

Let's Digitize!



Funds provided by



Purpose

The Highline Historical Society is pleased to make this digitization workbook available to anyone seeking to create a more permanent storage medium for their treasured personal photographs or for heritage institutions seeking to create their own digitization project.



First Sunnydale Band, 1905—Photo Highline Historical Society

Contents

DIGITAL PRESERVATION

- **Documentation....5**
- **Staffing....5**
- **In house vs. Outsourcing....6**
- **Modes of Capture....15**
- **Color Space....16**
- **Compression....16**
- **File Formats....17**
- **Metadata....18**

HARDWARE

- **Computers....7**
- **Monitors....7**
- **Scanners....8**
- **Camera Equipment....10**

SOFTWARE

- **Scanning Software....11**
- **Image Editing Software....11**

BASICS

- **Pixels....12**
- **DPI....12**
- **Bit Depth....13**
- **Resolution....14**
- **Tonal Range....14**

CREATING DIGITAL IMAGES

- **Master Files....19**
- **Service Files....19**
- **Derivative Files....19**
- **File Naming Conventions....20**
- **Watermarks....21**
- **Quality Control....22**
- **White & Black Points....22**
- **Visual Inspection...23**

ADDENDUMS

- I. Adding watermarks in Pastperfect**
- II. Bibliographical and Online Resources**

DIGITAL PRESERVATION FOR PHOTOGRAPHIC COLLECTIONS

The following are questions to ask before starting an institutional digitization project. There are many components of a digitization project that can go a long way toward ensuring a successful outcome.

- **What is your institution's overall strategic plan, technology plan and project workflow?**
- **What is your purpose?**
- **Does the project support your institution's mission?**
- **Who are you targeting?**
- **What are the physical characteristics of the collection?**
- **What is your time frame?**
- **Who will be responsible at different stages of the project?**
- **What is your digital capture experience?**

Documentation

Documenting the choices you have made for your project can be a key factor in the long-term success of digitization efforts. Good documentation can help you deal with digital collections made by your predecessors. You should document:

- **Local guidelines and benchmarks for image quality and resolution**
- **Resources that contributed to local practice guidelines**
- **Types of metadata captured**
- **File naming schemes**
- **Sustainability plans and procedures (storage, archiving, etc.)**

Staffing

Many collecting institutions have 100,000 or more images requiring digitization. Digitization projects require a combination of skills from a variety of staff with different areas of expertise. The following areas and skills may be important to manage a very long project:

- **Project management skills**
- **Knowledge of cataloging, registration methods or metadata schema**
- **Familiarity with conservation methods**
- **Understanding of photographic techniques and methods.**
- **Subject matter specialists (curators, archivist, scholars, librarians, etc.)**
- **Database development and administration skills.**

In House vs. Outsourcing?

In- house pros:

- **Development of digital imaging project experience by doing it (project management, familiarity with technology, etc.)**
- **More control over the entire imaging process as well as handling, security and storage of originals**
- **Requirements for image quality, access and scanning can be adjusted as you go instead of being defined up front**
- **Direct participation in development of image collections that best suit your organization and users.**

• In- house cons:

- **Large initial and ongoing financial investment in equipment and staff**
- **Implementing imaging process and technical infrastructure**
- **Limited production level**
- **Staffing expertise not always available**

Outsourcing pros:

- **Pay for cost of scanning the image only, not equipment or staffing**
- **High production levels**
- **On site expertise**
- **Less risk**
- **Vendor absorbs costs of technology obsolescence, failure, downtime, etc.**

Outsourcing cons:

- **Organization has less control over imaging process and quality control**
- **Complex contractual process (image specifications must be clearly defined up front, solutions to problems must be negotiated)**
- **Originals must be transported, shipped and then handled by vendor staff**
- **Possible inexperience of vendor with library/archival/museums**

HARDWARE

Computer

When selecting a computer take the following into account:

- **As much Random Access Memory (RAM) as your budget allows**
- **Purchase as much hard drive space as possible. Portable hard drives that attach via the USB port can be used to supplement the hard drive workspace on the PC being used for digitization**
- **Get a balance of reliable components, speed and storage**
- **For best results purchase a computer dedicated solely to digitization initiatives.**

Monitors

The monitor or display is the most-used output device on a computer.

Monitors are increasing in viewing size, color depth, bit depth and resolution pixels. Other areas of improvement include adaptive contrast enhancement, texture enhancements and color correction.

LCDs are smaller and lighter than CRT monitors.

LCD's are more energy efficient and are now the default option with most computers. They are attractive because they are thin and take up less space on a desk than the bulkier CRT units.

Scanners

There are a number of types of scanners you can consider

- **Flatbed Scanners** — most popular types of scanners used in libraries, archives and museums. Flatbed scanners can be used to scan 2-dimensional materials such as papers, some photographs, printed material, etc. This are not recommended for original art, including art prints. Transparency adaptors are needed when scanning 4"x 5" or 8"x 10" negatives and transparencies.



- **Slide/Film Scanners** —a dedicated film scanner has much higher quality scanning capabilities: resolution, color density, film handling and focusing. Slide/film scanners have higher dynamic tonal ranges and optical resolutions.



- **Large Format Scanners** — Can be useful when scanning maps, blueprints, architectural drawings, site plans, posters, etc. Institutions needing to digitize this type of material may want to consider outsourcing to a digital imaging vendor or using a digital camera for capture.



- **High-End Book Scanners — They allow for overhead copying of bound books and oversized and/or fragile materials that cannot be placed on a flatbed scanner.**
- **Book scanners include software which compensates for any distortion caused by the curve of the page when digitizing complete books.**
- **Wide-format Scanners — Wide-format scanners were developed to digitize large format materials such as engineering drawings and architectural blueprints and are frequently found in municipal engineering departments or local blueprint shops.**



The number of sensors in the array determines the optical resolution of a particular device. The optical resolution is normally expressed in scanner specifications as pixels per inch (PPI).

The optical resolution of any equipment you purchase should exceed the maximum resolution needed to accurately capture the types of material in your collections.

Make sure to select equipment based on its optical resolution and not the interpolated resolution since scanners with adequate optical resolution will produce more accurate scans.

Camera Equipment

- **Copy stand with lights** — A copy stand is a stand for mounting a camera, complete with base board and arms to hold lights. The stand enables you to take photographs of flat art, books and three-dimensional objects.
- **Mono stand** — A mono stand is a column with a movable cross arm, for mounting a camera.
- **Book cradle** — Book cradles are special supports for scanning bound documents.
- **Tripods** — A tripod is a three-legged stand for a camera, used to stabilize and elevate the camera.
- **Lighting** — The key to taking quality images with a digital camera is adequate, proper and even lighting.

In order to capture the appropriate image file size for creating a sustainable archive and appropriate for all output possibilities (print, web and multimedia), the camera should have a minimum of 10 to 14 megapixels. This will generally result in a master TIFF file of around 50 to 70 megabytes. If the project requires higher resolutions, a medium or large format digital capture device must be used to produce files from 200 to 500 megabytes.

SOFTWARE

Scanning Software

The link between your scanner and computer hardware is the software that both controls the scanner and passes scan information directly to your storage media, image editing software or both. Most scanners come equipped with software to manage this transfer for you. Don't use the scanning software's automatic settings. If available, opt for the "advanced" or "custom" settings mode, and disable features that alter the image in any way. The advanced settings mode will allow you to control both the image captured and the file formats generated by your scanner.

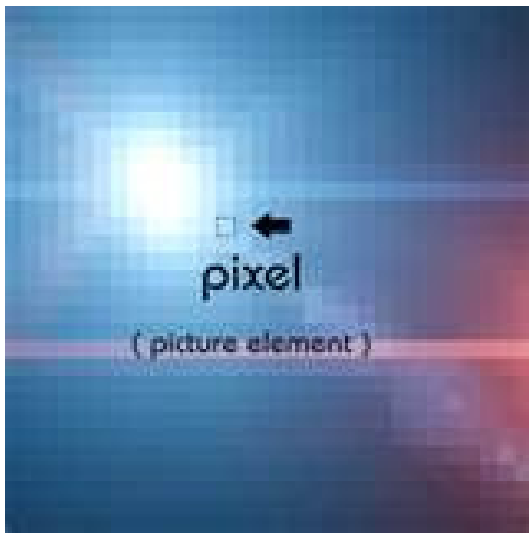
Image Editing Software

Once you have captured an image from your scanner and have saved an unaltered master file to your storage media and you need to manipulate surrogate copies you will use with image editing software. Project managers will need professional image editing software. Adobe Photoshop is the industry standard for the creation of surrogates which can be delivered via the web, print publication or for in-house uses such as exhibits.

BASICS

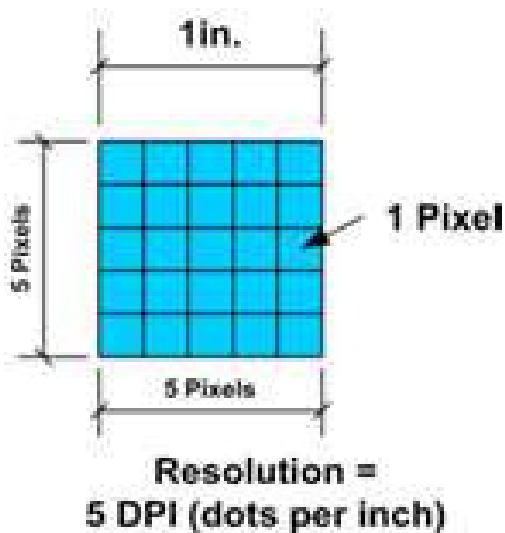
Pixels

Pixel is Picture Element, points of information which make up a digital image.



- **High resolution has many pixels per inch**
- **High resolution shows more detail**
- **Low resolution has few pixels per inch**
- **Low resolution shows less detail**

DPI

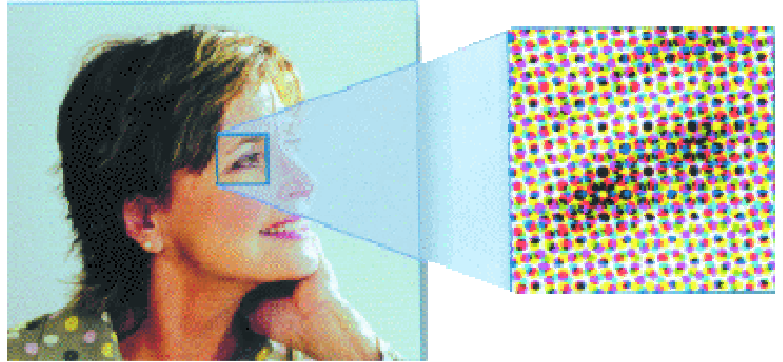


DPI: Dot per inch refers to output devices, the number of dots of ink per inch a printer puts on the paper or onscreen monitor.

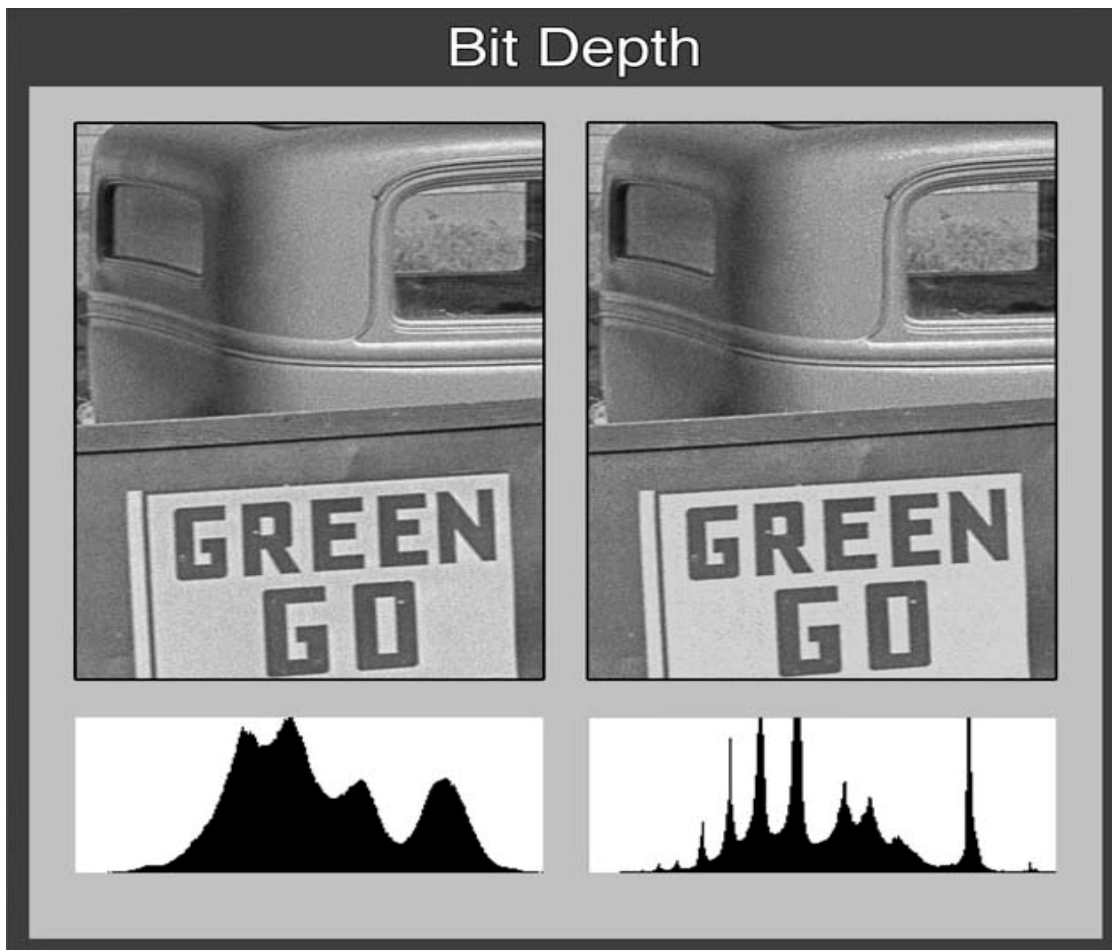
Normally this will be used for prints or published materials.

Bit Depth

Bit depth measures the number of colors (or levels of gray in grayscale images) available to represent the color/gray value in the original work.



Bit depth is the number of bits of digital information used to define one pixel. The higher the bit depth the greater the number of grays or colors fill a pixel



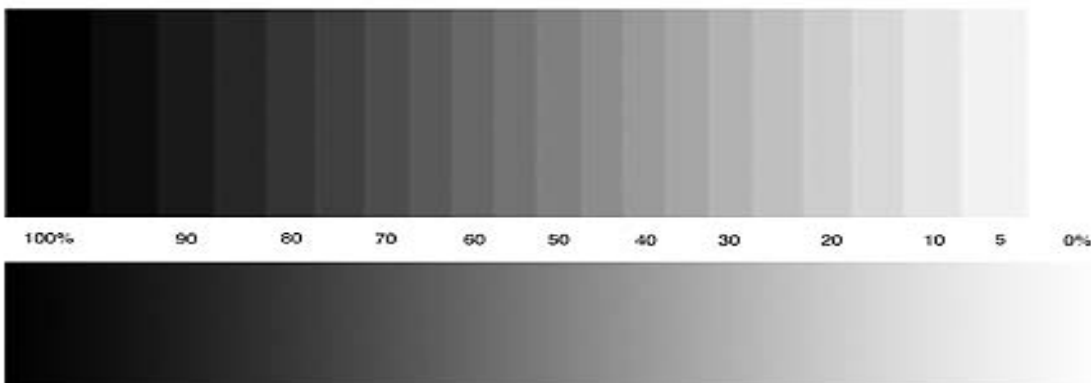
Resolution determines the quality of an image.



Resolution is the number of pixels per square inch on a computer-generated display; the greater the resolution, the better the picture.

Tonal Range

Put simply, the tonal range is the range of tones between the lightest and darkest areas of an image. An image with a wide tonal range will have both very dark (black) and very light (white) elements within the photograph. An image with a narrow tonal range will be limited in its tonal scope to mid ranges (varying shades of grays).

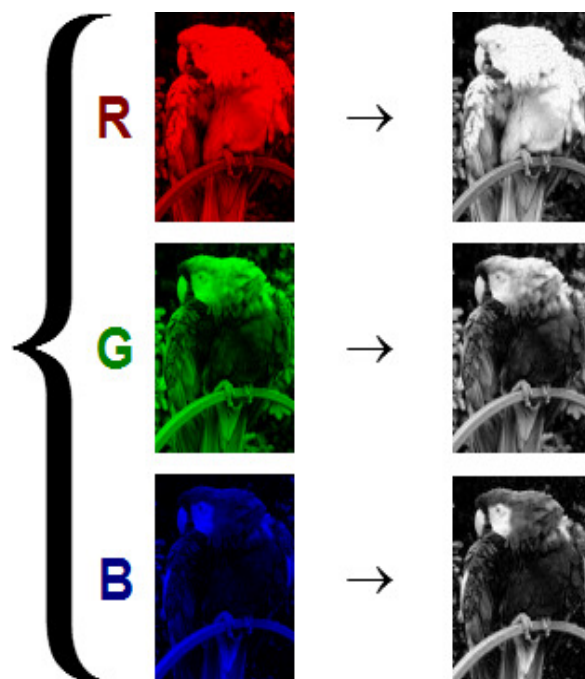
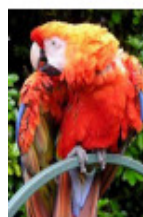


Four modes to capture

There are 4 different ways to register the images you are scanning:

- **Bitonal** — One bit per pixel representing black and white.
- **Grayscale** -Grayscale images are distinct from one-bit black and white images, which in the context of computer imaging are images with only the two colors, black, and white. Grayscale images have many shades of gray in between. Grayscale images are also called monochromatic, denoting the absence of any chromatic variation.
- **RGB** — Stands for "Red Green Blue." It refers to the three hues of light (red, green, and blue, for those of you that are a little slow), that can mix together to form any color. When the highest intensity of each color is mixed together, white light is created. When each hue is set to zero intensity, the result is black. TVs and computer monitors use RGB to create the colorful images you see on the screen. In print, however, the 4 colors -- cyan, yellow, magenta, and black (CYMK) -- are used to create color images.
- **CMYK** — Cyan-Magenta-Yellow-Black. CMYK is the four-color model used for printing standard colors.

The reverse is also possible: to build a full color image from their separate grayscale channels. By mangling channels, offsetting, rotating and other manipulations, artistic effects can be achieved instead of accurately reproducing the original image.



Color Space

Setting the color space is critical to digital capture and can be set within your photo editing software prior to digital capture, unless you are using a digital camera and capturing a RAW file. A device color space simply describes the range of colors, or gamut, that a camera can see, a printer can print or a monitor can display.

Compression

Certain amount of information is discarded during the compression process. This is called “lossy”.

Although the discarded information may be invisible to the human eye, a loss of quality occurs.

Lossy compression formats also introduce generational loss – each time a lossy image is manipulated or edited, the quality of the image decreases.

Generational loss is one of the reasons master and service master images are not stored using compression.



File Formats

- **RAW Files-** Image file that contains unprocessed data. The proprietary nature of these files is of concern for the long-term preservation and access of these digital files. After processing or editing and before use, RAW files must be converted to an open standard format such as JPEG or TIFF.
- **DNG**—In response to concerns over the future support for proprietary RAW files, Adobe has produced the DNG (Digital Negative) format. This format provides a common platform for information about the file and adjustments to the image. Because of this, cataloging applications can see metadata that has been entered in Photoshop, and these programs can see the image adjustments made in Photoshop.
- **TIFF** — Tagged Image File Format (TIFF) is the format of choice for archival and master images. It is a flexible, highly portable, widely accepted, open standard image format and considered the professional image standard. Due to the large file size of TIFF images, they are not suitable for web delivery.
- **JPEG 2000** — JPEG 2000 is a wavelet-based standard for the compression of still digital images. It was developed by the ISO JPEG committee to improve on the performance of JPEG while adding significant new features and capabilities to enable new imaging applications. Beyond image access and distribution, JPEG 2000 is being used increasingly as a repository and archival image format.
- **JPEG** — Joint Photographic Experts Group. JPEG image files allow online access to full screen image files because they require less storage and are therefore quicker to download into a web page.

Metadata

Metadata is defined as information created and maintained to describe the content and context of a unit of digital information or a digital object.

Most common metadata schemes include the following sets of information:

- **Descriptive Metadata-** Metadata that describes content and context of a digital object in a way that is similar to how the library catalog describes a book.
- **Administrative Metadata-** Metadata that is like a medical chart for a patient, describing what software and hardware is necessary to view the files, what migrations of this data have occurred in the past, and what errors resulted from past migrations.
- **Structural Metadata-** Describes relationships between multiple digital files, such as page order in a digitized book or manuscript. For example, is required to present or maintain the correct order of chapters and pages of a book.
- **Preservation metadata-** Includes elements of descriptive, technical, and structural metadata specially needed for preservation. Like pg 14 format

Some metadata is created automatically by software and/or hardware when digital information is created (e.g., digital cameras often add the date of the photo to the image file). Other metadata must be created by librarians or other technical specialist upon inspection of the digital information. Sometimes metadata is created by individuals who add descriptive information when they contribute digital information to a website; this is known as “author-supplied” metadata.

Master Files

Good digital imaging projects scan a high-quality master or archival image and then derive multiple versions in smaller sizes or alternative formats for a variety of uses.

There are compelling preservation, access and economic reasons for creating an archival-quality digital master image: it provides an information-rich, research-quality surrogate.

The master image file format should be:

- **Nonproprietary / open source**
- **Large color space**
- **Uncompressed**

The recommended format is Image File Format (TIFF).

Service Master Files

The service master is an optimized working copy of the master file that can be used as a source for all subsequent derivatives. Service masters are also used to create print publications.

Derivative Files

Derivative files are created from the service master and are used for general Internet or network access.

File Naming Conventions

Systematic file naming will help you obtain system compatibility, interoperability and demonstrate ownership of the digital asset.

File naming recommendations include:

- **Use lowercase letters of the Latin alphabet and the numerals 0 through 9.**
- **Begin each file name with a two- to three-character acronym representing the institutional name followed by a second two- to three-character acronym representing the department or unit name (when applicable).**
For example: am000421.tif (Art Museum, digital object, number 421, TIFF)
- **Follow the institutional and departmental acronyms with an object ID. The object ID consists of any unique numbering scheme already in use to represent the object or, if no such number exists, a short description representing the item.**
- **For example: lhs_9800_b5f2i26.tif. (Lamamie Historical Society, Accession Number 9800, Box 5, Folder 2, item 26)**
- **File names should be limited to 31 characters, including the three character file extension**

When selecting a file naming convention, think long-term. Select a system that will outlast staff involved in the current project. Consider the number of files your institution will ultimately be managing.

Remember human error – if technicians will manually be assigning file names, how simple or easy will it be to make a mistake? File names do not take the place of metadata. Keep them simple and straightforward.

Watermarks

A watermark is information stored in or on a digital image, which allows image creators to store copyright or branding information to images, audio and video files and documents. Watermarks are applied to images in the hope of reducing misuse or unauthorized distribution of images.

There are two types of watermarks in use for digital image files; visible and electronic:

- **Visible watermarks are applied on top of the image, very much like a seal is applied to an official document. Often these watermarks consist of the name of the institution that owns the file, that institution's official seal or some other identifying logo.**
- **Electronic watermarks are imbedded in the image file, and they are invisible. They usually use a numeric code licensed by an electronic watermarking firm.**



The numeric code is specific to the institution that owns the files. The cost of watermarking varies. Visible watermarking is virtually free, but invisible watermarking can be costly.



Numerous factors play an important role in the final outcome of a digitization project. Original condition of materials, quality and maintenance of equipment, staff training and external lighting are some factors that can influence the quality of images.

Images should be inspected while viewing at a 1:1 pixel ratio or at 100 percent magnification or higher. Quality is evaluated both subjectively, by project staff (scanner operator, image editors, etc.) through visual inspection, and objectively, in the imaging software.

White & Black Points

Optimum placement of whites and blacks is best observed through the histogram, although the image itself will be examined without the histogram on the screen. It is important to look at the number value assigned to the brightest highlight and the darkest shadow.

Highlights should not read a number value higher than 247, and shadows should not be less than 7 or 8. If these numbers are exceeded, the scan must be redone. This is particularly vital if the original image has a short dynamic range.

The white and black points must not be set on 0 and 255, as this will stretch the dynamic range of the image, creating gaps in the histograms, and thus unusable scans.

Visual Inspection

Things to look for during archival master visual inspection may include:

- **Image is the correct size**
- **Image is the correct resolution**
- **File name is correct**
- **File format is correct**
- **Image is in correct bit depth and color mode**
- **No loss of detail in highlight or shadows**
- **Even tonal values**
- **Correct focus**
- **Not pixilated**
- **Excessive dust spots or other objects**
- **Image not cropped**
- **Image not rotated or reversed**
- **Correct color balance**
- **Histogram:**
 - No spikes or clipping**
 - No tonal values lower than 9 or higher than 247**

If levels and curves are adjusted in the service master, then you should check the histogram again.

Adding Watermarks to an Image in PastPerfect

The ability to attach up to 999 digital image files to each of your PastPerfect records is enabled with the optional Multi-Media Upgrade. Additionally, the Multi-Media Upgrade for PastPerfect 4.0 allows for the attachment of audio files, video files, Microsoft Word documents, Microsoft Excel files, PDF files, web pages and more to each catalog record. In order to view some of these attached files, you will need to have the related software installed on your computer.

The attachment of digital image files to your records may assist with the ability to identify items, perform research and diminish the need for handling, which will help preserve your items. So, while adding images can increase the usability of your data, watermarking your images can keep them secure.

1. Find the record containing the image to be watermarked.
2. Click on the “Image Management” button, located below the thumbnail of the image.
3. Click on the “Watermark” button in the lower right corner of the Image Management screen.
4. Click on the “Setup Watermark” button.
5. The “Text shade” combo box contains two choices: grey or white. This will provide the color base for your text watermark.
6. “Visibility %” allows you to adjust the intensity of the watermark, using the up/down arrows. Because the watermark is based on shading, even if the percentage is at 100, you may need to apply the watermark a few times, before it is at the desired opacity.
7. The “Font” combo box allows you to choose between Arial or Times New Roman.
8. “Bold” and “Italic” are check boxes. Placing a check in the box will apply that style to the watermark.

- 9. If you wish to do so, you may enter custom location coordinates for your watermark. It should be noted that x2 and y2 must be greater than x1 and y1 and “Location of Watermark” on the Watermark screen must be set to “Custom.”**
 - 10. The “Text of Watermark” field can contain a maximum of 125 characters. The height of the letters will typically depend on the number of characters used as well as the width of the area to which the watermark is applied.**
 - 11. The “Remove spaces before and after text” check box will remove any extra spaces surrounding the text. Due to the fact that the height of the letters is partially dependant upon the number of characters in the “Text of Watermark” field, the removal of spaces before and after the text can result in a larger watermark.**
 - 12. When you have set up your watermark, please click on the “Exit” button to return to the Watermark screen.**
 - 13. Next, please choose the location of the watermark using the combo box in the middle of the Watermark screen. If you wish to use the custom coordinates entered on the Setup Watermark screen, “Location of Watermark” should be set to “Custom.” Also, the height of the text will be affected by the area allotted for the location of the watermark.**
 - 14. To preview your watermark settings, please click on the “Apply Watermark” button. You may need to click on this button several times before the watermark is at the desired opacity. The more you click on the “Apply Watermark” button, the more opaque the text will become.**
 - 15. Click on the “Cancel” button at the bottom of the Watermark screen, if you do not like the appearance of your watermark and wish to remove it.**
 - 16. If, following step 14, you are happy with the appearance of your watermark, please click on the “Exit” button.**
 - 17. Then click on the red “Save” button on the right, central area of your Image Management screen. Once you save the image, you cannot remove the watermark.**
 - 18. Click on the “OK” button when you receive the PastPerfect Message “Image replaced with modified image.”**
- .

ADDENDUM II

Bibliographical and Online Resources

BCR's Collaborative Digitization Program. "Questions to Ask Before Your Begin a Digitization Project." BCR. http://www.bcr.org/cdp/digitaltb/getting_started/faq.html (accessed May 21, 2008)

BCR's Collaborative Digitization Program. "Standards"
<http://www.bcr.org/cdp/digitaltb/metadata/standards.html> (accessed May 21, 2008)

BCR's Collaborative Digitization Program. "Legal Issues to Consider When Digitizing Collections" BCR. <http://www.bcr.org/cdp/digitaltb/laws/index.html> (accessed May 21, 2008)

Westerns States Digital Standards Group– A Cultural Heritage Collaboration "Digital Imaging Best Practices"

Chapman, Stephen. Handbook for Digital Projects: A Management Tool for Preservation and Access <http://www.nedcc.org/digital/intro.html>

Pastperfect E- Newsletter

Images: Microsoft Art clipping



Highline Historical Society

P.O Box 317

Seahurst, WA 98062

Web site: www.highlinehistory.org

E-mail: editor@highlinehistory.org

E-mail curator: jjnmckay@msn.com

Funded with the assistance of

