

Online Northwest 2001

EIGHTEENTH
ANNUAL CONFERENCE

EXECUTIVE SUMMARIES

Sponsored by the

INTERINSTITUTIONAL LIBRARY COUNCIL
ONLINE COORDINATORS COMMITTEE
Of The
OREGON UNIVERSITY SYSTEM

Kyle Banerjee, OSU
Colleen Bell, UO
Susana Flodin, OIT
Theresa Gillis, EOU

Art Hendricks, PSU
Janeanne Rockwell-Kincanon, WOU
Steve Teich, OHSU
Dale Vidmar, SOU

CONTENT: A Model for Collaborative Database Building

Trevor James Bond, Special Collections Librarian
Manuscripts, Archives, and Special Collections, Washington State University Libraries
Pullman, Washington

Alan Cornish
Systems Librarian, Washington State University Libraries
Pullman, Washington

The Washington State University Libraries have used the CONTENT multimedia software suite as a tool in serving an array of image collections to users via the Internet. Several issues relating to the use of the CONTENT software are described in this summary document. Additional information on the use of CONTENT at the WSU Libraries can be obtained at URL <http://www.wsu.edu/~cornish/content/info.html>.

An important strength of the CONTENT software is its ability to support the remote building and administration of databases. The CONTENT Server stores database images and other objects (such as sound and video files) within defined collections and makes the collections available over the Internet. The CONTENT Acquisition Station software can be used to add objects to a collection, to edit or delete existing collection objects, to add, modify, and delete collection fields, and to perform other administrative tasks.

This decentralized database building capability offers a number of opportunities to libraries. For example, an academic library can use CONTENT to both serve image collections and to enable faculty members to serve their personal collections over the Internet. Using the Acquisition Station software, a faculty member working from his or her workstation can perform routine database building and administration tasks directly, without working through the library as an intermediary. In this scenario, the library's involvement in online collection building is in strictly substantive areas, such as administering the CONTENT server, installing the Acquisition Station software on the faculty workstation, and in assisting the faculty member with questions on the CONTENT software and metadata structure. This flexibility allows for a new level of collaboration between partners.

Another strength of the CONTENT product is its search and retrieval capabilities. Multiple search clients are available for providing access to CONTENT collections. Through the use of a Query Builder Wizard, CONTENT databases can be embedded into fully customized HTML documents. The Query Builder generates HTML coding for predefined queries; drop-down lists, and searches entry boxes, which can be pasted into HTML documents. This contextual search client provides a high degree of control over the presentation of CONTENT collections and permits multiple customized interfaces to be built for a single collection. In addition, an HTML search engine can be used to provide access to CONTENT collections. Sites that show the functionality of these two CONTENT search clients are below:

Contextual search client

<http://www.wsulibs.wsu.edu/holland/masc/xmatsura.html>

(Washington State University Libraries Manuscripts, Archives, and Special Collections: Frank S. Matsura Image Collection)

HTML search engine

<http://content.wsulibs.wsu.edu/cgi-bin/go.exe?CISOROOT=/buildings>

(Washington State University Libraries Manuscripts, Archives, and Special Collections: WSU Buildings Image Collection)

A final capability of CONTENT, which is important to the WSU Libraries in database deployment, relates to collection organization. For metadata, CONTENT uses the 15 element Dublin Core description set as a default template. Collection administrators can create custom field names relevant to a collection, with the option of mapping a field to a Dublin Core element. The custom field names will appear for searches performed in the specific collection. The Dublin Core field names will appear for searches across multiple CONTENT collections. Using the CONTENT Collection Administration software, fields can be designated as searchable or non-searchable, and visible or hidden, and fields may also be repeated. Additionally, a Template Creator function in the Acquisition Station software can be used to minimize the repetitive entry of metadata (by, for example, entering a text string that will be used as the default field value for all future objects added to the collection from the Acquisition).

Although all 15 elements of the Dublin Core, as well as local fields, are available for database construction, it is up to the project manager to determine how much (or how little) metadata to include in each image description. At WSU, we wrote descriptions as robust as we thought feasible; employing all relevant Dublin Core fields in the hope that once the project was finished there would be no need to revisit the image metadata. This decision, of course, meant significant costs in research time and editing. To a large extent the available collection descriptions will dictate the effort required cataloging the objects. Although the Matsura collection was fully processed with good access at the collection level, a closer examination of the photographs revealed that individual descriptions were poor, sometimes inaccurate, or more common, simply not available. Once the indexing began (after 330 hours of initial research), we spent on average 30 minutes per image entering the initial metadata, revising titles, adding notes, subject, and genre terms. This estimate also includes time spent editing. For a current project in collaboration with the University of Washington to describe early Washington maps, we anticipate less time required in describing individual maps, where the basic descriptive data is available on the object. However, the difficulties in scanning oversized maps, compressing the files (via MrSid software), delivering them online, and site design will most likely offset any time saved in metadata creation.

To maintain a CONTENT database, a server with a storage capacity adequate for the size of the CONTENT database(s) is required. Acquisition stations must run on PC machines (a Mac version is in development) and support a current version of Internet Explorer. The costs of scanning projects, either in-house or through a vendor, should also be considered. CONTENT requires JPEG service images and will also save a full-resolution TIFF image. Version 3.0 will support MrSid files in lieu of the TIFF image.

The capabilities, management issues, and costs of CONTENT that are described in this summary have been applied at the Washington State University Libraries in online collection building. The Libraries currently serve collections created by the WSU Libraries, Ellensburg Public Library, and Gonzaga University's Foley Library as part of the 2000 Washington State Library Digital Imaging Initiative.

A full description of the CONTENT software suite and current license prices can be obtained at Center for Information Systems Optimization WWW site, at URL <http://content.engr.washington.edu/>.