Getting Started with the CONTENTdm Catcher

This guide provides instructions for working with the CONTENTdm Catcher. Catcher enables you to batch load and edit the metadata of digital collections using a standard XML-based method so that you can run a batch process unattended.
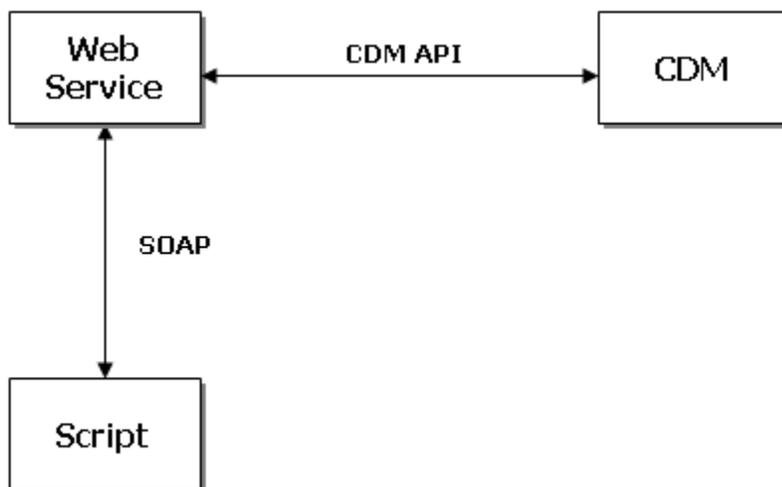
Overview

CONTENTdm Catcher is a Web service that supports batch requests to add metadata-only records to a CONTENTdm collection, edit existing metadata fields in a record, or delete a record. When processed through Catcher, the existing reference URL for an item will remain unchanged. Catcher provides an ideal mechanism for the addition of a batch of metadata records from another digital repository to CONTENTdm collections.

Catcher is a SOAP (Simple Object Access Protocol) Web service, intended for the advanced CONTENTdm user, that processes CONTENTdm data provided in an XML format. Interaction with this Web service is through a script that sends data in XML provided by the user. The initial release of Catcher will support requests to add or edit metadata in single items, and requests to delete single items. Later releases will support addition and replacement of digital files, along with the ability to add, edit or delete metadata in compound objects.

Catcher does not enable you to add or delete collections, only the metadata in items. Use CONTENTdm Administration to create and manage collections, including defining fields.

The following diagram illustrates the interaction between components:

- A script forms a message that the Web service understands.
- A script wraps the data into a message to send the Web service.
- A reply is provided to the script by the Web service.

Recommended steps to using Catcher:

1. Review all requirements and best practices detailed in this guide.
2. Organize all the metadata for items to be added (note: when editing items, make sure all fields with controlled vocabulary contain approved terms).
3. If editing or deleting items from collections, identify the Item ID (the same as the

dmrecord value or CONTENTdm number) for each item.
4. Review the sample scripts for how to write your own script.
5. Write a script for the transactions to submit.
6. Run your script.
7. Check the initial status of the transactions. A confirmation receipt containing a unique transaction ID will be returned for each valid request to edit, add or delete an item. Invalid requests will return an appropriate error message.
8. Approve items (approval required for adding items only) and build the index in CONTENTdm Administration.

Requirements

Check that you have the following before using the Catcher Web service:
- CONTENTdm Server version 6.x or later.
- Public access to your CONTENTdm Server. If your server is protected by a firewall, then you must setup a rule to allow access for the Catcher service.
- A script that sends data in canonical XML format and receives SOAP messages. The script may be written in a programming language of your choosing.
- The user credentials submitted in the script should have the appropriate permissions for the transactions you need (add, edit or delete). For more information about user credentials in your script, see the detailed code comments in the sample scripts. For more information about CONTENTdm user permissions, see **Working with Users**.
- If you are adding new items to a new collection, make sure you have completed setting up your collections. For more information, see **Collection Configuration**.

Best Practices

It is highly recommended that you back up all collection data prior to processing any script through Catcher.

Keep the transaction ID responses from each script as a reference for the action taken with each item. The transaction ID is useful for any troubleshooting.

Before processing any script transactions, it may be useful to run the getCONTENTdmHTTPTransferVersion call to verify authentication with your CONTENTdm Server. If the expected string is not returned as described in the sample script, this should be resolved prior to processing actual add, edit or delete requests.

Controlled vocabulary terms will not be validated by the Web service. If a record field contains controlled vocabulary, a warning message is provided, and the transaction request will proceed. You may review controlled vocabulary terms for a collection field by running the getCONTENTdmControlledVocabTerms call identified in each sample script. If you want to review the fields with controlled vocabulary, there is a getCONTENTdmCollectionConfig call that may be run, which contains the vocab element for each field.

Connecting to the Web Service

As noted in the requirements, the Catcher service requires public access to your CONTENTdm Server. If your server is protected by a firewall, you must setup a rule to allow access for the Catcher service IP address. The Catcher Web server IP range (132.174.2.*) 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.

For more information about connecting to the service, review the sample scripts' detailed code comments.

Scripting

The Catcher Web service requires that you write a script that sends data in canonical XML format and receives SOAP messages. The script can be written in any programming language.

You will want to verify that your script reflects the current WSDL version. The WSDL version is identified in the URL as follows:

https://worldcat.org/webservices/contentdm/catcher/6.0/CatcherService.wsdl

Note that your scripts must include the following information:

Required	Example / Notes
CONTENTdm Server URL	http://yourcollection.yourorgname.org
CONTENTdm Server license code	123456789abcdefghijk A valid 20-digit CONTENTdm Server license code to match the server URL.
Collection alias	/pNNNNqdc Existing CONTENTdm collection alias name.
Login credentials	Valid CONTENTdm user name & password.
Item ID number per record	Required to uniquely identify items to edit or delete. This is the dmrecord number or the CONTENTdm number.
Title field per record	Required to add items.
Add, edit, delete request for each record	Add: Adds a new record to a collection. Edit: Edits existing metadata field(s) of an existing record in a collection, or adds new metadata field(s) to an existing record in a collection. Delete: Deletes an existing record from a collection.

	For more information, see the sample scripts.

It's recommended that each script contain a 5-minute minimum timeout length. When processing larger batches of records, it may be useful for troubleshooting to have the script capture additional error details and to retry any failed transaction because of connectivity or SOAP-related issues.

The following date formats are supported by Catcher and should be used in your script: yyyy-mm-dd, yyyy-mm and yyyy.

For more information about writing your script, review the sample scripts with detailed code comments:
- **CatcherSample.java**
- **CatcherSample.cs** (in .zip)
- **CatcherSample.php** (in .zip)

Viewing Transaction Status

Each transaction in your script returns a receipt from the Catcher Web service, indicating the initial status of the process requested (edit, add or delete). A transaction ID is assigned per transaction; one ID for each item in your script. This initial status only indicates the action taken by the Web service, and does not provide a status of the final outcome of the transaction.

Final success or failure of each transaction is available through a report in CONTENTdm Administration. This Catcher Transaction Activity report is accessed from CONTENTdm Administration by clicking the **Collections** tab and then **Reports**. This downloadable report displays each transaction ID, the date and time processed, final status, request type and a short description.

Frequently Asked Questions

Q: How do I know if the requests in my script are being processed by the Web service?

A: You will immediately receive an indication of the initial status of each requested transaction for an edit, add or delete process. Each valid request will also be assigned a unique transaction ID. Here's an example of what you will see reported by the Web service:

```
Transaction ID: dcscatchws01dxdu:248:14078347143541
Delete initiated.
```

Q: What action do I take after my script has completed processing?

A: Check the approval queue in CONTENTdm Administration for items added through the Catcher Web service. Approve the items and build the index. Items edited or deleted through the Web service are automatically approved, so simply build the index to update your collection. Edit transactions with items containing unapproved controlled vocabulary terms will not be automatically approved. These items will require review in CONTENTdm Administration.

Q: Where do I find the Item ID for an item?

A: The Item ID is part of the URL of the item located after "id/". For example:

```
http://yourserver/cdm/singleitem/collection/name/id/nnn
```

The value of the Item ID should be reflected as the value of the dmrecord field for an item in your script. The CONTENTdm number of an item is also equal to the Item ID and dmrecord values. The export option in CONTENTdm Administration is an additional way to locate the CONTENTdm number of an item. For more information, see **Exporting a Metadata Subset**.

Q: What is the expected relative transaction time for each request type?

A: An Add request will be the slowest to process and a Delete request will be the quickest. An Edit request is most likely medium to short in duration to process.

Q: Will I be able to follow the regular workflow in CONTENTdm Administration for approving and indexing very large batches of records?

A: Yes, but performance of these processes may be noticeably impacted. You may want to consider an alternate workflow for approving and indexing tens of thousands of records all at once as described in this **Support FAQ**.