LEVELS OF SERVICE FOR IMAGE DIGITIZATION

William LeFurgy's 2002 article, "Levels of Service for Digital Repositories," presented a broad and useful classification scheme to characterize emerging services for digital preservation.

LeFurgy's minimal, enhanced, and optimal classes of service describe "inevitable differences in serviceability" according to attributes of the source materials and the technological infrastructure of the repository.

This same concept applies to digitization. Spectrums exist for the complexity, quality, and effort required for digitizing historical works. With images, the distinctions among technique, quality, and format choices in traditional copy photography still apply in the digital world.

Functional requirements for reproductions dictate the minimum thresholds for staffing (skills), equipment, and quality control. Librarians cannot reasonably expect to receive 8x10-inch negative quality at 35mm negative prices.

This chapter describes the program components necessary to produce discoverable, sustainable, and usable collections of surrogate digital still images for legacy collections of photographs, prints, and other pictorial works.

Assuming that a library has already made appropriate program investments in digital library technology (Chapter 1) and digitization program management (Chapter 2), the baseline level of service for image digitization must be configured to yield a digital image *collection*. In this context, *service* refers to the standards, systems and skill (staff expertise) brought together to serve production goals for throughput and quality.

In other words, omit any single baseline component and the production operation will not have the ability to create image and metadata products that meet the following criteria:

- The surrogate is appropriately cataloged and discoverable, and the descriptive metadata are stored in a well-supported system.
- The surrogate digital file can be opened and rendered as an image.
- The surrogate is appropriately named to be identified by some type of inventory control mechanism (ranging for a printed list to a complex database).
- All versions of surrogate files are appropriately stored, identified, and documented (with administrative metadata) for ongoing management.
- The surrogate can be reliably delivered by the library's (or partner organization's) designated applications for networked access.

Fulfillment of these minimum criteria—whether measured against local or community definitions of what is appropriate or good (see, for example, NISO's *Framework of Guidance for Building Good Digital Collections*)—are presumed to offer the potential for sustainability.

NISO's *Framework of Guidance for Building Good Digital Collections,* www.niso.org/framework/frameworkforum.html

Levels of service above-the-baseline are required for any project that explicitly states requirements for pictorial quality, for search and discovery with a controlled vocabulary, and for persistence. As every camera user knows, taking pictures is different from taking consistently good pictures.

Regardless of what is behind the lens—film, a CMOS chip, or a CCD array—all products of light-lens photography start from the aggregate quality of the composed image before its capture: its illumination, focus, exposure, framing, and point of view. In short, the human component to photography remains strong.

Baseline production services

Provided that systems are in place to store cataloging data and images, and to make the catalog(s) and images Internet accessible, key image digitization tasks (and their attendant standards, specifications, and best practices) requiring infrastructure are:

- Production of descriptive metadata (cataloging)
- Production of images (scanning)

Cataloging and digital imaging infrastructure neither needs to be deep nor expensive to produce Internet-accessible digital image collections. Many libraries and historical societies have produced digital images collection with one or two scanners linked to standard office computers, image processing software, a database software for cataloging and linking, and a Web front end for indexing, searching, and image delivery.

When methodically planning a program of any size, the manager will encounter tradeoffs in costs, quality, and sustainability in nine production arenas of image digitization, as presented below. The challenge is to configure systems, staffing, and procedures that reflect and support program priorities over time.

Selection strategy

From the perspectives of the scanning technician and the data manager, homogeneity yields affordability.

Source material of the same format, size, complexity, and condition can be handled in the same way and scanned to one specification. Setup time—per collection, per batch, and per image—is a meaningful cost component in digitization. Thus, selection strategies that can minimize differences among physical attributes of source materials without undermining project goals for usability promote highest-production, lowest-cost digitization workflows.

By the same token, output digital image files of the same format, compression type, color space, number and type of related images, and type and extent of administrative metadata present only one profile for the data manager to process for production of deliverables or to monitor for obsolescence.

Unfortunately (from the perspective of digitization), cultural heritage collections tend to be heterogeneous. Even collections of one source format, such as mounted 35mm color slides, tend to span several years and represent either many photographers or many projects from a single photographer. Differences in film stock, original quality of photography and film processing, and effects of aging are also common.

Ease of digitization should not be ranked ahead of intrinsic values (artistic, documentary, evidential, intellectual, etc.) when selecting images to digitize. Stakeholders in any project should be aware, however, of the tradeoffs between homogeneity, which promotes high production and relatively low cost of digitization and sustainability, and heterogeneity, which increases costs by increasing numbers of production batches and preservation profiles.

Rendering intent

To scope technology requirements, staffing, digitization methods, and costs, service providers need to know the librarian's intent for the surrogate images.

This intent is not necessarily easily addressed by inspecting the item in hand—particularly photographic negatives and aged (toned) black-and-white prints. Do not assume that delegating decisions of rendering intent to the photographer or scanning operator (or even riskier, to the scanning software) will achieve the quality expected by key project stakeholders.

Franziska Frey and James Reilly have provided the best explanation of pictorial rendering intent in *Digital Imaging for Photographic Collections: Foundations for Technical Standards* (1999). They present four options:

1. To match the appearance of the image
2. To represent the photographer's intent (by correcting exposure or film processing errors) to achieve a pleasing reproduction
3. To reconstruct the original appearance of the image (by removing artifacts from aging, such as fading and color shifts)
4. To represent the original scene (by removing inherent biases to many films)

Frey and Reilly present the caveat that if original images are properly exposed and processed, option #1 could potentially be executed with a largely automated workflow. The other three pictorial rendering intents require manual intervention, informed by subjective judgments of the photographer or technician processing digital images.

Collection review is one component of the digitization workflow that receives comparatively little attention compared with discussions of cataloging and scanning, yet detailed knowledge of the quality and condition of the sources is essential to deciding which intent(s) should be met through digitization.

If curators and collection managers lack the time to review sources to specify rules of capture for each batch, the default intent should be to match the source. In theory, this baseline approach lends itself to automation (low cost) and implicitly conveys to the user that the surrogate purports to be a good copy rather than a faithful restoration.

Specifications development

Developing specifications takes time, requiring the application of a methodology that provides realistic forecasts of timelines and costs for each project. This important set of baseline skills should be cultivated within the digitization operation, whether the library manages digitization or outsources production.

Experienced practitioners increasingly advocate starting at the end to develop good digitization specifications for each project.

By working backward from the delivery application known to satisfy user expectations, the catalogs and cataloging tools in place for discovery, and the storage systems designated to manage digital masters and surrogates, the digitization project team can derive *contextual* best practice specifications for metadata and digital image formats.

As discussed in Chapter 2, "Manage Tradeoffs," there are many standards to choose from. Beginning from the user's and the data manager's perspectives, one is more likely to sketch good project workflows than by inferring best specifications for digitization from format classifications or even inspections of source materials—particularly if these decisions are made without considering the downstream technologies for storage and delivery.

Good specifications result from curators (collections owners), the digitization project manager, and the service providers (catalogers, digitizers, and Information Technology data managers) working together to consider options, *then* balancing cost and quality considerations against the attributes of the source materials and the pictorial rendering intent(s) prescribed for the surrogates.

Overhead to outsourcing

Developing and documenting specifications requires more work when digitization is outsourced. Typically, digital imaging and technical metadata production are the two production activities that might be outsourced in image digitization. (Management, selection, materials preparation, conservation, and cataloging activities almost always take place in house.)

Although outsourcing might reduce digital imaging costs, and potentially increase image quality, *the costs of vendor identification and specifications development will be higher* for each project in outsourcing than managing all digitization in house.

When outsourcing digitization, allocate adequate time to find candidate vendors, then write requests for information (RFIs) or requests for proposals (RFPs). Institutions must not only articulate in detail the policies and procedures for materials handling and digitization in these documents, they must also—under advisement of senior administrators or legal counsel (when available)—ensure that institutions' rights and liabilities are adequately stated and protected.

When producing RFPs and Agreements, organizations should declare fully the rights of ownership of content produced in digitization, to state the business relationship (work for hire and termination clauses) between library and service provider during the project, to specify project-wide criteria and responsibilities for errors and remedies, invoicing and payment, and, for each batch, to present detailed specifications for materials handling, disposition, and delivery of surrogates and originals back to the library.

Digitization workflows involving curatorial departments and digitization providers *within* the library, on the other hand, tend to require much less oversight and documentation.

Regardless of whether images will be digitized by internal or external service bureaus, curators and the digitization project manager need to decide whether or not to invest time in preparing and reviewing samples before signing off on specifications and embarking on production.

As part of the RFP process, many service providers are willing to receive sample source items—which should be *representative* of the collection as a whole—to produce sample digital images (at no charge to the library) as part of their proposal.

These samples can be extremely useful in revealing what technology is capable of (for a given price), thereby helping to manage quality expectations among the principal stakeholders in a project before it begins. Furthermore, these sample products facilitate selecting the service provider most likely to offer the best balance of quality, cost, and service.

Metadata production

As noted above, cataloging, with digital imaging, is conventionally viewed as one of the two major cost centers of image digitization. Although pictorial images can be digitized without cataloging, a digital image collection cannot be created and delivered without metadata.

With respect to sustainability, descriptive metadata is especially valuable because it fosters use of the digital collection. Among preservation experts, there are two schools of thought regarding longevity:

- The first one correlates use with longevity, speculating that, over time, digital items and collections that are not used will be more susceptible to obsolescence than those that are.
- The second one correlates digital image quality with longevity, hypothesizing that files with greater inherent pictorial and functional quality will receive stronger mandates and support for preservation.

Each of these theories underscores the vulnerability of digital surrogates to loss and the importance of engaging users (or other stakeholders) to champion sustaining some collections over others.

Descriptive metadata are essential to developing user communities. One of the main cost-benefit questions for every project is, How much metadata should be created for this image collection?

Although standards and best practices for descriptive metadata are beyond the scope of this report, note that an image digitization program requires expert capabilities to review options and to develop fit-for-purpose metadata specifications for each project with regard to:

- Granularity (item-level cataloging, group-level cataloging, both?)
- Vocabularies (controlled, standards-supported?)
- Extent
- Format
- Distribution of data (local system, union catalog, or both)

Beyond production of descriptive metadata to facilitate use (and indirectly contribute to sustainability), stakeholders in each image digitization project must also consider the value of including (or costs of excluding) workflows to produce administrative, preservation, and technical metadata for images.

The general rule is to obtain as much free technical metadata as possible from scanning and image processing software. These devices can populate image headers or associated XML files with rich metadata documenting the systems and settings used for image production.

They cannot, however, automatically record imaging rationales (see "Rendering Intent"), descriptions of source materials, ownership, rights of access, or other information fundamental to management and sustainability.

Making images requires investments in materials preparation and digital imaging; making image collections dictates investing in metadata. The extent of such investments depends, for each project, on the following:

- Local decisions regarding minimum requirements for discovery and access (based largely on needs of current and future users)
- Documentation of provenance, ownership, and rights of access for management
- The structural and technical metadata needed to fulfill present or anticipated future mandates for data packaging and long-term storage (see below).

Digital imaging

Scanners and digital cameras are often the first devices that managers think about purchasing when ramping up digital image production capabilities. In planning the digital imaging component of a production operation, however, librarians are prudent to consider whether the library should develop systems or purchase services.

Cultural heritage materials can be digitized safely and well by service providers who work within the library or at their own production studio. Thus, librarians might begin planning with the proposition that outsourcing is preferred to developing services. In other words, embrace hiring expert imaging vendors unless vetted, supported reasons exist not to.

If valuable or fragile materials will be digitized, collections do not have to leave the institution to be scanned. The National Archives and Records Administration's *Electronic Access Project* exemplifies a project workflow in which vendors are contracted to bring their staff and equipment to the collections, and all digitization is done on site (within the library).

Cornell University Library's *Louis Agassiz Fuertes Collection* typifies the more common bring-the-collections-to-the-vendor workflow. In the case of *Fuertes*, film surrogates rather than original drawings were shipped in batches to the vendor's facility.

Outsourcing imaging rather than developing in-house digitization services, of course, is not always the best solution. Under one or more of the following criteria, program managers might be motivated to invest in the space, staff, training, and equipment for their own digitization service:

- The library already administers or has access to a photography studio that lives within the broader organization (such as a university), and has the means to expand or reconfigure the studio to provide digital as well as (or instead of) traditional photography services.
- The library can save costs because their staff (or volunteers) receive comparable or lower wages than the vendors' employees.
- The library selects materials that must be digitized in-house and decides to invest in developing expertise that can be leveraged for other services. By growing local expertise in digital image production—even if the approach results in higher per-item costs for digital products—the library positions itself to digitize collections and to obtain cost-recovery revenues through

fee-for-service over-the-counter digitization for library patrons. Some libraries have even extended these services beyond their own organization to compete with other vendors to digitize cultural heritage collections.

- The library receives a gift of equipment, or funds to purchase digitization hardware and software.
- The library values a staffing model in which all members of the digitization team—manager, conservator, cataloger, digitizer (photographer or scanning technician), system and data administrators, public service librarians—share the core values of the library that has provided preservation and access services for the source collection.

These and other considerations might justify investing in equipment rather than services. Recipients of Library of Congress/Ameritech grants in the mid-1990s, such as the Nebraska State Historical Society, have been able to sustain in-house digitization operations seeded by an initial project grant.

As discussed in the "Above Baseline Services for Consistent Quality" section (and Chapter 5), no standard template exists to configure an image digitization workstation or studio.

A *production studio* capable of producing digital images (to any quality level) would have the following:

- Secure space
- Skilled staff
- Scanner(s)
- A service agreement for all scanners
- Adequate server space for production and file management, scanning software, image editing/quality control software
- Technical metadata production software
- A quality control program
- The means to transfer all data to the appropriate storage location or media following production.

Quality control

Quality is relative, not absolute. Judgments of quality for metadata and images reflect an institution's value in standards, consistency, accuracy, and accountability.

The primary question to address in planning baseline digitization is, What quality metrics will be used?

Means to define and assess quality for metadata and images are both objective and subjective. Software can be used to validate data structure and integrity—have all required fields been populated in a database? with the correct values? encoded in the correct way?—as well as technical conformance to image format standards.

Semantic checks (of completeness and accuracy) of data and checks of any pictorial components of rendered images are subjective, however, requiring review by trained specialists.

Presuming that no library would establish an image digitization program without some measures and checks of quality, plans to configure baseline operations for digitizing must allocate sufficient resources for quality control.

In practice, this necessitates training for staff, creating or purchasing software for validation (in some cases, these programs are free), and allocating sufficient monies in project budgets for time spent on quality control in production.

Higher-level strategies for quality control require purchases of calibration hardware and software for image review, and targets and related instrumentation and software to measure system performance.

Data tracking, assembly, and packaging

The need for systems to track and monitor data, and to assemble and package data into batches relates partly to scale and partly to decisions about where master and delivery images will be stored.

Source materials need to be moved through digitization workflows: preparation, conservation, cataloging, and digitization. Several factors influence requirements for tracking the status of items during production: the location of digitization studio, the value of the originals, policies for insurance (when applicable), accountability to users.

Methods of tracking can range from annotating printed batch inventory control sheets that accompany material through each stage of production, to inclusion of status fields in custom-built project databases.

The key production challenge? Digitization yields multiple files. In some cases, what is scanned is saved as the digital master. In others, raw camera data are saved temporarily and passed on to a secondary processing workflow that yields the masters.

Master files are routinely processed to generate one or more derivative or delivery files: thumbnails to accompany catalog records, and one or more larger images for viewing or printing.

Some source materials require multiple surrogates. File sizes for masters can range from single to hundreds of megabytes per file in image digitization workflows.

Libraries that create digital collections must adopt (or develop) methods to disambiguate versions of files, then to relate these versions in some meaningful way at the end of a workflow. The Metadata Encoding and Transmission Standard (METS) provides one schema that an organization can adopt to define these relationships (in XML) according to a community standard. Alternatively, a library can define its own meaningful naming scheme for files and file directories.

Metadata Encoding and Transmission Standard (METS), www.loc.gov/standards/mets

In either case, post-scanning software will need to be configured to assign and check file names and to provide any other metadata needed to manage related files (that comprise a single logical object) stored in single or multiple locations following digitization. Discipline and vigilance are critical in tracking files and assembling objects, particularly when an operation is managing multiple projects.

Repository services such as the OCLC Digital Archive specify how data must be packaged to be accepted for deposit. In some cases, as with OCLC's service, the repository provides tools to create compliant deposit packages for each batch that is submitted; production is the responsibility of the depositor.

Short- and long-term storage

Following production, many libraries have written and presumably will continue to write data files to offline disks (such as CD-ROMs) rather than to RAID systems (online media) with robust backup systems, and administration by an information technology group.

With appropriate controls, this offline strategy can manage risk adequately for brief periods (one to five years). *This approach, however, is not viable*, where librarians intend to keep files and media playable and error free for decades.

Saving costs, of course, is the main reason to adopt this approach to storing high-value assets, such as digital image masters.

The National Archives and Records Administration's Special Media Preservation Laboratory's "Storage Recommendations" in their *Technical Guidelines* (June 2004) provide sensible caveats to producing disks for short-term storage:

- Use high-quality or archival quality CD-Rs (such as Mitsui Gold Archive CD-Rs)
- Purchase new, brand-name media
- Produce two or more copies
- Assign and use a CRC (cyclic redundancy checksum) or MD5 (message digest) checksum to measure and monitor data integrity
- Handle and store media appropriately

Even if media need to be usable only for short periods (one to five years) librarians are prudent to store disks in a cool-storage repository if one is available or in a climate-controlled media storage environment. Store separate copies in separate locations.

In sum, sustainability depends on managed storage of all library assets—metadata, master images, delivery images. As is the case with digital imaging, a high-level question to address in configuring baseline capabilities for image digitization is: should the library develop systems or purchase services?

In most cases, libraries will have purchased the database and storage systems to produce and manage all metadata for digital image collections. As discussed in Chapter 1, decisions about where to store master and delivery versions of digital collections are informed by policies regarding ownership, budgets, and local capabilities to define and fulfill preservation obligations.

Delivery

Format requirements for delivery systems profoundly influence choices of image formats, image sizes, numbers of images per original work, and metadata formats in digitization.

Although budgets to create and sustain these systems will likely be administered within the library's digital library rather than digitization operation, delivery systems indirectly drive the budgets for metadata and digital imaging services.

Delivery systems are the most visible component of baseline image digitization infrastructure. When scoping image digitization, a library must evaluate carefully the features of given delivery systems vis-à-vis the genres of collections and needs of users.

Can one system be used for searching and display of map images as well as images of art and architecture? Would scientific images need separate systems? Can any of these functions be purchased as services rather than additional local systems to configure and maintain?

Storage and delivery systems are the major infrastructure components which distinguish baseline *programs* from opportunistic (yet often effective), project-based infrastructure to sustain digital image collections.

Above-baseline services for consistent quality

The baseline services described above for image digitization are quality agnostic. With these nine components in place for production digitization, a library could produce a digital image collection (or several), but not necessarily to any stated level of quality.

In *Digital Imaging for Photographic Collections* (1999), Franziska Frey and James Reilly surmised, "[I]nstitutions often have unrealistic expectations about digital projects." Grant applications, project reports, presentations, and other documents reveal a continuation of this trend in recent years: infrastructure is rarely synchronized with stated production and quality objectives.

Too often, libraries seize on a given set of specifications—such as sampling rates, file format, and color space choices for still images—as a guarantor of good quality, without accounting for the systems and staff that create images.

In the current definitive handbook for producing raster images from archival materials, the National Archives and Records Administration's experts in the Special Media Preservation Laboratory caution, "[W]ithout a good technical foundation and experience for production staff, there can be no claim about achieving the appropriate level of quality as defined in these guidelines." (See, Puglia, Reed, and Rhodes, *NARA Technical Guidelines for Digitizing Archival Materials for Electronic Access,* June 2004.)

Manufacturers and resellers of equipment for the graphic arts community heavily promote the perception that their systems will be simple to use and will reliably produce quality images. Frey and Reilly countered this claim in 1999 by asserting, "…successfully digitizing a photograph collection requires as much experience as conventional reformatting."

As levels of quality increase, so must levels of service—where service is comprised of systems, skill (staff and training), and application of standards.

The matrix on the next page presents some of the meaningful attributes of source, quality, and quality control that force a program to increase its investments in digitization operations above the baseline.

Low-effort strategies

The easiest way to hit the ground running with a new image digitization program is to limit selection to sources that fit a single scanning device, and require the digital reproductions to be of legible quality.

The manufacture and distribution of scanners is, of course, market driven. In the 1990s, libraries benefited from the number of specialized systems—drum scanners, slide scanners, flatbed scanners (with adapters to accommodate transparencies and plates), and high-end digital cameras—marketed to consumers and to graphics arts professionals.

Matrix of Attributes* for Above-baseline Layers of Service for Image Digitization

	Low effort	Medium effort	High effort
Source type(s)	Reflective media <14" or 35mm slides	Negative film, plates, reflective media >14"	Media in poor condition or of poor quality; out-of-gamut colors in source; three-dimensional objects
Digital quality	Legibility	Fidelity (dimensions and details)	Fidelity + accurate tonal or color reproduction
Quality control	Check of pictorial attributes only	Check of pictorial and digital attributes	Check of *matching* pictorial attributes; format validation

* excluding descriptive metadata

As early as 2001, however, experts were forecasting consolidation of products and the discontinuation of certain classes of devices. Peter Wolff of photograpical.net wrote in November 2001, "We are perhaps seeing the very last film scanners in these times. In a few years…only people with large archives of negatives and transparencies…will use scanners." (Librarians might later include the postscript, "…if they can obtain them.")

Industry trends, over which libraries and cultural heritage institutions exercise almost no influence, offer good-news and bad-news. On the one hand, high-end flatbed scanners (such as the Heidelberg Topaz line) that achieved high quality and high production have disappeared.

On the other hand, current scanners in the $600 range (such as the Epson Perfection 4870 Pro) are reportedly achieving comparable levels of quality to the previous generation, high-end flatbed scanners in the $10,000 to $50,000 range.

Libraries seeking to minimize both effort and cost in their image digitization operations should scope their workflows within *all* the following parameters:

- **Scanning:** Have library staff digitize all materials in house (do not outsource), using moderately priced flatbed or 35mm slide scanners. Avoid use of digital cameras for copy photography—thereby eliminating needs to purchase and configure separate components, particularly lighting, to achieve levels of quality comparable to graphics arts flatbed scanners.
- **Selection:** Restrict digitization to source images of dimensions and formats that fit the scanner(s) in use.
- **Digital image quality:** *Make no claims for accuracy* in the digitized images. Consistently meet the goal of creating and distributing legible surrogates.
- **Quality control:** Limit quality metrics and quality control to subjective methods. Require technicians to verify that each image corresponds to its source; that it is complete; that it meets project tolerances for borders and skew; that it conforms to specifications for rendering intent; that, overall, its tones, colors, and details are legible (without necessarily being faithful).
- **Technical metadata:** Assign checksums to each file. Produce and store metadata that adequately documents ownership, rights, and access for each image.

Medium-effort strategies

Levels of service for image digitization fall into the baseline + medium-effort category when *any* of the following needs must be accommodated (even if only in one project):

- **Scanning:** Some or all digitizing is outsourced. Alternatively, one or more digital cameras are purchased for the library's digitization operation.
- **Selection:** Images are selected for intellectual (content) value, regardless of their formats, age, dimensions, quality, and condition. Digitizing must fit the source.
- **Digital image quality:** The library *makes any claim of fidelity to the source*. The quality objective is raised from simply creating copies to creating copies that meet any pictorial criteria for goodness.
- **Quality control:** In addition to the subjective methods described in the low-effort service configuration, technicians would be required to have sufficient visual literacy to compare copies with sources. (Objective methods of quality control, such as use of technical targets, would only be introduced if staff were appropriately trained.)
- **Technical metadata:** In addition to checksums and administrative metadata, some preservation or technical metadata would be mandated in the workflow—whether stored internal or external to the image files—and verified to be accurate and complete.

Digital cameras are applealing for copy photography, but they require considerably more effort to configure, use, and maintain than flatbed scanners. Large two-dimensional artwork can be digitized, as can film and other transmissive formats when custom light tables are used. Direct digital photography of any three-dimensional material, of course, requires a camera.

Unfortunately, unlike scanners, digital cameras are not self-contained systems. Studio cameras must be configured from many component parts, including: camera back (CCD array) and software, camera body, lens, copy stand, cradles or other systems to position media on the copy stand, light stands, and lighting with appropriate filters to shield ultraviolet light.

Digital cameras not only require greater effort to configure, use, and maintain, but they are also considerably more expensive (at this writing) than flatbed and slide scanner counterparts. The manager tasked to write an equipment budget should consider the following:

- Consumer cameras are inappropriate for copy photography and will not yield the quality needed to reproduce a wide range of cultural heritage materials.
- Professional cameras generally fall into two classes: relatively affordable ones marketed to photojournalists and prosumers; expensive ones marketed to studio photographers in advertising, fashion, and other studio-based industries.
- The apparatus needed to run the camera—computer, lighting, copy stand, camera body, lens—can easily double the price of a camera back, although these items, like cameras, range from affordable to expensive. Compare, for example, the features and prices of copy stands from two companies:

  —Tarsia Technical Industries Inc. (TTI) copy stands

  —Bogen Imaging camera stands

Tarsia Technical Industries Inc. (TTI), www.ttind.com

Bogen Imaging, www.bogenimaging.us

- Service agreements are strongly recommended to support production digital cameras. Whenever a program makes meaningful investments in systems and the people who run them, it needs to consider carefully how much down time it can afford when equipment fails or needs periodic repair. If production must be ensured, then service agreements are essential.

Any digital camera being used for copy photography requires photographic skill, if not actual professional credentials, simply to achieve good focus and exposure.

Image quality in digital cameras is also sensitive to ambient lighting conditions. Even a medium-effort level of service should include some means to control, if not necessarily optimize, the environment in which the camera is placed.

Given the amount of overhead associated with selecting a vendor and preparing specifications and agreements, outsourcing digitization also increases the effort to establish and maintain services.

Finally, quality requirements that mandate any element of fidelity to the source require greater skill from digitization staff, if not greater quality in hardware and software. As noted in the baseline requirements for rendering intent, requirements to represent photographer's intent or to reconstruct appearances of original images or original scenes demand greater skill from technicians and greater clarity in specifications.

Steven Puglia and other digitization experts regularly observe that not all software is alike, and that, in general, higher-priced scanning and image processing programs outperform lower-priced products. LinoColor (now discontinued), for example, purportedly did a better job for comparable imaging tasks than Adobe Photoshop—at a much greater cost. Photoshop, in turn, generally does a better job than less expensive programs.

At this level of service, an organization sould achieve modest levels of image quality and have the systems to accommodate a broader range of source formats in production digitization.

High-effort strategies

Two desires often stated in image digitization, but rarely supported by the configuration of digitization services, are:

- To match surrogates to sources (including color matching)
- To produce device-independent masters (masters not yet finalized for rendering on a specific type of device, such as a CRT monitor)

CRT: Cathode ray tube

Both of these goals are laudable, but managers establishing program foundations for such image digitization must be realistic about the level of effort required to achieve them.

Levels of service for image digitization fall into the baseline + high-effort category when *any* of the following needs must be accommodated (even if only in one project):

- **Scanning:** Same requirements as medium-effort, but with emphasis on hiring staff with appropriate credentials and skills.
- **Selection:** Same requirements as medium-effort, plus institutionalized policy to incorporate conservation review into the materials preparation workflow.

- **Digital image quality:** The library *makes any claim of fidelity to the source*. The photographers and technicians make any attempt to manage color or pictorial quality in masters is mandated to be device-independent.
- **Quality control:** Same requirements as medium-effort, with particular emphasis on using objective metrics and methods; insistence on subjective review being undertaken on calibrated devices in a calibrated viewing environment.
- **Technical metadata:** Mandatory production of technical metadata for preservation.

Investments in high-effort strategies yield consistently accurate reproductions. Accurate copy photography, first and foremost, requires staff with a good technical foundation in imaging.

Questions to consider in hiring staff (or reviewing qualifications of vendor's staff) are: Have they made images? Do they have the vocabulary to describe and evaluate quality? Do they have formal training? Have they been tested for color blindness? (Men are statistically more prone to color blindness than women.)

Matching surrogates to sources

Any evaluation of rendered image quality—that is, inspection of digital images on screen or in print—requires normalization of the viewing environment and device calibration. Without these, no one can reliably claim to re-create or match at any later point in the image life cycle the rendering that was seen when quality was finalized.

Normalization of the viewing environment

This task requires modest effort, but not necessarily great expense. One standard provides guidance in this arena:

- ISO 3664: 2000 *Viewing conditions – Graphic technology and photography Controlled Illumination of Object*

In brief, ISO 3664 specifies eliminating outside light—digital photography and digital image quality control should be done in rooms (or booths) without windows—and painting the walls black or neutral gray. Munsell notation N8, "Standard Gray Neutral 8 Vinyl Latex" paint, may be purchased from Graphic Technology, Inc.. (Suppliers of conventional brands of paint also may be able to mix Munsell N8.)

Graphic Technology, Inc., www.gtilite.com

The standards further specify keeping room lights at a low level—as everyone who watches movies does to appreciate the full quality of the images—and to eliminate reflections and glare on the computer monitors being used to inspect images.

Device calibration

Calibration, the basis to reliable inspection of displayed images, requires specialized hardware, software, and skill. Again, an ISO standard provides guidance:

- ISO 12646 *Graphic Technology – Displays for Colour Proofing – Characteristics and Viewing Conditions*

  This standard specifies the gamma, color temperature, and luminance levels for monitor calibration and their surrounds for direct comparison of images to originals (soft proofing).

Three devices are required to support soft proofing: a quality graphics arts monitor, a monitor calibrator and associated software, and a viewing booth to illuminate the source material (or a viewing table for transparencies).

A library needs these three tools for scanning and for quality control. Thus, outsourcing does not excuse an operations manager from needing to purchase and maintain these devices.

Calibration tools being used today by high-end museum and library photography studios include:

Professional graphics displays (monitors)

- Barco Reference Calibrator (CRT), www.barco.com
- Eizo Nanao ("Eizo") series (LCD), www.eizo.com
- Sony Artisan series (GDM-C520K) (CRT), www.sony.com
- NEC-Mitsubishi Diamond Plus and Diamond Pro series (CRT), www.necmitsubishi.com

Display calibrators (hardware)

- ColorVision Spyder, www.colorvision.com
- MonacoOPTIX^{XR} Color Calibrator, www.monacosystems.com

Viewing booths

- GTI Graphic Technology, Inc., Color Viewing Stations, www.gtilite.com
- GretagMacbeth Viewing Station ("Judge II"), www.gretagmacbeth.com

Color management

Color matching is the most challenging task in image digitization. Projects that demand (or even seek) this level of quality require skilled staff and complex systems. Despite the promises of color management software, decisions about color matching are still largely made by skilled observers, not algorithms in easy-to-use software.

Several system components, however, are essential to color management. Selected examples of tools being used by museum, library and archives studios include:

Digital Camera and Display Monitor Profiling Software

- ColorVision OptiCAL, www.colorvision.com
- GretagMacbeth ProfileMaker Pro, www.gretagmacbeth.com
- Monaco Profiler, www.monacosystems.com
- Pictographics InCamera Professional, www.picto.com

Spectrophotometer

- GretagMacbeth Spectrolino and Spectroscan T, www.gretagmacbeth.com
- Monaco X-Rite spectrophotometers, www.monacosystems.com

The *NARA Technical Guidelines* provide expert commentary on the uses and limitations of these instruments, as well as the standards and accompanying technical targets used for assessing the performance of imaging systems.

Like databases used for metadata production, the tools used for digital imaging cannot in and of themselves produce high quality products.

Accurate descriptions of images and faithful digital image reproductions are produced by dedicated, well-trained professionals. Tools at their disposal may improve production and help to ensure consistency, but, as yet, they are unable to automate the process of interpreting rendering intents from source images, then creating high-quality copies.

Producing device-independent digital masters

Many specifications and image digitization workflows in use by libraries for the past decade reveal two misconceptions about how master images should be optimized for sustainability.

The first is believing that unprocessed, uncompressed image files made at the camera are ideal for archiving. The second is assuming the scanning workflow should be designed to produce high-resolution images that look pleasing (for example, modestly sharpened) on computer monitors before being saved as masters.

One drawback of editing high-resolution images until they look pleasing on screen—even if this editing is done with a calibrated system (which is rare)—is the increase to production time.

The chief risk to finalizing quality for display is to introduce irreversible losses of quality in the master images. Good masters may not necessarily look pleasing when opened (displayed), but they should be optimized for automated production of a variety of image products—for display, printing, digital migration—according to present and future needs.

Studio managers at the University of California, Berkeley, the Library of Congress, NARA, Harvard University and other institutions that have made multiyear investments in digital imaging all subscribe to the principle of bringing master images to a "common rendition" in the workflows designed to produce sustainable images.

Significantly, these renditions are defined *objectively*—by aim points on targets, for example—not purely subjectively by reviewing and editing displayed images.

Each of these studios uses multiple imaging devices and regularly encounters source images of varying attributes and challenges. Specific choices of color spaces, "aim points" for tone reproduction, and sampling rates (resolution) in scanning vary among these practitioners, but the underlying philosophy is the same: normalizing data without finalizing it for output yields masters optimized for *batch* processing.

Image files brought to a common rendition *and identified by their administrative metadata as having common properties*, are highly amenable to production of future deliverables and automated migration paths in digital archiving programs. Some institutions explicitly label these files as production masters as one means to describe their provenance and purpose.

Technical metadata

High-effort strategies incorporate the use of standards-compliant XML tools to generate, validate, and store technical metadata for digital still images. The standards and tools being developed, distributed, and used by museum, library and archives studios include the following:

- NISO Z39.87-2002/AIIM 20-2002, *Data Dictionary—Technical Metadata for Digital Still Images*, Draft standard for trial use, www.niso.org/standards/standard_detail.cfm?std_id=731
- *MIX: NISO Metadata for Images in XML Schema*, www.loc.gov/standards/mix
- JHOVE (JSTOR/Harvard Object Validation Environment, http://hul.harvard.edu/jhove

JHOVE provides functions to perform format-specific identification, validation, and characterization of digital objects. The current version has modules to validate 11 formats, including arbitrary bytestreams. JHOVE is made available for download by JSTOR and the President and Fellows of Harvard College under the GNU Lesser General Public License (LGPL).

Summary

Digitization programs committed to producing standards-compliant metadata and digital images need to invest in the education and training—to instill technical foundations of cataloging and imaging—that guide the use of the tools (technology) that will be replaced relatively frequently.

Producing consistently good metadata records and images—the building blocks of digital image collections—challenges staff to learn objective as well as subjective quality metrics and methods, to value the short- and long-term benefits of normalizing data, to incorporate standards into workflows, and, finally, to document processes and participate in efforts to establish good community practice.