

# “Providing Reliable End User Access To Your Digital Content”

*Russ Hunt*

OC LC PICA Product Specialist  
7<sup>th</sup> Floor Tricorn House, 51-53 Hagley Road, Edgbaston, Birmingham, B16 8TP, UK  
[r.hunt@oclc-pica.org](mailto:r.hunt@oclc-pica.org)

## Abstract

OC LC PICA Digital & Preservation Resources provide libraries with the flexibility to preserve and manage digital collections via microfilming, digitising, collection management tools and archiving services.

This demonstration aims to highlight the benefits of two of our tools that enable users to publish collections electronically:

- **ActivePaper Archive** by Olive Software features processes and software that allow for the publication of digitised versions of microfilmed newspaper collections as fully searchable online archives.
- **CONTENTdm** is a versatile, easy to use digital collection management suite allowing users to organize, describe, publish and search digital collections online.

Both software tools use XML for their underlying structure and delivery, providing standardised, reliable, long term access by end users to valuable digitised collections published online.

## 1 CONTENTdm – Complete Management of Digital Collections

CONTENTdm is a complete digital collection management system allowing users to import digital objects, create metadata descriptions for those objects, index, store, query and publish digital collections, providing web access to special collections.

A distributed architecture enables multiple contributors to collaborate on publishing rich digital collections online. Different departments within a single organisation or multiple organisations can combine resources to make digital collections accessible via standard web browsers.

Photographs, Letters, Manuscripts, Diaries, Postcards, Maps, Video files and Audio recordings can all now be made available to end users in fully searchable collections delivered to standard web browsers.

Textual documents, printed and handwritten, can have full text search capabilities added. Favourites Functions allow end-users to save links to digital objects and return quickly to them later. Online Slide Shows can be created from Favourites files and with the use of the CONTENTdm PowerPoint Plug-in presentations can be created from the contents of an online collection (This is how the software is being used at the Kings College London HumSlides site). Zoom & Pan features allow users to examine detail of large online documents such as maps.

### 1.1 Demonstration Sites

This demonstration will aim to show the range and diversity of materials published on the web from an end-user perspective. We will take a look at how various organisations have published collections of slides, posters, maps, images, books and audio files and how CONTENTdm’s end user tools allow interaction with those collections.

Utah Fire Insurance Maps - <http://www.lib.utah.edu/digital/sanborn/index.html>

The Sanborn Map® Collection consists of a uniform series of large-scale detailed maps, dating from 1867 through 1969 and depicting the commercial, industrial, and residential sections of cities

Cleveland Memory - <http://www.clevelandmemory.org>

Thousands of historical photographs, as well as a growing collection of ebooks documenting the history of greater Cleveland and the Western Reserve region of north-eastern Ohio

American Journeys - <http://www.americanjourneys.org>

Contains more than 18,000 pages of eyewitness accounts of North American exploration. From the sagas of Vikings in Canada to the diaries of mountain men in the Rockies 800 years later.

International Poster Collection - <http://lib.colostate.edu/posters/>

The International Poster Collection holds poster entries from the Colorado International Invitational Poster Exhibition (CIIPE)

Digital Library of Appalachia - <http://www.aca-dla.org/>

The Digital Library of Appalachia seeks to provide online access to archival and historical materials related to the culture of the southern and central Appalachian region using the collections of thirty-three member libraries, archives, and museums associated with the Appalachian College Association

## **2 Olive Software ActivePaper Archive – Publishing Historic Newspaper Collections Online**

Newspapers are a unique historical record, detailing everyday life on a day to day basis

Newspapers are intrinsically ephemeral. Published on poor quality paper, poor quality ink (even in the 21<sup>st</sup> century most of the ink transfers itself to the reader !). They are here today, gone tomorrow and were not produced to last.

Newspapers are trapped in user un-friendly formats. Either contained in cumbersome bound volumes or on reels of microfilm. Microfilm is an excellent long-term preservation format but in terms of users finding what they want from a microfilm archive, it is not the most user friendly of mediums.

The ActivePaper Archive process takes microfilmed newspaper collections and creates a fully searchable archive built on XML standards that preserves the look and feel of the original publication while allowing users to search not only across multiple issues of a title but also across multiple titles within an archive.

Digital page images are created from microfilm. Olive's unique distillation software is utilised to identify the typographic elements of each page image. At the broadest level, the distillation software is looking to identify articles, photographs/pictures, advertisements, and tables. Within each of these elements, the distillation software looks for more detailed elements such as headline, byline, paragraph, tagline, cut line, etc.

The location of each of the identified elements on the page is saved along with a surrogate image of the element in an XML repository. Once the distillation software has identified all of the elements on the page, each of the elements is subjected to optical character recognition. The text generated from this process is associated with the page element in the XML repository.

The XML repository is a data set that includes XML tagging of each newspaper issue, its typographic elements and their coordinate locations on the page. It also includes surrogate images for each page, each element of the page, and the ASCII text associated with each element of the page.

The XML repository is made available to users using XSL Style Sheet technology via a standard web browser from an ActivePaper Archive server. This server also includes the "librarian module" giving capabilities for adding metadata to articles, creating special clipping files of articles on specific topics, and adding thesaurus/subject terms to articles. This manual enhancement option allows further optimization of the archive to suit your user community.

### **2.1 Demonstration Sites**

British Library Online Newspaper Archive – <http://www.uk.olivesoftware.com>

Brooklyn Daily Eagle - <http://eagle.brooklynpubliclibrary.org/>

The Scotsman Digital Archive - <http://archive.scotsman.com>