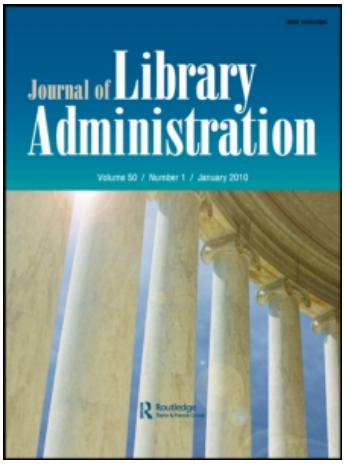

Journal of Library Administration
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t792306901

Digital Collections: History and Perspectives

Greg Zick[a]
[a] Global Engineering, OCLC Online Computer Library Center, Inc., Dublin, OH, USA

To cite this Article Zick, Greg(2009) 'Digital Collections: History and Perspectives', Journal of Library Administration, 49: 7, 687 — 693
To link to this Article: DOI: 10.1080/01930820903260465
URL: http://dx.doi.org/10.1080/01930820903260465

Journal of Library Administration, 49:687–693, 2009

ISSN: 0193-0826 print / 1540-3564 online
DOI: 10.1080/01930820903260465

Routledge
Taylor & Francis Group

Digital Collections: History and Perspectives

GREG ZICK

Global Engineering, OCLC Online Computer Library Center, Inc., Dublin, OH, USA

ABSTRACT. *The author reviews the history of early efforts by libraries to build digital collections, the development of CONTENTdm software at the University of Washington, and OCLC's current directions in helping libraries build and manage digital collections.*

KEYWORDS digital collections, emerging standards, digital images, metadata, Internet, large data sets, primary source material, special collections

INTRODUCTION

Today, the result of a search in WorldCat often includes the full digital file of a large selection of primary source material. This primary source material in digital form is accessible through your browser. The available digital content is rich and diverse—from peer-reviewed images of cells for the study of cell biology to multiple fully transcribed and searchable handwritten diaries from the U.S. Civil War held in collections of multiple libraries. Since it is in digital form, the primary source material can be reviewed, read, researched, retrieved, and included as part of a rich resource for scholarly research and learning for all ages. The breadth and depth of digital collections available is extensive. Other examples are:

Amelia Earhart search and rescue report (1937) from the George Putman collection of Amelia Earhart papers, Purdue University
http://earchives.lib.purdue.edu/cdm4/document.php?CISOROOT=/earhart&CISOPTR=3011&REC=3

Senator Joe McCarthy: Audio Excerpts, 1950–1954—An audio recording of McCarthy defending his methods at a veterans' gathering in Rhinelander, Wisconsin, in 1952 at the Marquette University Raynor Memorial Library

Address correspondence to Greg Zick, OCLC Online Computer Library, Inc. 6565 Kilgour Place, Dublin, OH 43017-3395, USA. E-mail: zick@oclc.org

http://digitalmarquette.cdmhost.com/cdm4/item_viewer.php?CISOROOT=/p128701coll0&CISOPTR=3&CISOBOX=1&REC=18

Journal of the First Voyage of Columbus—American Journeys: Eyewitness Accounts of Early American Exploration and Settlement, Wisconsin Historical Society
http://content.wisconsinhistory.org/cdm4/document.php?CISOROOT=/aj&CISOPTR=4213

Letter from Maya Angelou to Alex Haley—The Voyages of Alex Haley: Notebooks and Memoirs of an African-American Writer, Broward County Library
http://digilab.browardlibrary.org/cdm4/document.php?CISOROOT=/alexhaley&CISOPTR=2342&REC=10

Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11 (a peer-reviewed article published in the journal *Nature* that was coauthored by geneticist Mario Capecchi, who was a joint recipient of the Nobel Prize in Medicine for 2007 for this line of genetics research)—USpace Institutional Repository, University of Utah
http://content.lib.utah.edu/cdm4/document.php?CISOROOT=/ir-main&CISOPTR=520&REC=2

Jane Eyre: Historical image character portrait of Katharine Hepburn in the title role (1936–1937)—Theater Photography from the Theresa Helburn Collection, Bryn Mawr College
http://triptych.brynmawr.edu/cdm4/item_viewer.php?CISOROOT=/Helburn&CISOPTR=792&REC=4

Lesson Plan: Depression Photo Essay (Grades 9–12)—Teaching with Digital Content, University of Illinois–Urbana-Champaign; Chicago Public Library Special Collections Library; Early American Museum; Illinois Heritage Association; Illinois State Library; Lakeview Museum of Arts and Sciences; Lincoln Home National Historic Site; McLean County Museum of History; Museum of Science and Industry; Mystic Seaport
http://images.library.uiuc.edu/projects/tdc/lessonplans/DepressionPhoto.html

Access to these rich collections of digitized primary source material is the result of many individual library efforts supported by software tools and services offered by Online Computer Library Center (OCLC) and others. It is a long-term goal of OCLC and its members to build a global digital repository that aggregates the metadata of digital items with links and access to the items. This article discusses the origins of the digital collections tools, how the evolution of technology and information science has combined to provide technology solutions, and how the activity of building digital collections has moved from small initial "special" projects within the library to large and at times regional efforts with a longer-term focus on digital access and preservation of these unique and valuable digital resources.

HISTORY

The initial efforts of building digital collections were fueled by the widespread and rapid growth of the Internet. The sharing of digital information in the form of computer text files rapidly extended to images, documents, and other media. During the 1990s, significant advances in computer technology created a wide user base for this information channel. The decade saw a significant improvement in the color quality and resolution of computer displays, a rapid increase in central processing unit (CPU) power, a substantial increase in disc storage capacity, and an ever increasing improvement in network speed—each generation of technology costing less than its predecessor. Easily accessible and relatively affordable software tools to support Web sites were widely used, and the first wave of local collection Web sites sprouted up in many libraries. Typically, these collections were comprised a set of scanned photos posted within a locally designed Web homepage and various navigation designs. While these efforts attracted considerable interest due to their novelty and creativity, there were a number of problems.

First, there were no established standards or set of best practices for organizing this new data medium. Finding particular items was an intellectual challenge. The collections were created in separate silos of effort. They lacked a cohesive connection for comprehensive searches, Web store site maintenance, and consistency in implementation. Second, sustainability was a continual problem. These projects were typically implemented as a special effort by staff that held other full-time jobs. After completing the first instance, they typically returned their focus to these primary jobs. Support and evolution of the Web site were a great challenge because of staff changeover and the lack of common tools. Finally, the existing software for large data sets was designed for text and numbers. The software did not handle images well and was usually priced for the commercial market and beyond the budget constraints of most libraries.

In the mid 1990s, I was director of a research laboratory at the University of Washington in Seattle, whose focus was on nontraditional database applications in medicine. Our group was studying how to handle the mixed data resulting from the relation type of patient name, address, insurance company, etc., and of the visual type with text such as an X-ray or CT scan with the associated text diagnosis. Implementing a mixed data type data solution allowed us to search large repositories of medical images for specific types of patients who were diagnosed with a specific condition.

I was invited to speak at an IBM library forum to present this work as an example of future directions for library systems. My presentation covered the evolving technology, the features of our research systems, and the opportunities of future information systems for library organizations. The second speaker at the forum expressed great concern about the limitations of technology in the context of how difficult it was for end users and

professional librarians to implement and utilize these new Internet-connected computer systems. The issues included a lack of understanding of metadata descriptions for the digital data, rapidly evolving technology, difficulty in curating collections because of changing staff and complex tools, the challenge of ever-changing standards, and a lack of search options for multiple and diverse collections.

Over the course of several additional follow-up discussions I identified an opportunity to use the technology from our research efforts and the knowledge of the library community to develop a new kind of software that would handle these challenges. The design included these objectives: simple to use, customizable for local requirements and recognition, internally robust, powerful text search engine, open metadata definition. The software would also support metadata cross walking to allow cross-collection searching while retaining the individual richness in the collection metadata description.

BROADER GOALS FOR THE SOFTWARE

An early challenge was how to identify and support the full range of digital content type found in a special collections library. We wanted to support the full range of special collection content types—images, photos, maps, postcards, historical newspapers, documents, and books. It was also important to provide the ability for each collection coordinator to define and use metadata that was specific to their collection while providing a collection administrator-controlled crosswalk to a common metadata scheme like the Dublin Core.

We were very interested in an architecture that would encourage sharing and support collaborative work in the digital workflow. We achieved this by developing a desktop application—called the acquisition station—that could be used by multiple catalogers working on multiple collections on the same server at the same time. This implementation was successfully used—first by the University of Utah and then by the State Library of Louisiana (LOUIS)—to initiate and support multiple institution digitization and collection building. We called this software suite of tools CONTENTdm.

This version of the CONTENTdm digital collection software was developed by a team at the University of Washington that included Craig Yamashita, Lawrence Yapp, Joe Tavares, Geri Ingram, and Jill Fluvog. The first digital collection implemented on CONTENTdm was done in partnership with the University of Washington Libraries. In 1996, the Libraries launched a campuswide effort to transform the unique scholarly collections of faculty and libraries into a rich multimedia digital archive. The Libraries partnered with the College of Engineering's Center for Information Systems Optimization to develop a high-performance image archive complete with robust

metadata capabilities, multiresolution scanning, and a powerful search engine. These features were needed to bring new life to special collections.

About 100,000 digital images were transferred from video discs to the CONTENTdm system. The images were a collection of early Seattle vaudeville and entertainment posters. The online access to these images generated great interest. Over the next 5 years we continued to improve the software with a variety of new features, and as more libraries heard about our efforts they began to request copies of the software. In 2001, the technology was transferred from the university to a new company named DiMeMa (for digital media management). Over the next 5 years the CONTENTdm user base grew from a handful of institutions to over 300 U.S. libraries. In 2002, DiMeMa partnered with OCLC to provide access to libraries and joined OCLC in 2006, as the leading provider of digital collection management software.

CURRENT STATUS

OCLC now has a Digital Collection Services group that provides services for libraries and other cultural heritage organizations to create, curate, and manage digital collections. The services help organizations create and showcase digital collections on the Web and provide long-term archiving of digital master files. The group also works with librarians and other professionals to develop and evolve new and best practices for digital collections.

The CONTENTdm Digital Collection Management Software supports such standards as Dublin Core, JPEG2000, OAI (Open Archive Initiative), Harvesting, PHP (Hyptertext Preprocessor), and API (Application Programming Interface). The software can handle a full range of primary source material types, including images, photographs, posters, postcards, maps, documents, journals, diaries, theses, dissertations, newspapers, yearbooks, and audio and video files.

Another key component of supporting the full cycle of digital collection curation is digital preservation, whose challenges include an increasing volume of digital materials with limited resources.

To address the need for digital preservation in the library community, OCLC has implemented a Digital Archive Service that provides secure managed storage for digital originals and master files. The ISO-9001 certified service integrates disaster recovery into the library's workflow for building digital collections.

This service provides Automated Monitoring, which includes manifest verification, virus checking, fixity check (digital fingerprinting), and format verification. A user of the archive receives regular reports that include storage use and growth, file types, and a record of an access or disseminations.

The digital collections can reside on a CONTENTdm Server, either installed locally or on an OCLC-hosted server. Digital items can be added from

anywhere using the CONTENTdm Project Client, which can be distributed among staff and collaborating partners, through the Connexion client using Connexion digital import, or through a Web browser using a simple Web form. Collections can also be managed remotely over the Web.

The digital collections can be searched via the Web using standard Web browsers by any number of end users, unless a library elects to restrict access. With its ability to handle any file type, CONTENTdm can serve as an institutional repository to bring collections together into a cohesive and accessible Web-based environment. For organizations seeking opportunities for broader collaboration in developing and sharing collections, the Multi-Site Server is an option for cross-collection searching on multiple servers.

CONTENTdm is scalable, enabling a library to upgrade and increase capacity without new software installations. Organizations with small, unique collections to statewide groups with hundreds of collections have selected CONTENTdm for their projects. While most users run the CONTENTdm software "out of the box," it also has an API that allows for custom development. The open architecture supports extensions, and the Web interface is fully customizable.

CONTENTdm also facilitates discovery of items in special collections through WorldCat. Metadata for collections can be added to WorldCat to make digital items more visible to searchers on the Web through WorldCat.org and WorldCat Local. Those searchers who go to sites such as Google and Yahoo! Search also will discover WorldCat results for library-owned items among their results.

GOING FORWARD

Recent surveys on the impact and direction of digital library efforts have identified some significant trends. Libraries are increasing their efforts to digitize "reborn" paper material in addition to setting up institutional repositories to capture the born digital material published by their patrons such as electronic theses and dissertations. These efforts focus on digitization first to establish the digital collections with planning for digital preservation a future effort to be determined.

The increase in the digital collection activities within a growing number of libraries also supports the evolution of this type of effort from a special project limited by time and resources to an integral part of the libraries' activity. Best practices are emerging that support more open and flexible metadata creation, the use of social networking and tools to leverage group interest, and a strong interest in the aggregation of the digital collection metadata to support cross-collection searching of these rich resources through sophisticated and robust tools. A common objective is to create, contribute to, and use a universal search and access source. Our experience has demonstrated

that every item is special to someone, the aggregate is special to most and the integration with other content is valuable to all.

As digital collection creation and maintenance (curation) becomes mainstream in the library, it is also becoming the focus of more traditional measures of library activity, including assessment and evaluation, both from the collection manager's perspective and the user's point of view, and sustainability, that is, how to move the creation of digital collection resources from a grant-funded, project status into a program with dedicated operating funding.

To respond to these trends, OCLC plans to extend the tools that support digital collection efforts in the library to increase the efficiency of workflow and to reduce the cost of these efforts. Capture and cataloging of digital content will be enhanced through new software tools that support use of terminology services, metadata templates, and batch processing. The CONTENTdm software is fully Unicode compatible, and has extended the support for electronic archival description (EAD) Finding Aids. A new digital collection Gateway service will support collection administrator control of harvesting digital collection metadata into WorldCat. This harvested data will be discoverable through the Open WorldCat.org and the local institutions' view in WorldCat Local.

While there is much work still to be done to create and preserve these valuable digital collections, it is clear that these efforts by the library organizations involved are adding great value to scholarly research and interest at all levels.