

## MANAGING COSTS

Digitization costs must be managed to set aside some of the institution's revenues for sustaining content and, more fundamentally, purchasing, maintaining, and replacing the systems that mediate between the content and users.

The assumptions that high-level administrators make about sustainability influence decisions about how operational funds for digitization are best allocated.

Associating use with sustainability favors adoption of baseline, fit-for-purpose technologies and project specifications that achieve high throughput and build critical masses of material. Associating quality with sustainability favors strategies to invest in skilled staff and high-quality equipment (whether owned by the library or by the contractor).

Associating managed storage with sustainability favors investing in local digital repository or repository-like infrastructure, or purchasing these services from a trustworthy entity that allows the owning library to maintain full control over its assets.

### Tradeoffs: Quality and production

Digitization requires a team. Sustainable collections emerge from programs with capabilities to plan and manage projects, oversee materials preparation, cataloging, digitization, quality control, storage, and delivery workflows. When factoring portions of systems administrators' and public services librarians' time to provide ongoing services for digitized material, the administrator quickly recognizes that a digitization *program* is not a one-person operation.

Controlling and limiting digitization costs begins with limiting staff costs. Are volunteers available to do some of the heavy lifting in digitization? Students or interns? Will the library scope the digitization program narrowly so that it can effectively be administered by one player/manager? Are prices for outsourcing meaningfully lower than costs of creating similar products in house?

Limiting product quality requirements to the baseline levels described in Chapters 3 and 4 also controls costs. Above-the-baseline services that increase quality not only require greater effort to establish and maintain, but also greater capital. Price lists from vendors providing digitization services shed light on the proportional differences between approaches to digitization.

Scan a typescript manuscript page on a flatbed scanner to produce a 1-bit digital image and expect to pay \$0.15 to \$0.50 per page for a digital master and minimal structural or administrative metadata; scan the same page to facsimile color at a high-end digital camera and pay \$5 to \$20 for a higher-quality master.

The highest rates of throughput—at item, batch, and project levels—are achieved:

- When libraries select materials with existing catalog records or other online descriptive metadata

- When materials are permitted to be autofed during scanning (with slide scanners, microfilm scanners, autofeed/flatbed scanners, or book scanners with robotic page-turning capabilities)
- When quality requirements for post-digitization processes facilitate automation—for example, acceptance of uncorrected OCR-generated ASCII for text used for indexing and full-text searching; of structural metadata that encodes page sequences, but not page numbers; of delivery images produced in batch by a script from an image processing program
- When middleware is used to generate delivery objects dynamically on request
- When users are satisfied with legible surrogates of baseline functionality, rather than products that purport either to be examples of the best that technology can do, or faithful reproductions of sources
- When capable and dedicated staff manage projects and each phase of the digitization and post-digitization operations

### Tradeoffs: Quality and sustainability

The best time to obtain documentation of ownership, rights, provenance, and overall project rationales and methodologies is when digital objects are created. Building in sustainability begins with establishing specifications and workflows for metadata production, and developing or obtaining tools that automate as many of these processes as possible.

With respect to the quality of the source material (its housing, handling, and disposition), descriptive metadata, digital images, text for searching or display, storage, and delivery, the adage *you get what you pay for* holds true—to a point.

The image scientist Don Williams notes that, inevitably, increasing sampling rates (such as resolution) increases noise and produces empty pixels or empty magnification. Running page images through multiple OCR engines will not, at some point, yield higher percentages of accurate retrieval to search queries.

Investments in quality of descriptive, structural, and administrative metadata must be justified by users and uses—whether these are today's or tomorrow's users. For many digital processes, quality can be (and perhaps should be?) enhanced above an initial baseline only when demanded by users.

Few enhancements require redigitizing material. The Library of Congress Preservation Directorate's policy for phased delivery proposes eight strategies in which surrogates that meet minimal preservation requirements today can be "expanded or elaborated on at any time, by any party."

### Tradeoffs: Control and sustainability

Terms of ownership and control need to be studied carefully in cooperative programs or fee-for-service repository management relationships.

To receive benefits of preservation and ongoing delivery, to what extent does the institution that owns the original works cede or share control of the

Source: [www.diglib.org/standards/imqualrep.htm](http://www.diglib.org/standards/imqualrep.htm)

Source: [www.loc.gov/preserv/prd/presdig/presphase.html](http://www.loc.gov/preserv/prd/presdig/presphase.html)

marketing, distribution, and future processing of the digital surrogates—particularly if a partnering entity proposes to offer any of these services for free?

If the bits are to remain in local control, on the other hand, than more of the digitization program's operational resources must be allocated to data management and digital preservation activities (however they might be defined).

Managing this tradeoff shapes decisions about whether to centralize and bundle services within one library, or to divide them among multiple operations or organizations.

### Considerations for building in-house digitization studios

As with any new library operation, administrators must somehow find space, capitalize start-up for systems and staff, including training, and then develop a business model (or identify a revenue stream) for ongoing program support.

Two justifications to build the case for an in-house studio are to protect and control the handling of special collections (and other unique material), and to bundle services.

#### *Protect collections*

The rule of thumb that should guide security requirements is that no collection should be transferred to an environment that introduces greater risks than the storage environment from which it was retrieved.

*Studios* are immediately recognizable and distinguished from "scanning operations" or "scanning areas" by the configuration of space. If designed to accommodate special collections, studios are secure, clean, and adequately furnished. Any studio with a digital camera or any type of film scanning service should also be meticulously clean.

In addition to specifying controls over the macro environment to protect source materials—security measures, tolerances for fluctuations of temperature and relative humidity, air circulation and filtration, and light levels—librarians also should dedicate sufficient space for storage and movement of material.

The Library of Congress report, "Conservation Implications of Digitization Projects," specifies having "...table space six times the size of the largest object to be scanned to have enough free space to allow for the proper assembling of materials before capture and landing space after capture."

#### *Bundle services*

The scope of digitization services for image and text digitization should guide the space planning and budgeting for the workstations themselves, as well as the requirements for network bandwidth and production storage space (either on a file server, ideally, with back up or on portable or removable media).

Bundled services reduce management overhead and improve project oversight. Since selection, materials review, cataloging and finding-aid production, and quality control are not going to be outsourced, consider how many of these operations could be placed near one another to facilitate communication among the project members.

A studio configured for scanning, metadata production, and quality control will need more workstations, or certainly more software, than the studio dedicated solely to flatbed scanning of printed and photographic material.

Among the activities that could be outsourced, the library needs to decide how many of the following tasks it would own and manage in its studio:

- Scanning
- Digital photography
- Quality control of images
- Structural metadata production
- Administrative metadata production
- Optical character recognition (OCR)
- Keying
- Mark up
- Assembly and packaging
- Deposit (to digital repository) services

Keying or OCR workstations, and structural and administrative production stations are generally the least expensive to configure: a standard office computer often suffices.

Excepting robotic book scanners that automatically turn pages (and are priced well above \$100,000 per machine), workstations with high-end digital cameras are the most expensive to configure and support: plan to spend—\$50,000 and up; also be prepared to need to purchase a Macintosh computer (even if you are an all Windows library) for some digital cameras.

Workstations for flatbed scanning may cost as little as \$1,000 above the cost of the PC for the scanner and appropriate image processing software; prices for these workstations, however, can easily rise when levels of service require use of calibrated monitors, color management, and portable hard drives for additional storage.

Book scanners and digital cameras need separate spaces (for example, sheltered by partitions or curtains) with dark walls and ceilings to achieve consistent quality at optimal levels. Unless the library already has a photographer on staff, a consultant is highly recommended to help an organization plan space, select equipment, then install, configure, and calibrate equipment if medium- or high-effort strategies are to be supported over-and-above baseline services.

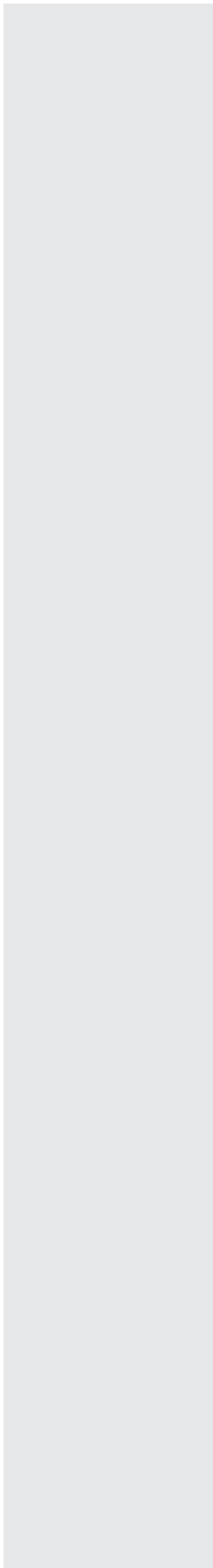
Resellers of the calibration and color management equipment cited in Chapter 3 are logical experts to turn to for consulting or for referrals, as are the sources in Appendix A.

The Tanner and Deegan report, "Exploring Charging Models for Digital Cultural Heritage," presents strategies and models to obtain revenues to help support digitization (primarily digital photography) efforts in-house.

Studios should, above all, be *used*. Develop a studio capacity that meets—but neither falls short of, nor exceeds—the capacities demanded by the library's image and text digitization programs. Purchase service agreements for all equipment.

Colleagues in museums, particularly members and member institutions of the Museum Computer Network (MCN), are other useful consultants to bridge gaps

Museum Computer  
Network (MCN),  
www.mcn.edu



in knowledge and experience. Many of the MCN member institutions have had photography studios in place for decades.

Creating a studio, like so many aspects of digitization, provides wonderful opportunities for outreach and collaboration. A commitment to build an effective, sustainable program, with managers open to receiving expert advice wherever it can be obtained, instill confidence and trustworthiness in the products created by its digitization initiatives.