Guide to the CONTENTdm Catcher

Find information about requirements, best practices, scripting, and frequently asked questions about the CONTENTdm Catcher.

This guide provides instructions for working with the CONTENTdm Catcher. Catcher is a web service in your CONTENTdm site that enables you to add and edit metadata using a standard XML-based method.

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 any record, or delete a record. Catcher doesn’t alter the Reference URL for items you edit.

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 items and requests to delete single items.

Catcher does not enable you to add or delete collections, only to edit 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 and transmits the message using SOAP protocol to the Web service.
- The Catcher Web service interprets the request and takes the action requested.
- 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.
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

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 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 administration.

Best Practices

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.

Edits will not succeed unless all fields with controlled vocabulary are compliant. 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 vocabulary element for each field.

**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

Before writing your script, you’ll want to gather this information:

- **CONTENTdm Server URL** – Locate this using your CONTENTdm Website URL by adding `/utils/diagnostics` on your CONTENTdm Website URL to display it. For example, the CONTENTdm Sandbox site
  - https://sandbox.contentdm.oclc.org/utils/diagnostics shows this CONTENTdm Server URL
  - http://server10010.contentdm.oclc.org:8888
    Note: Catcher only operates on http today. There is no support for https.
- **User name** – A user name that has permissions to work with the collection you wish to manipulate.
- **Password** – The password for the user name.
- **CONTENTdm License Code** – Locate this in CONTENTdm Administration in the Server tab under About.

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:

- Java application download a text file
- PHP script download a .zip file with PHP text file inside

**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. In the Sandbox URL example the Item ID is 244.

https://sandbox.contentdm.oclc.org/digital/collection/textiles/id/244

The value of the Item ID should be reflected as the Value of the dmrecord Field for an item in your script. The export option in CONTENTdm Administration is an additional way to locate the CONTENTdm number of an item. For more information, see Export metadata.

**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.