DIGITAL LIBRARY
FEDERATION

NEW YORK UNIVERSITY

REPORT TO THE DIGITAL LIBRARY FEDERATION
FALL, 2003

Table of Contents

I. Collections, services, and systems

A. Collections

Online Audio Reserves

NYU has digitized all of the language tapes used for foreign language instruction, and is currently making them available through an electronic reserve system. The audio files are available as streaming media delivered by Real Server 8, and access is limited to students enrolled in the various language courses offered each semester.

Archives of Irish American

The Archives of Irish America has been building upon a pilot project started in 1997 to survey and collect materials related to the New York Irish community. Under a multi-year grant from the Irish Institute of New York in memory of its founder Paul O'Dwyer, material was retrieved from spare rooms, basements, attics, and garages in the metropolitan New York area. Selected primary source materials are being digitized and made available online.
http://www.nyu.edu/library/bobst/research/aia/

Database of Recorded American Music

Working in partnership with New World Records (NWR) on a project funded by the Andrew W. Mellon Foundation, NYU is developing a database system capturing descriptive, administrative and structural metadata regarding NWR's entire catalog of music down to the audio track level. The database is linked to both high and low bit rate (MP3 and Real, respectively) streaming versions of all NWR recordings. A web-based search interface for the database and streaming media has been developed, with the database available to the general public and access to the streaming media currently limited to NYU faculty, students and staff. The system will eventually be made available as a licensed service. Enhancements to the database, including revisions to the user interface, a transition to the Internet2 Shibboleth authentication framework, and use of MPEG4 audio formats, are continuing.
http://dram.nyu.edu/

Encoded Archival Description

The University Archives, Fales Library & Special Collections and the Tamiment Institute Library & Robert F. Wagner Labor Archives are engaged in a collective process to bring all online archival finding aids in compliance with EAD 2002 and make them available in HTML format using dynamic XSLT transformation.
http://dlib.nyu.edu/divlib/bobst/archives/findingaids/

B. Services

Ask a Librarian

"Ask a Librarian" provides both e-mail and real-time chat electronic reference service to the NYU community, as well as a searchable set of the most frequently asked reference questions.
http://library.nyu.edu/ask/

Electronic Reserves

Bobst Libray has implemented Docutek's Eres electronic reserve system to provide faculty with greater flexibility in making reserve reading materials available to their students.
http://eres.library.nyu.edu/default.asp

C. Systems

Sun Fire Cluster

In order to provide a highly reliable technological infrastructure for NYU's Digital Library, collections and software services for the Digital Library have been migrated on to a cluster of Sun Fire servers. The primary server for the Digital Library is a single domain on a Sun Fire 15K from Sun Microsystems with 24 processors and 48 gigabytes of main memory; this domain can fail over to a separate domain hosted on a Sun Fire 12K with 12 processors and 24 gigabytes of memory. Both the primary and secondary servers have access to 10 terabytes of disk storage provide by a set of Sun StorEdge T3 arrays. The servers are running Solaris 2.8 and are providing database services using both Oracle 9i and MySQL along with streaming media services provided by RealServer 8 and Darwin Streaming Server.

DigiTool

NYU Libraries and Information Technology Services staff are working with Ex Libris, Ltd. and Ex Libris (USA), Inc. to implement their digital library system, DigiTool, within NYU's Digital Library Sun Fire cluster environment. A prototype installation has been completed, but is not yet publicly available.
http://www.aleph.co.il/dtl/index.html

II. Projects and programs

A. Projects

Infrastructure for Rich Media Education Environments

The Infrastructure for Rich Media Education Environments (IRMEE) is a university-wide effort to develop a technological infrastructure for the storage, organization and retrieval of rich media assets which will support both faculty and students' needs for creation and use of complex hypermedia narratives for educational purposes and librarian's needs to ensure long-term preservation of digital assets. A pilot effort to develop IRMEE and use it in conjunction with surgical training at the NYU School of Medicine is currently in planning stages.

Projects in progress 2003:

Afghanistan Digital Library

This project intends to make available both on readable electronic media and on the internet the entire publishing output of Afghanistan from 1871 (the earliest printed book) to 1930, searchable by title, author, subject, and date. The project will begin with the earliest books and proceed chronologically. At some point, a decision may be made to expand the scope to include rare newspapers, journals, and government documents.
http://afghanistandl.nyu.edu/

Hemispheric Institute

This recently initiated project will digitize and make available 200 hours of video documenting Latin American indigenous and avant-garde performance held in archival collections.
http://hemi.ps.tsoa.nyu.edu

Political Communication Web Archiving Project

Working under the auspices of the Center for Research Libraries and in cooperation with Cornell University, the Internet Archive, Library of Congress, Stanford University and the University of Texas, NYU Libraries is leading a technological investigation into effective methodologies for the systematic, sustainable preservation of Web-based political communications.
http://www.crl.edu/content/PolitWeb.htm

Metadata Encoding & Transmission Standard (METS)

Working with the Digital Library Federation and the Library of Congress, NYU continues to take a lead role in the development of METS, an XML format for the standardized encoding of digital library objects. METS has already been adopted by a variety of other institutions and projects, including the Fedora project at Cornell University and the Library of Congress Audio/Visual prototyping project.
http://www.loc.gov/standards/mets/

III. Specific Digital Library Challenges

Automating technical metadata collection

NYU is focusing much of its digital library efforts on audio and video resources. As we begin to address these new media types, we are encountering the same issues around capturing technical metadata that other libraries have encountered in working with text and still image resources. We need to have standardized element sets and formats for recording this metadata; none currently exist. We also need to automate the production and recording of this metadata to the greatest extent possible.

Descriptive metadata costs

Generating accurate, reliable descriptive metadata has proved the most expensive, time-consuming tasks on most of our digital library projects. Much of the material is from special collections, and existing item-level metadata is weak or non-existent. The scale of many projects preclude using library cataloging staff to create descriptive metadata; we simply do not have sufficient staff in-house, and the costs to place sufficient numbers to work would be prohibitive.

Build-it-yourself vs. Off-the-shelf

While we have gone out of our way to try to implement new digital library projects in a manner which allows for significant re-use of both database design and code, developing digital library applications in-house is expensive and leads to a series of on-going maintenance costs. The advantage is that such systems allow the digital library team to ensure conformance with both relevant standards and local practice to a degree which would not be possible with any commercial 'digital library' application or asset management system. Employing commercial off-the-shelf application would reduce development costs and free programmer time for improving end-user interfaces and supporting digitization work flows, but existing commercial systems are lagging the digital library community by several years in terms of their understanding and implementation of relevant standards and features. We seem to face a continual choice of getting wanting we want and paying dearly for the privilege, or using a commercial application and getting far less than we want or need.

Preservation worthy video: hope you brought your checkbook

We have encountered both technical and financial issues in our efforts to start producing preservation worthy digital video. At this point, we believe that 'preservation worthy' digital video should use no or lossless compression, should employ 4:4:4 sampling, and should use a standard, non-commercial format for storage. The Motion JPEG 2000 standard (ISO/IEC 15444) may provide a storage format meeting the technical requirements, but software providing an implementation of this standard and which meets our requirements for digital capture and editing is difficult to find. The costs in creating a workstation for digital capture of video which fulfills our requirements are also quite high. When one includes the costs for the workstation itself, a disk array capable of absorbing data at the rate required for uncompressed real-time video capture, a video capture card which will support 4:4:4 sampling, and the necessary equipment for calibrating and monitoring video signals (time base correctors, waveform monitors, vector scopes, signal generators, etc.), it is fairly easy to spend $200,000 on a single capture workstation. Standard definition video, if stored uncompressed, can consume 120 GB for an hour of video, leading to rather exorbitant storage costs if kept on disk.