

Out of the Archives

Scanning and Digital Files

By Jim Cascino, Jordan Radke, and Maddie Shovers

“Out of the Archives” is a column that brings to light the world of professional archiving, providing a regular forum to share selections from the Center’s collections and tips for maintaining your own photographs. Whether you are a photographer, collector, or avid fan, it is important to organize and preserve the materials you create or collect. In this second installment, we will take a look at the basics of scanning and digitizing photographs along with how to handle those digital files. Please get in touch with Jordan Radke, jordan@railphoto-art.org, if there are any topics you would like us to cover in the future.

Who we are

The Center’s collections, comprising some 200,000 photographs, form the basis of our Railroad Heritage Visual Archive. The team in Madison, Wisconsin, consists of Jordan Radke, Archives Manager, graduate archival intern Maddie Shovers, intern Aviva Gellman, and volunteer John Kelly. We also partner with Lake Forest College,

working with Anne Thomason, Archivist, along with graduate archival interns Jim Cascino and Colleen O’Keefe, to process and maintain Center materials housed in the college’s Archives & Special Collections in the Donnelley and Lee Library. Scott Lothes, Center president and executive director, and the Collections & Acquisitions Committee of the board of directors provide oversight.

What we do

In keeping with the Center’s mission of preserving and presenting significant images of railroading, with the Railroad Heritage Visual Archive we seek to securely house collections and make their contents accessible. We adhere to established archival principles to ensure safety and accuracy. Our work as archivists includes:

Preservation. One of the Center’s main objectives is properly preserving our collections. This includes appropriately caring for and handling our materials by using archival-safe supplies, and providing a controlled environment where our collections are housed.

Processing. Processing materials is a long, tedious, and detailed endeavor.

Organizing a collection appropriately sets up the rest of the processing work that includes any digitization and metadata entry. This work is essential to the long-term care and future accessibility of a collection.

Arrangement and Description. To maintain quick and easy retrieval of our materials, we organize every collection down to its individual items, if possible, given the time and resources available to us.

Accessibility. Finally, the Center will make sure that users have access to our processed collections. We create detailed finding aids to describe each collection and share images electronically through our websites and many social media outlets.

Railroad Heritage Visual Archive updates

In our Madison office, Jordan Radke has completed digitizing more than one third of the John F. Bjorklund Collection. Scanning and processing work is complete for the Leo King Collection thanks to Maddie Shovers. John Kelly continues work on the Glenn A. Oestreich Collection. At Lake Forest College, Jim Cascino has completed processing the black and white photographs in the Fred M. Springer Collection (color slides were finished in Madison at the end of 2014), along with processing the Hal Lewis Collection and finishing up metadata entry for the Wallace W. Abbey Collection. Meanwhile Colleen O’Keefe has begun digitizing and metadata entry on the Ted Rose Collection and digitizing the Sam Breck Collection. Photographs from these collections and more are posted regularly to our social media accounts on Facebook, Flickr, Twitter, and LinkedIn.

Archival terms

Like railroading, archiving has its own language. Some terms used in this column appear below, as defined by the Society

Canadian Pacific Railway steam locomotives nos. 3716 and 2860 double-heading on BC Rail. The train is shown here southbound between Clinton and Kelly Lake, British Columbia, Canada, on October 25, 1988. Scanned from a 35mm color slide. Photograph by Hal Lewis



of American Archivists (SAA). See a full glossary of archival and records terminology at: www2.archivists.org/glossary

Format. A structure used for the interchange, storage, or display of data.

Format Migration. The process of copying data from one type of storage material to another to ensure continued access to the information as the data structure becomes obsolete.

Master. An item from which duplicates are to be made.

Edited Master. The final copy of a recorded production that is ready for broadcast or duplication.

Scanning and resolution

After you have organized your archive into a cohesive collection (see the first installment of this column beginning on p. 4 in the Fall 2015 issue), the next step is to scan your materials to create digital files. Even if you do not plan to directly use your files digitally, scanning prints, negatives, slides, and other photographic formats can be an important form of preservation. You will not only have multiple copies in various formats, but you will be able to put them online, if you wish, and you can crop or edit any photograph using photo editing software such as Adobe Photoshop or GIMP, a free graphics editor.

There are numerous types of scanners available. At the Center, we use an Epson Perfection V700 Photo Flatbed Scanner. A flatbed scanner can be used for photographic prints and documents, which are generally placed on a piece of glass beneath a lid. It is crucial for negatives to have a bright light source so that they scan properly; so a scanner such as the V700 with a backlit lid or Transparent Media Adapter is necessary. Such scanners often include trays that hold transparent format photographs in place for more manageable scanning. Besides flatbed scanners, there are also film scanners, which are specifically for negatives and slides (we use a Minolta DiMAGE 5400 at the Center), and sheet-fed scanners that are best used for documents and prints. Top manufacturers

Scanner Examples							
Scanner Brand	Scanner Type	Lowest Price	Optical Resolution	Bit Depth	Scanning Area	Scanning Options	Negative Slide Trays
Canon CanoScan LIDE 120	Flatbed	\$69.99	2400 PPI	48 color/ 16 b&w	8.5" x 11.7"	Reflective	No
Epson Perfection V370	Backlit Flatbed	\$99.99	4800 PPI	48 color/ 16 b&w	8.5" x 11.7"	Reflective, Transparency	Yes
Mustek A3 1200S	Flatbed	\$157.99	1200 PPI	48 color/ 16 b&w	11.7" x 16.5"	Reflective	No
HP Scanjet G4050	Backlit Flatbed	\$199.99	4800 PPI	96 color/ 16 b&w	8.5" x 11"	Reflective, Transparency	Yes
Plustek OpticFilm 8100	Slide Scanner	\$302.82	7200 PPI	48 color/ 16 b&w	35mm	Transparency	Yes
Epson Perfection V800	Backlit Flatbed	\$739.99	6400 PPI	48 color/ 16 b&w	8.5" x 11.7"	Reflective, Transparency	Yes

include Epson, Canon, HP, Plustek, and Mustek. An easy-to-use and good-quality flatbed scanner for reflective scanning of prints can be purchased for around \$100, but you will need to spend more for a transparency-capable scanner for digitizing slides and negatives. (See the chart above.)

Organization will simplify scanning. The Center's collections are organized in archival binders and sleeves, which make it easier to scan in pieces. When you purchase a scanner, there will be software included, such as Epson Scan, which will help you to organize your scans. Scanning software has settings that can be tailored to specific needs in order to produce a high quality scan. First, choose the document type (print, positive film, color negative film, or B&W film). Next, select the image type (grayscale or color) and bit depth, which is the ability to analyze and reproduce the full color spectrum (grays for grayscale; red, green, blue for color). Lastly, the most important setting to determine scan quality is the resolution, expressed as PPI (pixels per inch) or DPI (dots per inch).

If you are planning on printing your images, scanning at the proper resolution will make a great difference in the quality of your prints. This is why understanding the difference between PPI and DPI is key.

The image resolution is the PPI, which is an *input* resolution that determines how many pixels per inch are shown on your computer screen. The printer resolution is known as DPI, an *output* setting that means dots per inch or how many ink dots your printer can place in an inch. When a photograph is scanned, it gets translated into picture elements, or *pixels*, and when a photograph is printed, those pixels get translated into ink dots. DPI is commonly used interchangeably with PPI in reference to scanning images, but do not be mistaken, technically, DPI refers to printing resolution and PPI refers to image resolution. In any case, spend some time getting familiar with your scanner, its software, and its settings before you start any digitization project.

When you are scanning, you will typically change resolution settings depending on the format and size of the image. Essentially, smaller images such as slides and negatives need higher-resolution scans in order to produce larger prints. For example, if you want a twelve-inch wide print at 300 DPI (a common printing resolution), you will need an image that is 3600 pixels wide (12 x 300 = 3600). Using the desired amount of pixels you can do some simple math to understand what image resolution you will need. A 35mm slide is

Scanning Recommendations*						
Prints						
Size of Original	File Format	Resolution	Bit Depth		File Size	Scan Time
4" x 5"	TIFF Master	900 – 1200 PPI	Black & White	8-bit grayscale	up to ~15 MB	~ 1 min.
			Color	24-bit RGB	up to ~40 MB	~ 1 min.
5" x 7"	TIFF Master	600 – 900 PPI	Black & White	8-bit grayscale	up to ~15 MB	< 1 min.
			Color	24-bit RGB	up to ~40 MB	< 1 min.
8" x 10"	TIFF Master	300 – 600 PPI	Black & White	8-bit grayscale	up to ~15 MB	< 30 sec.
			Color	24-bit RGB	up to ~40 MB	< 30 sec.
Negatives						
Size of Original	File Format	Resolution	Bit Depth		File Size	Scan Time
35mm	TIFF Master	2400 – 4000 PPI	Black & White	up to 16-bit grayscale	up to ~30 MB	> 3 min.
			Color	up to 48-bit RGB	up to ~80 MB	> 3 min.
Medium (120, 220)	TIFF Master	1500 – 2400 PPI	Black & White	up to 16-bit grayscale	up to ~30 MB	> 2 min.
			Color	up to 48-bit RGB	up to ~80 MB	> 2 min.
4" x 5"	TIFF Master	900 – 1200 PPI	Black & White	up to 16-bit grayscale	up to ~30 MB	~ 1 min.
			Color	up to 48-bit RGB	up to ~80 MB	~ 1 min.

*Approximates, recommended results may vary depending on scanner models, settings, and print sizes.

*Approximates, recommended results may vary depending on scanner models, settings, and print sizes.

1.42 inches wide and would need at least a 2535 PPI scan ($3600 \div 1.42 = 2535$). The same goes for a medium format negative, such as 120, which is 2.25 inches wide and would need at least a 1600 PPI scan ($3600 \div 2.25 = 1600$). If you want to be able to make large prints from your scans, it is a good habit to scan at a higher PPI the first time instead of going back and re-scanning every time you need a larger print. Doing so also gives you the best archival quality for your digital images. It is always a good idea to take note of the settings that you use for each type of photograph, as the settings will change often and it is easy to forget what you previously used.

There are various approaches to scanning photographs depending on the size of a collection, time available, and intended purpose. If you are interested in creating master files for your archive and are more concerned about quality rather than time spent, it would be beneficial to spend more time scanning your photographs at a high resolution (see the chart above for recommendations). If you have a large collection and want scans only for ease of

searching and sharing, you may want to scan at a lower resolution to save time. Finally, if you do not want to make the scans yourself and can afford to spend more money, you could send your photographs to a third party, such as a ScanCafe (www.scancafe.com), to scan your photographs. Whether you decide to send your photographs to a third party or scan everything yourself, creating digital files is the primary way to preserve your images in lasting file formats.

Digital files and storage

One important factor to consider when undertaking a digitization project is the preservation of these files. A major tenant of the archiving profession is to preserve materials as closely to their original form as possible, and digitization is no exception. To this end, there are two ways in which we can achieve this in a digitization project: sustainable file formats and smart storage.

We recommend saving your scans as TIFFs (.tif) for your master files. This is a file type that is “lossless,” as opposed to JPEGs (.jpg) that are “lossy.” The major difference between these two file formats

comes down to size, and what is done to achieve this. Saving your scan as a JPEG will compress or digitally *remove* data and pixels that are redundant, in order to provide a smaller file size. For example, if a scanned image has 15 shades of the color gray, JPEG compression will simplify the tonal palate by eliminating some of these shades. Because TIFFs are uncompressed, they will take up a lot more storage space—several-hundred megabytes or more, depending on PPI as well as other factors—but will provide you with an image that is much more similar to the original and is of much higher quality. However, many web applications, including social media sites, do not support the TIFF format for display and most have a size limit for image uploads. We recommend making JPEG copies as edited masters from your TIFF master files for these purposes, as you can always make as many copies as is necessary, and you can edit them without altering the masters.

A crucial, ongoing step in any digitization project is adhering to an organizational scheme when creating your digital files (see the Fall 2015 issue). Especially with a large collection, it can be difficult to keep track of which file corresponds to which image without having to manually look through them. Organizing your digital files and knowing where copies are located is crucial. Therefore, maintaining these files with an organizational scheme in line with your physical collection can save time and resources in the long run. For example, the photograph King-07-102-001 (collection-box-page-image) will have a digital file named King-07-102-001.tif that can easily be tracked down in an Excel spreadsheet and its corresponding digital folder. Although deciding on how you would like to organize your physical and digital collection is a personal choice, the most important rule is to stick with it. Making changes to your organizational scheme, especially while in the process of digitizing, will take a lot of correcting, could result in confusion, and may cause you to misplace or mislabel files.

In terms of storage, the adage “more

is always better” is certainly true. Sudden hardware and software failures can happen, which means that digital objects can be susceptible to loss. Two of the best options for digital storage are cloud servers and external hard drives, and each comes with pros and cons. Cloud servers are the more convenient option, as they can be accessed almost anywhere, by nearly any device, provided you have a working Internet connection. Sites such as Google Drive, Dropbox, and Amazon S3 are all safe and secure, and offer a relatively decent amount of storage for free (15 GB, 2 GB, and 5 GB, respectively), and they all allow you to upgrade your storage space for a monthly fee. The downside to these types of cloud servers is that your files are being managed and stored by a third party and, as a result, are out of your hands the moment you upload them.

External hard drives, by contrast, come in a variety of different sizes, price points, and models. Most plug into your computer’s USB ports and can be used to store almost any file types. The biggest decision to make with external hard drives is whether to purchase a HDD (hard disc drive) or a SSD (solid-state drive). HDDs are much cheaper in price, but because the data is written onto them via a magnetic disc, there are more moving parts which can fail. SSDs, on the other hand, do not have these magnetic discs, which makes them less prone to failure, and much quicker for uploading and retrieving files. One medium we do not recommend using are optical discs such as CDs and DVDs. While these were once thought to be a relatively stable and permanent means of storage, this has not proven to be the case; the Council on Library and Information Resources reports that these types of discs have a life expectancy of just five to ten years *before* being written on.

All of this is to say that there are plenty of storage options available and which one(s) to use depends on the size of your digitization project and how much you are willing to spend. We recommend storing your files through several different means,



as the last thing you want to do is lose all of your hard work! The idea of digital obsolescence should also be considered when dealing with both file format and storage options. This refers to the tendency of hardware, software, and machinery to become antiquated, rendering data or files associated with it difficult to migrate and/or use. To prevent this, we recommend keeping your files in a sustainable format, or one that is not in danger of becoming obsolete. The Library of Congress maintains a list of sustainable file types, which may help in determining whether you should consider format migration.

Tips for scanning and digital files

- Research different stores and vendors to find the right scanner for your needs
- Set goals for your scanning to make work less tedious

Alaska Railroad’s Whittier Shuttle train switching cars in its namesake town of Whittier, Alaska, in the early 1970s. Scanned from a 120 black-and-white negative. Photograph by Leo King

- Be sure to check and test the settings of your scanner so you get the best scan the *first* time
- Generally, a higher-resolution scan produces a higher-quality image
- Create master files with TIFFs and edited master files with JPEGs
- Create multiple copies of digital files and store them in multiple places
- Be aware of what storage methods are in danger of digital obsolescence

Further Reading and Resources:

Learn more and see a list of links to online resources on the Center’s website at:
www.railphoto-art.org/ota/