Wisconsin Heritage Online
Digitization and Scanning Glossary

access image
A file derived from a **digital master file**, used to make a **digital** object accessible. An access image is resized and compressed to reduce storage requirements and make online delivery faster. *See also* **derivative image**.

batch processing
The automated application of a given process to multiple files, e.g. resizing or renaming.

bit
The smallest unit of computer data, denoted by a single binary value, either 0 or 1. Eight bits make up one **byte** in most computer systems.

bit depth (1-bit, 8-bit, 24-bit)
The number of bits used to describe the color value of each **pixel** in an image, which in turn dictates the number of colors available to a given media file, monitor, or other device. An 8-bit image has 256 possible colors. A 24-bit image has approximately 16 million. Also known as **color depth**. *See also* **dynamic range**.

bitmap (or bitmapped image, raster image)
An image made up of a given number of **pixels**, each with a specific color value, laid out in a grid. Ideal for reproducing photographic representations, because a sufficient quality and quantity of pixels can give the appearance of a **continuous tone** image. **Resizing** will affect apparent image quality. For instance, enlarging an image involves enlarging each pixel, which entails a reduction in **resolution**.

born-digital
Creations originally generated in **digital** form, rather than copies or surrogates of analog originals. They exist entirely in a digital environment. Examples include software, Web pages, hypertext fiction, and digital art.

CD-ROM (Compact Disk, Read-Only Memory)
A storage disk for computer files. Gold-reflective-layer CDs are most recommended for long-term storage. A standard CD-ROM stores approximately 650 megabytes of data. *See also* **DVD-ROM**.

CMYK (Cyan, Magenta, Yellow, Black)
CMYK is a subtractive **color model**, using a mix of cyan, magenta, yellow, and black inks to reproduce a range of colors. CMYK is the most basic color process used in print.

calibration
The comparison of the specifications of **image capture**, processing, display or output devices to a known **standard** to determine, and perhaps correct, any deviation or error. *See also* **color management**.

capture resolution
The number of pixels per inch (ppi) that a **scanner** or **digital camera** is capable of capturing, or the number of pixels per inch captured when a particular image is digitized.

channels
The separate color components used by various **color models**. By default, **RGB** images have three channels: red, green, and blue; **CMYK** images have four: cyan, magenta, yellow, and black. Extra or "alpha" channels (8-bit) can be added to describe, for example, levels of transparency, or be used as masks that allow or restrict the output of color.

clipping
Clipping can cause a thresholding problem with the image data if you adjust the endpoints (black and white) too far in on the **histogram**. The resultant image will have missing values in the highlight/shadow areas.

color chart or **target**
A **calibration** target consisting of a matrix or spectrum of colors set to a known **standard**. Color charts can provide reference points to ensure accuracy of color capture and to calibrate output devices. *See also* **gray scale**.

color correction
The process of adjusting color values in an image to match the original or a reference image in order to compensate for the normal shifts and biases incurred during digitization and subsequent image processing.

color depth *See* **bit depth**.

color management
The practice of **calibrating** all devices in the **image capture**, processing, and output chain to ensure the fidelity of **digital** image files to the objects they represent. A color managed workflow can be achieved by manually calibrating devices using **color charts** and **ICC** profiles followed by regular reassessment.

color profile
A file containing data that describes a particular device's **color space** in relationship to an ideal, theoretical, device-independent color. Profiles facilitate the conversion of images from one color space to another and form the foundation of color management systems (CMS). Most profiles are generated using **ICC** standards.

color space
A three-dimensional geometric representation of the colors that can be discerned and/or created by a particular color model. May also refer to the range of possible colors that can be produced by a particular output device--such as a monitor, color printer, photographic film, or printing press--or the color description abilities of a particular file format; may also be known as a color gamut. Usually described by a color profile.

compression
The reduction of image file size for processing, storage, and transmission. The quality of the image may be affected by the compression techniques used and the level of compression applied. There are two types of compression: **Lossless compression** reduces the storage space needed for an image file without loss of data. If an image has undergone lossless compression, it will be identical to the image before it was compressed. Primarily used with bitonal images. **Lossy compression** reduces the storage space needed for an image file, but it discards information. If an image that has undergone lossy compression is decompressed, it will differ from the image before it was compressed, even though the difference may be difficult for the human eye to detect. Depending on the amount of compression, a considerable visual loss can occur.

DVD-ROM (Digital Versatile Disk, Read-Only Memory)
A type of write-once, read-many (WORM) disk used to store and distribute large amounts of **digital** data on low-cost, optically recorded media. The DVD-ROM is a newer and more densely packed form of storage than the **CD-ROM**. DVDs are equivalent to gold CDRs, and are physically stable. A double-sided, dual layer disk can store up to seventeen gigabytes of data, but the more densely packed the data is, the more vulnerable it is to degradation.

decompression
The process of retrieving compressed data and reassembling it so that it resembles its original form before compression.

derivative image
An image produced from another image. It is usually created through some kind of automated process. Techniques used to create derived images include sampling to a lower resolution, using lossy compression techniques, or altering an image using image processing techniques. *See also* **access image**.

derivative master *See* **service master**

digital
Electronic technology that generates, stores, and transmits data in terms of a limited number of discrete states, most commonly as binary data in which two possible states, positive or nonpositive, are represented by 1 or 0, respectively. Because there are only two possible values, the accuracy of binary digital data at any given point is relatively easy to test, and therefore digital technology facilitates the creation of accurate copies. *See also* **digital image**.

digital archival master
The raw, original image captured by the scanning process and/or an image created and managed to optimize longevity and future usefulness. File naming, file formatting, **color space** selection, **capture resolution**, and similar specifications should be based on documented **standards**. Archival masters may be used as the source for **service masters** and then **access images.** *See also* **digital preservation**.

digital artifact
An error introduced into an image during capture or digitization, formatting, **compression**, or other transformation processes. An error can be introduced by dust or dirt on the original or on the scanner, or scan lines, etc. Artifact also refers to the perceptible degradation of an image after a **lossy compression** schema has been used.

digital assets
A collection of computer files that contain intellectual content (images, texts, sounds, video) and/or descriptive metadata of the content and its digital format. They represent an investment for the depositor and an information resource for the researcher.

digital back
Digital backs resemble microfilming stands. The source material is placed on the stand and the camera is moved up or down in order to fit the material into its field of view, which allows for the scanning of a range of broader range of sizes of materials. A digital back can also be called a copystand scanner. *See* **digital camera, flatbed scanner**, **transparency scanner**.

digital image
A computer file which, when used in conjunction with the proper software, will display an image on the computer screen or print out to a digital device such as a laser printer. Each pixel is assigned a tonal value (black, white, a shade of gray, or color) and is represented digitally in binary code (zeros and ones). The term "image" does not imply solely visual materials as source material; rather, a digital image is simply a representation of whatever is being scanned, whether it be manuscripts, text, photographs, maps, drawings, blueprints, halftones, musical scores, 3-D objects, etc.

digital object
Data (the content or "essence" of a **digital** file) and the metadata describing it, regarded together as a single entity. Also known as a digital asset, an information object, or an information package. May also refer to **born digital** objects.

digital preservation
The managed activities necessary for ensuring the long-term retention and usability of digital objects.

digitization
The conversion from printed-paper, film, or some other media, to an electronic form in which the source material is represented as either black and white dots, or color or grayscale pixels. Also known as analog-to-digital conversion and **image capture**.

dots per inch (dpi)
A measurement of the scanning resolution of an image or the quality of an output device. dpi expresses the number of dots a printer can print per inch, or that a monitor can display, both horizontally and vertically.

dynamic range
The number of colors or shades of gray that can be represented by a pixel. The smallest unit of data stored in a computer is called a bit. Dynamic range is a measurement of the number of bits used to represent each pixel in a digital image. 1-bit or bitonal means that a pixel can either be black or white. Bitonal imaging is good for black and white images, such as line drawings and text. However, scanning in grayscale rather than bitonal may produce a richer image. 8-bit color or 8-bit grayscale means that each pixel can be one of 256 shades of color or one of 256 shades of gray. 24-bit color means that each pixel can be one of 16.8 million colors. *See also* **bit depth**.

file format
The specific way digital information is made and stored by the computer. Not all software applications can read and/or manipulate all file formats. *See* **GIF, JPEG, TIFF**.

file size
The file size of an image is proportional to its resolution. The higher the resolution, the bigger the file size. File size is not the same as image size. *See* **image size**

flatbed scanner
An **image-capture** device resembling a photocopy machine. The object to be scanned is placed facedown on a glass plate, and a **CCD** (charge-coupled device) array that passes beneath the glass captures an image of the object by **sampling** it at regular intervals.

flatten
In Photoshop, the Flatten function merges the original image and any adjustment layers that were created. The file is now back to the size it was before editing. You will no longer able to make corrections to the separate layers.

format
A specification for organizing data. **Digital images** (and their associated metadata) may be presented in a number of formats depending on **compression** schemes, intended use, or **interoperability** requirements. Some image formats are broadly decipherable, while others may only be accessible to certain application programs. *See* **GIF**, **JPEG**, **JPEG2000**, **PNG**, **TIFF**.

GIF (Graphics Interchange Format)
A common graphic file format on the web. Used by online services and Web browsing software, GIFs contain information compressed into a relatively small file size. It uses LZW compression, which makes it good for color and grayscale images. GIF is useful for limited-palette, monochrome, or **thumbnail** images.

grayscale
An image type that uses black, white, and a ranges of shades of gray. The number of shades of gray depends on the number of bits per pixel. The larger the number of shades of gray, the better the image will look, and the larger the file will be. *See* **color chart** or target.

histogram
Illustrates how pixels in an image are distributed by graphing the number of pixels at each color intensity level. This can show you whether the image contains enough detail in the shadows (shown in the left part of the histogram), midtones (shown in the middle), and highlights (shown in the right part). A low-key image has detail concentrated in the shadows; a high-key image has detail concentrated in the highlights; and an average image has information distributed across the histogram. An image with full tonal range has a high number of pixels in all areas, without gaps. *See also* **combing**.

image capture
Using a scanner or other device to create a digital representation or electronic photograph of an image.

image file format
When a page is scanned, the page can be stored in a number of file types. The type should be chosen based on the desired use of the image, and the software that will be used. Different file formats commonly use different methods of compression as well, and some types of images compress better using some formats rather than others.

image file size
The amount of computer storage space a file requires; usually measured in kilobytes (K) or megabytes (M, MB, mgs or "megs"). An image file that is 5 x 7 inches, 8-bit grayscale (as in a black and white photo), resolution 300dpi, is 3MB in size.

image manipulation
Making changes (such as tonal adjustments, cropping, moiré reduction, etc.) to an image using image-processing software.

image resolution
The number of **pixels**, in both height and width, making up an image. Generally, the higher the number of pixels, the greater the image's clarity and definition. *See* **resolution**.

image size
Describes the actual physical dimensions of an image, not the size it appears on a given display device.

interpolated resolution
The **resolution** at which a device is capable of capturing or generating an image using **interpolation** (i.e., using data that has no authentic relation to an original). *See* **optical resolution**.

interpolation
An estimation of a value within two known values. A means by which a device can exceed its **optical resolution** capacity by inserting new **pixels** in between those derived by sampling the original. It can improve apparent picture quality; however, interpolated images tend to look blurred when they are enlarged.

JPEG (Joint Photographic Experts Group)
JPEGs are helpful in allowing access to full screen image files on-line because they require less storage and are therefore quicker to download into a web page. A compression algorithm for

condensing the size of image files. JPEG supports **lossy compression**, which means that it will throw away some detail to achieve better compression.

JPEG2000
A file **format** that uses wavelet compression to allow both **lossy** and **lossless compression.** It can provide scalable images from a single compressed file. There are commercial implementations available, and open-source implementations are in development. A separate standard to **JPEG**, and generally able to provide reasonable image quality at higher compression ratios (some sources cite ratios of 2:1 with lossless compression and up to 200:1 with lossy compression).

lossless compression
Reduction in file size without loss of information, achieved by storing data more efficiently. A **bitmapped image** that has undergone lossless compression will be identical to the original uncompressed image when decompressed. The **GIF**, **TIFF**, **PNG**, and **JPEG2000** image formats allow lossless compression, which cannot shrink file size to the extent possible with **lossy compression**.

lossy compression
Reduction in file size that involves permanent loss of information. Algorithms selectively discard data in order to attain a greater size diminishment than is possible with **lossless compression**. Entails a decrease in quality, but this is often imperceptible with image files, depending on the level and type of compression employed. The **JPEG** and **JPEG2000 formats** allow lossy compression.

monitor resolution
May be used interchangeably with **screen resolution** or may indicate the maximum possible resolution of a computer monitor. Higher monitor resolution indicates that a monitor is capable of displaying finer and sharper detail, or smaller **pixels**.

noise
Unwanted data or unidentifiable marks developed in the course of scanning, processing, or data transfer that are not in the original.

optical resolution
The **resolution** at which a capture device, such as a **scanner** or **digital camera**, is capable of capturing **pixel** values based on actual samples taken from an original to construct an image. Optical resolution is the true measure of the capture capacity or quality of a scanner, as opposed to **interpolated resolution**.

output resolution
The detail and clarity produced by the **resolution** of an image based upon the dimensions with which the image will be displayed or printed.

pixel
"Pixel" is short for picture elements, which make up an image, similar to grains in a photograph or dots in a half-tone. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it. Pixel size, frequency, and color determine the accuracy with which photographic images can be represented. The greater a pixel's **bit depth**, the greater the number of different shades or colors it can represent.

pixels per inch (ppi)
Pixels per inch (ppi) or pixel density is a measurement of the **resolution** of a computer display, related to the size of the display in inches and the total number of **pixels** in the horizontal and vertical directions. This measurement is often referred to as **dots per inch**, though that

measurement more accurately refers to the resolution of an output device, i.e., a printer. PPI may also be used to describe the resolution of an image scanner or digital camera.

PNG (Portable Network Graphics)
A patent-free file format for **lossless compression** of images that provides some additional features that improve the ability to control image appearance over the **GIF** format.

printer resolution
The maximum density of dots per inch that a printing device is capable of producing or the density of dots per inch used in a particular printing.

quality control
Techniques to ensure accuracy and high quality are maintained through various stages of a process. Quality control during **image capture** might include comparing the scanned image to the original and then adjusting colors, orientation or checking for artifacts.

RGB (Red, Green, Blue)
An additive **color space** or system for representing the color spectrum using combinations of red, green, and blue. Used in video display devices, it is the standard color system for most digital imaging devices and applications.

refreshing
The transfer of digital files to new media on a regular basis. This is the most important part of an institution's long-term commitment to digitization. A **digital preservation** strategy will protect against the possible degradation of digital content due to storage media decay by copying digital information held on a particular storage medium to a new medium, of the same or different type, while keeping the digital information itself in the same format.

resampling
May be used interchangeably with **resizing**. Alternatively, refers to changing the number and values of **pixels** in an image, technically by creating a new, empty **bitmap** of the desired dimensions and using the original image pixels as the basis from which to work out the values for each new pixel using various algorithms. Resampling generally involves **interpolation** and should be used cautiously. *See also* **compression**.

resizing
May be used interchangeably with **resampling**. To change the size of an image by reducing or increasing the resolution and/or the square measurement of the file. It is always preferable to scan an image at the size needed rather than to try to increase the size or resolution later. It is not possible to add meaningful data to an image after it is scanned.

resolution
The number of pixels (in both height and width) making up an image. The more pixels in an image, the higher the resolution, and the higher the resolution of an image, the greater its clarity and definition. Resolution is expressed for the scanner and for the screen as **pixels per inch (ppi)**; for the printer as **dots per inch (dpi)**. The more data per inch (pixels, dots) the higher the resolution of the image and the better looking the image will be. *See* **capture resolution**, **effective resolution**, **image resolution**, **interpolated resolution**, **monitor resolution**, **optical resolution**, **output resolution**, **printer resolution**, **screen resolution**.

retouching
Used in image-editing programs to crop or repair overexposed or underexposed photos or eliminate red-eye.

scanner
A device that breaks down an image and records it as a digital file for computer editing and display. There are many types of scanners, such as flatbed scanners, digital backs, slide scanners, microfilm scanners, transparency scanners and digital cameras.

screen resolution
Sometimes used interchangeably with **monitor resolution**. Otherwise refers to the number of **pixels** shown on a computer monitor screen. Screen or display resolution is variable and may be set to a number of default settings, such as 800 x 600 or 1024 x 768.

service master
A high-quality "working" image file that is derived from an **archival master** image file, then subjected to some form of processing, such as **color correction**. May also be known as a submaster. Typically used as the source from which smaller, lower-quality **access images** intended for transmission over the Internet are derived, while archival masters remain unmanipulated.

sharpening
Sharpening enhances the definition of edges in an image. Whether your images come from a digital camera or a scanner, most images can benefit from sharpening. The degree of sharpening needed varies depending on the quality of the digital image and output device.

sRGB
sRGB is a **color space** that most monitors and graphics cards support by default. sRGB was developed to match the color space of a typical computer monitor. It's the default color space of Windows XP, Web browsers, numerous image browsers, image editors and other software on the Windows platform. sRGB is a small gamut colorspace and you will shortchange your images where a wider gamut is supported. **Use Adobe RGB, not sRGB.**

TIFF (Tagged Image File Format)
An industry standard image file **format** developed for scanning. A file storage format implemented on a wide variety of computer systems, usually used for archival scans. TIFFs can be read by both Macintosh and PC/Windows applications. TIFF is widely used as a format for storing uncompressed, or **losslessly compressed**, **digital image** data, though it also supports several compression algorithms. TIFF is in wide use for high-resolution scanning, image archiving, and editing applications.

thumbnail
A proxy image, generally scaled to a much smaller size, used to represent a parent image in circumstances where loading the original is undesirable. Often used on the Web to display tables of smaller images with links to a larger view, preserving bandwidth and enabling the display of more images in the same area for the purposes of browsing.

transparency scanner
A **scanner** specifically designed to capture images from film or transparent media.

watermark
A unique identifier added to a content file, such as an image, which can be visible or invisible to viewers. The mark, which could be a statement, symbol, or hidden encoding, is designed to persist through processing and serve as evidence of ownership in order to deter piracy. Watermarks are only embedded into derivatives (access) images.

zooming
To make an image appear larger (zoom in) or smaller (zoom out) by re-displaying the image at different resolutions. Higher resolutions will make the image appear larger and easier to read.

Definitions selected and adapted from:

California Digital Library. Digital Library Building Blocks Glossary

Collaborative Digitization Program. Digital Imaging Best Practices Glossary.

J. Paul Getty Trust. Besser, Howard. Introduction to Imaging Glossary.

North Carolina ECHO (Exploring Cultural Heritage Online). Glossary.

Princeton University. Princeton Imaging Glossary.

Scott, Alyce. WebJunction, 2006. Glossary of Scanning and Digital Imaging Terms.