

Russ Hunt*

OCLC Tools Managing & Preserving Digitised Map Libraries

Keywords: Digital collection management; digital collection access; digital preservation;
long term preservation.

Summary

This paper explains how Cartographic Heritage institutions can develop digital map libraries, tools to aid their creation and remote access to them, Support preservation of digitised collections and use ICT solutions to help with teaching and diffusion of digital map collections using the CONTENTdm Digital Collection Management Software and the OCLC Digital Archive service from OCLC, the Online Computer Library Centre, a non-profit, membership, computer library service and research organisation dedicated to the public purposes of furthering access to the world's information and reducing the rate of rise of library costs (<http://www.oclc.org>).

Introducing OCLC tools, CONTENTdm and the OCLC Digital Archive, to assist management, presentation and preservation of digitised map libraries.

Founded in 1967, OCLC Online Computer Library Centre is a non-profit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing the rate of rise of library costs. It now serves more than 60,000 libraries in 112 countries and territories around the world who use OCLC services to locate, acquire, catalogue, lend and preserve library materials.

OCLC and its member libraries cooperatively produce and maintain WorldCat - the OCLC Online Union Catalogue. WorldCat is a global network of library content and services that uses the Web to let your institution be more connected, more open and more productive.

The International Cartographic Association Commission's Terms of reference ask for the following area's to be addressed:

- Development of digital map libraries and tools to aid their creation and remote access to them.
- Support for preservation of digitised collections
- ICT solution to help with teaching and diffusion of digital map collections

A number of OCLC's products and services address these aims, 2 of which I would like to talk about specifically today.

The two products I would like to talk about are CONTENTdm and the OCLC Digital Archive.

CONTENTdm allows users to Store, manage and give access in a web environment to digitised collections. It is already being used by a number of organisations to present digitised map collections via the web.

The OCLC **Digital Archive** service is a new service providing a cost-effective, managed storage environment for digital master files that fits in with the several workflows for acquiring digital content.

- For users of **CONTENTdm** the Digital Archive service is an optional capability integrated with the various workflows for building collections. Master files are secured for ingest to the Archive using the CONTENTdm Acquisition Station, the Connexion digital import capability, and the Web harvesting service.
- For users of other content management systems the Digital Archive service provides a low-overhead mechanism for safely storing master files.

* Product Specialist OCLC [russ.hunt@oclc.org]



What Is CONTENTdm ?

- **A complete software solution specifically designed for web presentation of digitised special collections**
- **In excess of 400 licensed users Worldwide (over 1000 users)**
 - USA, Netherlands, UK, Ireland, Germany, Belgium, France, Spain, Australia, Canada
- **Organisations of all types and sizes**
academic libraries, public libraries, historical societies, cultural heritage organisations, and other collaborators
- **Showcase a wide range of digital/media types**
photos, documents, newspapers, audio and video

I'd like to take a look at the two services in more detail

CONTENTdm was originally developed by librarians at the University of Washington in the USA as a system for presenting their digitised special collections on the internet.

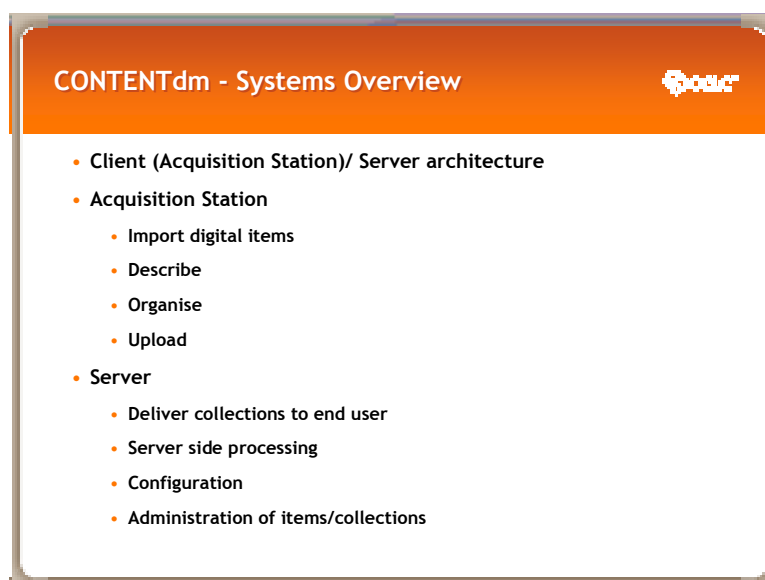
As the software became more popular among academic institutions a separate company, DiMeMa Inc, was created to further develop and maintain the software.

In 2003 OCLC entered into an agreement with DiMeMa to market and support the software and at the start of 2007, OCLC acquired DiMeMa Inc.

There are currently around 400 licensed users worldwide. Due to a number of consortia using the software there are around 1000 organisations using the software to present digitised collections.

Organisations of all types and sizes are using CONTENTdm from large research and National libraries right down to small local history groups.

The types of materials being presented cover a wide range of digital resources from images to books, journals and diaries alongside audio and video material and, of course maps.



CONTENTdm - Systems Overview

- **Client (Acquisition Station)/ Server architecture**
- **Acquisition Station**
 - Import digital items
 - Describe
 - Organise
 - Upload
- **Server**
 - Deliver collections to end user
 - Server side processing
 - Configuration
 - Administration of items/collections

CONTENTdm works on a Client/Server basis.

The Acquisition Station client is installed on local workstations and allows users to interact with the CONTENTdm server in many ways.

The Acquisition Station is the place where you will import digitised items, add metadata for those items, build compound objects (virtual objects where a number of items are connected together to always be

displayed as a single entity. i.e. all the pages of an atlas), prepare areas on the server to where the collections will be uploaded and apply access re-strictions if required.

The Acquisition Station also gives you access to the servers administration area. The servers most important function is to deliver your digitised collections to end users via their web browsers. In the server administration area you will also define and configure collections and their metadata structures, administer those collections and the items they contain, generate reports on the usage of your collections. There is also some server side processing handled here for PDF format documents.

CONTENTdm has been helping libraries and cultural heritage institutions to deliver their digitised collections on the web for almost 10 years.

Users find that CONTENTdm is a stable system both in terms of its software reliability and in terms of it having the backing of a world renowned organisation like OCLC.

The software is scalable, both in terms of the number of items that can be stored and delivered and in terms of the number of concurrent users it will handle. We have users presenting a few hundred items using CONTENTdm up to one user delivering in excess of 5 million items and reporting never having had any problems with concurrent staff or patron users overloading the system.

CONTENTdm features such as the ability to build compound objects, its integrated OCR functionality and the JPEG2000 extension have enabled the building of rich interactive digital collections.

The ability to apply customised and multiple interfaces to your digital collections and CONTENTdm's security options ensure that you can control how your collections are viewed and used. If required access restrictions can be applied from the collection right down to an individual item level and access can be granted to an object or collections metadata while restricting access to images.

Usage report generation allows you to see how your collections are being used and by whom.

As CONTENTdm is fully web compatible full remote collection building and administration is possible via the Acquisition Station client of using standard web browsers.

CONTENTdm offers flexible Metadata options in order for collection builders to be able to best describe collections.

The default metadata scheme is unqualified Dublin Core. A template for VRA Core is also included (VRA Core is a data standard for the cultural heritage community that was developed by the Visual Resources Association).

All metadata schemes within CONTENTdm are configurable. Up to 200 collections may be defined per server and each of those collections can feature its own metadata template of up to 100 fields. There exists therefore the option to apply metadata schemes to individual collections that are most suitable for the materials within the collection.

The ability to include EAD finding aids is also part of the software's metadata abilities.

A look at some current implementations of CONTENTdm in relation to map collections

l'Institut Cartogràfic de Catalunya (ICC)

Started using CONTENTdm early in 2007. Currently making available collections containing around 5000 items and growing.

The American Geographical Society Library make around 300 map images available of the city of Milwaukee via a CONTENTdm installation at the University of Wisconsin in Milwaukee.

From their home page users can choose to run pre-prepared searches based on City, State or country. Users can also perform keyword searches from the home page. A link is available to an advanced search page where users can search across all fields, by selected field or by proximity.

Metadata descriptions of individual map images are quite detailed and secondary searches can be started by clicking on the hyperlinked words and phrases in the metadata description.

Maps have been scanned and uploaded in JPEG2000 format, this allows for CONTENTdm's zoom and pan functionality to be used so that users can see areas of the maps in great detail and easily move around images when zoomed in to high levels of magnification.

The Sanborn Fire Insurance Maps collection at the University of Utah contains around 1300 items pertaining to maps created by D.A. Sanborn in 1866 which assisted fire insurance agents in determining the degree of hazard associated with a particular property.

This is one collection in the University's extensive CONTENTdm digital collections, which contain just over 5 million digital items.

Alongside the functionality we have already seen CONTENTdm allows users to compare images side by side while still being able to zoom and pan around images. This can be a very useful feature when viewing images of maps

CONTENTdm's link with OCLC's WorldCat Union Catalogue allows digital collection builders to deposit metadata from their digital collections in WorldCat thus giving global exposure to your digital collections via related services including the freely available WorldCat.org website and search engines such as Google and Yahoo.

Metadata is harvested from CONTENTdm servers via the OAI-PMH protocol. Patrons searching in places such as WorldCat.org or Google can then be linked directly into your digital collections from search hits at these sites.

CONTENTdm is easy and intuitive to use,

It supports all your digitised assets. Digital files of any type can be uploaded to CONTENTdm and it will display any file that can be natively viewed in a web browser (JPEG, GIF, JPEG2000 etc.) along with any file that can be delivered from a web browser using the appropriate plug in (Windows Video files, Real media, DJVu files etc.). The interfaces and metadata schemes within CONTENTdm are fully customisable. The software is based on standards commonly used in the library and cultural heritage worlds.

Look at options for long term storage of digital master files

Digitisation itself creates a level of preservation, for the original item. Digital presentation of materials reduces the risk to sometimes fragile and valuable original materials through constant handling.

When creating digitised collections it is important to create master files in open formats that offer the possibility for long term storage and preservation as well as the ability to migrate to new formats as they become available in the future.

The OCLC Digital Archive Service provides a secure storage environment enabling easy management and monitoring of the health of your master files and digital originals.

The importance of preserving master files grows as digital collections grow. The collection manager needs a workflow for capturing and managing master files that "fits in" with the acquisition of both digitized and born-digital content while not outpacing their capability to manage these large files.

The **Digital Archive** service provides a managed storage environment for digital master files that fits in with several workflows for acquiring digital content. It can be used in tandem with CONTENTdm or as a standalone repository for master files.

INGEST - Items and collections can be ingested to the Digital Archive in a number of ways.

Users of the **CONTENTdm Acquisition station** can prepare master files for delivery to the Archive using its Full Resolution capability.

With this capability enabled the Acquisition station creates Archival Volumes of master image files as collection builders prepare digital objects for uploading to a CONTENTdm server. Metadata relating the master image files to the CONTENTdm collections are kept on the CONTENTdm server for easy management.

Archival Volumes are kept in the Acquisition station until they are ready to be copied to a removable storage device.

An electronic shipping manifest is then created and device and manifest are shipped to OCLC where our technicians load the Archival Volumes into the Archive.

If another content management system is in use content is prepared for delivery to the Digital Archive by copying it to a removable storage device and creating an electronic shipping manifest. Once the device is shipped to OCLC technicians load the content in the Archive.

MONITORING - From the time content arrives in the Digital Archive OCLC systems begin inspecting it to ensure that what is sent to us is the same thing you'll get back in the future.

On the day content is ingested to the Digital Archive OCLC systems perform the following quality checks:

1. **Manifest verification** – to verify that the files sent us match what's listed in the shipping manifest. We'll notify you of any anomalies we find.
2. **Virus check** – to check each file received by OCLC for viruses using industry standard security software.
3. **Fixity check** – to create a 'fingerprint' for each file copied to the Digital Archive. Each file is scanned and a unique "fixity key" is created for the file. If anything is altered within the file a subsequent fixity scan will generate a different key so we and you will know if there's been a change.
4. **Format verification** – to verify that the file is the format indicated by its file extension (e.g., .tif file extension for TIFF format files).

Once OCLC's systems have gathered this information it is added to the health record kept for each file in the Digital Archive.

OCLC's automated systems revisit the Virus check, Fixity check, and Format verification periodically so up to date reports on the health of collections can be made. OCLC provides updated reports for all collections via a personalised archive report portal each month.

CARE & HANDLING - OCLC's Digital Archive service is a specially designed system in a controlled operating environment dedicated to the ongoing managed storage of digital content. OCLC has developed specific systems process and procedures for the Digital Archive tuned to the management of data for the long term.

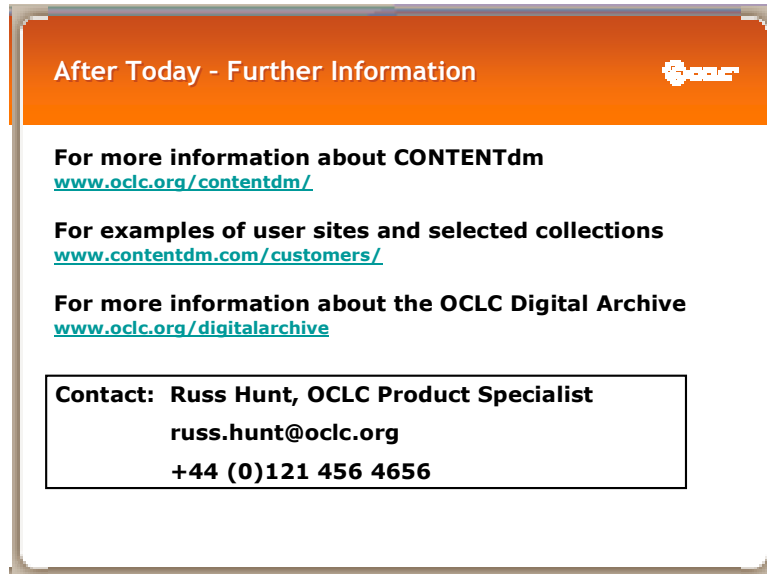
- **Systems management** – the Digital Archive exists in a managed systems environment. Many of the tasks that your staff would perform if you kept physical copies of your master files are automated and many others are regular procedures. For example, OCLC monitors the lifecycle of the disk storage devices and regularly replaces them. As content is migrated to new devices we ensure the integrity of each file being migrated.
- **Physical security** – the Digital Archive system is in OCLC's limited access operations facility secured by a badge reader system and monitored 24x7 by systems operators, security guards, and CCTV cameras.
- **Data security** – OCLC employs a dedicated information security team. The team is part of the review process for applications, systems, and procedures.
- **Data backups** – Content in the Digital Archive is backed up using industry standard procedures. At any point in time there are 6 copies of the content of the Digital Archive at offsite facilities and one copy onsite. Short term data loss (less than 24 hours) is minimized by utilizing redundant, self-healing storage devices.
- **Disaster recovery** – OCLC's facilities and systems are protected from minor outages with uninterruptable power supplies, fire suppression systems, and redundant servers and network feeds. OCLC has specific recovery procedures for the Digital Archive system in the event of a major disaster.


DISSEMINATION – Retrieving copies of archived content is as easy as sending an e-mail. You can request a dissemination of your content by Archival Volume, or by entire Collection. When we receive an e-mail from an authorized representative of an institution we copy the requested files to a removable storage device and create an electronic shipping manifest for the package. The storage device and the shipping manifest are sent to the requestor.

CONTENTdm Collection administrators can track Archival Volumes and Collections using the CONTENTdm Admin interface. CONTENTdm Collection Administrators can retrieve single files online to support users who request high resolution or original copies. Also, "full resolution" archived content from CONTENTdm collections can be configured for online access through the full resolution hyper-

link in the CONTENTdm item viewer display, Files up to 1GB in size can be disseminated through the online mechanism.

Further information at...



After Today - Further Information 

For more information about CONTENTdm
www.oclc.org/contentdm/

For examples of user sites and selected collections
www.contentdm.com/customers/

For more information about the OCLC Digital Archive
www.oclc.org/digitalarchive

Contact: Russ Hunt, OCLC Product Specialist
russ.hunt@oclc.org
+44 (0)121 456 4656