

# Batch Editing CONTENTdm Records

Anneliese Dehner

Some of the content in this documentation has been adapted from OCLC's Support & Training [CONTENTdm Help files](#).

Version 1.2, January 2018

## Table of Contents

[Overview](#)

[Requirements, permissions, and limitations](#)

[Batch export](#)

[Compound objects](#)

[Cleanup exported data](#)

[Prepare for batch import](#)

[Prepare your CONTENTdm settings file](#)

[Prepare your .csv import file](#)

[Batch import](#)

[Review & re-index](#)

## Overview

This document is intended to guide CONTENTdm users through the process of batch exporting a tab-delimited text file of metadata, cleaning up that metadata, and then batch importing it back into CONTENTdm. The batch import process, described in this document, uses Jamie Little's [Pitcher gem](#) and [CONTENTdm's Catcher web service](#) to batch edit existing CONTENTdm metadata. The import process imports a .csv file, containing edited CONTENTdm metadata, then overlays the existing CONTENTdm metadata. When a CONTENTdm record is edited with this batch import process, the existing reference URL and record ID for an item remain unchanged.

## Requirements, permissions, and limitations

Please review the following requirements, permissions, and limitations before using this batch import process.

### *System requirements*

- **Minimum CONTENTdm version required:** 6.x

- **Hosting compatibility:** Compatible with both self-hosted and OCLC-hosted CONTENTdm systems.
- **Self-hosted CONTENTdm systems:** The server hosting CONTENTdm must be publicly accessible. If the CONTENTdm server is protected by a firewall, then you must setup a rule to allow access for the Catcher service IP address. The Catcher Web server IP range (132.174.\*.\*) should be designated as trusted sites. Port 81 is used by the CONTENTdm server on Linux, and port 2012 is used by the server on Windows.

### *Permissions*

- **CONTENTdm user credentials** are submitted by a script in this batch import process, and these credentials (CONTENTdm username and password) should have permission to approve items and build indexes, edit metadata, run reports and export metadata.

### *Limitations*

- **One field per record:** This batch import process imports one field in a CONTENTdm record at a time. Your .csv file for batch import will be formatted with one record per row and each row will contain value(s) to populate one CONTENTdm field.
- **Controlled vocabulary:** When editing a CONTENTdm field for which controlled vocabulary is enabled, make sure the new field value is an approved term before importing.
  - Alternatively, you can disable the controlled vocabulary feature before batch importing, and then enable controlled vocabulary again when the import is complete, rebuilding the vocabulary with the imported values.
  - It may be possible to use CONTENTdm's approval process to "approve" the imported records containing controlled vocabulary terms. However, we have not tested this yet.
- **Locked records:** When a CONTENTdm record is locked, its metadata cannot be edited. Unlock records in CONTENTdm before editing their fields with this process.
- **Required fields:** When a CONTENTdm record contains empty required fields, its metadata cannot be edited. Set the field's "required" property to "no" before editing. Then set the field's "required" property back to "yes" after re-indexing your collection.
- 

### **Batch export**

To export a CONTENTdm collection's object and page level records as a tab-delimited file, follow the steps below and refer to CONTENTdm's [Exporting to Tab-delimited text files documentation](#) for more information.

1. Login to your CONTENTdm Administration.
2. Click on the "collections" tab.

3. Select a collection from the “Current collection” drop-down menu at the top of the page, and click on the “change” button.
4. Click “export”, and you will see an Export Metadata page, like the screenshot below.

CONTENTdm Administration

admin home | server | collections | items

profile | fields | website | reports | export | view collection | help

Current collection: Jackson Collection | change

**Export metadata**

Choose the method used to export metadata.

☒ **Tab-delimited.** Exports data as a tab-delimited text file.

☒ Return field names in first record

☐ **XML.** Exports data as XML.

☒ Standard Dublin Core XML

☐ CONTENTdm Standard XML

☒ Include only the full text field from page-level metadata

☐ Include all page-level metadata

☐ Custom XML

☐ Include all page-level metadata

☐ **OCLC SiteSearch.** Exports data that can then be loaded to an OCLC SiteSearch database.

☒ Dublin Core. Produces the SGML and DTD files that can be used to load a defined Dublin Core database.

☐ Custom. Define tags used in the exported SGML file, which will then be used with the DTD file to load a custom SiteSearch database.

next

powered by CONTENTdm® | ©2001–2017 OCLC. All rights reserved. | to top

5. Select “Tab-delimited”.
6. Check “Return field names in first record”.
7. Click “Next”, and the Export successful page will open.

server16094.contentdm.oclc.org

CONTENTdm Administration

admin home | server | collections | items

profile | fields | website | reports | export | view collection | help

Current collection: Jackson Collection

**Export successful**

Right-click on the link and choose **Save Target As** to save the file.

Tab-delimited text file | export.txt

powered by CONTENTdm® | ©2001–2017 OCLC. All rights reserved. | to top

8. Your tab-delimited text file will appear as a link that reads “export.txt”.
9. Download your tab-delimited text file by right-clicking on the “export.txt” link.

## Compound objects

If you would like to remove the page-level records from your exported tab-delimited text file follow the steps below:

1. Import your tab-delimited text file into OpenRefine. (Refer to “[OpenRefine Metadata Cleanup](#)” for information about using OpenRefine.)

jackson xlsx    [Permalink](#)

---

**Facet / Filter**    [Undo](#) / [Redo](#)

### Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

## 110 records

Extensions: undefined | RDF

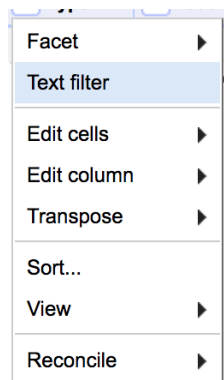
Show as: rows records    Show: 5 10 25 50 records

[<](#) first < previous **1 - 50** next > last >

file	Subject	Description	Creator	Type	CONTENTdm nu	CONTENTdm fil	CONTENTdm file path
Day By Day 1				text	0	1.pdfpage	/p15300coll1/image/1.pdfpage
Day By Day 2				text	1	2.pdfpage	/p15300coll1/image/2.pdfpage
Day By Day 3				text	2	3.pdfpage	/p15300coll1/image/3.pdfpage
Day By Day 4				text	3	4.pdfpage	/p15300coll1/image/4.pdfpage
Day By Day 5				text	4	5.pdfpage	/p15300coll1/image/5.pdfpage
Day By Day	Proverbs 27	Sermon	Jackson, John H.	text	5	6.cpd	/p15300coll1/suapp/6/index.pdf
e-fold cord 1				text	6	7.pdfpage	/p15300coll1/image/7.pdfpage
e-fold cord 2				text	7	8.pdfpage	/p15300coll1/image/8.pdfpage
e-fold cord 3				text	8	9.pdfpage	/p15300coll1/image/9.pdfpage
aefold Cord	Proverbs 27	Sermon	Jackson, John H.	text	9	10.cpd	/p15300coll1/suapp/9/index.pdf
ninent Christian worker 1				text	10	11.pdfpage	/p15300coll1/image/11.pdfpage
ninent Christian worker 2				text	11	12.pdfpage	/p15300coll1/image/12.pdfpage
ninent Christian Worker	Proverbs 10	Sermon	Jackson, John H.	text	12	13.cpd	/p15300coll1/suapp/12/index.pdf
sion-the child's thought of the parent 1				text	13	14.pdfpage	/p15300coll1/image/14.pdfpage
sion - the Child's Thought of the Parent				text	14	15.cpd	/p15300coll1/suapp/14/index.pdf
temper 1	Proverbs 4	Sermon	Jackson, John H.	text	15	16.pdfpage	/p15300coll1/image/16.pdfpage

In the “CONTENTdm file name” column, you will see filenames with the extension .pdfpage. These rows are your page-level records.

- Click the drop-down arrow at the top of the “CONTENTdm file name” column.



- Click “Text filter” and a “CONTENTdm file name” text filter will display in the “Facet / Filter” panel at left.


jackson.xlsx
[Permalink](#)

Facet / Filter
Undo / Redo 2

Refresh
Reset All
Remove All

☒ CONTENTdm file name

page

☐ case sensitive
☐ regular expression

88 matching records (110 total)

Show as: rows records
Show: 5 10 25 50 records

▼ All	▼ Title	▼ Subject	▼ Description	▼ Creator	▼ Type
Facet	By Day 1				text
Edit rows	Star rows				text
Edit columns	Unstar rows				text
View	Flag rows				text
7. A threefold	Unflag rows				text
8. A threefold					text
9. A threefold	Remove all matching rows				text
11. A prominent Christian worker 1					text

- OpenRefine: Refer to “[OpenRefine Metadata Cleanup](#)” for information about using OpenRefine.
2. Your data will be arranged in tabular format, one column for each CONTENTdm field and one row for each record. You can safely remove the columns of data that you won’t be cleaning up. **You must not delete the ‘CONTENTdm Number’ column however.** The data in this column is required for the CONTENTdm batch import process.
3. Edit your data.

## Prepare for batch import

### *Prepare your CONTENTdm settings file*

The settings file is a .yaml file, containing your CONTENTdm credentials. The script in this batch import process will use your credentials, as entered in the settings file, to communicate with CONTENTdm's Catcher web service and your CONTENTdm system. You only need to enter your credentials in the settings file one time and you can re-use this file whenever you batch import with this process. The settings file will not be stored online, it will be deleted immediately after it is read by the script.

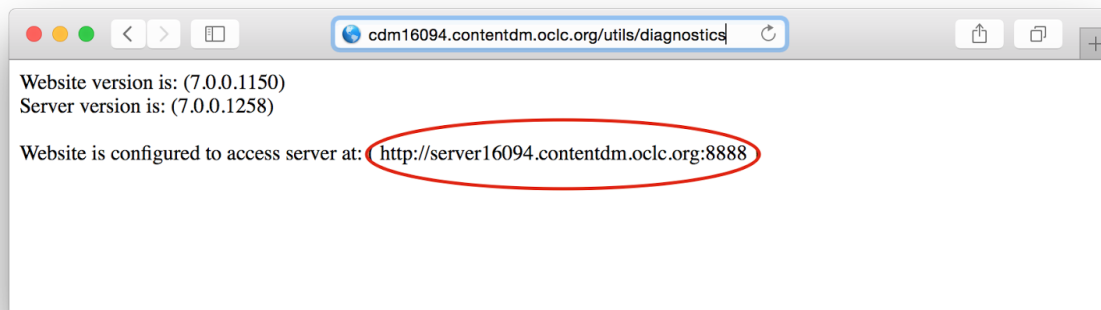
Follow these steps to prepare your CONTENTdm settings file:

1. Download the settings file template here: [CONTENTdmSettingsTemplate.yaml](#)
2. Open the settings file template with a text editor (eg. Notepad, TextEdit, etc.) and you should see the following:

```
cdmurl: http://contentdmurl:81
username: myusername
password: mypassword
license: contentdm-license-number
```

3. Find your server URL by visiting your CONTENTdm diagnostics page:

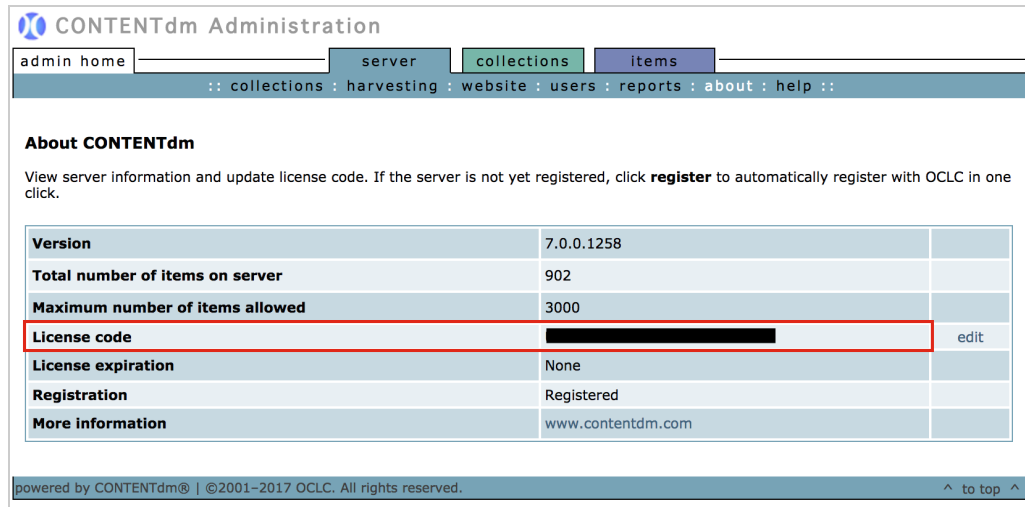
[http://\[CONTENTdm Site Url\]/utils/diagnostics](http://[CONTENTdm Site Url]/utils/diagnostics)



The page will look like the screenshot above. It will display your CONTENTdm versions and server URL. Replace “http://contentdmurl:81” in the settings file template with the CONTENTdm server URL you see in your CONTENTdm diagnostics page.

4. Replace “myusername” in the settings file template with your CONTENTdm username. See “Permissions” in the “Requirements, permissions, and limitations” section above for information about the user permissions required to complete the batch import process.

- Replace “mypassword” in the settings file template with your CONTENTdm password.
- Find your CONTENTdm license code by logging into your CONTENTdm Administration, clicking on the “server” tab, and then clicking “about”.



**CONTENTdm Administration**

admin home **server** collections items

:: collections : harvesting : website : users : reports : about : help ::

### About CONTENTdm

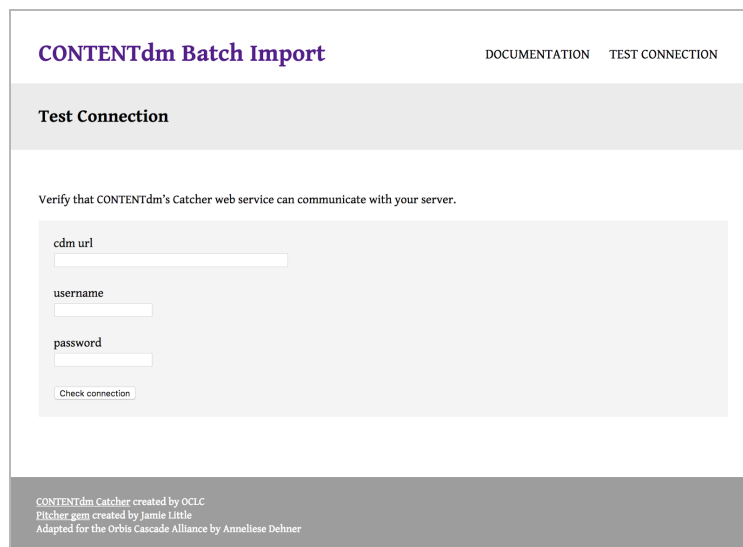
View server information and update license code. If the server is not yet registered, click **register** to automatically register with OCLC in one click.

<b>Version</b>	7.0.0.1258	
<b>Total number of items on server</b>	902	
<b>Maximum number of items allowed</b>	3000	
<b>License code</b>	[REDACTED]	<a href="#">edit</a>
<b>License expiration</b>	None	
<b>Registration</b>	Registered	
<b>More information</b>	www.contentdm.com	

powered by CONTENTdm® | ©2001–2017 OCLC. All rights reserved. [^ to top ^](#)

The About CONTENTdm page will open, and you will see your license code in the table presented to you. Replace “contentdm-license-number” in the settings file template with your CONTENTdm license code.

- Save your CONTENTdm settings file as **UsernameContentdmSettings.yml**
- Login to <http://yourdomain.com/testconnection> (username: admin password: admin) and you will see a Test Connection page like the one below:



**CONTENTdm Batch Import** [DOCUMENTATION](#) [TEST CONNECTION](#)

### Test Connection

Verify that CONTENTdm's Catcher web service can communicate with your server.

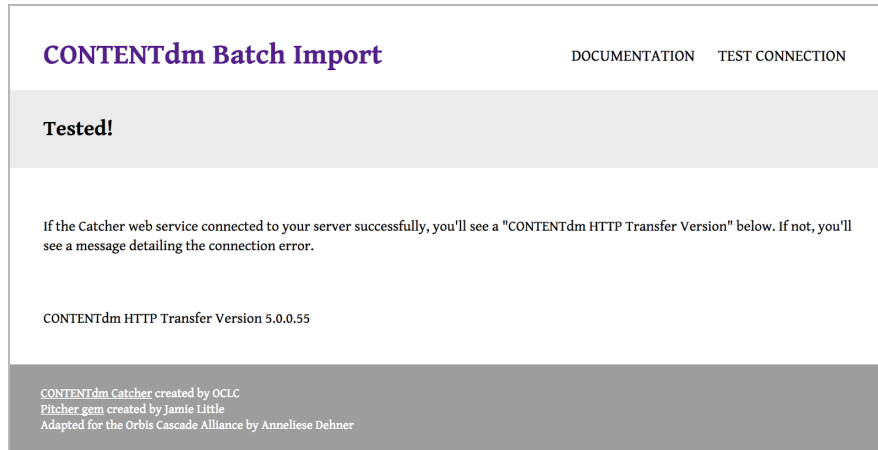
cdm url

username

password

CONTENTdm Catcher created by OCLC  
 Pitcher gem created by Jamie Little  
 Adapted for the Orbis Cascade Alliance by Anneliese Dehner

- To test your connection, enter your cdm url, username, and password. Click “Check connection” and you will see a Tested! Page like the one below:



When the Catcher web service connects to your server successfully, you'll see a CONTENTdm HTTP Transfer Version on this page. If the web service failed to connect to CONTENTdm, then you will see a message detailing the connection error.

If you encounter problems connecting, revisit “[System requirements](#)” in the “[Requirements, permissions, and limitations](#)” section above.

## *Prepare your .csv import file*

### About the .csv import file format

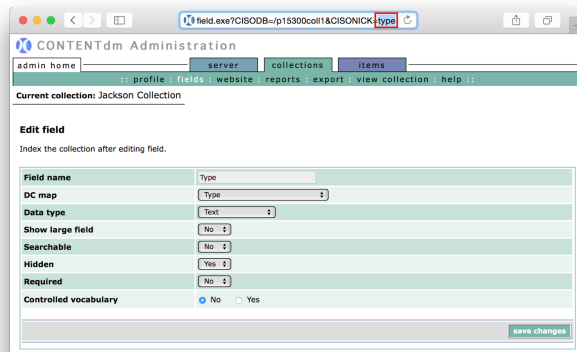
The .csv import file, used in this batch import process, must be formatted in a specific way. The downloadable .csv import file template and “Populate the import file template” section below will guide you in preparing your data for batch import. After preparing your import file, your data will be arranged in rows. Each row will contain data to update one CONTENTdm field in one CONTENTdm record. This data will be present as five comma-separated values: cdmnumber, field, value, collection, and action.

The five values included in the import file are described here:

- **cdmnumber:** The CONTENTdm number is the CONTENTdm id of the record you are editing.
- **field:** The nickname of the field you are editing. This may not be the same as the field label you see in your CONTENTdm Administration. You can find the nickname for a single field by following these steps:
  1. Login to your CONTENTdm Administration.
  2. Click on the “collections” tab.
  3. Select a collection from the “Current collection” drop-down menu at the top of the page, and click the “change” button.
  4. Click “fields”.



5. Click “edit” in the row of the desired field
6. View the URL in your browser address bar. The nickname will be the value after “CISONICK=” as indicated in the screenshot below.



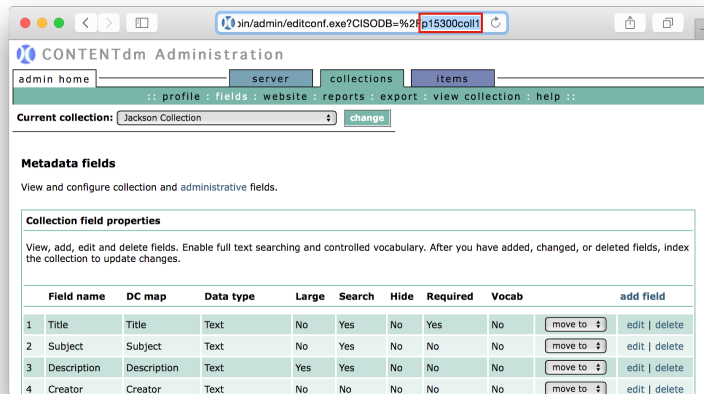
Or you can find the nicknames for all fields in a collection by viewing your Get Collection Field Info page:

[http://\[CONTENTdm Administration Site Url\]/dmwebservices/index.php?q=dmGetCollectionFieldInfo/\[collection alias\]/xml](http://[CONTENTdm Administration Site Url]/dmwebservices/index.php?q=dmGetCollectionFieldInfo/[collection alias]/xml)

```
<field>
  <name>
    <![CDATA[ Type ]]>
  </name>
  <nick>
    <![CDATA[ type ]]>
  </nick>
  <type>
    <![CDATA[ TEXT ]]>
  </type>
  <size>
    <![CDATA[ 0 ]]>
  </size>
  <find>
    <![CDATA[ BLANK ]]>
  </find>
  <req>
    <![CDATA[ 0 ]]>
  </req>
</field>
```

- **value:** This is the cleaned-up value that will be imported into a CONTENTdm field. This value will overwrite the CONTENTdm field’s existing value.
- **collection:** This is the collection alias, or handle, for the record you are editing. The collection alias is probably different from the collection name you see in your CONTENTdm system. You can find the collection alias for a collection by following these steps:
  1. Login to your CONTENTdm Administration.
  2. Click on the “collections” tab.

3. Select a collection from the “Current collection” drop-down menu at the top of the page, and click the “change” button.
4. View the URL in your browser address bar. The alias will be the value after “CISODB=%2F” or after “CISODB=/" as indicated in the screenshot below.



Or you can find the collection aliases for all collections by viewing your Get Collection List page:

[http://\[CONTENTdm Administration Site Url\]/dmwebservices/index.php?q=dmGetCollectionList/xml](http://[CONTENTdm Administration Site Url]/dmwebservices/index.php?q=dmGetCollectionList/xml)

```
<collections>
  <collection>
    <alias>p15300coll1</alias>
    <name>Jackson Collection</name>
    <path>/cdm/sites/16094/data/p15300coll1</path>
    <secondary_alias>p15300coll1</secondary_alias>
  </collection>
</collections>
```

- **action:** This is the action you are requesting for each record. The action will always be “edit”, which edits the existing metadata field of an existing record in a CONTENTdm collection.

Download the import file template

1. Download the import file template here: [.csv import file template](#)
2. Import the file into Google Sheets. If needed, refer to Google Sheets [Import data sets & spreadsheets documentation](#) for more information.

Populate the import file template

Follow the steps below to format your file for import:

1. Import your cleaned up tab-delimited text file (This is the cleaned up version of the file you exported from CONTENTdm) into Google Sheets. When Google Sheets asks whether to convert text to numbers and dates, select “No”. If needed, refer to Google Sheets [Import data sets & spreadsheets documentation](#) for more information about importing.

- Open the import file template in Google Sheets. You will be moving data from your tab-delimited text file into this template, so both Google Sheet files should be open simultaneously.
- In your tab-delimited text file, copy the cells in the “CONTENTdm number” column. Paste these cells into the import file template’s “cdmnumber” column.

	A	B	C	D	E
1	cdmnumber	field	value	collection	action
2					edit
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					

	E	F	G
	Type	CONTENTdm number	CONTENTdm field
	text	0	1.pdf
	text	1	2.pdf
	text	2	3.pdf
	text	3	4.pdf
	text	4	5.pdf
John	text	5	6.cpd
	text	6	7.pdf
	text	7	8.pdf
	text	8	9.pdf
John	text	9	10.cpd
	text	10	11.pdf
	text	11	12.cpd
John	text	12	13.cpd
	text	13	14.pdf
John	text	14	15.cpd
	text	15	16.pdf
	text	16	17.pdf
	text	17	18.pdf
	text	18	19.pdf

- From the columns of cleaned up data in your tab-delimited text file, choose one column to import.
- Copy the cells in this column. Paste these cells into the import file template’s “value” column.

	A	B	C	D	E
1	cdmnumber	field	value	collection	action
2					edit
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					

	D	E	F
reator	Type	CONTENTdm number	CONTENTdm field
	text	0	1.pdf
	text	1	2.pdf
	text	2	3.pdf
	text	3	4.pdf
	text	4	5.pdf
ackson, John	text	5	6.cpd
	text	6	7.pdf
	text	7	8.pdf
ackson, John	text	8	9.pdf
	text	9	10.cpd
	text	10	11.pdf
ackson, John	text	11	12.cpd
	text	12	13.cpd
ackson, John	text	13	14.pdf
	text	14	15.cpd
	text	15	16.pdf
	text	16	17.pdf

- In the import file template, fill the “collection” column’s cells with the collection alias of the collection you exported from CONTENTdm. See “collection” in the [“About the .csv import file format”](#) section above for information about collection aliases.
- In the import file template, fill the “field” column’s cells with the nickname of the CONTENTdm field you will be editing. See “field” in the [“About the .csv import file format”](#) section above for information about field nicknames.
- In the import file template, fill “action” column’s cells with the word “edit”.

9. The import file template should be populated with five columns of data, arranged left to right, with the column headers: cdnumber, field, value, collection, action. Double check that your populated import file template resembles the spreadsheet below:

	A	B	C	D	E
1	cdnumber	field	value	collection	action
2	0	type	Text	p15300coll1	edit
3	1	type	Text	p15300coll1	edit
4	2	type	Text	p15300coll1	edit
5	3	type	Text	p15300coll1	edit
6	4	type	Text	p15300coll1	edit
7	5	type	Text	p15300coll1	edit
8	6	type	Text	p15300coll1	edit
9	7	type	Text	p15300coll1	edit
10	8	type	Text	p15300coll1	edit
11	9	type	Text	p15300coll1	edit
12	10	type	Text	p15300coll1	edit
13	11	type	Text	p15300coll1	edit
14	12	type	Text	p15300coll1	edit
15	13	type	Text	p15300coll1	edit
16	14	type	Text	p15300coll1	edit
17	15	type	Text	p15300coll1	edit
18	16	type	Text	p15300coll1	edit

10. If you have numerical data in your spreadsheet, format this data as plain text by following these steps:
- Select the cells with numerical values.
  - Click on “Format” in the Google Sheets navigation menu.
  - Hover over “Number” in the drop-down menu.
  - Select “Plain text” in the fly-out.
11. Download your spreadsheet as a Comma-separated values (.csv, current sheet) file.

## Batch import

Use your CONTENTdm settings file and your .csv import file to batch edit fields in CONTENTdm records. Follow the steps below to batch edit:

- Login to your CONTENTdm Administration.
- Verify that the records you will be editing are unlocked. If they are locked, unlock them. To view locked items, follow the steps in CONTENTdm’s [Lock Administration documentation](#).
- Login to <http://yourdomain.com> and you will see the CONTENTdm Batch Import file upload form, which is shown in the screenshot below.

username: **admin**  
password: **admin**

**CONTENTdm Batch Import** DOCUMENTATION TEST CONNECTION

**Edit CONTENTdm Records**

CONTENTdm Batch Import is a web interface for the gem Pitcher, a library to interact with CONTENTdm's Catcher web service. You can use this batch import process to do bulk updating of metadata via CSV file. Please read the [instructions for file preparation](#) before uploading files.

UPLOAD A DATA FILE (.csv)  
Choose File No file chosen

UPLOAD A CONTENTdm SETTINGS FILE (.yaml)  
Choose File No file chosen

Import

CONTENTdm Catcher created by OCLC  
Pitcher gem created by Jamie Little  
Adapted for the Orbis Cascade Alliance by Anneliese Dehner

4. Click “Choose file” to browse for your .csv import file and select it for upload.
5. Click “Choose file” to browse for your CONTENTdm settings file and select it for upload.
6. Click “Import” to upload your files and initiate the import process.

### *Import Report*

The import process run time depends on the number of edits in your batch. To monitor the progress of your import, open a new browser tab or window, and type the URL below into the browser address bar:

**[http://yourdomain.com/logs/response-\[YOUR USERNAME\].txt](http://yourdomain.com/logs/response-[YOUR USERNAME].txt)**

This is your import report. Reload the report web page to update the report’s content as your import progresses. (There is a screenshot of a sample import report and information about its contents on the next page of this document)

### *Failed Imports*

If your import is **not successful**, you will see an error page entitled “Uh oh...”

There are several reasons your import failed. Review the structure of your .csv file. Review the systems requirements. And if you’re really stumped, contact Anneliese ([anneliese@anneliesedehner.com](mailto:anneliese@anneliesedehner.com)) for assistance.

### *Successful Imports*

If your import is **successful**, you will see an Imported! page like the screenshot below:

CONTENTdm Batch Import

DOCUMENTATIONTEST CONNECTION

Imported!

Uploaded Files

The following files uploaded successfully:  
data: jacksonpcc.csv  
settings: AdehnerContentdmSettings.yml

Import Report

To review details for each update in your batch, visit your import report here:  
[View Report](#)  
  
You can save and download your import report here:  
[response-dehner.a.txt](#)

Next Steps

Login to your CONTENTdm Administration.  
Refer to the documentation [Batch Editing CONTENTdm Records](#), and follow the steps in the "Review & re-index" section.

The Imported! Page gives you feedback about your import:

- The names of the files you successfully uploaded. These files will be deleted from the server immediately after the script reads them.
- A link to a downloadable report for this import. Click the link to download your report. When you open the report, you will see a log like the screenshot below:

```
CONTENTdm Record Number:5
Transaction ID:cdmcatchws02pxdu:116:42319782178455
Edit initiated.

CONTENTdm Record Number:9
Transaction ID:cdmcatchws01pxdu:115:29731881075290
Edit initiated.

CONTENTdm Record Number:12
Transaction ID:cdmcatchws01pxdu:115:29731882323683
Edit initiated.
```

For each edit in your batch of imports, you will see the CONTENTdm record number of the edited record and the response from the Catcher web service. “Edit initiated” means the edit began successfully. If a record is locked or the Catcher web service encountered any other problems, you will see an error message following the CONTENTdm record number.

## Review & re-index

The Batch Import script begins editing your CONTENTdm records immediately after you click the “Import” button. CONTENTdm will lock each record as the Catcher web service begins editing the record’s data.

### *Review items in Lock Administration*

CONTENTdm locks a record when the record is edited, and the record remains locked until its collection is re-indexed. Review your locked records by following the steps below:

1. Follow the steps in CONTENTdm’s [Lock Administration documentation](#) to view the records that have been locked by CONTENTdm. **Do not unlock these records before re-indexing.** If you unlock these records before re-indexing your collection, the records will not be updated.
2. Compare the records in CONTENTdm’s Lock Administration to the successful edits in the report (response-[username].txt) you downloaded from the Batch Import tool. The records in your Lock Administration should match the successful edits in your report.
3. If they do not match, then there was a communication problem between OCLC’s Catcher web and CONTENTdm. Submit a service request with OCLC Support, and explain this discrepancy.

### *Approve items*

If your library uses CONTENTdm’s Approval queue, check the approval queue in CONTENTdm Administration for items requiring review. To approve items, follow the steps in CONTENTdm’s [Approving Items documentation](#).

### *Re-index collection*

The final step in this import process is to re-index your CONTENTdm collection. Indexing will add your new item metadata to the CONTENTdm database so that items can be searched and browsed. To re-index your CONTENTdm collection, follow the steps in CONTENTdm’s [Building the Index documentation](#). The re-indexing process will unlock and update your records.

After re-indexing you will see your new field values in your CONTENTdm records.