Digitizing special collections using the CONTENTdm software suite

Trevor Bond
Manuscripts Archives and Special Collections
Washington State University Libraries

Alan Cornish
Systems unit
Washington State University Libraries

Introduction

The Washington State University Libraries offer an array of online image collections to searchers via the Internet. This article describes the CONTENTdm software suite, a tool that has been used by the Libraries since 2000 to enhance its online collection building efforts, and provides a brief overview of collection building at WSU (1). CONTENTdm has several strengths that will be reviewed in this article, including its support for distributed collection building and multiple search clients and the ability to integrate CONTENTdm with other imaging products.

Architecture of CONTENTdm

Figure 1 illustrates the collection building model using the CONTENTdm software suite. The Server software can be run on Windows NT, Linux, and Sun Solaris platforms. The CONTENTdm Server stores objects (images or other multimedia files) in defined collections; using a Web server package such as Microsoft Internet Information Server on the NT platform as a supporting application, the Server makes CONTENTdm collections accessible to users over the Internet.

The Acquisition Station software is a key component of CONTENTdm. Using the Acquisition Station software, collection builders can add objects to a collection either individually or in batches. A Collection Administration software component is accessible within the Acquisition Station software. It enables a collection builder to perform the following tasks:

- Add new items. This allows objects submitted from the Acquisition Station software to be reviewed and added to the server.
- Build text index. This command builds the database index so that any changes made to a collection (such as adding or removing images, making changes in metadata) are reflected in user searches.
- View/edit collection field properties. This feature allows collection fields to be added and deleted; allows the administrator to set whether or not a field is searchable and whether or not a field appears in records.
- Global Search and Replace. This enables the database administrator to globally edit metadata at the collection level.
- Export Metadata. This command takes all of the metadata describing objects in a database and exports the information in an XML or text file - an important feature when planning for the sustainability of a digital project.
• Start the CONTENTdm Document Editor. This feature allows for the creation of compound documents (such as scrapbooks or books), post card views (provides access to the front and back of an image), and picture cubes (a multi sided cube that may be rotated).

Figure 2 provides a screen shot of the View/edit collection field properties screen from a WSU Libraries CONTENTdm collection, the Frank S. Matsura image collection. CONTENTdm uses the 15 element Dublin Core description set as a default template. However, custom field names, which can be discipline specific or collection specific, may be designated for collection fields (as shown by the Negative number, Studio location, and Photographer fields in the Matsura image collection). The custom names will appear for searches performed in a single CONTENTdm collection, while the Dublin Core field names will appear for searches across multiple CONTENTdm collections, such as a search across all collections served by an institution. Fields may be repeated and fields not mapped to the Dublin Core template may also be added for local administrative needs.

Referring again to Figure 1: A CONTENTdm license includes one copy of the Server software and multiple (up to 50) copies of the Acquisition Station software. This architecture enables distributed collection building. The Acquisition Station software, along with access to the Collection Administration software, can be established on any Windows machine with access to the Internet (2). For academic libraries using this approach, a faculty member working from his or her workstation can perform routine collection building tasks such as adding new images to a collection and editing metadata without contacting the library that runs the CONTENTdm server. In this scenario, the library's involvement in online collection building is in substantive areas, such as administering the CONTENTdm server, installing the Acquisition Station software on the faculty member's workstation, and in assisting the faculty member with questions on the CONTENTdm software and metadata structure. Truly collaborative digital projects may be undertaken across campus or around the country using CONTENTdm.

Providing access to CONTENTdm collections

Multiple search clients can be used to provide access to CONTENTdm collections. Below are URLs for the three supported clients:

Contextual search client
http://www.wsulibs.wsu.edu/holland/masc/xmatsura.html
(WSU Libraries, Manuscripts, Archives, and Special Collections: Frank S. Matsura Image Collection)

HTML search client
http://content.wsulibs.wsu.edu/cgi-bin/go.exe?CISOROOT=/avery
(WSU Libraries, Manuscripts, Archives, and Special Collections: Frank Fuller Avery Image Collection)

Java search client
http://content.wsulibs.wsu.edu/javaclient/clipping.html
(WSU Libraries, Pacific Northwest Clippings Collection)

Each search client has specific strengths giving libraries flexibility in establishing front-ends to their online collections. The contextual client is created through a Query Builder tool that generates HTML coding which can be pasted into custom designed HTML documents. There are three choices when using the Query Builder tool: generating a simple search box that will search any or all fields in the database, building a drop-down box containing multiple searches any of which the user may select, or creating a
single predefined query. The same tool may be applied to multiple image collections. We have employed these predefined searches to create links from icons, such as an image of a self portrait of Frank Matsura, to a search of Matsura’s self portraits, thereby creating an intuitive database interface. This is also an effective way to integrate the images in a collection with historical information. In several digital collections (the Frank Fuller Avery and the Early Washington Maps), we have included a historical time line with embedded predefined searches to relevant objects in the database. So for example from the Avery collection timeline, “1905: McLaughlin Agreement,” affirms the U.S. Government's $1.5 million payment for the cession of the northern half of the reservation (1891-2). If a user clicks on “McLaughlin Agreement,” two portraits by Avery from the signing ceremony will appear (http://www.wsulibs.wsu.edu/holland/masc/xaverytime.html).

The Query Builder tool also provides the option to select the display templates for the initial results list (a series of one or more thumbnails), as well as the full size image screen that appears when one of the results is selected. In the metadata that appears beneath the full-size image, CONTENTdm provides the option to have the text in any field of the database appear as hyperlinks. If a user clicks on one of these highlighted terms, a new search will be initiated. Links may also be provided in the metadata to take users to other resources related to the object, such as a finding aid to an archival collection. The contextual client therefore enables access to an online collection that is then integrated into web pages that have the same look and feel as an existing library or departmental page. In addition, multiple web pages can be created for access to a single collection, each customized for a specific audience (for example, a collection front-end can be created for K-6 students). The HTML and Java search clients provide search engine access to CONTENTdm collections. The Java search client provides a Workbox feature that enables HTML documents containing selected objects from a collection to be generated on the fly (3). With the HTML and Java clients, users may also browse all of the vocabulary used in a given database and double click on any term to have it appear as a search option.

Integrating CONTENTdm with other imaging products

In building an imaging program at the WSU Libraries, we have selected the CONTENTdm software suite as the primary supporting product. However, additional software tools have been employed in order to enhance CONTENTdm collections, most significantly MrSID imaging tools for the development of an online map collection (the Early Washington Maps collection, which can be accessed at URL http://www.wsulibs.wsu.edu/holland/masc/xmaps.html) (4).

Importantly, CONTENTdm's Acquisition Station software 3.1 supports the MrSID format, along with TIFF, as a full-resolution image format. A collection builder can import a MrSID image file into the Acquisition Station and the software will automatically generate two images: a best-fit JPEG compression service image, which serves as the base (full-screen) image in a CONTENTdm collection, and a JPEG thumbnail image. The result in a manageable workflow for collection building and quality MrSID and JPEG map images.

To enhance the online map collection, the WSU Libraries have employed LizardTech's MrSID Image Server to deliver dynamic, clickable full-resolution images (5). To create collection items, maps in MrSID format are linked to the MrSID Image Server product. A unique URL is available for linking to a map on the Image Server. This URL is entered into a CONTENTdm metadata field; in the case of the Early Washington Maps collection, the Full Resolution field stores the URL and provides the bridge between CONTENTdm and the MrSID Image Server. Importantly, both the CONTENTdm contextual client and the MrSID Image Server allow librarians to design custom HTML template documents. As a result, Early Washington Maps collection searchers will access web pages with the same design (look and
feel) whether they are reviewing the detailed metadata stored in CONTENTdm or the dynamic map images served by MrSID Image Server.

Finally, to improve access to CONTENTdm collections, the WSU Libraries uses the Innovative Interfaces (III) Imaging and Document Management product to create links between III INNOPAC MARC catalog records and online collections. As shown in Figure 3, a detailed catalog record has been created for the Frank Fuller Avery Image Collection in the WSU Libraries INNOPAC integrated library system. The III Imaging product allows a thumbnail image to be embedded in the catalog record; a searcher clicking on the thumbnail will be taken to the external URL for the online collection. The online collection is presented in a web browser frame so that it is easy for searchers to return to the INNOPAC catalog after working in the online collection. The III Imaging product is also employed to embed images of historical maps to their corresponding MARC records. Using the custom map thumbnail images created by CONTENTdm, these icons placed in the MARC record take users to the high-resolution MrSID files where they may be studied in detail. Once at the MrSID site for the Early Washington Maps project, users, if they feel so inclined, have the option to return to the maps database to view similar objects thereby entering the Early Washington Maps CONTENTdm site through a back door.

Collection building at the WSU Libraries

The initial impetus for purchasing CONTENTdm was the desire to provide online access to images drawn from the Historical Photograph Collections in Manuscripts, Archives, and Special Collections (MASC). The first digitization project in MASC started in 1999 prior to the purchase of CONTENTdm and resulted in the WSU Buildings Image Collection. In 2000, additional larger digitizing projects began in MASC with the receipt of a series of Library Service and Technology (LSTA) grants distributed by the State Library of Washington, which helped the WSU Libraries purchase servers and software and provided much needed funds for temporary employees. This allowed us to put together project teams with scanning technicians and graduate students specializing in public history as photograph researchers. The first LSTA pilot grant came as part of a state wide Digital Imaging Initiative for $17,000, with which the Libraries purchased a server and a CONTENTdm license and created the Frank Matsura Image Collection a database of 1,600 images and a 72-page scrapbook (http://www.wsulibs.wsu.edu/holland/masc/xmatsura.html). As part of the pilot project to test the feasibility of using CONTENTdm as the tool for regional collaborative image database building, the WSU Libraries also agreed to provide technical support and house collections on a local server built by the Ellensburg Public Library and Gonzaga University. In the summer of 2000, WSU MASC and the Map Collection and Cartographic Information Services department at the University of Washington (UW) Libraries received an LSTA grant, $49,733, for a project to scan early maps of Washington and together create a searchable database. Using a variety of methods for scanning (UW out-sourced their maps to a Seattle area vendor for scanning while WSU MASC performed in-house scanning), we built the database with UW adding images and metadata in Seattle to the CONTENTdm server located in Pullman at Washington State University. Current projects in MASC include adding an additional 150 maps to the Early Washington Maps project, providing full text, virtual use, copies of early Pullman directories and other WSU documents, and maintaining a partnership with the WSU Marketing Communications to house their digital photograph file in a CONTENTdm database.

Under the terms of the partnership agreement between WSU MASC and WSU Marketing Communications, the Libraries will provide CONTENTdm workstations, training, metadata support, and server space for the database. WSU Marketing Communications is adding to the database digital images (both scans and photographs from digital cameras) taken by the campus photographers. WSU Marketing Communications is also entering metadata and designing a contextual client for the collection in consultation with the Libraries. After the images are removed from the active digital photograph, they
will be transferred to MASC. Over the last thirty years, photographers from WSU Marketing Communications has donated their visual materials to MASC and these gifts form the core of MASC’s extensive campus visual collection. This project also represents the first time MASC will actively support and maintain digitally born collections.

Conclusion

The ultimate goal of these digital projects is to create powerful tools for librarians to provide improved reference service, while at the same time enhancing access to remote users. The assortment of imaging products described in this article allow the librarians at WSU to connect information in the Libraries' OPAC with image databases and collection finding aids hosted on the home page of Manuscripts, Archives, and Special Collections. Within the image databases such as the Early Washington Maps project, links to MrSID files and their corresponding image viewers allow users to access true digital surrogates where the finest details of cartographic information may be studied. As Teresa Grose Beamsley of the Henry Ford Museum noted,

The worth of institutional assets is no longer gauged by looking at the collections inventory appraisal. It is now redefined as the combination of the physical materials in the collections, the surrogates that satisfy a growing demand for visual information about them, and the text-based information that establishes their context and serves as the key to locating them (6).

This statement underscores the fact that the delivery of well-described online collections should be an important goal for libraries. The CONTENTdm software suite has served as the anchor in the construction of these collections at the Washington State University Libraries. While the emphasis thus far has been on building image collections (photographs, maps, and text), the software is capable of distributing collections in other media formats, such as sound and video. The distributed collection building model has enabled the Libraries to explore collaborative collection building efforts.
Notes

1. The CONTENTdm software suite is produced by Digital Media Management, Inc. of Seattle, Washington [URL http://contentdm.com/].

2. The CONTENTdm Acquisition Station 3.1 client runs on Windows 98, Me, NT, and 2000 computers. See URL http://contentdm.com/requirements.html for additional information.


4. MrSID imaging tools are produced by LizardTech of Seattle, Washington [URL http://www.lizardtech.com/].

5. LizardTech no longer supports the MrSID Image Server. Instead, the company now markets a more robust commercial product, Content Server, which provides the ability to view and manipulate MrSID files and additional features, such as administrative control over the printing and saving of image files from the server.

The CONTENTdm Acquisition Station software, with the Collection Administration tool software component displayed on the screen.

Figure 1: The CONTENTdm collection building model
Figure 2: The View/edit collection field properties screen, accessible through the CONTENT dm Collection Administration software.
Figure 3: Linkage to a CONTENTdm collection using III’s Imaging and Document Management Product