CONTENTdm and Content Pro: A Comparison and Evaluation

Elizabeth Byrne
San Jose State University, elizabeth.byrne7@gmail.com

Follow this and additional works at: http://digitalcommons.unl.edu/libphilprac
Part of the Library and Information Science Commons
CONTENTdm and Content Pro: A Comparison and Evaluation

Elizabeth J. Byrne

San Jose State University
Abstract

This paper examines two Digital Asset Management Systems (DAMS) to determine their suitability to provide access to digital photographic surrogates from a library’s archival photographic collection. The two DAMS examined are OCLC’s CONTENTdm (CDM) and Innovative Interfaces Inc.’s (III) Content Pro (CP). While both have been widely adopted, a literature review reveals that very little has been written about CP compared to CDM. Both use metadata schemes that are interoperable, reusable, and extensible and it is determined that both CDM and CP’s DAMS’ attributes recommend them. However, further work is needed to select one over the other. This includes consulting with institutions who have adopted each for further information, evaluating costs, and consulting with in-house technical services support.
Introduction

The purpose of this paper is to examine OCLC’s CONTENTdm (CDM) and Innovative Interfaces Inc.’s (III) Content Pro (CP) to determine which would be more suitable for a library’s archival collection’s digital photograph surrogates. The first issue is to determine which factors contribute to a good digital collection in a library. Next is to examine what general features would be most beneficial in a library’s Digital Asset Management System (DAMS). Both CDM and CP are widely adopted in libraries. Their vendors and the products are examined as well as how they are being used.

Literature Review

A literature review was undertaken to determine what has been written about DAMS selection, about the vendors III and OCLC, who is using CDM and CP, and how they are using them.

There are a few good resources on the topic of materials selection criteria. First is NISO’s A Framework of Guidance for Building Good Digital Collections (2007). It thoroughly covers the topic and emphasizes such necessary elements as workability for the user, interoperability, and sustainability. Another source of information is Mauthe and Thomas’ (2004) Professional Content Management Systems which has a chapter on content management system (CMS) hardware and software requirements. Krogh’s (2006) The Dam Book: Digital Asset Management for Photographers also has an excellent chapter on digital asset software and a helpful section on its evaluation.

Further, with respect to evaluating the metadata side of the equation, Zeng and Qin’s (2008) Metadata has a section that covers that topic. Also, a summary of the most commonly
used metadata schemes can be found in Park and Tosaka’s (2010) *Metadata Creation and Practices in Digital Repositories and Collections*.

Information about the vendors OCLC and III is mostly gleaned from their respective websites. OCLC’s documentation about their products is thorough and informs on features, collection building and management, end user experience, and system requirements, as well as who uses the products. On the other hand, III’s website is sparse, limited to overview pages of their products. It was necessary to search elsewhere to discover who is using III products as well.

There are some resources that show what organizations are using CDM and CP. On OCLC’s website there are direct links to CDM repositories. Also, Saginaw Valley State’s Zhanow Library has a page that lists CP and CDM libraries. Further, III has a press release which lists the first six libraries that are going live with CP IRX.

With respect to how CDM and CP is being used, Breeding (2010), who maintains the Library Technology Guides, has a paper discussing trends in integrated library systems, including OCLC and III. There is a discussion about a project of digital surrogates of three-dimensional objects at Oklahoma State University. This project uses CDM. Another CDM project, described by Higgins (2012), is one at San Jose State University. This paper is helpful because it describes the metadata scheme, image creation, cataloguing and further steps.

Martin (2011), in *Marrying Local Metadata*, describes how the library at University of Illinois at Chicago used metadata in CDM for a data dictionary in order to apply metadata consistently within digital collections. Unique metadata fields are the topic of discussion in Han, Cho, Cole and Jackson’s (2009) report which examines 21 digital collections hosted on CDM. They look at metadata mapping issues and interoperability. Also, OCLC’s *Best Practices for CONTENTdm and Other OAI-PMH Compliant Repositories* explain how metadata is to be
shared among repositories. The final paper by Jackson (2012) examines an archival digital photographic collection. They found that users preferred searching CDM’s interface.

The interesting thing that this literature review uncovered is that while more has been written about CDM, both CDM and CP have been adopted by many institutions.

**Discussion**

The Collection

The collection that the DAMS will organize is an incomplete set of digital surrogates of black and white and color photographs and slides as well as historic employee magazines and other archival documents. The archival documents relate to company history but do not include company records which are part of the company’s records management function. The whole archival collection is very large, including more than 100,000 images. Some images have already been digitized as they were either identified as commonly-requested items or had been requested for specific projects and digital surrogates made. However, the digital assets are stored on shared drives, personal computers, or on removable storage devices like CD-ROMs, memory sticks, and separate hard drives stored in various locations. So they are not being curated effectively and they are not accessible to most users, but rather require library staff to mediate discovery.

Currently, the library used III’s Millennium for its integrated library system (ILS). The enterprise that the library is part of has two main content management systems. One is IBM’s FileNet and the other is Microsoft’s SharePoint. SharePoint is meant to be used for document management including collaboration and workflows while FileNet has been used for more records management purposes such as serving as a repository for finalized documents. It is the author’s impression that these two repositories are the enterprise’s preferred repositories. Other data and document projects have been required to use these content management systems as they
are the enterprise’s sanctioned content management systems. However, the library could
legitimately make a business case for using an out-of-the-box digital asset management system
like CP or CDM if one of them can be shown to fulfill the library’s requirements and provide a
return on investment for the enterprise.

Criteria for Selecting a DAMS

Firstly, what is a good digital collection? NISO (2007) says that a good digital collection
“consists of objects that are selected and organized to facilitate their discovery, access, and use”
(p. 4). The collection principles they enumerate include that a collection “is curated” and “is
broadly available and avoids unnecessary impediments to use,” is interoperable and sustainable,
and “integrates into the users own workflow” (p. 4). When talking about the digital objects
themselves, NISO (2007) says that “a good object is exchangeable across platforms, broadly
accessible, and formatted according to a recognized standard” (p. 26).

Data and Metadata

The data will be stored either on SharePoint, FileNet, independent server, or a virtual
server. Mauthe and Thomas (2004) say that there is a “plethora of systems, applications, and
application components that interact with a CMS” (content management system) (p. 225). They
go on to talk about the necessity of system integration: that “data has to be shared across system
boundaries” (p. 227). In order for that to work, data and metadata has to be interoperable. Krogh
(2006) points out these points to remember when selecting a system: “you will probably not use
it forever”, “Make sure the software makes sense to you” and “test, test, test” (p. 199). So it is
important to use data in a format that can be migrated to the next version, a DAMS that is usable,
and metadata that is interoperable.
Zeng and Qin (2008) say that “the quality of metadata and its usefulness in retrieval form the two most important areas in metadata evaluation” (p. 263). They also say that good metadata is designed for “interoperability, extensibility, reusability, and simplicity” (p. 266). It is important that CP and CDM use an established metadata scheme so that the metadata will be interoperable and can be migrated. Park and Tosaka’s (2010) research showed that MARC coding and AACR2 used together are the most commonly used schemas and that DC was “the second most widely employed” (p. 114). CP uses Dublin Core (DC) metadata (Innovative, Content Pro, 2013) and CDM supports DC along with other metadata schemes.

The Vendors

OCLC is a library cooperative based on Dublin, Ohio, and its commitment is to “work together to improve access to the information help in libraries around the globe, and find ways to reduce costs for libraries through collaboration” (OCLC, Public Purpose, 2013). OCLC’s website has a lot of information about their services and products. OCLC (CONTENTdm Overview, 2013) says it is “a single software solution that handles the storage, management and delivery of your library’s digital collections to the Web. Further, digital items can be prepared in large batches using a windows-based, digital collection tool, a server for storage and editing, and a web-based discovery interface which can be customized OCLC (CONTENTdm Overview, 2013).

They provide a lot of information on their website to demonstrate how their product works. With respect to metadata and interoperability, they say: “OCLC’s adherence to commonly accepted standards allows CONTENTdm to be open and extensible, as well as provide functionality that meets a wide range of needs. CONTENTdm supports numerous industry standards including Unicode, Z39.50, Qualified Dublin Core, VRA, XML, JPEG2000,
OAI-PMH and METS/ALTO” (OCLC, *CONTENTdm: Overview, Features*, 2013). Some of the other attributes on that website include that CDM is scalable to fit the size of your current collection and can be extensible to allow for growth. Interoperability includes being able to import data from an existing DAMS using XML. They also say that workflow can be very flexible because it is web-based. They say that “digital items can be added from anywhere using the CONTENTdm Project Client, through the OCLC Connexion client or through a Web browser using a simple Web form” (OCLC, *CONTENTdm: Overview, Features*, 2013).

Innovative Interfaces Inc. (III) is based in Emeryville, California, and is known as a library software vendor. Their ILS Millennium and its successor Sierra are the products they are best known for. Content Pro and Content Pro IRX are their DAMS programs and Encore is a discovery layer which “gives Encore new function for the discovery of article-level content alongside local resources” (Breeding, 2010, p. 43). Further, Encore “takes advantage of the service-oriented platform to delivery operational and usage data for display and analysis” (Breeding, 2010, p. 43).

Who is using CDM and CP

OCLC’s website says that more than 2,000 organizations use CDM. They have direct links from their website to organizations’ CDM collections. The organizations include academic libraries, consortia, state libraries, public libraries, historical associations, and museums. It’s hard to determine who is using III’s CP. Also, there are not any periodical articles about CP projects or experiences that user have had. An examination of two repositories, one using CP and one using CDM, shows that their interfaces look similar to the end user. Appendix A shows a CDM search and the resulting screens and Appendix B shows a CP search with resulting screens. Both have thumbnails and have similar descriptive metadata which displays with the image.
How CDM and CP Are Being Used

Breeding (2010) says that:

New models of automation are beginning to take shape, challenging the traditional integrated library system (ILS) in an industry that has long favored incremental evolution. Some new products and projects remain poised to break free from traditional models, particularly discovery products that can deliver immediate improvement to library users. (p. 22)

He says that both OCLC and III are working to increase searchability of serials and this is accomplished through both of their discovery products. While it is interesting to look ahead to a time when metadata may be automatically harvested and semantics applied to it, we are not at that place yet.

On a University of Albany Wiki, Brown (2012) has provided a commercial system vendor comparison chart in a Microsoft Excel. It includes both CDM and CP and compares such things as metadata, data storage, system requirements and so on. The information about CP is sparser than for CDM. However, one thing that is pointed out is that for both, the data can be hosted on a local server or remotely by the vendor and possibly virtually. Also, Brown confirms that both CDM and CP are OAI-PMH compatible which would help with preservation of the data and interoperability.

There are many papers about implementations of CDM. Valentino and Shults (2012) have written an interesting paper describing how the University of Oklahoma Library has created of a digital collection of three-dimensional objects. CDM is capable to using a “picture cube feature to create the digital representation of the three-dimensional object” (p. 208) in six-photograph sets. In addition, Higgins (2012) describes cataloging images in a CDM digital collection at Martin Luther King, Jr. Library at San Jose State University. Higgins says that the library uses CDM “to describe, catalog, and publish collections of digitized and born-digital
items” and further that CDM serves as “a complete content management system for digital collections” (p. 7). The author says that the server is hosted and maintained by the library.

At the University of Illinois at Chicago, Martin (2011) writes about how the University consolidated its metadata into one data dictionary in CDM (p. 33). The university had done what many institutions have which was to create separate digital collections with different metadata. They used CDM to bring the collections together and unify the metadata using DC. The author points to the research of Han, Cho, Cole and Jackson (2009) in which it is pointed out that customization that CDM affords can be problematic. The customization of metadata could hinder harvesting in an OAI-PMH environment. They say that “locally defined unique fields, which potentially have substantial contextual information, could impede the interoperability of metadata since the contextual information in such fields as implemented could not be or was not being mapped to Dublin Core elements in ways to facilitate interoperability” (Han, Cho, Cole and Jackson, 2009, p. 216). Their recommendations include six ways that CDM users and curators can maintain metadata interoperability in spite of of local practices to enhance DC fields.

In a usability study of digital content finding aids, Jackson (2012) examines a CDM digital collection and how it is used. At the University of North Carolina’s Southern Historical Collection, rather than adding metadata on the items, which would have been time-consuming and therefore more expensive, it was deemed that finding aids could add the information needed to help users. The online finding aids would “mirror the researcher experience in the reading room and be both cost-effective and efficient” (p. 24). “In this way, contextual and hierarchical information would be available to the researcher without requiring extra metadata input or additional curation” (Jackson, 2012, p. 24). One thing Jackson (2012) points out is that in this
study, all users experienced “usability issues with CDM, particularly with the Advanced Search function and distinguishing which collections were being search (p. 68).

**Conclusion and Recommendations**

This research has shown that both CDM and CP have similar attributes as DAMS. They are both being used by academic libraries, public libraries, and historical associations. The intended collection is in a special library and is a historical collection so the types of implementations match the subject collection. Both CDM and CP use metadata based on DC which can be customized as needed. The collection that these products are being reviewed for does not share its records but, at the same time, metadata should be OAI-PMH compliant to allow for in-house discovery and preservation. DC metadata allows for this.

Research has shown that there is not very much information written about CP either on III’s website or by other authors in journal articles, books, or online. However, according to III and OCLC, both CP has been adopted by over 1,500 organizations and CDM by over 2,000. A review of secondary literature would not be sufficient to choose or rule out III’s CP because there is just not enough information available. It would be recommended that consultations with organizations with similarly-sized collections and similar goals for the digital collections could be undertaken to discover more about their experiences using CP. At the same time, it would be recommended to consult with a few CDM users to discuss their experiences for comparison purposes. There may be leverage to adopt CP because it is from the same vendor as the library’s current ILS.

Finally, consultation with IT support is required to determine which type of implementation would be permissible according to the enterprise’s IT policies. Questions such as where the data will be stored and what type of server is permissible would be paramount to
deciding whether to adopt CP or CDM. A virtual server, while convenient, may not be permitted, for example. Also, the purchase price for each DAMS together with annual service costs and support costs would have to be determined, compared, and budgeted for. Research is the first step towards selecting a DAMS for the library’s digital photograph surrogates. The next step will be to consult with CP and CDM users to discuss usability, consult with the vendors on costs and implementation recommendations, and finally to write a business case to secure budget funding.
References


Appendix A

Here are search screen visuals for CDM. The first image is the initial search result at University of Washington University Library website of its CONTENTdm digital collections: http://content.lib.washington.edu/index.html

Sample of search result for “Mask”.

Sample of details for record #2 selected above. The digital image is above and the descriptive metadata is displayed below.
American Indians of the Pacific Northwest Collection

Title
Tsimsian carved face mask, British Columbia, 1912
Photographer
Emmons, George Thornton
Date
1912
Notes
Carved mask of wide-mouthed face; seven 'spokes' extend from the face, with hunks of hair extending from the spokes. Displayed on wooden platform next to exterior wall.
Subjects
Still life photographs—British Columbia; Wood carvings—British Columbia; Tsimsian Indians—Arts & crafts; Masks
Location Depicted
Canada—British Columbia
Object Type
Photographs
Negative Number
NA3403
Digital Collection
American Indians of the Pacific Northwest Images
Collection
George Thornton Emmons Collection no. 131
Repository
University of Washington Libraries. Special Collections Division
Ordering Information
To order a reproduction, inquire about permissions, or for information about prices see:
http://www.lib.washington.edu/specialcollections/services/reproduction-info
Restrictions
http://content.lib.washington.edu/aipnw/copyrights.html
Transmission Data
Image/JPEG
Appendix B

Here are search screen visuals for CP. The first image is the initial search result at Arlington Public Library’s website (Arlington, Virginia) of its Content Pro digital collections:

Sample of search result for “airplane”. The display shows a thumbnail with the photo obscured partially by the title.

Then, when the image is clicked on there is a larger image displayed with the descriptive metadata below.
"Orville Wright and his World Record-breaking Aeroplane, Fort Myer, Va."

<table>
<thead>
<tr>
<th>Title</th>
<th>&quot;Orville Wright and his World Record-breaking Aeroplane, Fort Myer, Va.&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher</td>
<td>B.S. Reynolds Co.</td>
</tr>
<tr>
<td>Subject</td>
<td>Wright, Orville</td>
</tr>
<tr>
<td></td>
<td>Airplanes</td>
</tr>
<tr>
<td></td>
<td>Fort Myer</td>
</tr>
<tr>
<td>Description</td>
<td>Image of Orville Wright flying his airplane at Fort Myer, Virginia.</td>
</tr>
<tr>
<td>Contributor</td>
<td>Dorothy Daniel</td>
</tr>
<tr>
<td>Identifier</td>
<td>Airplane postcard</td>
</tr>
<tr>
<td>Source</td>
<td>Dorothy Daniel Postcard Collection</td>
</tr>
<tr>
<td>Type</td>
<td>Image</td>
</tr>
<tr>
<td>Format</td>
<td>Postcard</td>
</tr>
<tr>
<td></td>
<td>black and white</td>
</tr>
<tr>
<td></td>
<td>3 1/2 in. by 3 1/2 in.</td>
</tr>
<tr>
<td>Spatial</td>
<td>Arlington County (Va.)</td>
</tr>
<tr>
<td>Temporal</td>
<td>1900s</td>
</tr>
<tr>
<td></td>
<td>1910s</td>
</tr>
<tr>
<td>Rights</td>
<td>This image is subject to copyright. Unauthorized use of the images in the Local History Photograph Collections of the Arlington Community Archives is prohibited</td>
</tr>
<tr>
<td>Rights Holder</td>
<td>Virginia Room, Arlington Public Library</td>
</tr>
</tbody>
</table>