

DIGITAL LIBRARY FEDERATION

The New York Public Library

Report to the Digital Library Federation Spring, 2004

TABLE OF CONTENTS

- I. Collections, services, and systems
- II. Projects and programs
- III. Specific digital library challenges
- IV. Digital library publications, policies, working papers, and other documents

I. Collections, services, and systems

A. Collections

A Virtual Archive of Whitman's Manuscripts.

The New York Public Library is participating in "A Virtual Archive of Whitman's Manuscripts" as one of two-dozen sub-contracting institutions for the University of Nebraska's IMLS national leadership grant. NYPL has completed its first batch of images including 126 captures, about 15% of the total to be done, at 600 dpi, from the Berg Collection of British and American Literature.

Art and Architectural Images from Eastern Europe and Eurasia.

NYPL's "Art and Architectural Images from Eastern Europe and Eurasia" is an NEH Access and Preservation that includes the conservation and digitization of more than 5000 illustrations for web delivery in our larger digitization project, the NYPL Digital Gallery. Imaging underway with JJT Inc. as the onsite subcontractor.

B. Services

Email Reference.

The New York Public Library provides free online ask-a-librarian services. Each curatorial or research division has an email address posted and users submit questions directly.

C. Systems

Servers.

The Digital Library Program uses a combination of Windows 2000 and UNIX servers. In order to improve information retrieval capacity and server reliability, the Digital Library Program has acquired a SunFire V480 with 2 900MHz processors and 4GB RAM, and Sun StorEdge 3310 SCSI Array with twelve 73GB Disk Drives.

XML and Open Source Search Engines.

The Digital Library Program has recently developed a text search interface that queries and retrieves digital image metadata, via an open-source search engine -- the Apache Software Foundation's Jakarta Lucene (Release 1.3). Jakarta Lucene is a full-featured text search engine written in Java. The web application retrieves indexed XML files, and runs on Cold Fusion MX. Digital assets, primarily metadata, are stored in an Oracle 8i (Release 8.1.7) database, however our production site uses a subset of those data that are exported and converted to XML.

II. Projects and programs

A. Projects

NYPL.org Redesign.

The New York Public Library has engaged the design firm Razorfish.SBI to complete a redesign of the NYPL web site. Web designers and information architects have been given the task of recasting the vast resources on the NYPL web site in a more systematic and tightly organized structure.

Some interesting challenges have included:

- 1) creating a unified presence for a library that comprises two systems, Branch and Research,
- 2) improving access to library materials in a two-OPAC site, and
- 3) presenting resources to a cradle-to-grave audience with considerable variation in age, education and interest.

Over the next several months web developers across the library will implement the new design and information architecture.

Projects in progress 2003:

NYPL Digital Gallery

The New York Public Library will soon launch the NYPL Digital Gallery, the Library's new digital image database, developed to provide free and open online access to thousands of images from the original and rare holdings of the Research Libraries. Spanning a wide range of visual media, NYPL Digital Gallery will offer digital images of artwork, maps, photographs, prints, manuscripts, illustrated books and printed ephemera. Encompassing the subject strengths of the vast collections of the Research Libraries, the materials presented in digital form will represent the arts, humanities, social sciences, applied sciences and performing arts.

The Library gratefully acknowledges The Atlantic Philanthropies for its leadership support for the NYPL Digital Gallery. Additional support has been provided by generous grants from the Stavros S. Niarchos Foundation and AOL Time Warner.

In Motion: The African-American Migration Experience

The Schomburg Center for Research in Black Culture has been awarded a grant from the Institute for Museum and Library Services to create a web site entitled: *In Motion: The African-American Migration Experience*. The site will present and interpret thirteen migrations to, within, and out of the United States; from the transatlantic slave trade to the Western migration, the colonization movement, the Great Migration, and the contemporary immigration of Caribbeans, Haitians, and sub-Saharan Africans.

In addition to original essays on each migration being written by scholars eminent in the field, the site will use over 8000 images and 16000 pages of digitized text culled from the collections of the Schomburg Center for Research in Black Culture and other institutions. The site will also feature lesson plans and other materials being developed by educators to help make use of the material in the classroom.

The site will make extensive use of XML. All of the text resources will be marked up in TEI format, and other parts of the site will use XML written against in-house DTDs for presentation and configuration. Additionally, we will be using software developed using several open-source programming libraries, including the Lucene search engine API. The site will use the ColdFusion MX application server.

<http://digital.nypl.org/migration.htm>

III. Specific Digital Library Challenges

Mixed Media Collections

More and more collections are being gifted to the library that contain mixed media. In this case mixed media does not mean mixed, electronic media, but a combination of physical and corollary or complementary electronic material. The Library is in the process of making policy decisions about how to accession and describe these "mixed" collections and make them available to the public.

Collecting Born Digital Objects

Born digital materials coming to the library for permanent deposit are a new type of archiving responsibility. Library is exploring a retention policy for digital objects that will be based on taxonomy of content. The obligations and responsibilities for electronic materials being as varied and diverse as the physical holdings, the Library expects that it will develop a staged approach to electronic archiving.

IV. Digital library publications, policies, working papers, and other documents

CNI Presentation by John Parsons, Technical Project Manger for The African-American Migration Experience: *The New York Public Library Digital Gallery: An Open Source Approach Using XML and the Lucene Search Engine API*

The New York Public Library's Digital Library Program charge to provide access to hundreds of thousands of digitized images from the Research Libraries and to provide fast and easy search tools has been a technological challenge. NYPL has developed a low-cost architecture using XML representations of our image data indexed and searched using Lucene, an open source search engine API supported by the Apache Jakarta Project. This briefing examines the details of the architecture and elaborates on some of the benefits and problems that come along with it.

<http://www.cni.org/tfms/2003b.fall/abstracts/PB-newyork-parsons.html>