

## Digitization Glossary

### Mount Holyoke College Digitization Center

- **Archival Image**  
An image meant to have lasting utility. Archival images are usually kept off-line on a cheaper storage medium such as CD-ROM or magnetic tape, in a secure environment. Archival images are of a higher resolution and quality than the digital image delivered to the user on-screen. The file format most often associated with archival images is TIFF, or Tagged Image File Format, as compared to on-screen viewing file formats, which are usually JPEGs and GIFs.
- **Algorithm**  
The specific process in a computer program used to solve a particular problem.
- **Artifacts** Visual digital effects introduced into an image during scanning that do not correspond to the original image being scanned. Artifacts might include pixellation, dotted or straight lines, regularly repeated patterns, moiré, etc.
- **Backup**  
Never keep only a single instance of any digital image. It is always important to maintain at least two copies of every digital image created. Should anything happen to the original, you will still have another copy available. Backing up your scans is considerably cheaper than having to rescan the objects.
- **Batch Process**  
A method of processing digital images in large numbers such that minimal staff time is required. This becomes important when performing the same task on multiple image files.
- **Bit depth**  
The bit depth of an image refers to the number of bits used to describe the color of each pixel. Greater bit depth allows more colors to be used in the color palette for the image.
  - 1-bit color** is the lowest number of colors per pixel in which a graphics file can be stored. In 1-bit color, each pixel is either black or white.
  - 8-bit color/grayscale** has eight bits assigned to each pixel, providing 256 colors or shades of gray.
  - 24-bit color** has 24 bits assigned to each pixel, representing 16.7 million Colors. 8-bits are assigned to each of the red, green, and blue components of a pixel.

see also dynamic range.

- **Calibration**

The act of adjusting the color of one device relative to another, such as a monitor to a printer or scanner. Or, it may be the process of adjusting the color of one device to some established standard.

- **Capture Device**

These include flatbed scanners, drum scanners, film scanners, and digital cameras. They use electronic devices to capture images rather than photographic film.

- **CCD Array** Charge-Coupled Device array. Light sensitive diodes used in scanners and digital cameras that sweep across an image during capture and, when exposed to light, generate a series of digital signals that are converted into pixel values.

- **CMYK (Cyan, Magenta, Yellow, Black)**

One of several color encoding systems used by printers for combining primary colors to produce a full-color image. In CMYK, colors are expressed by the “subtractive primaries” (cyan, magenta, yellow) and black.

- **Color Correction**

The process of altering colors as they appear in a digital image or in print to ensure they accurately represent the work depicted.

- **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 is a process that 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.

**Lossy** compression is another process that reduces the storage space needed for an image file, but it discards information (often information that is "redundant" and not perceptible to the human eye).

There are both standard and non-standard compression techniques available. In general, it is better to employ a compression technique that is supported by standards, non-proprietary, and maintained over time. In selecting a compression technique, it is necessary to consider the attributes of the original. Some compression techniques are designed to compress text; others are designed to compress pictures.

- **Continuous Tone**  
Smoothly varying gradation of tones. Includes photographs, postcards and some original artwork, as well as graphic art that is produced with continuous tone-like attributes.
- **Derivative image**  
an image that has been created from another image. Usually involves a loss of information. Techniques to create derivative images include sampling to a lower resolution, using lossy compression techniques, or altering an image with image manipulation software during image processing.
- **Descriptive metadata**  
metadata primarily intended to serve the purposes of discovery, identification, and selection.
- **Digital Camera**  
A camera that does not contain any film but records the image as a digital object. The digital camera allows for the scanning of a range of differently sized materials. Resolution of digital cameras is usually fixed and is expressed as a pixel ratio.
- **Digital Image**  
An electronic photograph scanned from an original document, made up of a set of picture elements ("pixels"). 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.
- **Digitization**  
the conversion from printed paper, film, or other media formats to an electronic format where an object is represented as either black and white dots, color or grayscale pixels, or 1s and 0s.
- **Download**  
to transmit a file from one computer to another. Usually implies retrieving a file from a remote computer to a local one, or from a large computer to a smaller one. FTP is a commonly used command for this.
- **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. Also called bit depth.

- **Grayscale**  
An image type that uses black, white, and a range 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.
- **Image capture**  
using a scanner, digital camera, or other device to create a digital representation of an object.
- **Image manipulation**  
making changes (i.e. tonal adjustments, cropping, moiré reductions, etc.) to an image using image processing software; altering the image from its original digital capture.
- **Interpolation**  
a method of creating new data points from a set of discrete known data points. For digital objects, it refers to the practice of creating new bits from known digital bits. Used to decrease the number of pixels taken at capture. Not recommended for master image capture or the scan once methodology.
- **JPG, JPEG**  
Joint Photographic Experts Group. An 8-24 bit image file format that is best suited for photographs. It supports "lossiness", which means that it will throw away some detail in order to achieve better compression. It has variable amount of compression to vary quality and file size. It does not work well for text. Widely used as a delivery format.
- **JPEG 2000**  
An image format that provides the inclusion of metadata and structural elements for the image within the code stream.
- **Metadata**  
structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource.
- **Noise**  
data or unidentified marks picked up in digital capture or data transfer that do not correspond to the original.
- **OCR**  
Optical Character Recognition, a process that produces a page of text from an image file.

- **PDF**  
Portable Document Format, 4-64 bit depth. Uncompressed. Used mainly to image documents for delivery. Need plug-in or Adobe application to view. Adobe's Portable Document Format, the term Adobe uses to describe Acrobat files.
- **Pixel**  
short for picture elements, which make up an image. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it.
- **Pixel per inch (ppi)**  
A measurement of the scanning resolution of an image or the quality of an output device. PPI expresses the number of used to capture the digital image. Can also be expressed as pixel per inch.
- **Preservation**  
(As it relates to scanning) Digitizing an original photograph, document, or three-dimensional object is only a method of preservation if the digital file becomes the access tool and the original is no longer available for use. Although high resolution scanning (i.e., scan at the highest resolution possible appropriate to the type of media you are scanning) is recommended for all materials in order to achieve the highest quality possible and to ensure that information held in the original is not lost in the scan. However, the digital file, as of yet, should not serve as a replacement of the original for preservation purposes.
- **Project Management**  
Digitization projects require substantial planning. A project manager is responsible for coordinating components of the project including goals, budget, and the project team. The project manager should be familiar with all aspects of a digitization project and will be able to either make decisions about technical and procedural questions that arise or will be able to identify the person responsible for making those decisions. Without a project manager, there is no mechanism for making decisions that fit into the overall goals of the project and the budget.
- **Quality Control**  
Techniques ensuring that high quality is maintained through various stages of a process. For example, quality control during image capture might include comparing the scanned image to the original and, then, adjusting colors or tonal values, or checking for artifacts. Quality control is vital for a project's success. Both expertise about technical digitization issues and knowledge about collection materials being digitized are important, and project should seek to build adequate quality checks into project workflow.

- **Resolution**

the number of pixels (in both height and width) making up an image. The more pixels, the higher the resolution; the higher the resolution, the greater its clarity and definition and the greater the file size. Can be expressed as a ratio (640 x 480 pixels) or in terms of dots per inch (dpi). It is recommended that you use between 72 and 100 dpi for images that will be displayed on the screen, and 300 dpi for images that will print on common inexpensive printers.
- **RGB**

Short for Red, Green, and Blue. This is an “additive” method used to simulate natural color on computer monitors and television sets.
- **Scanner**

A device for capturing a digital image. There are many types of scanners, such as flatbed scanners, drum scanners, slide scanners, and microfilm scanners.
- **Skew**

during printing or scanning, the degree to which the page is not vertical. De-skewing is a process where the computer detects and corrects the skew in an image file.
- **Storage**

Digital images take up a large amount of space on any computer or network. The most common way to store them is on offline media. This includes CD, DVD or other optical media. You will need to determine which option is best for your needs and purchase the necessary equipment (e.g. CD or DVD recorder). Remember to make multiple copies or “backups” of all your images.
- **Technical metadata**

metadata primarily intended to document the creation and characteristics of digital files.
- **TIF, TIFF**

Tagged Image File Format, an industry standard image file format. Uncompressed, originally developed for desktop publishing. 1 to 64 bit depth, used mostly for high quality imaging and archival storage. Generally non-compressed and high quality, including large file sizes. Most TIFF readers only read a maximum of 24-bit color. Delivery over web is hampered by file sizes, although LZW compression can reduce these file sizes by 33%, it should not be used for archival material. It is unique in that it incorporates multiple compression techniques, allowing the user to specify the best format for a type of image, and that one file can contain multiple images.

- **Thumbnail Image (see also Access Image)**  
Small, low-resolution images often hyper linked to a larger version of the same image. Users can often view multiple thumbnail images on a single screen.
- **Watermarks**  
Bits altered within an image to create a pattern, which indicates proof of ownership. Unauthorized use of a watermarked image can then be traced.
- **Work Environment**  
If at all possible, a separate workspace should be allocated for the scanning station. Since scanning is such a specialized task, often involving the use of collection materials, it is a good idea to reserve enough space to adequately meet the needs of scanning staff.

**Glossary terms taken from the following sources:**

Besser, Howard and Jennifer Trant, Introduction to Imaging: Issues in Constructing an Image Database. The Getty Art History Information Program Imaging Initiative, 1995.

[http://www.getty.edu/research/conducting\\_research/standards/introimages/](http://www.getty.edu/research/conducting_research/standards/introimages/)

Colorado Digitization Program. "Digitization Glossary."

[http://www.cdpheritage.org/resource/introduction/rsrc\\_glossary.html](http://www.cdpheritage.org/resource/introduction/rsrc_glossary.html)

Cornell University Library. Research Department. "Moving Theory Into Practice: Digital Imaging Tutorial," 2000-2003.

<http://www.library.cornell.edu/preservation/tutorial/contents.html>

Kodak. "Digital Learning Center Glossary of Terms."

<http://www.kodak.com/US/en/digital/dlc/book4/chapter2/index.shtml>

Library of Congress. Metadata Encoding and Transmission Standard: Official Web Site.

<http://www.loc.gov/standards/mets/>