

6-2003

Building a New Infrastructure for Digital Media: Northwestern University Library

M. Claire Stewart

Northwestern University, cstewart@unl.edu

H. Frank Cervone

Northwestern University, fcervone@uic.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/librarianscience>



Part of the [Library and Information Science Commons](#)

Stewart, M. Claire and Cervone, H. Frank, "Building a New Infrastructure for Digital Media: Northwestern University Library" (2003). *Faculty Publications, UNL Libraries*. 386.

<https://digitalcommons.unl.edu/librarianscience/386>

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Building a New Infrastructure for Digital Media: Northwestern University Library

M. Claire Stewart and
H. Frank Cervone

The Northwestern University Library has been a pioneer in text and media digitization. From early efforts primarily focused on enhancing access to reserve material to current projects involving vast quantities of streaming media, in great part these projects have been the result of close collaboration between the library and other units on campus, particularly Academic Technologies. As the depth and breadth of digitization efforts have increased, so have the technological and organizational issues. This article examines the history of digitization efforts at Northwestern University as a context for exploring the emerging issues most libraries face as digitization enters a new era.

Northwestern University Library was an early pioneer in text electronic reserves, and has had a fully functioning service to digitize articles and book chapters for classroom use since 1995. The library also has an active digital library program, through which unique or rare pieces from the collection are digitized for delivery to the wide world of scholars. The Siege and Commune of Paris photograph digitization project, completed in 1995, is the earliest example, and the spectacular collection of Edward Curtis's early-1900s photographs, *The North American Indian*, is the latest.

Northwestern University was also a pioneering user of streaming media. Political Science professor Jerry Goldman, whose *Oyez Project* is now the authoritative site for Supreme Court oral argument audio materials, began using Real Audio when it was first introduced in the mid-1990s, and released the first all-streaming version of *Oyez* in January 1996. Other faculty projects soon followed, including fellow Political Science professor Ken Janda's *Videopaths Through U.S. Politics*. Janda's project was built around news archive footage from the *Video Encyclopedia of the Twentieth Century*, and was designed to give his American government class first-hand exposure to important historical events such as the Watergate scandal and Nixon resignation, struggles of the 1950s and 60s civil rights movement, and the Vietnam war. Building on the success of the Goldman and Janda projects, Northwestern secured permission in 1999 to digitize the entire *Video Encyclopedia*, and now serves all eighty-three hours of that important resource freely to the campus community as streamed MPEG-1.

Northwestern University Information Technology (NUIT) has been an active partner with the library on many technology projects and has been instrumental in assembling the systems and infrastructure to sustain their growth from experimental to production status. One of the most visible collaborations was the offering of a fac-

ulty boot camp, Technology in Learning and Teaching (TiLT), four times a year between 1993 and 2000. The four-day TiLT program introduced faculty to the technology to build instructional tools for their courses and provided a forum in which to discuss effective uses of technology in the classroom with their colleagues, campus library and technology specialists, and outside experts. As the years passed and the Internet became ubiquitous, Web-based instructional technologies became more prevalent. As a result, the focus of TiLT shifted to course Web-site development. Eventually, however, the need for this type of training decreased due to two factors: the increasing sophistication of the development tools, which made them easier to use, and the greater sophistication of new faculty in using information technologies. With the introduction of the Blackboard CourseInfo system, the focus of training shifted from being primarily technical to concentrating on the integration of course materials.

This shift from individually crafted course Web pages to a Web-based course management system allowed faculty, librarians, and information technologists to emphasize the content of these courses rather than the mechanics of building sites. This renewed the importance of Electronic Reserve, which expanded its services to include Blackboard delivery of scanned material and providing links to full-text articles in databases and journals that the library subscribes to electronically. Requests for digital selections of non-print media from the library's image and video collections and other campus collections, such as the Art History Slide Library, also began to increase. It became possible, for the first time, to begin to plan systems and services to deliver this rich media content around a secure mechanism. Once it was possible to restrict access to digitized material to particular students in a particular class, many concerns about copyright liability were alleviated.

The center of many of the new digitization services has been Digital Media Services (DMS), a unit of the Library's Marjorie I. Mitchell Multimedia Center (MIMMC), which has been operating as a do-it-yourself scanning and media digitization facility since 1995. This self-service approach, which paralleled the philosophy of do-it-yourself for Web page building before the introduction of Blackboard, artificially limited the number of faculty who were able to use digitized material in their classes. Those who had teaching or research assistants were in the best position because they were able to assign these students to digitize materials. Unfortunately, often the results of this activity were

M. Claire Stewart (claire-stewart@northwestern.edu) is Head of Digital Media Services and **H. Frank Cervone** (f-cervone@northwestern.edu) is Assistant University Librarian for Information Technology at Northwestern University, Evanston, Illinois.

uneven and unsatisfactory. Some faculty, truly passionate believers in the power of digital media, or with particularly media-rich courses, did invest their own time in digitization. For the most part, however, a lack of time and lack of access to digitization equipment prevented faculty from using many materials in digital form.

In addition, the library began to feel some concern about the stress repeated digitization was likely to place on library materials. Although the video collection housed in the MIMMC does not circulate, the tapes had begun to show signs of deterioration that would be exacerbated by regular recreation of digital media from these tapes. Similar fears were felt for other library collections.

Further issues requiring resolution related to a lack of consistency in applying digitization standards. Lacking expertise in media formats and standards, some faculty were saving images at incorrect resolutions or file formats. In many cases, needlessly high resolutions were used for image files. This burdened the course management system both during the media upload and download processes. In addition to being inconvenient because it forced students to wait an excessive amount of time for these large files to download, the misuse of media formats caused serious complications when many students tried to use these images concurrently which increased server utilization tremendously.

■ Creating a Synergy of Services

For these reasons, among others, the library expanded the role of DMS in January 2001 to offer drop-off digitization services, free of charge to all teaching faculty. This decision, however, was not made in a vacuum and was carefully considered. Adding this service coincided with the second phase of the library's renovation plan, a complete redesign of the second floor of the east tower of the main library building.

As a result of the redesign, the possibility of creating a new synergy between services provided by the various units to be housed in the area was possible. With the completion of the remodeling, a new entity, 2East, was born. Comprising 2East are the Academic Technologies (AT) unit of NUIT, DMS, and the collection management offices of the library.

For the two years prior to their move into 2East, AT had been renting office space off campus. Moving back to campus into the library was both a solution to a practical problem and a way to solidify a strong, effective campus partnership of the two largest pedagogical support units on campus.

Bringing AT into the physical library building increased cooperation, and as a result, joint work has been greatly enhanced. Staff members from AT frequently serve on com-

mittees, task forces, and ad-hoc work groups based in the library and vice versa. As an example, representatives from all three units in 2East are members of the library's Digital Library Committee. Through such joint activities, the opportunities to exploit the synergy between library-initiated and faculty-initiated digitization projects have increased exponentially.

Despite their physical and ideological proximity, the separate missions of the 2East units have been maintained. An example is the approach to providing services around the Blackboard Course Management System, which was tested in spring 1999 and moved into production the following fall. Once the new 2East facilities opened in January 2001, DMS began to accept the first drop-off projects destined for Blackboard delivery, while AT focused its attention on key areas such as accounts creation, training and support in using the course management system, project development, and providing live video Web casting services. This clear division of labor simplified the message to faculty, and the close physical partnership allowed the library and NUIT units to refer faculty and graduate students to each other and avoid some of the miscommunication that had plagued technology services in the past.

Once the initial issues related to occupying the new space were resolved, production schedules and procedures were established and demand for digitization services increased further. DMS began to produce significant numbers of digitized slides, photographs, and streamed audio and video for faculty. This was demonstrated by the fact that DMS made more than four thousand individual pieces of media available in streamed, digital format between January 2001 and December 2002. The proliferation of these media, which in most cases were partial or complete digital surrogates of materials in campus collections, highlighted some weaknesses in the campus infrastructure that urgently demand attention.

■ The Problems of a Wealth of Activity

For the most part, media files had been handled individually, and tracked using primitive tools, if at all. The shortcomings of such a system had long been apparent, but a server migration in mid-2002 demonstrated that, over time, the institutional memory had failed and many assets were orphaned, with no known owner, project, analog parent, or in some cases, any idea of the subject of the material. In another project, work with materials in the Music Library collection focused attention on the desire to build upon existing library catalog data to more fully describe digital surrogates. This was particularly

important for music materials, where uniform title issues are notoriously difficult.

Furthermore, security issues were becoming more complex and difficult to address. Streamed content is not housed on the course management server but rather on a separate streaming server. As such, Blackboard can serve as a secure gateway, but in order to safeguard the media, it is imperative that users be authorized on the streaming media server as well. This means that the user is forced to re-authenticate when leaving Blackboard upon requesting an actual streamed file.

An additional complexity is the problem of metafile creation and arbitrary rearrangement of digital media. Most streamed media platforms such as Real and QuickTime use a redirection process to deliver streaming media. When the user requests a streaming media file through the Web browser, the request made is not for the streaming media file, but for a metafile that redirects the request to the actual streamed resource. In addition, metafiles can be used in some cases to assemble multiple media in varying ways depending on the ultimate use of the media. Synchronized Media Integration Language (SMIL) will be a key technology in this area, and in order to make the best use of it an infrastructure must be in place to automate metafile creation and allow users to create and save custom groupings or arrangements of files.

Moreover, a more robust file storage system will be required to hold the increasing numbers of media files and to insure their integrity over the long term. Given the direction media use is taking, both AT and the library hope to abandon the practice of storing access files on a server, but storing the digital masters locally—often on optical media such as CD or DVD.¹ Ideally, spinning disks connected to the network will be large and secure enough to house both digital masters and the service files used for rapid access. Eliminating the barrier between digital archival materials and their surrogates will increase the likelihood that the masters will be well backed up, undergo periodic data validity checks, and be included in migrations, either from storage location to storage location or between formats or both. All of these will be essential in preserving these valuable digital assets indefinitely.

■ Preparing an Infrastructure for the Future

These and other artifacts of the overwhelming success of our collaboration have prompted the library and AT to investigate the issues related to media infrastructure planning. In the course of the investigation, five major areas of concern have emerged: federated searching,

repository structures, still images, streaming media, and asset and rights management. However, in an investigation such as this, a major concern must be the vagaries of the future: we have to consider future trends and what the requirements of future students and faculty will be as they become more familiar with the possibilities of digital media and as the technology itself advances.

Within the issues identified, a noticeable trend is the expansion of the realm of the issues to more traditional areas within the library. Perhaps the most obvious of these is federated searching.

Although primarily a concern for the library, the issue of federated searching has a significant impact on the teaching function of faculty, particularly in the online environment. Federated searching is the ability to aggregate the contents of a search that is performed across databases. The reasons why this particular issue is of critical importance to the larger academic community are complex.

As we know, libraries spend large percentages of their budget for electronic commercial content and this is a trend that will continue.² For students, faculty, and staff, this has resulted in a proliferation both of electronic resources and interfaces to those resources, which must be learned in order to navigate and find appropriate content. In addition, because most commercial e-resources are found in aggregator sites (such as EBSCOhost), complexity is added because the same journal article may be available from multiple vendors. Moreover, navigation is made more complex by the fact that having located a citation, users still have to search to find out if the article or monograph is available through the library, and if not, know to make an interlibrary loan request (and how to make this request.) An additional issue is the need to manage who has rights to content and off-campus access.

Federated searching will be the next strategic system for libraries because it has the potential to resolve these problems and leverages the cost of local and commercial content by providing the architecture and tools to manage access.³ As such, it may become a more important service the library provides to the academic community than the online card catalog.

Some of the critical issues a federated searching system must address include providing:

- a single interface that acts as a portal and helps end users discover which campus resources will provide the research and information needed;
- intellectual organization of categorizing electronic resources through a collection management function;
- multiprotocol searching to bring backretrieve content from these resources and to allow direct access to the native interface of another content provider or search engine;

- reference linking that allows lateral navigation from citations to full text, and from any content to other relevant services such as ILL;
- the capability of integrating and managing local content such as electronic reserves and institutional scholarly digital content (such as art slides, audio clips, or archival information); and
- hooks for integrating into systems that are pivotal to the educational mission of the institution, such as course reserve systems like Blackboard.

The second major issue that has been identified is related to repositories. A robust repository structure is needed because digital objects are more numerous, volatile, and mutable than “traditional” materials and digital objects depend on and are bound to a technical environment and infrastructure.⁴

A digital repository service provides for the storage and retrieval system of digital material within collections. These services and facilities include:

- an electronic storage facility within which the digital objects created or purchased reside;
- management of administrative and structural metadata associated with stored objects;
- preservation policies and procedures to insure the continued usability of stored objects, and delivery of an object to a registered or known software application (e.g., an online catalog, a Web browser); and
- a name resolution service, which is a comprehensive service for creating, maintaining, and resolving persistent identifiers which are location-independent names for network-accessible resources. Name resolution is the process of mapping from a given abstract name to a URL that represents a particular instantiation of the named resource.

There are also a handful of issues strictly pertaining to certain types of media. Still images present a challenge because they may represent several different kinds of information. This is often seen in digital library projects, where the digital facsimile of a text page must be retained along with the optical character recognition (OCR) recognized or rekeyed text. In many cases, the two modes—digital facsimile image and searchable text—must be presented to the user simultaneously. With other materials, however, the still image facsimiles represent other forms of data, including musical notation and traditional visual information, for which the technology to build a direct index of the contents does not yet exist or is not readily available. For projects dealing with images of this kind, the campus media infrastructure must be flexible enough so that when tools such as visual content mapping and automated music recognition technologies do become available, they can be integrated with a minimum of effort.

In addition to such specialized image indexing tools, an infrastructure designed to describe and store images must support tools for high-resolution browse. Many faculty, particularly those using images in lecture, wish to zoom in on details of paintings or oversize maps, making a zoom and pan tool a critical component. Some of these same users require tools that will allow them to save annotations about the images as a whole, or attach annotations to specific regions of an image. This is useful for maps and anatomical images, which are often highly complex, but is also useful for a host of other types of images.

The rapidly evolving nature of streaming media technologies makes careful planning absolutely essential. While compression algorithms and file types are changing constantly, users demand the highest visual quality that computing power and network connections can support. Much more so with digital audio and video than with digital text, the formats commonly used for desktop delivery today will almost certainly be obsolete in a short time frame, perhaps even within a few years. In addition, the library community has not yet agreed on a file format for long-term archival storage of digital video.⁵ With this uncertainty at both the high and low end of the spectrum, planning both for long-term retention and short- and long-term delivery of video becomes a nearly impossible task.

To help resolve this dilemma, Northwestern has been investigating server-side transcoding technologies. These systems store a high-quality, high bit-rate audio or video file (mostly likely MPEG-2 or MPEG-4 in the case of video) but transcode it to a small, lower-quality version at the server for delivery to the client. Such a scenario will avoid conflicts between different versions of media players, and gives users with slower network connections a version of the content that they may reliably play without loss of data. This approach introduces a host of new questions. Once requested and created, should service files be retained permanently? What storage format for video will yield the best possible balance between high quality and storage efficiency? Of the many pieces of auxiliary data, such as edit decision lists and raw footage that may accompany a finished film or video work, which are worthy of long-term retention?⁶

The last major issue that has been identified is rights and asset management. A robust, scalable campus media infrastructure must allow media creators and managers to assign and change the right to access shared materials in both a fine and a coarse manner. Some materials might, for example, be freely accessible at low resolution to all music faculty for preview purposes, but the high-resolution versions only accessible once assigned to a specific course. Translating complex relationships such as these into clear rules for interoperating campus information systems is a significant challenge. Nevertheless, a robust

and complete access management system must be in place to safeguard the university's valuable media assets. This is important both to assure copyright compliance and to give faculty and other media producers on campus a guarantee that their unique materials will not be vulnerable to theft on the Web.⁷

Northwestern University Library is engaged in a process of reevaluating its entire media infrastructure. This study will have implications ranging from determining the amount and type of storage media needed to sustain growth, to bridging traditional divisions between database subscriptions and local digital materials, to expanding the scope and complexity of core library collection development activities.

But this reevaluation is just part of a larger process within the library. This is just a manifestation of our engagement in a continuous process of rethinking the library and its services. The importance of this was outlined by Wilson

[For libraries] to be successful, we need to be essential to people and to stay ahead of client expectations. Library managers need to show vision and leadership on behalf of their clients, rather than simply respond to client feedback, when developing services and facilities. This means we need to continually rethink libraries themselves, rather than simply the services we provide.

In order to remain a vital part of our academic community, we must continually question what we do, how we do it, and for whom we are doing it.⁸

AT and other NUIT units are key partners in evaluating and implementing components of the new campus media infrastructure. Large-scale systems support has traditionally been provided by NUIT; the several large servers that comprise the library's current library management system, for example, are maintained by NUIT's Computing Services unit. AT is responsible for management of the Blackboard Course Management system, and therefore is central to all decisions made about systems that allow faculty to more easily find and integrate digital information into their classroom teaching. Already, AT is working to develop tools and utilities to build learning modules around reusable, sharable digital objects.⁹ The library selects and makes available a variety of materials in electronic formats: full-text articles, electronic books, and digitized media. Improving and extending intellectual access to these materials must be the result of close collaboration between the library and AT.

Conclusion

By working with faculty who were truly passionate believers in the power of digital media, the Northwestern

University Library has developed a rich infrastructure to support text and media digitization. Combined with the opening of a new shared office space within the library, known as 2East, where the AT unit of NUIT is colocated with the DMS department and the collection management offices of the library, the possibilities of collaboration in providing media services have been greatly extended.

While early efforts primarily focused on enhancing access to reserve material, current projects extend across disciplines and media formats and include vast quantities of streaming media. For the most part, these projects have only been possible as a result of close collaboration between the library and other units on campus, in particular, AT.

As the depth and breadth of digitization efforts have increased, so have the technological and organizational issues. Furthermore, the rapidly evolving nature of streaming media technologies makes careful planning absolutely essential. As a result, the Northwestern University Library is engaged in a process of reevaluating its entire media infrastructure in partnership with AT and other NUIT units. By working together, we will be better able to evaluate and implement various components for a new campus media infrastructure that will support the primary purposes of our institution: teaching and research.

References and Notes

1. Franziska Frey, "File Formats for Digital Masters," *Guides to Quality in Visual Resource Imaging* no. 5, Jul. 2000, Council on Library and Information Resources. Accessed Mar. 3, 2003, www.rlg.org/visguides/visguide5.html.
2. T. H. Hogan, "Drexel University Moves Aggressively from Print to Electronic Access for Journals (interview with Carol Hansen Montgomery, Dean of Libraries)," *Computers in Libraries* 21 no. 5 (May 2001): 22-27.
3. Brian E. C. Schottlaender, "The New Academic Platform: Beyond Resource Discovery," (paper presented at the ARL forum, "Collections & Access for the 21st-Century Scholar: A Forum to Explore the Roles of the Research Library," Washington, D.C., Oct. 20, 2001). Accessed Mar. 3, 2002, www.arl.org/forum/schottlaender.
4. R. Crow, *The Case for Institutional Repositories: A SPARC Position Paper*, (Washington, D.C.: ARL Scholarly Publishing and Academic Resources Coalition, 2002), 1-4. Available at www.arl.org/sparc/IR/ir.html.
5. "Video Technical Issues," *Audio Visual Prototyping Briefing Document*, in Digital Audio-Visual Preservation Prototyping Project, (Washington, D.C.: Library of Congress, 1999). Accessed Mar. 3, 2003, <http://lcweb.loc.gov/tr/mopic/avprot/avbrief3.html>.
6. Mary Ide, Dave MacCarn, Thom Shepard, and Leah Weisse, "Understanding the Preservation Challenge of Digital Television," *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, (Washington, D.C.: Council on Library

and Information Resources and the Library of Congress, 2002). Accessed Mar. 3, 2003, www.clir.org/pubs/reports/pub106/television.html.

7. Kenneth D. Crews, *New Copyright Law for Distance Education: The Meaning and Importance of the TEACH Act*, (Washington, D.C.: ALA, 2002). Accessed Mar. 3, 2003, www.ala.org/washoff/teach.html.

8. M. Wilson, "Understanding the Needs of Tomorrow's Library User: Rethinking Library Services for the New Age," *Australasian Public Libraries and Information Services* 13 no. 2 (June 2000): 81–86.

9. Jonathan Smith, "Learning Technologies," *2East eNewsletter*, 4 no. 3 (Dec. 2002). Accessed Mar. 3, 2003, <http://web.at.northwestern.edu/da/archives/12-10-02/article2.htm>.