

# Digital Archive Getting Started Guide

## Contents

<a href="#">Introduction</a> .....	page 1
<a href="#">Archive Processing and Organization</a> .....	page 1
<a href="#">Allowable Data Formats</a> .....	page 2
<a href="#">Ingestion</a> .....	page 3
<a href="#">Fees</a> .....	page 3
<a href="#">Processing</a> .....	page 3
<a href="#">Reports</a> .....	page 9
<a href="#">Dissemination</a> .....	page 10
<a href="#">Bulk Dissemination</a> .....	page 10
<a href="#">Single File Dissemination</a> .....	page 12
<a href="#">Support</a> .....	page 14

## Introduction

This document is a user guide for collection administrators using the Digital Archive. It contains descriptions of how to use each of the features of the Archive to manage your content.

There are three (3) primary functions of the Digital Archive:

1. **Ingestion.** The process of moving your content into the Archive.
2. **Reporting.** Getting feedback from the system in order to manage your content.
3. **Dissemination.** The process of getting copies of your content out of the Archive.

Information on how to use each of these features is covered in detail in the following sections of this document.

### ***Archive Processing and Organization***

The Digital Archive is a managed storage environment for your content. Your content will be stored in the format and directory structure in which you send it to OCLC for processing. You prepare to send your content to OCLC by creating an Archival Volume containing your content on a removable storage device, along with an electronic shipping manifest. As each batch of content is received by OCLC, an archive technician performs an Ingest operation. During Ingest processing, the Archive checks the content against the electronic shipping manifest that you

provide. The system also checks each file for viruses, verifies the file's data format using JHOVE, and creates a digital fingerprint for the file so we can do an independent "fixity" check to verify that no bits in the file have been altered. At the conclusion of Ingest processing an Archive Accession Report is created for you to review.

A batch of content is loaded into the Archive using multiple identifiers that allow you to uniquely identify individual files and organize your collections. Each of the identifiers below is part of a hierarchy, with Institution Name at the top of the hierarchy and File Name at the bottom:

- **Institution Name** (OCLC symbol). We use your OCLC symbol as the broadest identifier for all your content. We can uniquely identify all the files in the Archive that belong to you using your OCLC symbol.
- **Server URL** (CONTENTdm "host name"). We use Server URL to identify the source server from which content was extracted or the location where the access copy of the content can be found. Using the Server URL identifier allows you to group content in the Archive by the source server where associated access copies are kept.
- **Collection Name** (CONTENTdm "collection alias"¹). We use Collection Name in the Archive to allow you to group your content by collection. This identifier is unique within a Server URL.
- **Archival Volume Name** (CONTENTdm "volume ID"). We use Archival Volume Name to identify the batch of content you sent to OCLC. This identifier must also be unique within a collection.
- **File Name**. We use File Name to identify a specific file within an Archival Volume. This identifier must be unique within an Archival Volume.

This information is used in CONTENTdm servers to allow a unique link to a "preservation" file in the Archive to be tracked in the "access" file metadata associated with it.

While your data is stored in the Archive the system performs regular checks on the health of your content. The results of those checks are summarized in a File Integrity Report. You can use the results of this report to be assured that your content remains unaltered in the Archive.

At some point in the future you may wish to retrieve a copy of one or more of your files from the Archive. To do this, you should send a Dissemination Request via e-mail to OCLC. An archive technician will use your e-mail as a guide to copy the requested Collections or Archival Volumes to a removable storage device and ship them to you. During the Dissemination process a summary Dissemination Report is generated. You can also perform online requests for single files that are smaller than 1GB using the identifier elements listed above.

### ***Allowable Data Formats***

The Digital Archive will allow any data format type to be ingested and stored. The choice of archival data format depends upon your local practices and collection policies. One purpose of the Digital Archive is for you to be able to get back the exact file you sent to OCLC for archiving.

You may even include metadata files along with content files. If you do this you may want to create either a directory structure or a naming structure to associate metadata files specifically with content files. Metadata files will be treated as preservation objects like any other file.

## Ingestion

In order to get copies of your content into the Archive you must submit them to OCLC for Ingest processing.

### **Fees**

You are charged an Ingest Fee for ingest submissions. The Ingest Fee allows you to submit multiple Archival Volumes in one year up to a total of one terabyte of content across all submissions.

### **Processing**

There are three basic steps you must follow to submit your content to OCLC for Ingest processing:

1. Prepare content for submission to the Archive.
2. Create an electronic shipping manifest.
3. Package your content and ship it to the Archive.

The following sections describe each of these steps in detail.

## **Prepare content for submission to the Archive**

### **Using the CONTENTdm Acquisition station**

Preparing content for submission to the Archive with the CONTENTdm Acquisition station works automatically with the Full Resolution capability if you are building a collection from TIFF or JPEG image files. If you are building a collection using other file types, please see the directions below for “Using a non-OCLC content management system.”

If you are building a CONTENTdm collection using TIFF or JPEG image files, follow the directions for Managing Archival or Full Resolution Images at <http://www.contentdm.com/help4/acq-station/archiving.html>. The directions show you how to prepare a Staging Area for your Archival Volumes. When you start a new project on your CONTENTdm Acquisition Station remember to select **Have server generate volume name automatically**.

As you use the CONTENTdm Acquisition Station to upload display images, thumbnail images, and metadata to the collection on your CONTENTdm server, the full resolution image files (TIFF or JPEG) are stored in Archival Volumes in a temporary Staging Area.

Once an Archival Volume is “full” the Acquisition Station will notify you that the current volume is closed and a new volume is being created. After an Archival Volume is closed, you can create an electronic shipping manifest, copy the Archival Volume and the electronic shipping manifest to a removable media device, and ship the removable media device to OCLC for Ingest processing into your Archive.

Each file that is sent to the Archive for Ingest processing is tracked by your Institution Name (i.e., your OCLC symbol), your CONTENTdm server host name (without the ‘http://’), the Collection Name, the Archival Volume Name, and the File Name. You add all of this information to the electronic shipping manifest prior to sending it to OCLC.

### **Using a Non-OCLC Content Management System**

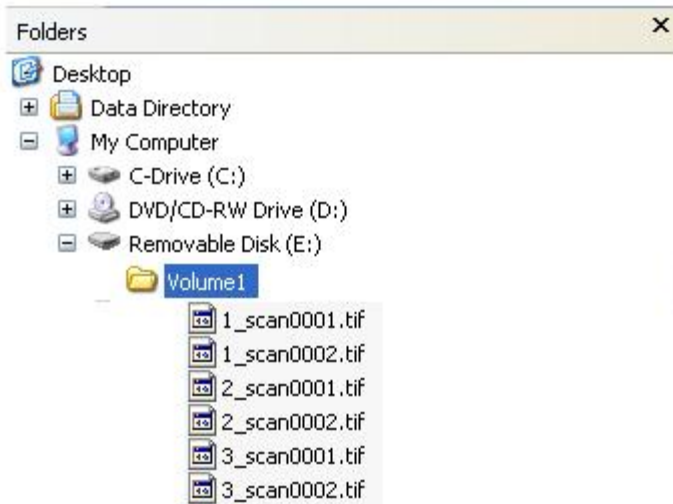
To prepare content for submission to the Archive you create an electronic shipping manifest, copy the content and the electronic shipping manifest to a removable media device, and ship the removable media device to OCLC for Ingest processing into your Archive.

Preparing content to send to the Archive requires that you put the content on a removable storage device such as a USB-connected disk drive or DVD. The content on the removable storage device is placed in the Archive in the same directory structure that was used on the removable storage device.

There are two types of file structures you can create on the removable media as you prepare your data to send to OCLC:

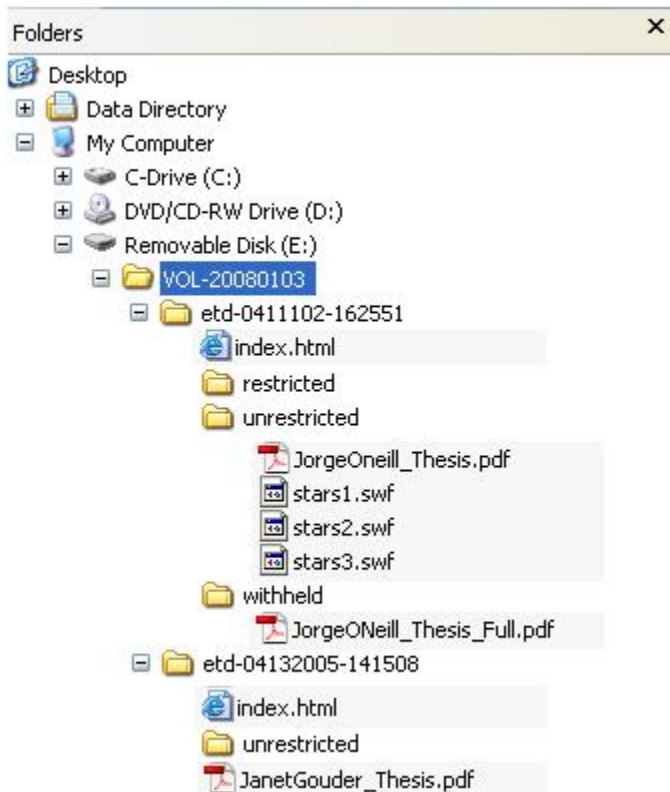
- Single directory structures
- Multi-level directory structures.

A single directory structure has all the files in a single Archival Volume directory on the removable storage device with no subdirectories. Example 1 shows six TIFF files within a single Archival Volume named Volume1 on a removable storage device.



**Example 1:** Single directory structure on removable disk E:

A multi-level directory structure includes subdirectories within the Archival Volume on the removable storage device. Example 2 shows various subdirectories and files within an Archival Volume named “VOL-20080103” on a removable storage device.



**Example 2:** Multi-level directory structure on removable disk E:

Each file you send to the Archive for Ingest processing is tracked by your Institution Name (OCLC symbol), your institution's URL (without the 'http://'), the Collection Name, the Archival Volume Name, and the File Name. You add all of this information to the electronic shipping manifest prior to sending it to OCLC.

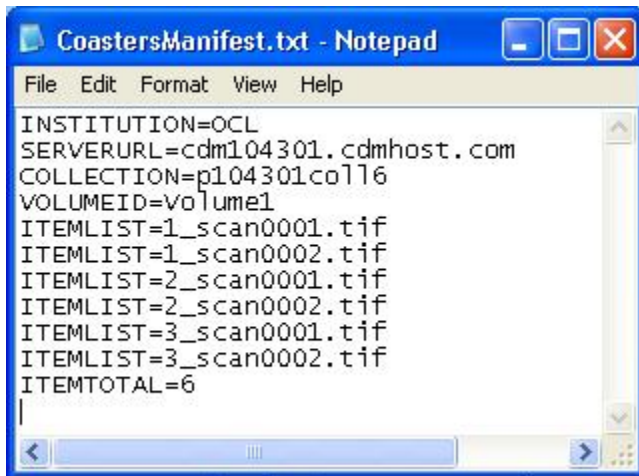
## Create an electronic shipping manifest

The electronic shipping manifest is a text file (format .txt) containing information identifying your institution and the list of files you are sending on the removable storage device. In addition, you provide information about how you want this content organized in the Archive and where to look in the directory structure of the removable storage device to find the Archival Volume with your content.

There are 6 required data elements in the electronic shipping manifest:

1. **Institution.** Your institution's OCLC symbol.
2. **ServerURL.** The full domain name of the server where corresponding access copies are kept (e.g., cdm104301.cdmhost.com), or simply the web domain name of your institution (e.g., www.columbia.edu).
3. **Collection.** The name of the digital collection to which this content belongs.
4. **VolumeID.** The Archival Volume name and also the directory name on the removable storage device where the content resides. The value of VolumeID should be unique within a collection. If you are using a content management system other than CONTENTdm we recommend naming volumes with a date stamp, e.g., "VOL-20080301" for a volume created on March 1, 2008.
5. **ItemList.** The file name of a content file within the Archival Volume. The file name should include a mime type extension, e.g., .pdf, or .tif. For Archival Volumes with a multi-level directory structure the file name should include the relative directory path of the file within the Archival Volume (see Example 4).
6. **ItemTotal.** The count of files within the Archival Volume. This should be equal to the number of occurrences of the ItemList element.

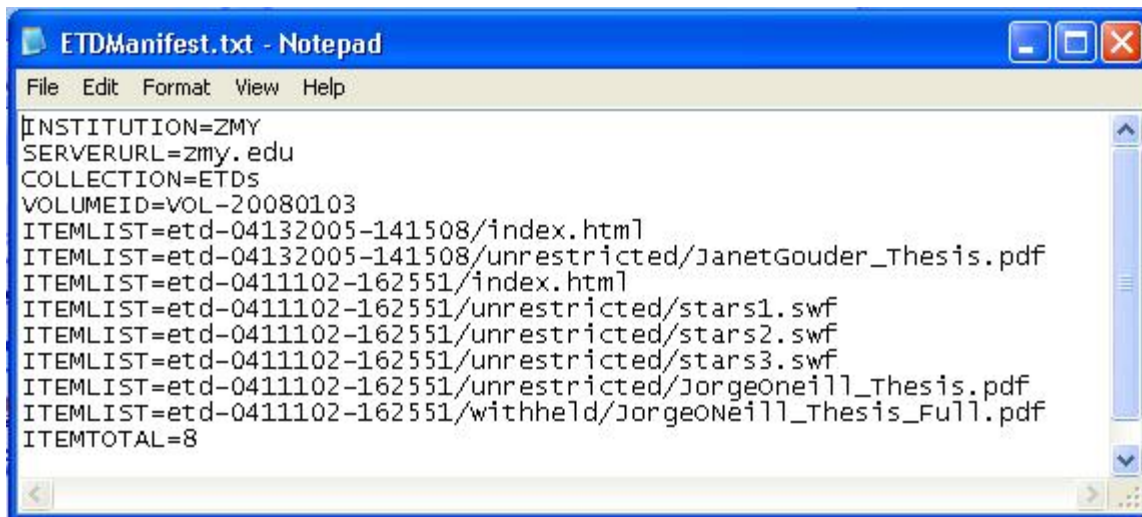
For Archival Volumes that have a single directory structure each ItemList element contains the name and mime type extension for each file submitted.



```
CoastersManifest.txt - Notepad
File Edit Format View Help
INSTITUTION=OCL
SERVERURL=cdm104301.cdmhost.com
COLLECTION=p104301co116
VOLUMEID=volume1
ITEMLIST=1_scan0001.tif
ITEMLIST=1_scan0002.tif
ITEMLIST=2_scan0001.tif
ITEMLIST=2_scan0002.tif
ITEMLIST=3_scan0001.tif
ITEMLIST=3_scan0002.tif
ITEMTOTAL=6
```

**Example 3:** Manifest file for a single directory structure Archival Volume (this is the type of Archival Volume created by CONTENTdm)

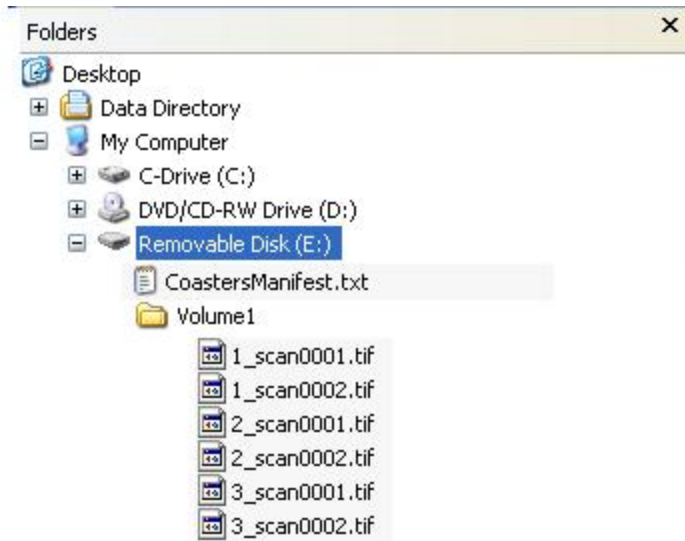
For Archival Volumes with a multi-directory structure each ItemList element contains the relative directory path of the file within the Archival Volume as well as the file name and mime type extension for each file submitted.



```
ETDManifest.txt - Notepad
File Edit Format View Help
INSTITUTION=ZMY
SERVERURL=zmy.edu
COLLECTION=ETDs
VOLUMEID=VOL-20080103
ITEMLIST=etd-04132005-141508/index.html
ITEMLIST=etd-04132005-141508/unrestricted/JanetGouder_Thesis.pdf
ITEMLIST=etd-0411102-162551/index.html
ITEMLIST=etd-0411102-162551/unrestricted/stars1.swf
ITEMLIST=etd-0411102-162551/unrestricted/stars2.swf
ITEMLIST=etd-0411102-162551/unrestricted/stars3.swf
ITEMLIST=etd-0411102-162551/unrestricted/JorgeO'Neill_Thesis.pdf
ITEMLIST=etd-0411102-162551/withheld/JorgeO'Neill_Thesis_Full.pdf
ITEMTOTAL=8
```

**Example 4:** Manifest file for a multi-directory structure Archival Volume

Once you've created your electronic shipping manifest place a copy of it on the removable storage device at the highest level allowable, at the same level as the directory of the Archival Volume.



**Example 5:** CoastersManifest.txt file on removable storage device

Save a copy of the manifest for your records and e-mail a copy to OCLC to notify them of the upcoming shipment of your content.

**Package your content and ship it to the Archive**

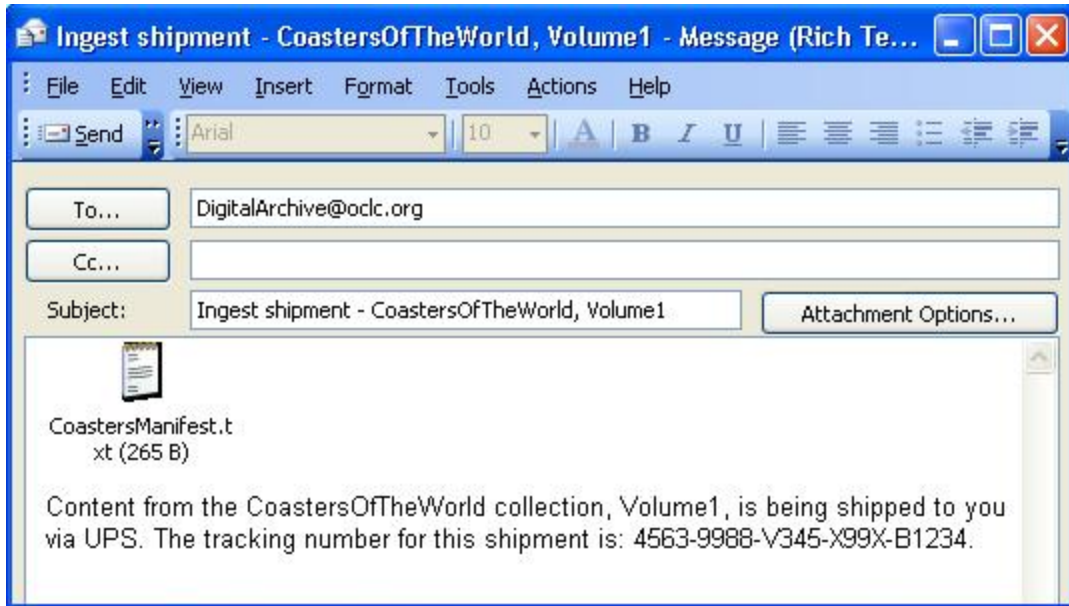
Once you're prepared your content and your electronic shipping manifest all that's left to do is to send the content and manifest to the Archive.

Pack your removable storage device according to instructions from your shipping service. Address your package to the Archive:

OCLC  
Attn: Digital Archive, MC 431  
6565 Kilgour Place  
Dublin, OH 43017

Once you've placed your shipping order with your shipping service, send an e-mail with the electronic shipping manifest file attached to the Archive to notify us that your package is on the way.





**Example 6:** E-mail notification of shipment to the Archive

Address your e-mail to [DigitalArchive@oclc.org](mailto:DigitalArchive@oclc.org) with the subject line:

**Ingest shipment - <collection name>, <volume ID>**

Attach the electronic shipping manifest document to the e-mail. If possible also include your shipping service tracking number for the package in the e-mail.

Once your content is received an OCLC archive technician will perform Ingest processes on your Archival Volume and load your content into the Archive. Once the Ingest completes, an Archive Accession Report will be available for you in the Archive's report subsystem.

## Reports

The Digital Archive produces a variety of reports to help you monitor the health of your content in the Archive. Reports for your content can be viewed at <http://worldcat.org/digitalarchive/reports>. You will be prompted for an authorization and password when you access this site. Use the OCLC Authorization which you had associated with your Digital Archive account when you placed your order for the Digital Archive. If you cannot remember which OCLC Authorization you had associated with your account, please write to [support@oclc.org](mailto:support@oclc.org) for help with your account.

Within the Digital Archive Reports web site your reports are organized in folders by report type. Within a folder for a specific report type, the reports are organized by year, then by month. For a given month, you'll see two copies of each report generated by the system, a copy encoded in XML and a copy encoded in HTML. The HTML copy is easily viewable online. The XML can

be used with one of the templates in the “Templates” folder to move the report information into a Microsoft Excel spreadsheet (<http://worldcat.org/digitalarchive/reports/templates>).

## Available Reports

There are four reports available for you:

**Accession Report.** Produced each time new content is ingested in the Archive. The Archive Accession Report provides a summary of Ingest processing on the content you’ve submitted to the Archive and a detailed report about the initial checking and verification performed on each file that was received.

**Activity & Storage Report.** Produced monthly, summarizes your entire collection and data activities. This report provides a summary of file health, storage usage, and content accesses.

**File Integrity Details Report.** Produced monthly, lists the details of exceptions found during routine scanning of content.

**Dissemination Report.** This report is produced each time a Dissemination request is processed. It includes summary information about the dissemination including number of files and amount of storage. The report also includes format, virus, and fixity details about each file being disseminated.

## Dissemination

To request copies of your content from the Archive you have two options:

- Bulk dissemination requests by e-mail
- Single file dissemination online, via file-by-file requests using HTTP for files smaller than 1GB.

### ***Bulk Dissemination***

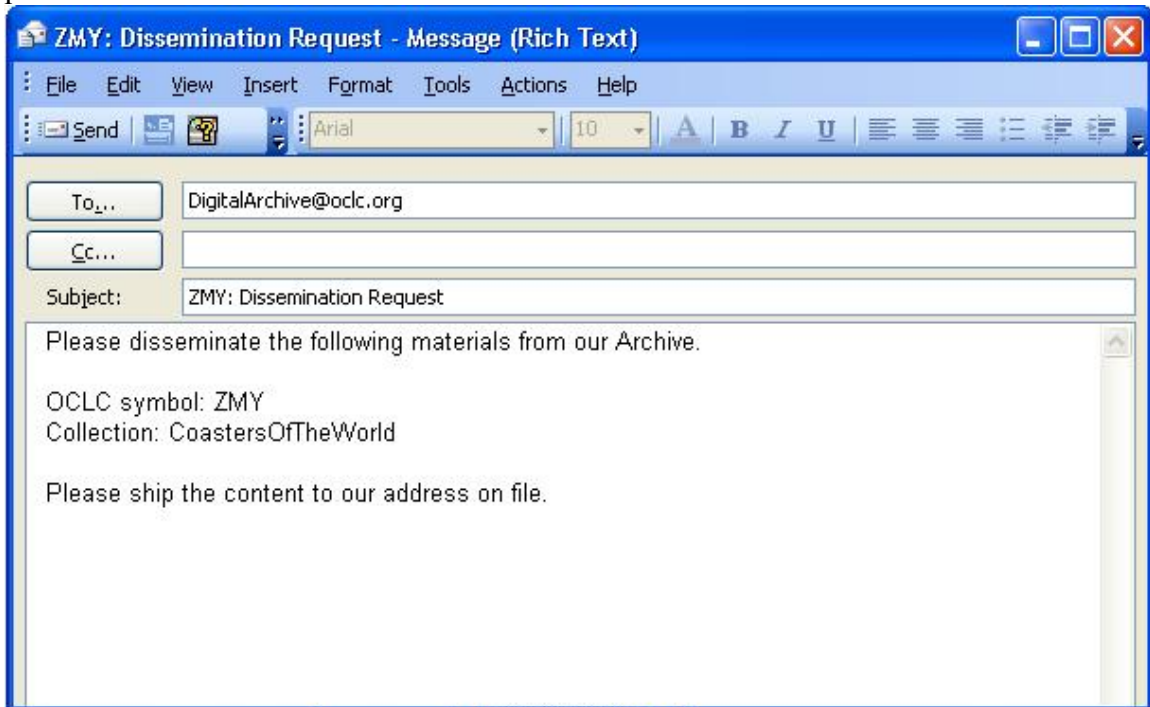
#### **Fees**

You are charged a Dissemination Fee and a Shipping and Media Charge for bulk dissemination requests. The Dissemination Fee allows you to request multiple disseminations in one year up to a total of one terabyte of content across all requests. Each time OCLC processes a dissemination request you will also be charged for the removable storage device and shipping costs.

#### **Preparing a Dissemination Request**

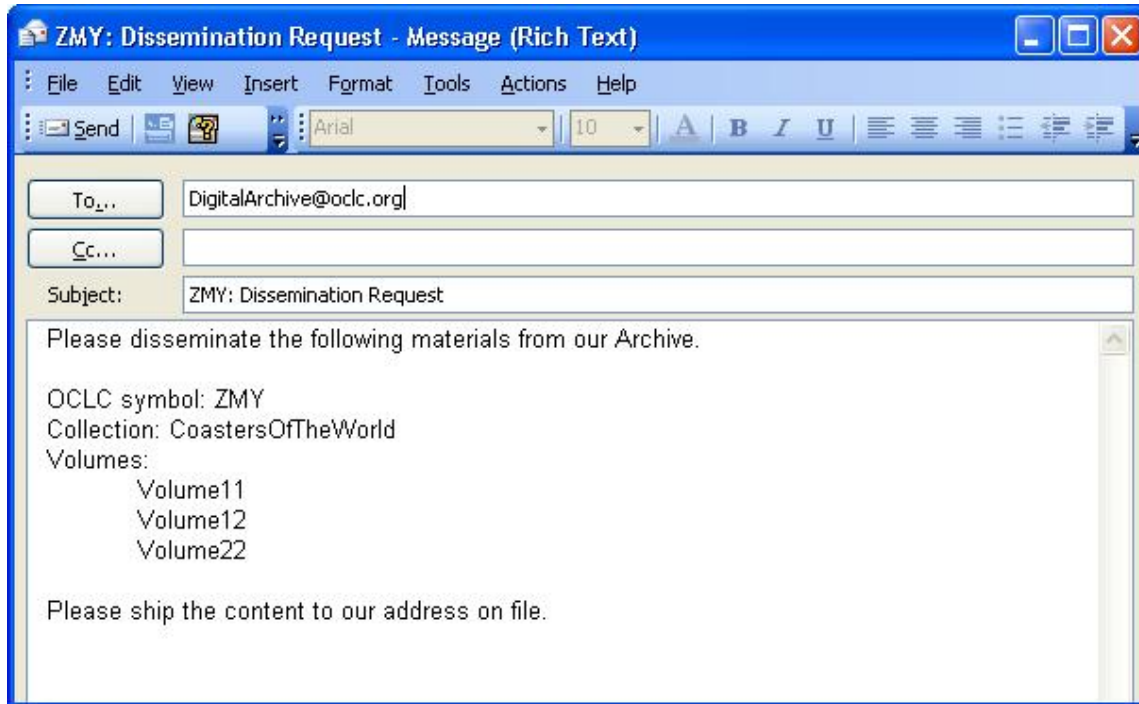
You can get copies of your content from the Archive by sending an e-mail to [DigitalArchive@oclc.org](mailto:DigitalArchive@oclc.org). An archive technician will process your request and copy the selected content to a removable storage device along with an electronic shipping manifest. At the completion of dissemination processing the removable storage device will be shipped to you and a Dissemination Report will appear in your Digital Archive Reports web site.

You can request disseminations by Collection Name or by Collection Name + Archival Volume. You can request dissemination of multiple collections or multiple archival volumes in a single e-mail request.



**Example 7:** Dissemination request – whole collection

When requesting a dissemination please include your OCLC symbol in the e-mail and any special shipping instructions. If no shipping address is supplied we will ship disseminated content to the address we have on file for your institution.



**Example 8:** Dissemination request – multiple volumes in a collection

## ***Single File Dissemination***

The Digital Archive can disseminate a single file up to 1GB in size to a web browser. In order to request a single-file dissemination you construct a URL based on the unique identifying elements of the file.

### **Fees**

There is no fee for single-file disseminations. These online disseminations are counted in the monthly Storage and Activity Report for your collections in the Archive.

### **Configuring a Single-file Dissemination URL**

The URL is constructed from these elements:

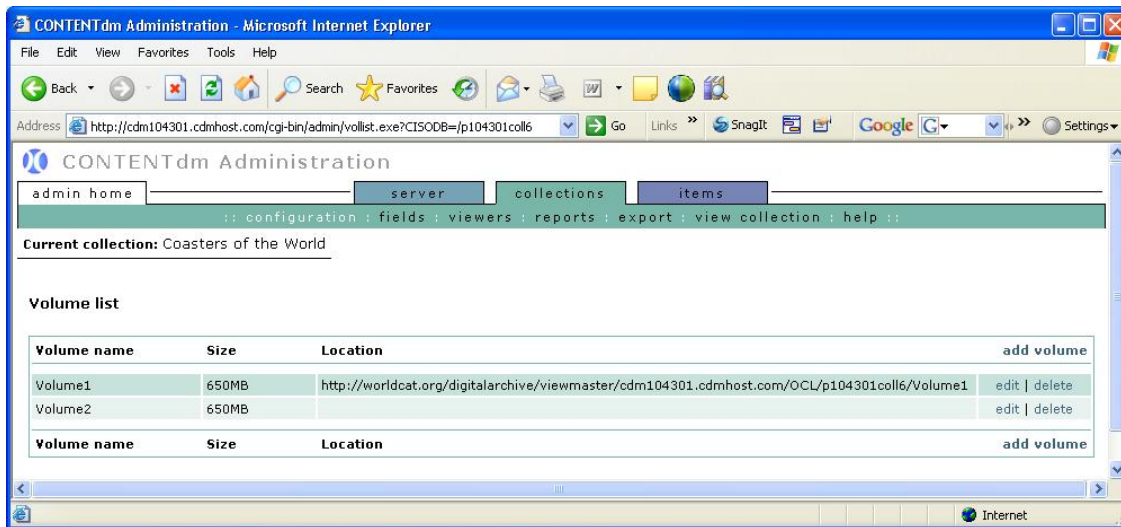
- ROOT → <http://worldcat.org/digitalarchive/viewmaster>
- Institution identifier → /<OCLC symbol> ,e.g., /ZMY
- Server name → /<server> ,e.g., /cdm104301.cdmhost.com
- Collection Name → /<collection name> ,e.g., /p104301coll6
- Volume ID → /<volume ID> ,e.g., /Volume1
- File Name → /<file name> ,e.g., /1\_0001scan.tif

When these elements are put together, the URL looks like this:

[http://worldcat.org/digitalarchive/viewmaster/ZMY/cdm104301.cdmhost.com/p104301coll6/1\\_0001scan.tif](http://worldcat.org/digitalarchive/viewmaster/ZMY/cdm104301.cdmhost.com/p104301coll6/1_0001scan.tif)

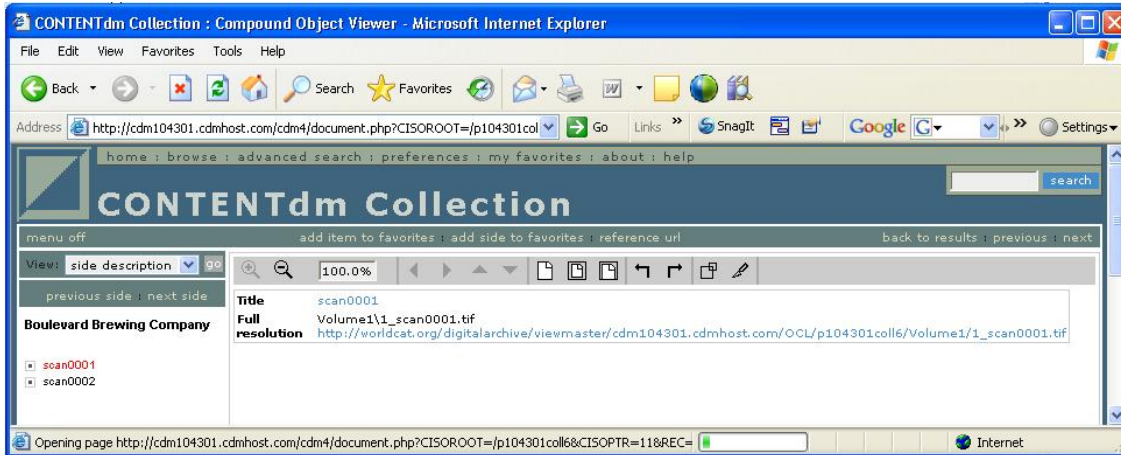
For CONTENTdm users the Digital Archive URL for single file dissemination can be managed in CONTENTdm Administration using Full Resolution settings and Volume Location settings. When you upload content to your CONTENTdm server in a collection with Full Resolution enabled, the Acquisition Station transmits both the VolumeID and the File Name to the **FullRes** metadata field for each file uploaded.

Using CONTENTdm Administration, you can edit the Full Resolution Volume List “Location” field and enter the first part of the URL described above (ROOT, Institution identifier, Server Name, Collection Name, VolumeID). Instructions for managing Volume List “Locations” can be found in CONTENTdm training materials under the heading **Moving Full Resolution Files to Permanent Locations** (<http://www.contentdm.com/USC/tutorials/full-res.pdf>). Entering this information once for each Archival Volume produces a URL for each file in the collection.



Example 9: View of CONTENTdm Administration with Volume List “Location” fields

If you’ve edited the full resolution volume location, and have set the Full Res metadata field to “viewable,” the metadata for the item in CONTENTdm will show a “live” hyperlink to the archived file.



Example 10: View of CONTENTdm Full Resolution Field in End-user Display

## Support

The Digital Archive is monitored 24x7. If you have support questions, please send an e-mail to [support@oclc.org](mailto:support@oclc.org) or call OCLC Customer Support at +1 800 848 5800 from 7:00am – 9:00pm U.S. Eastern Time.