Werk

Titel: Session I: Digitizing Multimedia Collections

Autor: Campbell, Laura

Ort: Graz

Jahr: 1995

PURL: https://resolver.sub.uni-goettingen.de/purl?514854804_0005|log41

Kontakt/Contact

Digizeitschriften e.V.
SUB Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen

info@digizeitschriften.de

Session I: Digitizing Multimedia Collections

LAURA CAMPBELL
Director of Library Distribution Services, Moderator

Digitization of the special collections in the Library of Congress will provide access to people who cannot travel to Washington to view the materials. About 80 percent of the special collections are housed in Library buildings in Washington, and many of the collections are unique. By offering electronic access, the Library can improve the educational and research benefits derived from these works.

During the last five years the Library has been conducting a pilot project to test methods for digitizing its special collections. Under the pilot, the Library has converted 220,000 items, including Mathew Brady Civil War photographs, early films of San Francisco before and after the great earthquake, political cartoons, and documents from the Continental Congress and Constitutional Convention.

During the tests, certain conversion challenges became clear. It is difficult to convert rare documents, because many cannot be sent through a document feeder or a book box for filming. Examples of other troublesome materials include 78 rpm records, oversized maps, and early video images. Creative solutions require time to develop, and the digitization process will not move quickly for materials such as these.

As the Library converts a large body of historical materials, some of the challenges that must be resolved include:

- Achieving a high level of productivity
- Preserving fragile artifacts
- Handling a variety of original formats
- Retaining the original
- Providing security for the original
- Verifying the quality of what is converted
- Keeping a preservation copy
- Adding finding aids and tagging to the material
- Adopting standards and developing a common method for conversion

The first goal of digitization is access, but the ways in which the public uses the information will affect the methods of storage and the finding aids that are needed. In a project at Cornell University, the staff compared usage of digital images versus digital text data from chemistry journals. Users most often chose to browse though the page images rather than the text keyed from the same, original page. This leads to questions about how library collections should be presented in the digital world. Should a page image be preserved or should access to the content itself be the primary goal. The answer is probably both, depending on the distinctive qualities of each item that is scanned into a computer system.

Access

There is great tension between providing electronic access to digital library collections and at the same time preserving them. The Library of Congress must find a balance. One participant suggested that today's generation will judge libraries by access and future generations will judge them by what they collected. The Library's goal is to provide more access to its collections while preserving them for future generations.

Costs

Digitization is costly; the Library cannot depend solely on its budget allocations to pay for the digitization of its collections; collaboration will be the rule. One participant suggested reducing library building costs by sharing the costs of digitization. Two different university libraries recently built new book stacks for about $20 and $30 per book. If both libraries owned the same book and could find a way to scan it legally and to share it, they would save storage space and cut down on future buildings costs. Such a strategy would work, however, only for books that could be destroyed because they had no artifactual value.

Some university libraries have projects under way to scan decades-old academic journals. Such materials lend themselves to digitization and destruction of the original paper copies.

Other libraries are working on developing "smart" optical-character recognition (OCR) or structured, editable, readable documents from scanned works. The Library of Congress has more than 300,000 reels of microfilm, including many presidential papers. Finding an efficient way - such as smart OCR - to convert that material to readable, searchable electronic text is of particular importance.

Another cost issue is determining whether the Library wants to spend money to clean up digital works (e.g., the fuzzy recordings offered by early sound technology). One participant suggested allowing individual users or customer organizations to clean up a work and send back the modified copy for the Library's collections. Audio tapes and artwork, for example, often have deteriorted and could benefit from restoration.

Local users also could play a role in creating specialized packages of Library information. Some are already developing very specialized CD-ROM multimedia packages in their homes. If the Library nurtured such efforts, this cottage industry could help the Library get its special collections out to the public.

Another way to manage digitization costs is to provide digitization-on-demand, much as organizations now provide print-on demand. Rather than digitize an entire collection, the Library could digitize items as requests are made. Such a strategy leads to other questions, however, such as who will play for the digitization and whether it could be performed quickly enough to meet the requester's needs. One participant suggested having a service that would require a requester to pay a nominal processing fee of say $25 for the digitization of a book on demand. A digital copy of the book also would go into the Library's collections. The costs of digitization are subsidized and the Library increases its digital collections based on usage. Obstacles to this approach include the high cost of digitizing little-used works. If the demand is low, processing fees are unlikely to equal actual digitization costs.

A similar approach would be to study usage patterns and digitize the most-used items. This approach raises questions of access, however. Would the usage of seldom used items increase if they were available over the Internet? The answer may be "yes." If so, basing digitization priorities on usage figures is misdirected.

The Library, therefore, needs several methods that work together. The first part of an adaptive, self-correcting system is scanning requested items. The second part is scanning materials that are deteriorating and have been determined to be worth preserving. Finally, a group of experts should crate a list of items in priority order for digitization. Based on the available budget, those items would then be scanned.

The consensus was that the Library needs to fund some digitization and seek alternatives to subsidize the remaining digitization process.

Cataloging

Cataloging is a major component of digitization costs. Should graphic or sound images be cataloged at the item level? If so, how can millions of items be cataloged so tht the cataloging is useful, cheap, and quick? If these formats are not cataloged, then how are they to be indexed?

Some participants' suggestions include:

- Making the graphic images available and letting users add their own ancillary data that could be used as low-level finding aids and would be distinguished in some way from those officially created by staff.
- Collaboration with telephone companies could further the goal of audiotape cataloging. AT&T's labs, for example, are working on speech recognition projects for cataloging voice mail messages so users can retrieve certain messages automatically. Such technology has the potential to be used for audiotape cataloging.
- Cooperation among government agencies that are working to develop cataloging techniques for digital libraries.

Searching

Digitizing information is an important step forward, but how will people find the information in the growing stream of data on the Internet? The digital collection of the Library of Congress must be usable and not just available. Searching tools are of vital importance in bringing these collections to the public.

As librarians and information scientists debate the best methods for searching electronic collections, it must be kept in mind that the digital document is a new entity, which characteristics of its own, including its evolution, usage pattern, interface, and demand. Searching tools must be flexible enough to change as the uses of documents change.