Texas A&M University Libraries
DuraCloud API

Presentation at TCDL 2015
27 April 2015
Several TDL members conducted a pilot of the DuraCloud instance at TDL
Questions arose about reporting and tracking
Looked to API for solution
Switch to Browser

View of Web Interface for TAMU-S3
Why use the API

The API allows us to enhance, extend and customize the interface. We can use it to automate processes and actually create curation tasks. We can quickly fill the gaps.
Very Good DuraSpace Documentation

The documentation at DuraSpace was a real help

- DuraCloud REST API
- DuraCloud REST API Examples using CURL
Some Terminology

- REST – Representational State Transfer
- API – Application Programming Interface
- cURL – command line tool for getting/sending URLs
Getting Started

- Need the host name
- Userid / Password
- SpaceID
- ContentIDs

The userid controls what you can see in the “cloud”.
What Spaces Can I See

curl -u ID:PSW https://duracloud.tdl.org/durastore/spaces

<?xml version="1.0" encoding="UTF-8"?>
<spaces>
  <space id="gads-test-duracloud-space" />
  <space id="image-gallery-test" />
  <space id="tamu-public" />
  <space id="tamu-s3" />
  <space id="tdl-public" />
  <space id="tpn-tamu-test" />
</spaces>
curl -u ID:PSW https://duracloud.tdl.org/durastore/tamu-s3

<?xml version="1.0" encoding="UTF-8"?>
<space id="tamu-s3">
<item>1898_Thesis_Ness_color.pdf</item>
<item>Feed/28171888b0001.zip</item>
<item>DI/HTFeed/28191888b0003.zip</item>
<item>DI/HTFeed/Generate_SIP.log</item>
<item>DI/HTFeed/JP.pl</item>
<item>DI/HTFeed/TessFork.pl</item>
<item>DI/HTFeed/compress.tif_jp2.pl</item>
<item>DI/HTFeed/feed_env.pl</item>
<item>DI/HTFeed/ingest_test.pl</item>
<item>DI/HTFeed/loop.txt</item>
<item>DI/HTFeed(validate_volume.pl</item>
<item>DI/OMG/OMG.zip</item>
<item>DI/OMG/OMG00101.xml</item>
<item>DI/OMG/OMG_files.txt</item>
<item>Scan_Metadata.docx</item>
<item>DI/TDL_DuraCloud_Msgrs.pdf</item>
<item>TPNtst/UpdatedLewis_0003.tif</item>
...
What Files Do I Have

curl -u ID:PSW https://duracloud.tdl.org/durastore/tamu-s3

<?xml version="1.0" encoding="UTF-8"?>
<space id="tamu-s3">
  <item>1898_Thesis_Ness_color.pdf</item>
  <item>1914_Thesis_H719.pdf</item>
  <item>DI/ArkAMGame.mp4</item>
  <item>DI/HTFeed/28171888b0001.zip</item>
  <item>DI/HTFeed/28191888b0003.zip</item>
  <item>DI/HTFeed/ChkSums.pl</item>

.....
What are the Attributes

curl -u ID:PSW https://...1898_Thesis_Ness_color.pdf

HTTP/1.1 200 OK
Server: Apache-Coyote/1.1
Set-Cookie: JSESSIONID=F3A...; Path=/durastore; Secure
x-dura-meta-content-file-modified: 2012-10-12T16:04:17.017
x-dura-meta-creator: mbolton
Content-MD5: ceca17436c9184bc7a0852e51ae44c2a
x-dura-meta-content-file-path: C:\TDLDuraCloud\1898_Thesis_Ness_color.pdf
x-dura-meta-content-file-created: 2012-10-18T10:54:59.059
ETag: ceca17436c9184bc7a0852e51ae44c2a
Content-Type: application/pdf
Content-Length: 32897707
Date: Mon, 26 Jan 2015 14:53:38 GMT
There has to be a better way

- I want more control over the process
- So, I will write a program to format the output
- I use PERL but any language will work
- I want to customize the output to include more information
Our first program will
- Make output easier to read
- And add file size and checksum to the listing
Show Code Sample

Show the GetChkSums.pl program
my $host="duracloud.tdl.org";
my $service="durastore";
my $spaceID="tamu-s3";
my $user="ID";
my $pword="PSW";
...
my $client = REST::Client->new();
$client->setHost( "https://$host/$service/$spaceID" );
$client->addHeader( "Authorization", "Basic ".encode_base64( $user.":".$pword) );
$client->HEAD(
    $contentID
);

if( $client->responseCode() >= 300 ) {
    print "Failed to read $contentID \n";
    next;
}

my @hdrs = $client->responseHeaders();
Format for Program

my $filename = $client->responseHeader('x-dura-meta-content-file-path');
my $chksum = $client->responseHeader('Content-MD5');
my $length = $client->responseHeader('Content-Length');
my $etag = $client->responseHeader('ETag');
my $creator = $client->responseHeader('x-dura-meta-creator');
## Checksum Report

<table>
<thead>
<tr>
<th>Filename</th>
<th>SpaceID</th>
<th>Length</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1898_Thesis_Ness_color.pdf</td>
<td>tamu-s3</td>
<td>31.37M</td>
<td>ceca17436c9184bc7a0852e51ae44c2a</td>
</tr>
<tr>
<td>1914_Thesis_H719.pdf</td>
<td>tamu-s3</td>
<td>7.02M</td>
<td>5d45cbd8df26278055ed3f3122f32119</td>
</tr>
<tr>
<td>DI/ArkAMGame.mp4</td>
<td>tamu-s3</td>
<td>606.89M</td>
<td>19fc98c471a93aa689c48c751be383cc</td>
</tr>
<tr>
<td>DI/HTFeed/28171888b0001.zip</td>
<td>tamu-s3</td>
<td>25.33M</td>
<td>bb8685744fbd21d5ae7a8384108ce0a</td>
</tr>
<tr>
<td>DI/HTFeed/28191888b0003.zip</td>
<td>tamu-s3</td>
<td>90.68M</td>
<td>7081c36db580fd55ed3b1f1f1145f52d</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What are the Attributes

curl -u ID:PSW https://...1898_Thesis_Ness_color.pdf

Server: Apache-Coyote/1.1
Set-Cookie: JSESSIONID=F3A...; Path=/durastore; Secure
x-dura-meta-content-file-modified: 2012-10-12T16:04:17.017
x-dura-meta-creator: mbolton
Content-MD5: ceca17436c9184bc7a0852e51ae44c2a
x-dura-meta-content-file-path: C:\TDLDuraCloud\1898_Thesis_Ness_color.pdf
x-dura-meta-content-file-created: 2012-10-18T10:54:59.059
ETag: ceca17436c9184bc7a0852e51ae44c2a
Content-Type: application/pdf
Content-Length: 32897707
Date: Mon, 26 Jan 2015 14:53:38 GMT
Add Metadata

- Put in something meaningful to my organization
- I want to monitor/manage my files

If you are adding metadata, you should have a plan. What do you want to accomplish?
Setting new metadata is a destructive process – that is, it resets all existing metadata in favor of the new data. So save your work. And, the API does not ask if this is OK.
$client->POST(
    $file, "", {
        'Content-Type' => 'text/plain',
        'x-dura-meta-content-name' => 'Files on TDL DuraCloud',
        'x-dura-meta-content-MyChksum' => $chksum,
        'x-dura-meta-content-owner' => 'mbolton',
        'x-dura-meta-content-customer' => 'Office of DI'
    }
);

die $client->responseContent() if( $client->responseCode() >= 300 );
Check the Data

- Add file and metadata
- Use cURL and PERL REST calls
- Can be list driven
my $curl=`curl -u ID:PSW -T $file https://.../$file`;

$client->POST(
    $file, ",", {
        'Content-Type' => 'text/plain',
        'x-dura-meta-content-name' => 'Files on TDL DuraCloud',
        'x-dura-meta-content-owner' => 'mbolton',
        'x-dura-meta-content-unit' => 'Digital Initiatives',
        'x-dura-meta-content-customer' => 'Michael.Bolton'
    }
);

die $client->responseContent() if( $client->responseCode() >= 300 );
Automate this process

- Run a job on schedule
- Look for an error
- Email me the results
Where to
Now

- Develop more robust interfaces
- Allow units to manage their content
- Go Web-Based