COMMITING TO CHANGE

As the world changes, so do libraries. People's information needs and information-seeking behaviors evolve, demanding that libraries keep pace with their audiences or lose them.

Technology changes are profound whenever new models improve convenience (but not necessarily quality). How many times have libraries acquired the same musical work—on various formats of vinyl, then cassette, then CD?

The same book in hardcover and paperback? How many images have libraries reformatted—from original negative or print to lantern slide, then to 35mm slide film, then to digital image—to remain consistent with classroom technology for display and teaching?

The transformation of printed catalog records and shelf lists to computer data constitutes libraries' most sweeping reformatting initiative to exploit new technologies to improve services.

Past practices to repackage and repurpose works—notably during the fast-paced 20th century—reveal a pattern that shows little sign of changing, particularly in America where industry manufactures obsolescence and people love to innovate.

Don't assume that any single format— no matter how high its intrinsic quality, no matter how popular, ubiquitous, and seemingly durable today—will meet demands of usability forever.

One should not become overly fond of the technologies or the products associated with today's projects. Over the long term, material should not be kept in formats that are either hostile to use, expensive to retain, or both.

Institutions are well-served to keep as many copies and versions of their works as they can afford: the current formats that serve today's audiences, the obsolete ones that document item provenance or keep a record of previous technologies.

For all machine-intermediated formats, it is important to anticipate change in the mediating technologies, user preferences, and, most importantly, user expectations. Unfortunately, the pace of change appears to be increasing, so interventions necessary to sustain material will become more frequent, driving costs of life-cycle management upward.

For multimedia objects (such as those with multiple parts of varying formats), failures of component parts threaten to make the entire object unusable, so these are the most challenging works to preserve.

Sustaining collections requires multiple commitments: to the content itself, its associated physical and technological environments, its mediating technologies, and its users. Sustaining content—and challenges to do this are considerable given the fragile nature of digital formats and media—is not enough to meet a full preservation commitment. Nevertheless, content preservation is the first priority.

Interfaces

How will new Web browsers or upgrades to current browsers affect perceptions of the quality of digital works and collections? How will the semantic Web influence modes of delivery?

Dale Flecker of Harvard University has speculated that today's methods of delivery may become quickly outmoded: tomorrow's users may not want to access collections through prepackaged, bounded websites. Roy Tennant and others have observed that people are already avoiding library catalogs when searching for electronic information.

For reasons both technological and sociocultural, information seeking behaviors on the Internet by teachers, students, researchers, and the public can change profoundly. And relatively quickly.

Fortunately, libraries and archives are deeply invested in the Web and, particularly through public services, have several ways to monitor user needs and user satisfaction. Digitization programs should heed findings from user studies—both formal and informal—and periodically review the front ends of publishers' databases and other information products.

From the user's perspective, the interface is a relevant component of content. More fundamentally, search engines and their protocols determine whether items can and will be discovered before being displayed.

Changing protocols for the Web, the Internet, and its associated technologies are important to monitor. Sustaining digital collections is largely an effort of ensuring compatibility—of metadata with search engines, content packaging and presentation with interfaces, and content relevance with user needs.

Delivery formats

Industry innovates and will create new Web-compliant formats that offer greater quality, functionality, and convenience. Libraries that have invested in digitization need to monitor these formats to reassess the value of existing products relative to new technologies.

The extent to which JPEG 2000 will be adopted for delivering digital still images, for example, will change the terrain in which images are being distributed and used. How will users (and collection managers) perceive today's static JPEG and GIF images when compared to the dynamic features of JPEG 2000-based applications that facilitate zooming, panning, and rotating images?

JPEG 2000's rich internal metadata structure also facilitates presenting (or embedding) associated textual information with the image as it is moved from one domain (such as the catalog) to another (the user's hard drive).

New modes of delivering still images, streaming audio or images, and text represent one of the key trigger events that precipitate reassessment, and potentially re-creation, of the delivery formats for existing digital collections.

Each digital library program must define and apply its own criteria to weigh the costs and benefits of producing new delivery formats against maintaining existing formats for its collections.

Because each format carries overheads of monitoring, maintenance, and rendering (by systems), one of the important policy question is whether to implement format decisions by project and genre (such as maps) versus by content type (such as still images).

Middleware

Two modes are available to produce and deliver digital objects in formats compatible with delivery applications: John Price-Wilkin has characterized these modes as just-in-case and just-in-time delivery.

In the former, digital objects (such as HTML files and PDF documents) are precreated, named, stored, and linked just in case someone clicks on a link to retrieve the object. In the just-in-time mode of delivery, the user's click to retrieve an item invokes an application which produces the same type of object on demand.

Libraries might invest in middleware—an application that mediates transactions and lives between the interface and the repository—to save costs of back-end storage by eliminating the need to produce and store multiple versions of their collections.

Alternatively, they might swap one middleware application today—that creates PDFs dynamically from stored master TIFF page images—for one that yields tomorrow's preferred format to render, navigate, and print multipage documents.

Finally, they might be attracted to applications that enable a user to choose among several formats—such as HTML, PDF, or e-book—according to whichever access device or use need the user has at a given time.

Archival formats

Digital masters are the primary asset of any given digital collection. These masters should be transformed as a last resort. Minimizing interventions is a good strategy to maintain integrity and also saves costs.

Two factors, however, might motivate owners of digital collections to transform their masters.

The first, of course, is obsolescence, which will be indicated primarily by events that signal the disappearance of the software infrastructure—applications and their associated terms and conditions of use—needed to process and render the files.

The second is economic. If potential cost savings are offered by switching formats, organizations might be motivated to undertake the larger jobs (relative to changing middleware or producing new delivery formats) of converting master files.

Assessing whether to compress master files is a good case in point. Existing digital masters might be stored in uncompressed formats, consistent with best practice recommendations to minimize the number of format attributes that need to be monitored for obsolescence.

But what if these files were nominated for transfer to a managed repository that prices its storage services according to size (for example, at an annual rate per GB)? In addition, what if open-source (or otherwise open) lossless compression algorithms were readily available?

Assuming that quality is fully maintained, would reduced storage costs for masters override the risks of potentially increasing obsolescence?

Processing costs also are important to monitor. Relative to the applications in place within an institution's digitization program and distributed among digital libraries and digitization service bureaus, do some formats lend themselves better than others to automated processing? If so, then organizations might transform formats that are not functionally obsolete to take advantage of new technologies that meet collection management and delivery needs.

The cost-benefit analyses pertinent to master formats for content files (images, text, audio) also apply to associated metadata. Transforming structural and descriptive metadata objects to new formats might be an easier task to program, but timing will, to some extent, influence the degree of difficulty.

Deciding when to reformat is an important management function of a library's digitization program. Like analog-to-digital conversion, digital conversions raise questions of good practices and methodologies.

Following practices of peer and leading institutions—particularly those that have digitized on a large scale and have large investments to protect—is one viable strategy. Organizations should adopt the safety-in-numbers approach with some caution.

The motivations underpinning one institution's decisions to transform formats or purchase new middleware applications can and do differ from another's. In some cases, institutions will invest in these activities primarily to sustain collections; in others, they might do so to improve usability or to save long-term costs to justify making additional near-term investments.