

Image Capture

Before embarking on image capture, the decision must be made whether to scan directly from the originals or to use photochemical intermediaries, either already in existence or created especially for this purpose. Photographic media are of proven longevity: black-and-white negatives can last up to two hundred years and color negatives for more than fifty years when stored under proper archival conditions. They can thus supply a more reliable surrogate than digital proxies. Moreover, there is some concern that the greater contact with the capture device (if using a drum or flatbed scanner) and lighting levels required for digital photography might be more damaging to originals than traditional photography, though this is changing as digital imaging technology advances. However, creating photochemical intermediaries means that both they and their scanned surrogates must be managed and stored. Moreover, when a digital image is captured from a photographic reproduction, the quality of the resulting digital image is limited both by the reproduction itself and the capability of the chosen scanning device or digital camera. Direct capture from an original work offers image quality generally limited only by the capabilities of the capture device. (See *Selecting Scanners*.) Note that different film types contain different amounts of information. For example, large-scale works, such as tapestries, might not be adequately depicted in a 35mm surrogate image but may require a larger film format (an 8-by-10-inch transparency, for instance) to capture their full detail.

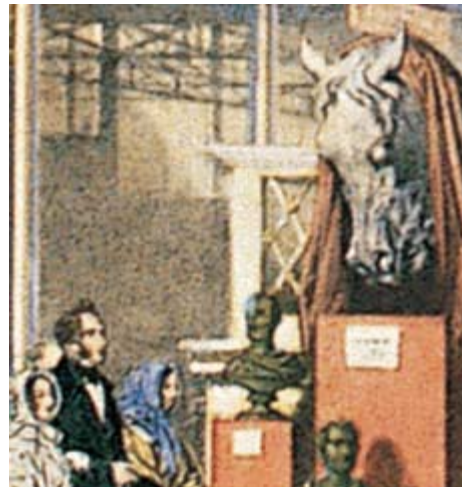
Digital image quality is dependent upon the source material scanned. The following images show the relative amount of detail found in a 4-by-5-inch transparency and a 35mm slide. Over three times as many pixels compose the same portion of the image when it is scanned from the larger format at the same resolution.



Fig. 12. Joseph Nash (English, 1809-1878), *The British Nave*. Color lithograph, 66 x 84.5 cm (26 x 33 ¼ in.), from: *The Great Industrial Exhibition of 1851* (London: Dickinson Brothers, 1851). Research Library, The Getty Research Institute, 960088



Detail of image scanned from a 4-by-5-inch transparency and enlarged by 353%



Same detail scanned from a 35mm slide and enlarged 1010%

Compare the people and various elements in each detail. A larger photographic intermediary produces a more nuanced image.

The quality of a digital image can never exceed that of the source material from which it is scanned. Perfect digital images of analog originals would capture accurately and fully the totality of visual information in the original, and the quality of digital images is measured by the degree to which they fulfill this goal. This is often expressed in terms of resolution, but other factors also affect the quality of an image file, which is the cumulative result of the scanning conditions (such as lighting or dust levels); the scanner type, quality, and settings; the source material scanned; the skill of the scanning operator; and the quality and settings of the final display device.

Digitizing to the highest possible level of quality practical within the given constraints and priorities is the best method of "future-proofing" images to the furthest extent possible against advances in imaging and delivery technology. Ideally, scanning parameters should be "use-neutral," meaning that master files are created of sufficiently high quality to be used for all potential future purposes. When the image is drawn from the archive to be used for a particular application, it is copied and then optimized for that use (by being compressed and cropped for Web presentation, for instance). Such an approach minimizes the number of times that source material is subjected to the laborious and possibly damaging scanning process, and should emerge in the long term as the most cost-effective and conservation-friendly methodology.

A key trade-off in defining an appropriate level of image quality is the balancing of file size and resulting infrastructural requirements with quality needs. File size is dictated by the size of the original, the capture resolution, the number of color channels (one for gray-scale or monochromatic images; three-red, green, and blue-for color images for electronic display; and four-cyan, magenta, yellow, and black-for offset printing reproduction), and the bit depth, or the number of bits used to represent each channel. The higher the quality of an image, the larger it will be, the more storage space it will occupy, and the more system resources it will require to manage: higher bandwidth networks will be necessary to move it around; more memory will be needed in each workstation to display it; and the scanning process will be longer and more costly. (However, remember that smaller, less demanding access files can be created from larger master files.)

Before scanning begins, standardized color reference points, such as color charts and gray scales, should be used to calibrate devices and to generate ICC color profiles that document the color space for each device in a digital media workflow. Color management is a complex field, usually requiring specialists to design and implement a digital media environment, and extensive training and discipline are required to maintain the consistent application of color quality controls. If such expertise is not affordable or available, color management systems that support ICC profiling are obtainable at a wide range of prices, as are color-calibration tools. Including a color chart, gray scale, and ruler in the first-generation image capture from the original, whether this is photochemical or digital, provides further objective references on both color and scale (fig. 13). Do not add such targets when scanning intermediaries, even when this is possible (a slide scanner, for instance, could not accommodate them), because doing so would provide objective references to the intermediary itself, rather than the original object.



Fig 13. KODAK Q-13 Color Separation Guide

To allow straightforward identification of digital images and to retain an association with their analog originals, all files should be assigned a unique, persistent identifier, perhaps a file name based upon the identifier of the original, such as its accession or bar code number. A naming protocol that will facilitate the management of variant forms of an image (masters, access files, thumbnails, and so forth) and that does not limit cross-platform operability (for instance, by using "illegal" or system characters) of image files must be decided upon, documented, and enforced.

Master Files

Archival master images are created at the point of capture and should be captured at the highest resolution and greatest sample depth possible (ideally 36-bit color or higher). These will form the raw files from which all subsequent files will be derived. After the digitization process, there is generally a correction phase where image data is adjusted to match the source media as closely as possible. This may involve various techniques, including color correction—the process of matching digital color values with the actual appearance of the original—and other forms of digital image preparation, such as cropping, dropping-out of background noise, and adjustment of brightness, contrast, highlights, or shadow, etc. The most common correction-phase error is to make colors on an uncalibrated monitor screen match the colors of the original. Great care must be taken to use standard technical measurements (such as "white points") during the correction process, which will create the **submaster**, or **derivative master**, from which smaller and more easily delivered access files are generated.

It will be up to the individual institution to decide whether to preserve and manage both archival and derivative masters, or only one or the other, for the long term. Constant advances in fields such as color restoration are being made, and if the original raw file is on hand, it may be returned to if it turns out that mistakes were made in creating the derivative master. However, it may be expensive and logistically difficult to preserve two master files. The final decision must be based on the existing digital asset management policy, budget, and storage limitations. Whatever the decision, editing of master files should be minimal.

Ideally, master files should not be compressed; as of this writing, most master images are formatted as uncompressed TIFF files, though an official, ratified standard that will replace TIFF is likely to appear at some point in the near future. If some form of compression is required, lossless compression is preferred. The files should be given appropriate file names so that desired images can be easily located.

Image metadata should be immediately documented in whatever management software or database is utilized. The process of capturing metadata can be laborious, but programs are available that automatically capture technical information from file headers, and many data elements, such as scanning device, settings, etc., will be the same for many files and can be added by default or in bulk. The masters should then be processed into the chosen preservation strategy, and access to them should be controlled in order to ensure their **authenticity** and integrity. (See *Long-Term Management and Preservation*.) It is possible to embed metadata, beyond the technical information automatically contained in file headers, in image files themselves as well as within a database or management system. Such redundant storage of metadata can serve as a safeguard against a digital image becoming unidentifiable. However, not all applications support such embedding, and it is also conceivable that embedded metadata could complicate a long-term preservation strategy.

Access Files

Generally, master images are created at a higher quality than is possible (because of bandwidth or format limitations) or desirable (for reasons of security, data integrity, or rights protection) to deliver to end users, and access images are derived from master files through compression. All access files should be associated with appropriate metadata and incorporated into the chosen preservation strategy, just as master files are. In fact,

much of the metadata will be "inherited" from the master file. Almost all image collections are now delivered via the Web, and the most common access formats as of this writing are JPEG and GIF. Most Web browsers support these formats, so users are not required to download additional viewing or decompression software. Each institution will need to determine what quality of access image is acceptable for its various classes of users and measure this decision against the cost in resources for image creation, delivery, and storage.

Web-based distribution using browser-supported file formats is the most common and broadly accessible way of distributing images, but it does impose certain limitations, especially if there is a desire to offer higher-quality, and therefore larger, images. For instance, delivering images to users accessing the Internet through a 56 Kbps modem will necessitate small (highly compressed) images to prevent those users' systems becoming overburdened. The general adoption and support of more efficient compression formats (see *File Formats*) and a wider adoption of broadband technology may go a long way toward remedying this situation. In the meantime, another option is to use a proprietary form of compression that requires special decompression software at the user's workstation. An example of such a compression system is MrSID, which, like JPEG2000, uses wavelet compression and can be used to display high-quality images over the Internet. However, the usual caution that applies to proprietary technology should be applied here: legal and longevity issues may emerge.

Another strategy is to provide smaller compressed images over the Web or some other Internet-based file exchange method but then require users to go to a specific site, such as the physical home of the host institution, to view higher-quality images either over an intranet or on optical media such as a CD- or DVD-ROM. This option may be a useful stopgap solution, providing at least limited access to higher-quality images. There may be other reasons for offering images and metadata on optical or other media besides (or instead of) over the Web, such as restricted Internet access in certain countries.