

HERITAGE BULLETIN

22

Tips, Ideas & More to Help Preserve Oregon's Heritage

IMPLEMENTING A DIGITIZATION PROJECT

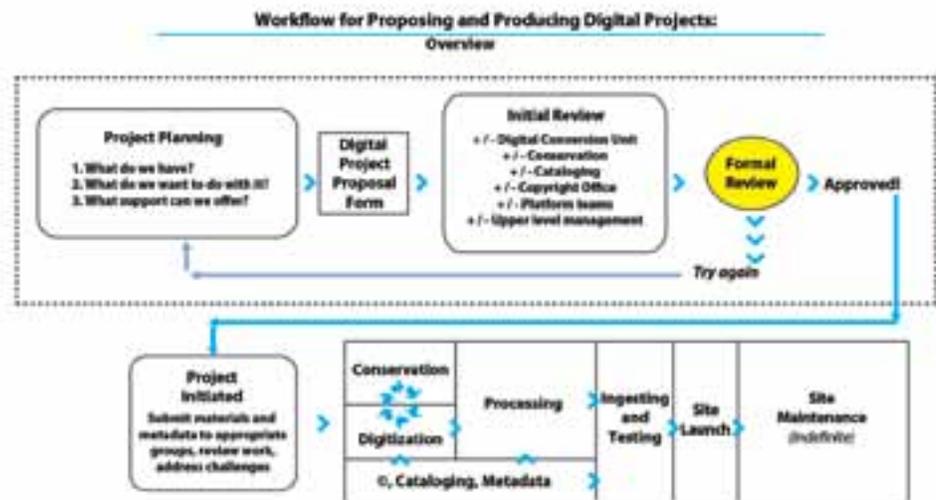
JULY 2015

Careful planning and defined policies are the first steps toward a successful digitization project. This type of project requires patience, organizational skills, knowledge of scanning technology and adherence to archival standards. This Heritage Bulletin assumes the reader possesses a basic understanding of the equipment and materials used to conduct a digitization project. It includes information about a step-by-step process known as a workflow and provides a helpful glossary of terms.

Creating a scanning workflow will improve the digitization process.

Developing a workflow for specific portions of your digitization project will reduce the time it takes to scan and keep the process consistent. There are different workflows to consider. Some scanning software has built in workflows. For example, it may include default settings for printed images that define color, resolution and the type of file desired for a saved digital surrogate. The settings for large format prints, glass negatives, film negatives, and slides will all be different.

Some software allows you to modify or create a workflow that describes the process for preparing the original materials for scanning, including best practices in handling and processing. By developing a workflow, you will become more proficient at scanning and processing hundreds of images in a single session.



Adapted from University of Michigan Library's *Starting a Digitization Project*.



Make a quality scan by adjusting software settings.

Creating a quality scan is dependent on many factors including the knowledge of the individual, and the capability of the hardware and software. Adjusting the scanning software's input and output options will affect the quality of the scan as well as the files in your collection. Observe the scanning software's interface carefully, noting options such as input, output, crop, filter, color, and preferences. Use the options panel to change the software's scanning settings. First, observe the scanning software's input options, or image adjustment options, for keywords such as "source", "mode", "media", "bits per pixel", or "resolution". Input and output options will vary depending on the software, but aspects such as "media", "bits per pixel" and "resolution" are consistent with most programs.

Many scanners automatically sense the "mode" option, but in some cases, you will set it manually. Mode refers to the type of scan. It could be paper-based (reflective) or cellulose-based (such as transparencies or negative film). Reflective media refers to items with continuous tone (photographic prints),



This sample input screen shows the user's "mode" options. Image courtesy VueScan.

options are set will depend on how much detail or quality your institution wants in its scans, as well as how much storage space (i.e., local or network storage) is available. The number of colors displayed will determine the number of bits assigned per pixel – the higher the bits per pixel, the greater the quality and file size. For high quality scans, set this option to 24-bits per pixel.

Two of the most important input options are "bits per pixel" and "resolution". The method in which these

Creating a high-resolution scan of a photographic image enhances our knowledge of the image. We can now examine a digital image at many times its original size, viewing details we haven't experienced before. Yet, it is not necessary to scan every image at the same resolution. In fact, your institution will increase or decrease resolution based on the type of image. For instance, a film negative possesses a higher photographic content than a print and necessitates a higher resolution scan. Before scanning a high-resolution image, evaluate each photograph individually based on its condition, quality, and historical value. Keep in mind that a high-resolution scan will also increase the file's size.

Weigh the costs, risks and benefits of scanning high-resolution images. For instance, a 1600 ppi scan of a glass-plate negative collection will result in amazing detail, but also lead to incredibly large file sizes. As an organization, ask the question, do we have the storage capacity to maintain over a terabyte of data indefinitely? To save storage space, consider scanning backs of prints (where writing is visible) at a low resolution, such as 300 ppi. By scanning the backs of prints, your institution can refer to the information during the indexing process. Consider adding this process to your workflow to be consistent. Heritage Bulletin 23 will discuss how to store, manage and curate the scanned digital objects.

Output options enable you to specify the types of files written while scanning, as well as the file names and their pathways. First, specify a default folder for file operations, and remember its file path. Keep in mind that the file name will also indicate a path to a subfolder of the default folder through a relative path. Output options for most photo scanners enable you to choose between creating TIFF, PNG, GIF and JPEG file formats. Determining which file format to select for scanning will depend on various factors such as the types of photos in the collection, costs and budget, and capacity of your local and/or networked storage. Consider scanning master files as TIFF (Tagged Image File Format) to create high quality, lossless files. Lossless compression ensures reconstruction of the original data from a file after it is compressed. However, lossless compression also leads to large file sizes. Your institution will determine its level of commitment and investment in local and/or networked storage options prior to scanning large quantities of TIFF images.



This sample output popup window screen demonstrates user options. Image courtesy VueScan.

Clean, scan and document the photograph.

Prior to the start of a scanning session, clean the workstation by wiping down all counters and tabletops. Remove unnecessary material, and lay out only what you need to complete the task. Make special plans for handling valuable or fragile materials. Certain photographic types such as glass-plate or film negatives are delicate and require careful handling during the scanning process. Wipe down the scanner's platen with a microfiber cloth, and use a small amount of glass cleaner when necessary. Don't spray the cleaner directly on the glass; apply it to your cloth. Be careful not to apply too much pressure to the platen and do not scratch the surface. Wear cotton or nitrile gloves while handling photographs. In fact, wearing latex free, disposable nitrile gloves allows more control and less clumsy handling. Clean each photograph by gently wiping away dirt and dust with a camelhair brush.

Document the analog (physical) material during the scanning process by taking notes. This will ensure future reference to the project and facilitate accurate indexing procedures. Purchase a notebook and document everything that is relevant to the scanning project including notes on the material, issues with the scanner, etc. Consider documenting the numbers related to the source material, including the number generated by the scanning software (i.e., 2013040-0001) and/or the accession number (i.e., 2013.101.001). Referring to your institution's collection management procedures, carefully write the software and/or accession number on the back of photographic prints using the prescribed tool (soft lead pencil or archival grade pen). For glass-plate and film negatives, document the numbers on the enclosure. Heritage Bulletin 23 will address the process of indexing digital images and managing representative information, such as descriptive, technical, and administrative metadata.

Seek help from the experts.

There is a wealth of information online regarding best practices and procedures for digitization and digital imaging technology. Most professional photographers possess knowledge in the processing, storage, compression, printing and display of digital images. For producing a quality digital representation of analog source material, seek imaging experts with knowledge of scanning technology and imaging editing software. For archival guidance, seek help from traditional institutions (see links at the end of this Bulletin).



Understand digitization concepts.

Bit Depth: The bit depth refers to the unique color information stored in an image. A higher bit depth means the image stores a greater number of colors, creating an image with precision quality.

DPI (Dots per Inch): DPI is a printer resolution measurement that defines how many dots of ink appear on the page when an image is printed.

Dynamic Range: The dynamic range relates to the tonal information (i.e., the difference in brightness between the lightest and the darkest detail) captured in an image. This term can also describe a scanner's ability to reproduce the tonal information of an image. With a small dynamic range, the dark parts of an image retain less detail and the bright parts are more likely to show just white, also resulting in less detail. A larger dynamic range retains less contrast and results in a better quality image.

GIF (Graphics Interchange Format): GIF (pronounced "jif") files are compressed, which means they are relatively small in size. They are also known for supporting animation, and are traditionally used as icons, symbols, or logos. The file extension is displayed as .gif.

Grayscale: The grayscale of an image ranges from the darkest possible shade (black) to the lightest possible shade (white), with varying shades of gray in between.

JPEG: JPEG stands for "joint photographic experts group" and provides lossy compression for digital photographic images. The file extension is displayed as .jpg or jpeg.

Lossy Compression: Lossy Compression is a data encoding method that compresses data by discarding (losing) some of it. Images become progressively coarser with the discard of the data. By contrast, text and data (i.e., bank records, text articles). It is wise to make a master lossless file from which to produce compressed files for different purposes (e.g., use a multi-megabyte file at full size to produce an advertisement in a glossy magazine; use a 10 kilobyte lossy copy for a small image on a web page).

Metadata: The simplest definition of metadata is data about data. Metadata comes in many different forms such as descriptive metadata, technical metadata and administrative metadata.

PPI (Pixels per Inch): Pixels per inch describes the amount of pixels on a computer monitor. PPI relates to the creation and display of images for onscreen use. DPI is often mistaken for PPI.

PNG (Portable Network Graphics): PNG file formats are similar to GIF's in that they small in size and incorporate low resolution images. PNG's are lossless which means that they don't lose quality during the editing process. The file extension is displayed as .png.

Resolution: A measurement of the numbers of pixels per inch, an image's resolution enables viewers to distinguish fine spatial detail. Resolution affects the level of detail contained within an image; a higher resolution results in the capture of more pixels.

RGB: RGB is a color model where a mixture of varied amounts of red, green, and blue produce a broad array of other colors.

TIFF (Tagged Image File Format): TIFF is the most common image format used today. This format uses a lossless data-compression technique, which reduces a file's size without losing data. The file extension is displayed as either .tiff or .tif.



Consult some resources.

- Association for Library Collections and Technical Services *Minimum Digitization Capture Recommendations*, 2014. <http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>
- Bibliographic Center for Research *BCR's CDP Digital Imaging Best Practices Version 2.0*, 2008. http://mwdl.org/docs/digital-imaging-bp_2.0.pdf
- Coyle, K. *Mass Digitization of Books*, (Preprint) 2006. <http://www.kcoyle.net/jal-32-6.html>
- Fulton, W. *A Few Scanning Tips*, (n.d). <http://www.scantips.com>
- Hamrick Software. *VueScan User's Guide Input Options*, 2013. <http://www.hamrick.com/vuescan/html/vuesc28.htm>
- Hughes, L. *The Price of Digitization: New Cost Models for Cultural and Educational Institutions*. A Digitization Symposium Presented by NINCH and Innodata, 2003. <http://www.ninch.org/forum/price.report.html>
- Kenney, Anne R. and Oya Rieger. *Digital Benchmarking for Conversion and Access. Moving Theory into Practice*. Mountain View, CA: RLG, 2000, pp. 24-60.
- Lee, S.D. *Digitization: Is It Worth It?* 2007. <http://www.infotoday.com/cilmag/may01/lee.htm>
- NISO Framework Working Group. *Framework of Guidance for Building Good Digital Collections*, 2007. <http://www.niso.org/publications/rp/framework3.pdf>
- Smith, A. *Why Digitize?* Council on Library and Information Resources, Washington, D.C., 1999. <http://www.clir.org/pubs/reports/pub80-smith/pub80.html>
- University of Michigan Library. *Starting a Digitization Project*. Digitization Workflow, 2013. <http://www.lib.umich.edu/digital-library-production-service-dlps/starting-digitization-project>
- Heritage Bulletin 21: Planning a Digitization Project http://www.oregon.gov/oprd/HCD/docs/Heritage_Bulletins/HB21_Digitization_Projects_Planning.pdf.

Seek guidance from traditional insitutions.

- City of Portland Archives <http://www.portlandonline.com/auditor/index.cfm?c=51816>
- Lewis and Clark College Archives <http://library.lclark.edu/specialcollections/contact.html>
- Linfield College Archives <http://www.linfield.edu/linfield-libraries/archives-special-collections.html>
- Oregon Historical Society <http://www.ohs.org/about-ohs/staff.cfm>
- Oregon State University Archives <http://scarc.library.oregonstate.edu/>
- Portland State University Archives http://library.pdx.edu/staff_directory
- University of Oregon Archives <http://library.uoregon.edu/speccoll/staff.html>
- University of Portland Archives <http://library.up.edu/default.aspx?cid=1517&pid=59>
- Willamette University Archives <http://library.willamette.edu/archives/about>

This bulletin was written by Matt Carmichael, MLIS, Archivist and Photography Collection Manager for the History Museum of Hood River County.

