Managing and highlighting collections using APIs

Andrew Bullen
Illinois State Library

Tomasz Kalata
BookOps / New York Public Library & Brooklyn Public Library
A Collection of Collections in CONTENTdm

A useful administration tool

Andrew Bullen
Illinois State Library
abullen@ilsos.gov
312-814-4386
A *collection of collections* is simply a CONTENTdm collection that houses data about other CONTENTdm collections.

Building a collection of collections— even if you only have one collection— adds another level of functionality.
A Village Grows - Fifty Years of Life in Elmwood Park

4
Local history images of Elmwood Park
dblCollections

Abraham Lincoln - Documents

5310
The Illinois State Library has a large collection of materials relating to the life, political career, and assassination of Abraham Lincoln. In addition to biographies of Lincoln, this collection includes contemporary accounts of the assassination and the trial of the conspirators; descriptions of Lincoln's funerals and the Lincoln National Monument; the restoration of New Salem; and Lincoln in Springfield.
dblCollections

Algonquin & Lake in the Hills Local History

4
The local history of Algonquin and Lake in the Hills
dblCollections

Alton Area Local History Collection

4
This is a collection of photographs and postcards related to the history of the Alton area. Includes topics such as historic buildings, downtown Alton street views, the Mississippi River, tombs and tombs, and select images from Godfrey, East Alton, Bunker Hill, Wood River, Grafton, Elsberry, and other nearby cities.
dblCollections

Architecture of McLean County

412
The McLean County Museum of History Print Collection has thousands of images of the built environment in the County, from the earliest cabins to construction of homes in the 21st century. The works of local architects such as Arthur L. Pittsboro, Paul O. Moritz, George Miller, Eugene Ascary and A.T. Simmons compose decades of architectural heritage in Illinois.
dblCollections

Arlington Heights History

4
Local history images of Arlington Heights
<table>
<thead>
<tr>
<th>Title</th>
<th>Arthur, Once Upon a Time - Local History Images of Arthur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>The community of Arthur, Illinois has a rich history, some of which has been recorded in the photographs in this collection. The first settlement of Arthur began about 1850. At that time, much of the area was swampland. Amish settlers arrived in 1865 and the Arthur area is now home to the largest Amish settlement in Illinois. In 1870</td>
</tr>
<tr>
<td>Creator</td>
<td>Arthur Public Library District: <a href="http://www.arthurlibrary.org/">http://www.arthurlibrary.org/</a></td>
</tr>
<tr>
<td>Identifier</td>
<td>api</td>
</tr>
<tr>
<td>Rights</td>
<td>Materials in this collection are made available by the Arthur P.</td>
</tr>
<tr>
<td>controlNumber</td>
<td>30020</td>
</tr>
<tr>
<td>subjectArea</td>
<td>4</td>
</tr>
<tr>
<td>cityName</td>
<td>Arthur</td>
</tr>
<tr>
<td>countyName</td>
<td>Douglas Moultrie</td>
</tr>
<tr>
<td>citationFields</td>
<td>create</td>
</tr>
<tr>
<td>Permissions</td>
<td>edit</td>
</tr>
<tr>
<td>OCLC number</td>
<td></td>
</tr>
<tr>
<td>Date created</td>
<td>2015-08-04</td>
</tr>
<tr>
<td>Date modified</td>
<td>2021-05-12</td>
</tr>
<tr>
<td>CONTENTdm number</td>
<td>19</td>
</tr>
<tr>
<td>CONTENTdm file name</td>
<td>20.jp2</td>
</tr>
</tbody>
</table>


INTRODUCTION

A Cookbook of Methods for Using CONTENTdm APIs

Andrew Bullen
Illinois State Library
312.814.4386
abullen@ilsos.net

CONTENTdm has a number of useful APIs for directly accessing information contained in its indexes and files. This site is intended as a practical reference guide to using many—though not all—of these APIs. The complete list of APIs can be found on the CONTENTdm support site (https://www.oce.org/support/services/contentdm/help/customizing-website-helper/customizations/contentdm_api_reference.en.html).

These programs use PHP.

Caveats

The test data I use is the data from the Illinois Digital Archives collection (http://www.idaillinois.org/), images and metadata contributed by institutions all across Illinois. Illinois Digital Archives is a service of the Illinois State Library and, especially, the office of Jesse White, Secretary of State and State Librarian.

All of these APIs can be used to display records as a JSON feed instead of HTML, by adding the parameter xml=1 to the end of the URL.
http://www.idaillinois.org/digital/custom/bySubject
---
header: true
footer: true
title: Collections in the Illinois Digital Archives
---

<div class="idaTables">
  <center>
    <h3 id="topOfList">
      <a href="#A">A</a>
      <a href="#B">B</a>
      <a href="#C">C</a>
      <a href="#D">D</a>
      <a href="#E">E</a>
      <a href="#F">F</a>
      <a href="#G">G</a>
      <a href="#H">H</a>
      <a href="#I">I</a>
      <a href="#J">J</a>
      <a href="#K">K</a>
      <a href="#L">L</a>
      <a href="#M">M</a>
      <a href="#N">N</a>
      <a href="#O">O</a>
      <a href="#P">P</a>
      <a href="#Q">Q</a>
      <a href="#R">R</a>
      <a href="#S">S</a>
      <a href="#T">T</a>
      <a href="#U">U</a>
      <a href="#V">V</a>
      <a href="#W">W</a>
      <a href="#X">X</a>
      <a href="#Y">Y</a>
      <a href="#Z">Z</a>
    </h3>
  </center>
</div>
function executeScript() {

    const queryString = window.location.search;
    const urlParams = new URLSearchParams(queryString);
    var requestedSubject = urlParams.get('subject');

    if (requestedSubject === null) {
        requestedSubject = 0;
    }

    var cdmStr = "/digital/bl/dmwebservices/index.php?q=dmQuery/p16614coll32/subjec^" + requestedSubject + "^all^and/title!descri!creato!identi!contro!subjec/title/500/1/0/0/0/0/json";

    const request = new XMLHttpRequest();
    request.open('GET', cdmStr);
    request.responseType = 'json';
    request.send();
    request.onload = function() {
        var obj = request.response;
        var recordCount = obj.pager.total;
        const table = document.querySelector("table");
        generateTable(table, obj, recordCount, requestedSubject);
        generateTableHead(table, requestedSubject);
    }
}

http://www.idaillinois.org/customizations/global/js/bySubject.js
function returnSubjects(subject) {
    let result = ({
        0: 'All Collections',
        1: 'Publications',
        2: 'Social Movements',
        3: 'History Makers',
        4: 'Community Histories',
        5: 'State History',
        6: 'Military History',
        7: '2020 in Review',
        8: 'Postcard Collections',
        9: 'Illinois Maps',
        10: 'Lincolniana',
        11: 'Oral Histories',
        12: 'Arts in Illinois',
        13: 'Institutional Histories',
        14: 'Industrial Histories',
        15: 'African-American Histories',
        16: 'Museum Collections',
        17: 'Genealogical Collections',
        18: 'Health Care',
        20: 'Transportation',
        21: 'Natural History'
    });
    var returnSubject = result[subject];
    return (returnSubject);
}
MILITARY HISTORY

Arlington Heights Military History
From: Arlington Heights Memorial Library
A collection of papers concerning the Civil War belonging to early Arlington Heights residents.
Subjects that this collection appears in: Military History.
ON Discovery

Charles Overstreet Collection
From: Florah Public Library
Charles Overstreet is a long-time citizen of Florah with a passion for photography. During most of his eighty years, Mr. Overstreet has used his camera to record images of history. During World War II, as a member of the U.S. Army, 252nd Field Artillery Battalion, he captured over 700 images of life as a soldier and unique snapshots of events of the war. Note: Please be aware that there are images (some graphic) of the atrocities committed at Ohrdruf, Germany in April 1945 in this collection.
Subjects that this collection appears in: Military History.
ON Discovery

DeWitt County World War I Collection
From: Waukon Public Library
Letters and service records of men and women of DeWitt County who served in World War I.
Subjects that this collection appears in: Community Histories, Military History.
ON Discovery

E.B. Johnson Collection. Grand Army of the Republic, Post 453
From: Elks Johnson Memorial Public Library District
The Grand Army of the Republic was an organization for Civil War Veterans that ran from 1866 until the death of its last member in 1956. Locally the organization was extremely active and the extensive collection of records left in the care of the Library provides a first-hand account of the period and the people that shaped it.
Subjects that this collection appears in: Community Histories, Military History.
ON Discovery
<table>
<thead>
<tr>
<th>Institution</th>
<th>Collection Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algonquin Area Public Library District</td>
<td><a href="http://www.idaillinois.org/digital/custom/byInstitution.js">Algonquin &amp; Lake in the Hills Local History</a></td>
</tr>
<tr>
<td>Arlington Heights Memorial Library</td>
<td><a href="http://www.idaillinois.org/digital/custom/byInstitution.js">Arlington Heights History</a></td>
</tr>
</tbody>
</table>
(function() {
    function executeScript() {

        var requestURL = '/digital/bl/dmwebservices/index.php?q=dmQuery/p16614coll32/0/creato!title!descri!identi!contro!subjec/creato!title/500/1/0/0/0/0/json';

        var request = new XMLHttpRequest();
        request.open('GET', requestURL);
        request.responseType = 'json';
        request.send();

        request.onload = function() {
            var obj = request.response;
            var recordCount = obj.pager.total;
            const table = document.querySelector("table");
            generateTable(table, obj, recordCount);
        }
    }
})()
ABOUT
Arlington Heights People & Places browse
From: Arlington Heights Memorial Library
Historical images of Arlington Heights people and places.
Subjects that this collection appears in: Community Histories;
undead

OAI Values
Root OAI: http://www.idallinois.org/oai/oai.php
Identify
Identifies the OAI provider:
http://www.idallinois.org/oai/oai.php?verb=identify
List sets
Lists all the collections (sets) on your server:
List metadata formats
Lists all the metadata formats on your server. The format “oai_dc” is required of all OAI providers.
List identifiers
Gives the header information only for all records:
For the next group of headers, use the “resumptionToken” value at the end of the results:
List records
List the records from one collection:
Lists the records from all collections (first group of records, in this case the first 1000):
For the next group of records, insert the “resumptionToken” value from the end of the results:
Get record
Get a particular record (e.g., item #22 in the collection):
(…adapted from an OAI guide from the University of Arkansas)
This XML file does not appear to have any style information associated with it. The document tree is shown below:

```
<field>
  <field>
    <title></title>
    <value></value>
    ...
  </field>
  ...
</field>
```

function executeScript(collectionName, recordNumber) {
    var cdmStr = "/digital/bl/dmwebservices/index.php?q=dmQuery/p16614coll32/identi^" + collectionName + "^all^and/citati!identi/citati/1/1/0/0/0/0/0/0/json";
    const request = new XMLHttpRequest();
    request.open("GET", cdmStr, true);
    request.send();
    request.onload = function() {
        if (request.status === 200) {
            const xml = request.response;
            var xmlDoc = xml.documentElement;
            var apiRecord = xmlDoc['records'];
            var citation = apiRecord[0]['citati'];
            if (citation) {
                var recordData = citation.split("|");
                var creatorLabel = recordData[0];
                var titleLabel = recordData[1];
                var mainRecordStr = "/digital/bl/dmwebservices/index.php?q=dmGetItemInfo/" + collectionID + "/" + recordNumber + "/json";
                var collectionName = myObj[collectionLabel];
                var institution = myObj[institutionLabel];
                var MLA = "<hr />
            
            <h3 style="color:#aa0000;">
                Cite this image:
            </h3>
            <p style="text-align: left;">
                MLA (v.8): " + author + ".

                + title + ", " + institution + ": " + collectionName + " (Illinois Digital Archives), " + dateOriginal + ", " + url + ". " + dateToday + ".</p>";
            
            http://www.idaillinois.org/customizations/global/js/citation.js
Questions? Comments?

Andrew Bullen
abullen@ilsos.gov
bookops-worldcat

Python wrapper around WorldCat Metadata API

Tomasz Kalata
BookOps | Metadata Manager
WorldCat Metadata API service

A read-write service that allows the user to search, create, update, and obtain full bibliographic records from the WorldCat database. The API provides also endpoints for maintaining library holdings and works with local bibliographic data.

Requires OCLC’s web service key (WSKey).

More info:

bookops-worldcat (v0.3.4)

- Python wrapper around WorldCat Metadata API
- Supports Python 3.7 and up
- Supports changes to Metadata API introduced in version 1.1 in May 2020:
  - OAuth 2.0 endpoints and flows
  - Uses Client Credential Grant and Access Token for authorization
- Interactions with the OCLC service are built with the popular Requests library ([https://docs.python-requests.org/en/master/](https://docs.python-requests.org/en/master/))
- It’s still in Alpha version ...
Search functionality

- Find member shared print holdings (/bibs-retained-holdings)
- Get a summary of holdings for known items (/bibs-summary-holdings)
- Brief bibliographic resources:
  - Search for brief bibliographic resources (/brief-bibs)
  - Retrieve a specific brief bibliographic resource (/brief-bibs/{oclcNumber})
  - Retrieve other editions related to a particular bibliographic resource (/brief-bibs/{oclcNumber}/other-editions)

Full bibliographic resources

- Retrieve a full bibliographic record (/bib-data)
- Get a current OCLC number (/bib/checkcontrolnumber)

Holdings

- Set and unset institution holding (/ih/data)
- Retrieve status of institution holdings (/ih/checkholdings)
- Set and unset institution holdings for a batch of records (/ih/datalist)
Virtual environment - an isolated environment for Python projects that prevents scripts interfering with the behavior of other Python applications running on the same system

create:
```bash
python3 -m venv myAppEnv
```

activate:
- ```myAppEnv\Scripts\activate.bat``` (Windows)
- ```myAppEnv/bin/activate``` (Unix/Mac OC)

More on virtual environments: https://docs.python.org/3/tutorial/venv.html
Setup: bookops-worldcat Installation

Make sure your virtual environment is activated

```
(myAppEnv) $ python -m pip install bookops-worldcat
```
Setup: Protecting WSKey

- Separate your WSKey from your code
- Do not commit your credentials to a version control system
- Store as a JSON file or similar and read it whenever needed

```json
1 {
2   "key": "my_WSKey",
3   "secret": "my_WSSecret",
4   "principal_id": "my_inst_principal_id",
5   "principal_idns": "my_ints_principal_idns",
6   "scopes": "WorldCatMetadataAPI",
7   "agent": "my_email@my_org.org"
8 }
```
Authorization & Authentication

```python
import os
import json

from bookops_worldcat import WorldcatAccessAuthToken

def get_token():
    creds_fh = os.path.join(os.getenv("USERPROFILE"), ".oclc/my_wskey.json")

    with open(creds_fh, "r") as file:
        creds = json.load(file)

    token = WorldcatAccessAuthToken(
        key=creds["key"],
        secret=creds["secret"],
        scopes=creds["scopes"],
        principal_id=creds["principal_id"],
        principal_idns=creds["principal_idns"],
        agent="tomaszkalata@bookops.org",
    )

    return token
```
from bookops_worldcat import MetadataSession

with MetadataSession(authorization=token) as session:
    response = session.get_brief_bib(oclcNumber="430840771")
    print(response.status_code)
    print(response.json())
MetadataSession: Features

- Provides convenient methods for making requests to various endpoints of the Metadata API
- Automatically requests a new token if the one provided happens to be expired
- Simplifies requests by circumventing some of the API's restrictions
  - OCLC numbers can be passed in any format (a string - with or without a prefix, or as an integer)
  - Allows updating library holdings for sets larger than 50 records (batch update)
- Inherits from requests.Session class retaining all of its behavior
MetadataSession: Session Arguments

`MetadataSession` object accepts following arguments:

- **authorization**: (mandatory) WorldcatAccessToken object
- **agent**: (optional) "User-agent" parameter to be passed in session’s request header (your email / your app name)
- **timeout**: (optional) how long to wait for the server to send data before giving up. The default value is 3 seconds
`MetadataSession` methods simplifies the user experience by keeping the same naming conventions for arguments as used in the Metadata API itself.

```python
with MetadataSession(authorization=token) as session:
    response_1 = session.search_brief_bibs(q="ti:zendegi AND au:greg egan")
    response_2 = session.search_brief_bibs(
        q="sn=40CC3B3F-4C30-4685-B391-DB7B2EA91455",
        inCatalogLanguage="eng",
        itemSubType="book-digital",
        orderBy="mostWidelyHeld",
        limit=1,
    )
```
Returned responses are standard `requests.Response` objects with all of their functionality.
Multiple requests within a single session

```python
from bookops_worldcat import MetadataSession
from bookops_worldcat.errors import WorldcatSessionError

oclc_numbers = [850933126, 850933139, 850933140, 850933141]

with MetadataSession(authorization=token) as session:
    for o in oclc_numbers:
        try:
            response_1 = session.holding_get_status(oclcNumber=o)
            data = response_1.json()
            holding = data['content']['holdingCurrentlySet']
            if not holding:
                response_2 = session.holding_set(oclcNumber=o)
        except WorldcatSessionError:
            pass
```
Writing retrieved full bibs to MARC21 file

```python
from io import BytesIO
from pymarc import XmlHandler, parse_xml
from bookops_worldcat import MetadataSession

def parse_xml_record(data):
    handler = XmlHandler()
    parse_xml(data, handler)
    return handler.records[0]

with MetadataSession(authorization=token) as session:
    response = session.get_full_bib(430840771)
data = BytesIO(response.content)
bib = parse_xml_record(data)
with open("bib.mrc", "wb") as out:
    out.write(bib.as_marc())
```

**pymarc** is an open source packages for working with MARC21

parses XML record from API response and converts it to a pymarc Record object

saves bib to a file in MARC21 format
Resources

bookops-worldcat:

Repository: https://github.com/BookOps-CAT/bookops-worldcat

Documentation: https://bookops-cat.github.io/bookops-worldcat/0.3/

requests package: https://docs.python-requests.org/en/master/

pymarc package: https://pymarc.readthedocs.io/en/latest/

Contact: tomaszkalata at bookops.org or klingaroo at gmail.com
thank you