

MANAGING DIGITIZATION

A fair measure of whether a library has, in John Price-Wilkin's terms, moved from project to production resides in its capability to undertake second and third projects, particularly of increasing complexity or scale.

Technical infrastructure is key to sustaining content. Managerial skills are essential to establishing and sustaining production operations.

Well-managed digitization operations have systems, policies, staff, processes, and procedures in place to carry out all phases of digitization efficiently and professionally.

A library planning to undertake multiple digitization projects should found its program on capabilities within the organization at large to:

- Manage risk
- Manage projects
- Manage tradeoffs

These skills should reside in-house. Many of the decision-making tasks in digitization should neither be outsourced nor delegated to consortia partners.

Library administrators and staff know their collections, their audiences, and their organization's tolerances for risk better than anyone. Even if much of the digitization work is outsourced, internal policies need to govern outcomes for original source materials and define baseline functional requirements for the digital surrogates.

Manage risk

Rights

Libraries committed to digitizing collections for enhanced access will inevitably select (or nominate) source materials that are not unambiguously in the public domain. Resources should be reviewed and experts consulted to understand risks and to develop policies and guidelines for investigating rights. (See, for example, links to clearinghouses and tutorials on copyright and fair use in the *NISO Framework of Guidance for Building Good Digital Collections*.)

Although good practice recommendations exist in this arena, the extent to which rights and permissions should be investigated is a local decision. As is the decision to limit access or provide open (that is, worldwide) access to surrogates.

Digitizing items that carry republication restrictions or otherwise expose an institution to potential risks not only increases costs for digitization but also for sustainability. Rights and permissions can change over time, so if they are gathered in today's digitization project, an organization needs to weigh the costs of recording and maintaining rights metadata against the risks of possible penalties from not doing so (including eligibility for repository services).

*NISO Framework of
Guidance for
Building Good
Digital Collections,*
[www.niso.org/
framework/
forumframework.html](http://www.niso.org/framework/forumframework.html)

METS Draft Rights Declaration Schema, www.loc.gov/standards/mets/news080503.html

Project RoMEO, www.lboro.ac.uk/departments/lis/disresearch/romeo

National Digital Library Program, <http://memory.loc.gov/ammem/techdocs/conserv83199a.pdf> and <http://memory.loc.gov/ammem/dli2/html/lcndlpl.html>

Producing rights metadata to manage risk

Complex and proprietary solutions have been developed by and for publishers to manage intellectual property and rights metadata. Librarians, in turn, are beginning to propose simpler draft schemas to facilitate recording and sustaining this class of administrative metadata. See, for example:

- Metadata Encoding and Transmission Standard (METS), *Draft Rights Declaration Schema*, 2003, which “allows the documentation of minimal administrative metadata about the intellectual rights associated with a digital object or its parts, and is to be used as an extension to the Metadata Encoding and Transmission Standard (METS).”
- Joint Information Systems Committee (JISC), Project RoMEO, *Rights Metadata for Open Archiving*, 2003.

A library does not necessarily need to adopt a specific rights metadata standard and have procedures in place before digitizing, but high-level administrators should be prepared to define and enforce tolerable levels of risk in this arena for any selection strategies that extend beyond materials in the public domain.

When implemented in production, such risk tolerance policies guide technical decisions pertaining to the extent of administrative metadata to gather, the standard (such as schema) to use in recording such information, the tools and processes to create the metadata, and the responsibilities to assign in keeping the metadata current.

They also raise questions of which metadata would be used (and by which systems) to limit access to objects or their component parts if risk is to be managed by restricting access to surrogates.

Care of source materials

Libraries committed to digitizing rare, unique, and otherwise special collections presumably will not discard these valuable assets following digitization. The library should—presumably at a high administrative level—dictate and enforce the transport, handling, and disposition policies for these materials throughout the digitization process.

The Library of Congress’ 1999 report, “Conservation Implications of Digitization Projects,” describes the policies and procedures it has instituted for conservation review and treatment of books, paper, and photographs selected for digitization. Recognizing that improper handling by staff and users is one of the greatest causes of damage to library and archival materials, the library mandated that its conservators would play active and ongoing roles in the National Digital Library Program (NDLP).

What constitute acceptable standards of care and best practices for materials transport, handling, and disposition are ultimately local decisions. Librarians must weight the risks of losing, damaging, or altering source materials against the cost savings offered by various digitization strategies.

By working with digitization project or production managers early in the planning stages, librarians are best assured of tailoring workflows to meet requirements for handling and security.

The overarching policy question is: What risks to source materials must be managed during digitization?

When implemented in production, responses to this question guide decisions pertaining to:

- Methods for transporting materials to and from the digitization studio
- The location of the digitization service bureau, the qualifications of its staff, and the level of security it offers for materials storage
- Procedures for assessing and treating materials before digitization (ranging from alteration, such as disbinding, to stabilization, to conservation)
- Insurance policies for materials transported off-site
- Permitted and prohibited methods of digitization
- Storage location and circulation policies for materials following digitization

Maintenance of surrogates

With rights management and care of originals, maintenance of digital assets is the third area to assess risk at a high level in the organization. Significant costs can be saved, for the short term, by storing as many files as possible on media other than spinning disks and keeping them out of professionally managed repositories.

Such advantages, however, must be weighed against the library's tolerance for risks of losing information—with the only viable remedies being to redigitize collections.

The British Library is undertaking careful analyses of life-cycle costs for its collections. The key question underpinning its cost modeling is: What lifespan is envisioned for the collection? Assigning specific numbers of years of assumed usable life—rather than using general statements, such as *in perpetuity*—facilitates in-depth cost modeling and risk analysis.

When implemented in production, the answer to the question of minimal lifespan guides decisions pertaining to:

- The criteria to measure and audit the quality of storage services will be asked to meet
- The number of versions of digital assets to be produced
- The number of copies of media and their respective locations, format of storage media
- The necessity for checksums, error checking, and other automated methods to monitor data integrity

Manage projects

A dedicated, energetic project manager is the most valuable asset to a digitization program. To ensure success for the program, give the project manager adequate time and resources.

With adequate lead time, the manager can write grant applications, plan production workflows and timetables, help to select materials, hire technicians, select service bureaus, and oversee batch-production workflows.

In consultation with technical experts, the manager also can assess standards and practices at other institutions (manage tradeoffs), then recommend and implement project specifications that balance quality and cost.

Highly-organized people make the best project managers. Strong communication, problem-solving, budgeting, and personnel management skills are most important. Knowledge of digital imaging terminology and systems also helps, although this expertise can be gained by attending workshops (see Appendix A) and learned on-the-job by a motivated employee.

The right time to hire the manager is after the key components of downstream technical infrastructure have been set up, but before any digitization products or services have been purchased and, ideally, before any collections have been selected for digitization.

Two staffing models have proven to work well:

- Modest- to large-size programs employ a full-time project manager, charged with planning, workflow coordination, financial accounting, communications, and reporting for each project in a multiproject program.
- Smaller programs employ a full-time project manager, who not only assumes all management responsibilities but also manages and even participates in one or more aspects of production workflow: preparation of materials, cataloging, scanning, or quality control.

The staffing model that does *not* work well is to make project management an add-on task to the work of a full-time librarian or archivist. This approach saps morale among the digitization team, introduces risks of projects running over budget or over schedule, and increases chances for products to be of deficient or uneven quality.

Even a single project of modest size and a single, relatively straightforward source format (such as 35mm slides) quickly yields many workflows and many products.

Production is typically divided into categories (based on source format and size), which often, in turn, are subdivided into batches. When production tasks are then divided among multiple departments—prep/review (or conservation), cataloging, and digitization—and even multiple organizations (such as with outsourcing), project oversight and tracking are necessary and time-consuming activities.

Because the products from digitization are highly visible, digitization projects may receive greater scrutiny than any previous library project. Particularly in the early stages of program development, investments in staff are more important to project (and program) success than investments in technology.

Despite the cost, full-time project or project/production managers should be viewed as one of the necessary technical specialist positions an organization needs to establish and sustain a digitization program.

Manage tradeoffs

Unfortunately, no sure-fire strategies exist to guarantee consistent quality in collections digitization. *Every project is unique.*

The technologies, specifications, and workflows that perform well for the first batch of books (let alone collection) may not work well for the next. Source materials vary in size, language, layout, content (such as extent of meaningful color), and condition; requirements for delivery also vary from acceptance of legible black-and-white page images in the correct sequence for some genres, to insistence on viewing color reproductions accompanied by full-text searching and go-to-page navigation for others.

Interrelationships among technology, personnel, management strategies and selection policies—not technology alone—ultimately govern product quality and project cost.

Libraries seeking to adopt single digitization solutions risk tailoring selection to technology rather than identifying source materials of the highest intellectual, artifactual, or informational value. They also risk under- or over-investing in quality vis-à-vis use requirements.

Organizations that accept tradeoffs as a given are well-positioned to assess all variables to arrive at cost-effective, fit-for-purpose workflows and specifications for the source materials most deserving of digitization.

Particularly when sustainability is a desired outcome, programs should systematically identify for each project:

- The audience(s) to serve
- The selection strategy(ies) appropriate to audience needs
- Digitization standards and best practices that can be well-handled by the systems which comprise the library's downstream technical infrastructure

Audience

Above all, digitization must serve the library's mandate to make its collections accessible.

Making collections Internet accessible is not the same as making them user accessible. "Knowing the library's users and uses," Howard Besser says, best empowers an organization to make good decisions about selection and digitization strategies for each project.

Audiences may be amateur or professional information users, children or adults, tentative or sophisticated users of technology, local or distant. Each of these pairings represents a meaningful distinction or a broad spectrum of information needs. Before emulating the policies and practices of a peer institution, ask whether its programs have been configured to serve comparable audiences and audience needs.

College and university libraries, for example, must heed patterns and preferences of classroom teaching, faculty research, and student learning. Knowing whether students need to print network-delivered items in addition to viewing them on screen can alter digitization and delivery strategies.

To succeed, digitization programs will rely in some measure on the goodwill of their users. They can earn this goodwill by tailoring selection and digitization strategies to meet (or come close to meeting) user needs, while also conforming to library best practices to facilitate discovery and interoperability.

Selection strategy

As with choice of audience, selection strategy calls for high-level decisions to set a digitization program's priorities. Tradeoffs often considered are:

- Quantity versus quality
- Part versus whole
- Public versus nonpublic domain

Programs dedicated to high production, particularly at low cost, will receive pressure from digitization service providers to select materials uniform in format, size, and condition.

Strategies to cherry-pick images and manuscripts from archival collections—or to photograph images from books—might create digital products that cannot easily or affordably be synchronized (linked) with their associated descriptive records.

Venturing into collections that may not reside in the public domain increases overhead for preparation, metadata production, and preservation (since rights metadata need to be sustained as well).

Every strategy is technically viable and can be justified according to institutional, audience, and funder priorities. For each project, the manager or management team should stake claims early, then articulate core values to all members of the production team.

Obtaining access to source materials, particularly special collections, can be greatly simplified if everyone in the organization is aware of the collection development priorities to be served by digitization, and can contribute his or her expertise to discussions of risk assessment and digitization strategy.

Digitization standards and best practices

The adage, “The great thing about standards is there are so many to choose from,” is especially true for digital libraries and digitization. Library digitization programs’ uses of standards to date reveal that:

- **Formats matter.** Virtually all digital collections being created to best practices incorporate format standards into specifications and workflows for metadata, digital still images, and text. (Although some formats are technically governed by specifications rather than standards, practitioners routinely refer to both as standards.)
- **Format standards have relative value.** In practice, the choice of best standard correlates strongly to the role of a given object rather than to its genre. Digitization workflows routinely yield multiple versions of surrogates: masters optimized for long-term preservation; production masters optimized for batch processing; and deliverables optimized for network distribution and use. Experts view tagged image file format (TIFF)—a digital still-image standard—as having poor attributes for delivery images (low value), but ideal ones for masters (high value). Don’t make the mistake of referring to TIFF as either good or bad for images.
- **To receive broad adoption, standards must be sufficiently easy to use.** Standards and their related instruments to measure quality of scanning systems, to manage color, to control viewing environments, and to calibrate devices (scanners, monitors, printers) are available, but their use in digitization projects is the exception rather than the rule. Museum studios and high-end studios in libraries and archives apply these standards, but only by having professional photographers and other highly skilled technicians on staff.
- **Accountability promotes and enforces compliance with standards.** The high value placed on accessibility and discovery has promoted widespread adherence to various descriptive metadata standards and to Internet protocols for searching and delivery. As funders and other stakeholders

place equal value upon persistence (storage) or quality, pressure will increase to demonstrate compliance with standards in these arenas as well.

Each digitization program must have the means to monitor and configure packages of standards just as they must have to configure the systems to produce, catalog, store, and deliver digital assets.

The number of categories of standards applicable to text and image digitization is growing, due in part to the introduction of digital repositories and digital preservation frameworks, but also from consumers' and professionals' growing concerns with image quality and interoperability across devices.

Like hardware and software choices, standards review and selection require time and skill. Practitioners have long advocated using standard formats for master images and text, for structural metadata, and, particularly for descriptive metadata. In some cases, such as the use of TIFF for raster still images, one standard predominates and the choice is deceptively easy.

Categories of Standards Applicable to Image and Text Digitization			
Metadata	File formats	Quality	Storage & naming
Administrative	Images	Digital cameras	Media
Descriptive	Text	Scanners	Enclosures
Preservation		Monitors	Environment
Structural		Printers	Signatures & identifiers
Technical		Viewing environment	Persistent names
			Digital archive

Remember, however, that the actual implementations of standards vary widely among institutions. Not all MARC catalog records, EAD finding aids, Text Encoding Initiative (TEI) XML files, or TIFF images are alike. The mere application of standards, in other words, will not ensure interoperability or quality.

Which standards and practices are best is largely a measure of *what* function(s) they are required to serve. For example, if costs must be minimized, then black-and-white (1-bit) imaging workflows are likely to be recommended over grayscale or color approaches.

When quality is paramount, however, then recommendations would likely be reversed. (Opinions vary as to which approach yields the highest accuracy in optical character recognition.)

When using standards to produce good digital objects, librarians should remember that content follows function. Digital products are dynamic, not fixed, so ensuring that colors, layouts, and context (the content) remain consistent from one copy to the next is not possible with digital as it is feasible with reprints.

Because users have their computer monitors set to varying resolutions, the physical width and height of each digital image or digital text page changes from device to device. Even the same television broadcast on many TV sets in an electronics department shows that color rendering varies from device to device. Aged monitors particularly show variations. Different printing technologies—laser, ink jet, pictograph—also render the same digital file in different ways.

When the systems that interpret and display content are upgraded, the look and feel of content can change, even though its format has been maintained. Librarians and library users accept this change as a given for library catalogs and catalog records. Similar changes are likely to occur when new versions of Web browsers and plug-in applications, such as Adobe Acrobat Reader, are released. As Robin Wendler of Harvard University notes, "data outlive systems."

Decisions about necessary functions (or behaviors, as described in Chapter 4) exert the greatest influence over the appearance of digital products, as well as the standards, technologies, and services required to produce them.

Requiring a digital copy of a photograph to be of good quality on screen is different from requiring the same quality on screen and in print. Another difference is stipulating that users not only be able to study a fixed image but also be able to zoom in on details without losing image quality.

Text functions span an even wider spectrum. Requirements can range from creating microfilm-like black-and-white digital photographs of each page, and presenting the pages in sequential order (on screen or in print), to producing high-quality color reproductions (when needed) accompanied by fully searchable text and delivered with buttons or other interfaces that permit the user to go to designated pages or to designated sections.

Functional requirements for the uses of digital reproductions drive the choices of standards, technologies, skills, and costs.

The appearance of the digital products when delivered largely follow the decisions made about the underlying functions, as well as the design of the surrounding interface (such as a Web browser) and viewing environment at large—both of which the librarian will have little influence to control with networked delivery of digital reproductions.

Given the number of standards for digitization and the number of permutations in combining them, best practices for digitization are highly contextual.

Model workflows established in noteworthy digitization projects of the 1990s, for example, apply only to specific circumstances of source materials, handling policies, and technologies.

A text digitization workflow that requires disbinding books to use a flatbed scanner to make standards-compliant black-and-white reproductions may not be an appropriate fit to all text digitization needs.

Best practices are instituted in digitization programs as local policies and processes for:

- Project planning and management
- Materials selection and handling
- Cataloging
- Digitization and quality control
- Digital object assembly, storage, and delivery
- User services.

Peer institutions are not necessarily going to have the practices and infrastructures to evaluate and possibly emulate. Rather, organizations of any size that produce and deliver the types of digital products appropriate to your audience (users and uses) are the best sources for best practices.

Good program foundations are built on standards. Evaluating options and tradeoffs against an institution's priorities and resources is essential to choosing best standards and technical specifications for a digitization program.

Priority setting by senior administrators facilitates this task. Stated parameters for system integration, cost, quality, persistence, and interoperability give the digitization team clear direction in choosing best standards—one of the first assignments they should undertake in building a program.