Intermediate DSpace: Metadata Imports & Exports

Today’s training topics include...

- Batch imports via Web UI
- Exporting Items, Collections, & metadata
- Batch edits & metadata cleanup
- Moving/mapping Items en masse
- Fun with spreadsheets!
Batch Imports Via Web UI (User Interface)
Why batch import?

Batch importer is good for uploading:

- Large number of Items all at once
- Large Bitstreams
- Anything where you can front-load your metadata and organization
  - Lets you bypass the 3 screens’ worth of Web Submit UI...
  - ...but you don’t get the 3 screens’ worth of guidance.
Batch importer: right sidebar, under “Content Administration”

Import Batch Load (ZIP)

Select a collection

Collection:
UT-RGV Training Community > Laura's Demo Collection

Select the collection you wish to submit an item to.

Choose File No file chosen

Upload Simple Archive Format ZIP
Try it!
Batch upload via web UI

1. Navigate to the Training Documents collection from last time and open up the Item “Intermediate DSpace”.

2. Click on the sample_collection zip file and save it to your machine.

3. On the right sidebar in DSpace, scroll down to “Content Administration”.
   - Click on “Batch Import (ZIP)”.

4. Use the pulldown menu to select your own collection.
   - Sorry, they are in some mystery order

5. Click “Choose File” and select the zip file you just downloaded.
Exporting from DSpace

- Items (zip files)
- Communities & Collections (zip files)
- Metadata (comma-separated text, viewable as a spreadsheet)
Try it!

Exporting metadata

Find “Export Metadata” in the sidebar on Community, Collection, or Item pages.

1. Go to your Collection
   - it should have the “Photographs” items you batch loaded

2. Click “Export Metadata”

3. Save the file, which will have a name like 123456789-XXXX.csv
   - The “XXXX” is the number of that collection, community, or item; 123456789 is the repository’s number.

4. Double click to open the file -- it should come up in Excel already in columns.
Result of a metadata export: UT-RGV Training Community

| id | collection | dc_acqisit | dc_contribute | dc_contributor | dc_date | dc_date_type | dc_description | dc_identifier | dc_language | dc_publisher | dc_subject | dc_title | dc_type | dc_type | dc_type | dc_type |
|----|------------|------------|---------------|----------------|---------|--------------|----------------|---------------|-------------|--------------|------------|----------|---------|---------|---------|---------|---------|
| 7070 | 123456789/5256 | 123456789/5 | McElfresh, Laura: 83 | 2009 | Chomp, rc Graham | apirates of dogs play | http://hdl.handle.net/123456789/5263 | Dogs | Snow day! | Image |
| 7082 | 123456789/5256 | (Unknown): seeb89b1 | 2015 | Pictures of puppies who look it! | http://hdl.handle.net/123456789/5265 | Dogs | Not my dogs | Image |
| 7031 | 123456789/5256 | Trask, Amy Kane: ba | 2015 | Emmie at Christmas in Tulsa | http://hdl.handle.net/123456789/5264 | Cats | Emmie at Christmas | Image |
| 7028 | 123456789/5256 | McElfresh, Laura: 83 | 2003-12 | Pictures of Graham, age approx | http://hdl.handle.net/123456789/5262 | Cats | Baby Graham | Image |
| 7028 | 123456789/5256 | McElfresh, Laura: 83 | 2004-01 | Bed, crate in which we see that Graham | http://hdl.handle.net/123456789/5261 | Cats | Cammi and Graham | Image |

Column headers are Dublin Core fields (only the ones we used)

Notice use of double pipe || to repeat a field/include multiple values

“id” is the item’s number

“collection” can have multiple values: owning collection (first listed) &
Metadata Cleanup & Batch Edits

- Edit metadata fields -- fix typos, standardize CaPiTaLiZaTiOn
- Add new metadata fields
- Remove unnecessary metadata fields (if needed)
- After editing the spreadsheet, upload it back into DSpace (“Import Metadata”) and approve the changes.
1. Look at the spreadsheet you exported earlier. Make any changes you like.
   ○ Example: the “Photographs of…” Items all have author “O. Nonymous”. Try changing it to something else.

2. Save your changes. Be sure to keep the spreadsheet in .csv format.

3. Under “Content Administration” on the right sidebar, click “Import Metadata” and select your .csv file.

4. Review your changes to make sure it’s what you wanted.

5. **BE SURE TO SCROLL DOWN & CLICK ON “APPLY CHANGES”**
Success! (Now scroll down and click “Apply Changes”.)

**Notice**
Upload successful

### Import Metadata

Pending changes are listed below for review

<table>
<thead>
<tr>
<th>Changes pending for item 7030 (123456789/5263)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add: (dc.description[en_US])</td>
</tr>
<tr>
<td>Remove: (dc.description[])</td>
</tr>
<tr>
<td>Remove: (dc.description[en_US])</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes pending for item 7032 (123456789/5265)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add: (dc.description[en_US])</td>
</tr>
<tr>
<td>Remove: (dc.description[])</td>
</tr>
</tbody>
</table>
...or fix your metadata and try again.

Notice
Unknown metadata element in heading
dc.nonsense

Import Metadata

No changes were detected

Return
...and did I mention, click on “Apply Changes”? 

Notice
Upload successful

Import Metadata

Pending changes are listed below for review

Changes pending for item 7030 (123456789/5263)

Remove: (dc.date.issued[])

Apply changes

THIS BUTTON HERE

SCROLL DOWN & CLICK IT

OTHERWISE DSPACE WON'T SAVE YOUR WORK

Return
Moving & Mapping Items

- We move items by changing the Collection number.
- We map items by adding Collection number(s) for the mapped Collection(s), separating the Collection numbers with a double pipe.
  - Leave the owning Collection first in the listing.
Try mapping some items into “Mapping Collection” within the UT-RGV Training Community.

Choose which items you’d like to map. Under “collection” in the metadata spreadsheet, append ||123456789/5284 to each entry.

Save your spreadsheet & import the metadata.
You know the drill. ("Apply Changes").

Notice
Upload successful

Import Metadata

Pending changes are listed below for review

Changes pending for item 7032 (123456789/5265)

Map to collection 123456789/5284 (Mapping Collection)

Apply changes

Return
Spreadsheet-Wrangling Techniques

Tips, tricks, questions

- Delete lines &/or columns you’re not editing
  - Especially since DSpace does have a limit on length of import spreadsheets
  - Documentation says 1,000 lines; I thought it was more like 200

- **CTRL+H** (Find/Replace) is your friend!

- Sorting your spreadsheets can help you edit more efficiently
Thank you!

Laura Kane McElfresh
Texas A&M University at Galveston
mcelfrel@tamug.edu