
http://www.loc.gov/pictures/item/2011636437/
Fireplace at Kensington Manor, an 1855 Renaissance Revival villa on the Kensington Plantation, Eastover, South Carolina. http://www.loc.gov/pictures/item/2011634100/
Pierce Arrow detail, Pioneer Auto Show, 1931. Murdo, South Dakota.
http://www.loc.gov/pictures/item/2010630556/
http://www.loc.gov/pictures/item/2010720140/
Dining room of the Reynolds House, Winston-Salem, North Carolina.
http://www.loc.gov/pictures/item/2011632180/
She often photographs stained glass and mosaics. Stained glass window, Boston Avenue United Methodist Church, Tulsa, Oklahoma.
http://www.loc.gov/pictures/item/2010630187/
Courtroom ceiling, Federal Building and U.S. Courthouse, Providence, Rhode Island.
http://www.loc.gov/pictures/item/2010718939/
Stain glass in a church in Bretton Woods, New Hampshire.
http://www.loc.gov/pictures/item/2011636463/
Untitled at Robert C. Byrd Federal Building, Charleston, West Virginia.
http://www.loc.gov/pictures/item/2010719857/
Singing Windows stained glass, designed by J&R Lamb, located in the University chapel at Tuskegee University, Tuskegee, Alabama. http://www.loc.gov/pictures/item/2010637794/
Paramount Theatre Building detail, Oakland, California. 
http://www.loc.gov/pictures/item/2010630219/
Many of the buildings she photographs include murals in the interiors. Lobby Painting, Elbert Parr Tuttle U.S. Court of Appeals, Atlanta, Georgia. http://memory.loc.gov/phpdata/pageturner.php?type=contactminor&cmIMG1=/pnp/pplot/13700/13737/01499t.gif&agg=pplot&item=13737&caption=3
http://memory.loc.gov/phpdata/pageturner.php?type=contactminor&cmIMG1=/pnp/pplot/13800/13818/01689t.gif&agg=pplot&item=13818&caption=3
Carol is particularly interested in photographing things she knows will not last forever such as exterior murals. Indian figure mural, La Veta, Colorado. http://www.loc.gov/pictures/item/2011630374/
http://memory.loc.gov/phpdata/pageturner.php?type=contactminor&cmIMG1=/pnp/pplot/13800/13816/01586t.gif&agg=pplot&item=13816&caption=2
Mona Lisa Mural, Columbus, Ohio. http://www.loc.gov/pictures/item/2010630104/
Virgin of Guadalupe, patron saint of Mexico, is portrayed on this mural in the form of a candle. Located next to the Avenida Guadalupe Association on Guadalupe Street on the West side of San Antonio, Texas.
Mural, Mission District, San Francisco, California.
http://www.loc.gov/pictures/item/2010630201/
Oregon History Center, Portland, Oregon. http://www.loc.gov/pictures/item/2011635070/
Mural, Mission District, San Francisco, California.
http://www.loc.gov/pictures/item/2010630200/
She also frequently photographs sculptures. Statue of Good and Evil, part of the Peace Fountain outside the Cathedral of St. John the Divine, New York, New York. http://www.loc.gov/pictures/item/2011633730/
Sculpture "Protagoras" at the Burger Federal Building, St. Paul, Minnesota.
http://www.loc.gov/pictures/item/2010720928/
Birmingham Botanical Gardens, Birmingham, Alabama.
http://www.loc.gov/pictures/item/2010639303/
http://memory.loc.gov/phpdata/pageturner.php?type=contactminor&cmIMG1=/pnp/pplot/13800/13826/01835t.gif&agg=pplot&item=13826&caption=2
“Rockman” statues, U.S. Courthouse, Minneapolis, Minnesota.
http://www.loc.gov/pictures/item/2010720747/
End of the Trail statue, Waupun, Wisconsin.
http://www.loc.gov/pictures/item/2011630626/
Iron horses and cactus near Sedona, Arizona.
http://www.loc.gov/pictures/item/2010630137/
Porter Sculpture Park, Montrose, South Dakota.
http://www.loc.gov/pictures/item/2010630520/
"Pheasants on the Prairie," Enchanted Highway, Regent, North Dakota.
http://www.loc.gov/pictures/item/2010630885/
Spoonbridge and cherry art, Minneapolis, Minnesota.
http://www.loc.gov/pictures/item/2011630868/
The first groups of photos we received were documentation of US government buildings which we added to our catalog in a group-level record, following the model of her analog work.
Carol has been systematically donating her work to the Library of Congress since 1992. She promised LC her entire archive and started scanning selected work and donating the film along with a scan, done to LC specifications. We cataloged these at a group level. As you can see in this catalog record for GSA buildings in Missoula, Montana we have cataloged all 14 images in a single record.
Clicking on the image icon brings you Carol’s scans of the 14 transparencies.
* The captions for these individual images were not searchable in the catalog and were only available through the group level record.
In 2002, Highsmith, an early adapter of technology, began talking to us about submitting born digital-work. The first groups of photos we received were documentation of US government buildings done for the GSA.
We added them to our catalog in a group-level record, following the model we used for her analog work.
Photographs of the John N. Harwood U.S. Court of Appeals Building in New Orleans, Louisiana

Title: Photographs of the John N. Harwood U.S. Court of Appeals Building in New Orleans, Louisiana
Creator(s): Harwood, John N., 1940-
Date created: 1940
Medium: 17 photographs : digital, TIFF files, color.
Summary: Photographs show exterior and interior views of courthouse in New Orleans, Louisiana. Details include entrances, doors, ceilings and sculptural decoration.
Reproduction Number: LC-DIG-pga-14930 (original digital file)
Rights Advisory: No known restrictions on publication.
Call Number: LOT 1390 (ONLINE) [PDF]
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA.
Notes:
- Arrangement: Arranged by photographer-assigned number. The corresponding digital id numbers used for reproduction requests range from LC-DIG-ppmsca-15700 to -61735.
- LOF title devised by Library staff.
- Visual copies are derivatives of original digital files.
- Item captions supplied by photographer.
- Photographed as part of an assignment for the General Services Administration.
- Forms part of the Harwood Archival.
Subjects:
Format:
Highsmith gave us a spreadsheet with item level-captions which were displayed with the individual images but were not directly searchable. We embedded our group level descriptions with relevant keywords to provide subject access.
In 2010 Highsmith began her 21st Century America project to systematically document the United States. She began transmitting hundreds of photos at a time to the Library. These compelling, copyright free images called for item level description to fully expose them to the public. The group level description method just wasn’t enough but we didn’t have the resources to craft item level catalog records for the thousands of photographs we were anticipating. We needed to devise a way to efficiently move this flood of popular pictures quickly and efficiently into our catalog at an item level.

First we needed to improve our means of getting the images onto the Library’s servers. Highsmith had initially been delivering the files on DVDs. She tried using FTP sites but found that it was too slow to upload images while she was in the field. She now provides us with an external hard drive of hundred images at a time (she just delivered 1600 photos of Texas yesterday) and we upload them using an ingest system the Library has devised to load large quantities of digital data. This ingest system has features that check the image files for problems, can be programmed to send the files to the proper location on the server, and provides accurate records of what we received. Once the files are loaded, we begin preparing them for service. The original files are not manipulated in any way. Once we load them we make working copies and we also make 6 different derivatives to fill the various functional requirements of our online catalog. All of these files are maintained on Library servers which rely on redundancy and physical dispersion to insure that the files are safe.

Once the files are loaded and accessioned into the Libraries collections, we face the challenge of making the images available in our catalog with their associated metadata. As a general rule, born digital photographs have a header – a string of information created by the camera when a picture is taken. This information is an integral part of the image file. The header contains basic information about the type of camera used, exposure data, and a date and time stamp along with other technical data. Depending on the camera it can also include more detailed information like GPS coordinates. Once the picture is taken, there is a wide variety of software readily available to edit these headers. Highsmith edited the
headers of her photos to include her name, caption information, location and keywords.
Here you can see a few of some of Highsmith’s metadata as it displays in PhotoShop. You can see there are other tabs in photoshop to display other metadata.
For the first group of born-digital photos Carol sent us, we cut and pasted her metadata from the headers into a spreadsheet. This was slow and tedious work that wasn’t sustainable. So we started exploring metadata extraction tools — software that can copy out information from headers. By typing in “metadata extraction tool” into Google you will find there are several of them available for free. We chose to use EMET. As seen from the EMET website “EMET is a free image metadata extraction tool intended to facilitate the management and preservation of digital images and their incorporation into external databases and applications. EMET was created by ARTstor through funding from NDIIPP.” It requires AdobeAir 2.0 software, which is also a free download. NDIIP is a program coordinated by the Library of Congress so we chose EMET because it was approved for use behind the Library of Congress firewall.
Once you download EMET, you launch the application.
Select the image files from which you want to extract the metadata. We extract metadata from about 350 photos at a time because EMET performs better with smaller data sets.
Once you click on “Extract Data” EMET goes to work.
It returns the metadata in a rather imposing Excel spreadsheet with about 200 columns. We were only interested in a few relevant fields in the spreadsheet which we can easily isolate and extract. These fields include Highsmith’s file name, caption, date, location and keywords. We then add selected default information to the spreadsheet including the libraries sequential call number for the image. Excel is great for quickly adding sequential numbers.
We readily export the data from Excel into Access. Access allows us to add additional default information. It also provides us with an efficient environment for identifying odd or missing data.
For example, in this case there was no caption information so we devised a caption and added a note.
Access proves to be an efficient vehicle for reporting out data with associated tags. We can therefore readily create a MARC export that we can upload into Minaret.
Minaret is a powerful MARC based software that gives us flexibility to add default administrative data with associated catalog tags. It is in Minaret where we add data that is common across all of the images such as the photographers’ name, the media, the rights advisory, and credit line. We move the data into a MARC environment because MARC is currently the markup language we use to stay compatible with the other LC resources.

Once we complete the export of data from Excel to Minaret the result is something that looks amazingly like an item-level MARC catalog record.
To complete the process, we upload the records from Minaret into the Library’s central Voyager online catalog.
The data highlighted in yellow is what the Library defaulted in to the record.
The data highlighted in pink comes from Highsmith.
We present that catalog record, along with a thumbnail image and access to higher resolution versions of the photos in the Prints and Photographs Online Catalog (PPOC).
The catalog is available at www.loc.gov/pictures.
Highsmith is a featured collection on the site so it is easy to limit your search to just images from the Highsmith archive.
The site also provides extensive information about the collection which is currently being updated.
Thanks to this efficient workflow the Library now provides over 25,000 images online. This year alone we added more than 5,800 images to the catalog, spending less than half the time of one person. We are proud to make these images available to the public very soon after Highsmith takes the photos. But we made some obvious compromises to pull this off.

The most obvious is subject access. Carol fills her image headers with both location information and subject keywords. These keywords are an excellent representation of what the photographer wanted to convey when she took the picture. However, the keywords are all in natural language and do not come from any controlled vocabulary. Looking at the Highsmith collection profile at the bottom left of the screen you can see that she has identified 1,328 different geographical locations
and 7,143 keywords that are listed as “subjects” in the catalog.
The keywords are a mix
of personal names
proper nouns
and pre-coordinated concepts.
But looking at one of the item records, you can see that her keywords frequently provide lots of hooks for the researcher looking for images. Example, “Title, date and keywords” also Highsmiths’ notes adding style keywords.
I mentioned earlier that Highsmith has dedicated all rights she has in the images to the public. But shortly after we put up Highsmith’s photographic survey of Alabama we received a friendly note from the Jesse Owens Memorial Park and Museum in Danville, Alabama. It seems that the Luminary Group manages the Jesse Owens publicity rights and that the International Olympic Committee held copyright on the Olympic Rings.
We eliminated access to the high resolution images of the statue and changed our restriction notice.
Carol M. Highsmith

Rights and Restrictions Information

Prints and Photographs Division, Library of Congress, Washington, D.C., 20540-4730

Carol M. Highsmith's photographs are in the public domain.

Access: Subject to P&P policy on serving originals.

Reproduction (photocopying, hand-held camera copying, photoduplication and other forms of copying allowed by "fair use"): Subject to P&P policy on copying, which prohibits photocopying of the original color photographs.

Publication and other forms of distribution: Ms. Highsmith has stipulated that her photographs are in the public domain. (See P&P Collection Files.) Photographs of sculpture or other works of art may be restricted by the copyright of the artist.

Credit Line: Library of Congress, Prints & Photographs Division, photograph by Carol M. Highsmith [reproduction number, e.g., LC-USZ62-123456]
We also revised our overall Highsmith restriction statement to include the statement “Photographs of sculpture or other works of art may be restricted by the copyright of the artist.”
One of the complicating factors related to rights is that once we code Highsmith images as public domain they can be copied by anyone. A number of commercial firms systematically crawl our rights-free images and sell them in various iterations on sites like Amazon. Generally, the more people use and enjoy images from our holdings, the happier we are. But this commercial use of Highsmith images has occasionally had unexpected consequences. For example, the “Big Boots.”

The artist, Bob Wade, contacted Highsmith. He noted that although he had formally copyrighted his sculpture, pictures of his boots were being sold on Amazon, that he was not credited in the picture and that the title of his sculpture was incorrect. This is an issue that will come up with photos that people take as they travel through the landscape. Frequently the name and title of a public artwork are not readily available to the photographer as she travels through the built environment.
In this case, we took down the high resolution images, added a note providing credit to the artist, and noted that the image carried restrictions.
These experiences led to an initiative to systematically go through the archive and attempt to identify the names of artists and titles of murals, sculptures and paintings. For example,
in the catalog record for this photograph of a mural Library staff have changed the caption to include the name of the artist and the title of the mural. We noted where the additional information came from. This additional work provides artists with credit for their work and makes it more likely that an artist can identify photos that they feel infringe upon their copyright. Art researchers benefit from the added access points. Everyone wins.
While I have provided the last few examples as examples of issues currently facing visual resource libraries, I want to emphasize that these are exceptions – a mere hand full of items requiring tweaks in a collection of over 25,000 images. The work the Library is doing in co-operation with Carol Highsmith is making available an archive that art researchers can mine. I invite each of you to explore the collection, and keep coming back as we add more and more images.