

# **DIGITAL ASSET MANAGEMENT SYSTEMS**

**Prepared by Richard W. Boss**

Libraries increasingly seek to offer digital assets as well as print and audio-visual materials to their patrons. These digital assets may include full-text files, PDFs, still images, audios, and videos that have been created electronically or have been converted to machine-readable format for preservation or to provide electronic access that is location independent. There are a number of products available that have been specifically designed to enable organizations to load digital assets, create associated metadata, and provide users with search tools to access the assets.

“Digital asset management” is sometimes called “digital collections management” or “digital media management.” The last is actually a sub-category of the first and is used specifically for the management of audio and video assets. All three terms have sometimes been confused with “electronic resources management.” Digital access management (DAM), the term used in this TechNote, deals with the management of digital files; electronic resources management (ERM) deals with the management of subscriptions to electronic resources supplied by others. ERM is the topic of a separate TechNote.

## **History**

Digital asset management began in the first half of the 1990s as a tool for use by large businesses and organizations. The first systems were developed in-house by large companies with archival requirements. By the middle of the decade, there were commercially available products from companies such as IBM and Content Management Systems, but they were aimed at Fortune 1000 companies that needed to archive a large number of documents.

In 1993, the Luna Imaging, Inc., ([www.lunaimaging.com](http://www.lunaimaging.com)), with financial support from the J. Paul Getty Trust and Eastman Kodak Company, began developing Luna Insight as a digital

asset management product to import, store, manage, and display collections of still images, audios, and videos. The target market was museums, not libraries.

In 1994, North Plains ([www.northplains.com](http://www.northplains.com)) introduced TeleScope, a highly scalable software package that could be used by small organizations as well as large, and by those that wanted to make a wide variety of formats available rather than archiving a single document type. However, there was no support for standards such as MARC. Dublin Core, EAD, and Z39.50.

When the Library of Congress began its American Memory Project (<http://memory.loc.gov/ammem>) and the University of California began development of its California Digital Library ([www.cdlib.org](http://www.cdlib.org)) in 1996, there were still no products that were suitable for libraries.

By 2001 there were some 98 products available and more than 5,000 organizations had implemented digital asset management programs, but there were still no products that targeted the library market. For that reason, the majority of libraries that implemented programs were large academic research libraries that had the financial and human resources to adapt the available products to their needs. Only a year later, OCLC and several vendors of integrated library systems launched development efforts on digital asset management systems suitable for libraries of all sizes and types.

### **Products Designed for Libraries**

Endeavor's ENCompass for Digital Collections, ExLibris' DigiTool, Innovative Interfaces' MetaSource, OCLC's CONTENTdm, Sirsi's Digital Media Archive, and VTLS' VITAL were all in general release by 2005. Hundreds of libraries implemented digital asset management in the next few years. However, by the fourth quarter of 2009, the number of products specifically designed for libraries had gone down because of the merger of Endeavor and ExLibris.

As of 2009, libraries had available ExLibris' DigiTool ([www.exlibrisgroup.com](http://www.exlibrisgroup.com)) Innovative Interfaces' Media Management ([www.iii.com](http://www.iii.com)), OCLC's CONTENTdm ([libservices@oclc.org](mailto:libservices@oclc.org)), SirsiDynix's Hyperion([www.sirsidynix.com](http://www.sirsidynix.com)), and VTLS' VITAL ([www.vtls.com](http://www.vtls.com)). SirsiDynix had embarked on a major rewrite of its Hyperion product and expected a new generation product to become available by mid-2010.

All of these products have the ability to import, catalog, edit, store, search, retrieve content and metadata (data about the data), and generate statistics. They generally do not include hardware and software for digitization. It is assumed that libraries will do that in-house on their own equipment or will contract it out to a firm that specializes in digitization.

### **Importing Documents and Metadata**

The ability to import digital documents is a standard component of all of the products. The documents can be downloaded from a digital storage medium, by FTP, or directly from any scanning device that conforms to the TWAIN standard for interfacing software applications and imaging devices. The documents can be images (GIF, JPEG, PDF or TIFF), plain-text, HTML, audio, or video. Generally only a few parameters are required to load documents. It is also possible to load metadata that already exists for the documents being imported.

As the scale of digital asset management projects has increased, there has been a shift away from the manual loading of records one at a time to bulk loaders that can import thousands or tens-of-thousands of documents with a single set of load commands.

### **Cataloging**

There are several cataloging options, including MARC, EAD, TEI, and Dublin Core, but the most widely used is Dublin Core. Several products offer the option of cutting and pasting

information from the actual document to facilitate cataloging. Most of the systems do not limit the number of fields of metadata that can be created. At a minimum, the metadata will include the name of the author, title, language, the source, date, condition, access restrictions, and content description. The metadata can be linked to a single document, also known as an object, or to many.

After the documents have been cataloged, virtual collections can be created. A single document can appear in several virtual collections.

## **Editing**

Digital content can be edited, including reducing high-resolution JPEG or TIFF documents to lower resolutions or creating thumbnails. Metadata can not only be created, but edited on a digital content management system.

## **Storage**

Storing digital files demands substantial computer resources. For example, a collection of 250,000 documents accessible to 40 concurrent users requires at least 100 GB of disk storage and a multi-processor server rated at 2.3 GHz or better, and 8.0 GB of RAM. If more than 25 percent of the documents are images, audio, and video, an even more robust hardware platform is required. Libraries that want to avoid investing in server hardware and software can utilize the hosting service of any of the vendors that focus on digital asset management for libraries.

## **Searching**

Searching can be done directly against the digital asset management system or through an interface from a library's patron access catalog, portal, or Web server. A patron access catalog or Web server would be only a gateway to the digital asset management system, but a portal could be used to simultaneously search the documents on the digital asset management system, but also

against other electronic resources within the library or on the Web. Because they have been developed for libraries, the systems support Z39.50 in addition to XML and HTTP gateways.

Searching may be limited to specific documents, media types, individual collections, multiple collections, or may encompass all collections.

Searching can be done either against both the full-text of the documents and the associated metadata, or against just the metadata. Non-text documents searching is, of course, limited to searching the metadata. In most cases, searches can be natural language, Boolean, proximity, wildcard, truncation, date and date range, and pattern. Pattern searching makes it possible to find words with incorrect or missing characters either in the search term or the body of the text.

Most of the system return merged result sets. Some also return relevancy-ranked results.

All of the systems include an image viewer for JPEG and TIFF files, enabling image manipulation and printing.

Increasingly, the products allow data harvesting via OAI-PMH, the Open Archives Initiative Protocol for Metadata Harvesting.

Security generally includes patron authentication to control access to the documents based on the profile of the documents and the users. For example, some documents may be subject to copyright, access restrictions required by donors, or the desire of a library to charge for printing or downloading. Some categories of users may not qualify for access because they are not registered borrowers, are not adults, or are delinquent borrowers.

## **Statistics**

All of the systems include a number of reports that summarize loading, cataloging, editing, and searching activity.

## **ILS Interface**

Most of the products cannot only function as a standalone system, but can also be interfaced with an integrated library system.

## **Standards**

There are a number of standards relevant to digital asset management. The following are described on the Library of Congress Web site ([www.loc.gov/standards](http://www.loc.gov/standards))

- [MARC 21](#) - Representation and communication of descriptive metadata about information items
- [MARCXML](#) - MARC 21 data in an XML structure
- [MODS \(Metadata Object Description Standard\)](#) - XML markup for selected metadata from existing MARC 21 records as well as original resource description
- [MADS \(Metadata Authority Description Standard\)](#) - XML markup for selected authority data from MARC21 records as well as original authority data
- [EAD \(Encoded Archival Description\)](#) - XML markup designed for encoding finding aids
- [METS \(Metadata Encoding & Transmission Standard\)](#) - Structure for encoding descriptive, administrative, and structural metadata
- [MIX \(NISO Metadata for Images in XML\)](#) - XML schema for encoding technical data elements required to manage digital image collections
- [PREMIS \(Preservation Metadata\)](#) - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials
- [TextMD \(Technical Metadata for Text\)](#) - XML schema that details technical metadata for text-based digital objects

- [ISO/DIS 25577](#) - Information and documentation -- MarcXchange
- [ISO/DIS 20775](#) - Schema for Holdings Information
- [ALTO \(Analyzed Layout and Text Object\)](#) – Schema for word positions and physical layout for metadata
- [Z39.50](#) – Supports information retrieval among different information systems
- [SRU/SRW \(Search and Retrieve URL/Web Service\)](#) - Web services for search and retrieval based on Z39.50 semantics
- [CQL \(Contextual Query Language\)](#) - Formal, user-friendly query language for use between information retrieval systems

Other relevant standards include XML, MPEG2, JPEG2000, TEI, Dublin Core, OAI-PMH, and PREMIS, Information about any of these standards can be found on the Web.

## **Costs**

The cost of digital asset management systems is related both to the number of documents and number of concurrent users supported. An entry-level system supporting up to 10,000 documents and five concurrent users can cost as little as \$15,000 for hardware and software, and as little as \$4,000 a year in annual support. Libraries that require a very small system may wish to consider a hosted service. Most of the vendors mentioned in this TechNote will mount the documents on a hosted server at the vendor's site for an annual subscription fee of as little as \$2,500.

As the size of the system increases, it becomes more cost effective to purchase the system. A system supporting 250,000 documents and 40 concurrent users will cost a minimum of \$70,000, plus approximately \$11,000 per year.

## **Sources of Information**

Currency is extremely important when seeking information about digital content

management. The Web sites of the vendors cited in this TechNote are highly current. Also highly current is *D-Lib Magazine* ([www.dlib.org](http://www.dlib.org)). It characterizes itself as the “magazine of digital library research.” The Digital Library Federation ([www.diglib.org](http://www.diglib.org)), a consortium of “libraries and related agencies that are pioneering in the use of electronic information technologies to extend collection and services” also has highly current information on its Web site. Another useful source is the Digital Libraries Initiative (<http://dli.grainger.uiuc.edu/national/html>), a group that is developing standards and best practices.

A somewhat older document that is still useful is the National Information Standards Organization’s ([www.niso.org](http://www.niso.org)) 2004 “A Framework of Guidance for Building Good Digital Collections.”

The largest digital asset management program undertaken by a library is that of the Library of Congress (<http://www.loc.gov/library/libarch-digital.html>). Two good examples of digital collections maintained by public libraries are the image collections of the Chicago Public Library (<http://www.chipublib.org/index.php>) and the image, text, audio and video collections of New York Public Library ([www.nypl.org/digital](http://www.nypl.org/digital)).

Final revision, October 28, 2009