Batch Editing Digital Commons Records

Produced by the Digital Collections Working Group of the Content Creation & Dissemination Team
Janet Hauck, Whitworth University (chair)
Anneliese Dehner, Metadata Applications Librarian, Alliance (author of this document)
Laura Zeigen, Oregon Health & Sciences University
Julia Simic, University of Oregon
Theodore Gerontakos, University of Washington
Jodi Allison-Bunnell, Alliance (ex officio)

Content adapted from Reference and How-to Guides created by bepress, specifically Batch Upload, Export, and Revise.

Version 1.0, March 2017

Table of Contents

Overview
Permissions
Batch export
Cleanup exported data
Cleanup issues unique to Digital Commons

Overview

The batch export and batch revise processes described below edit existing Digital Commons metadata. The batch revise process imports an Excel file, containing edited Digital Commons metadata, then overlays the existing Digital Commons metadata. When a Digital Commons record is edited with this batch revise process, the existing reference URL and record ID for an item remain unchanged.
Permissions

If you can see the Batch tools in the left sidebar of Digital Commons, then you have adequate permissions to complete this batch export process. If you do not see “Batch Revise Excel” in the left sidebar, your local bepress administrator can assign these permissions to you. You can also contact your bepress rep, and request these permissions.

Limitations

- **Controlled lists**: Some fields, like “Document Type” and “Disciplines,” rely on controlled vocabulary lists. For these fields, specific values are required. If any fields in your spreadsheet use controlled lists, make sure to only use controlled values in these fields.

- **Modified records**: A record will fail in the batch revise process if it was edited, in Digital Commons, after you generated and downloaded your spreadsheet of records. This happens because Digital Commons considers the record in your spreadsheet to be outdated, a previous version of the record. In this case, you will need to regenerate and download the spreadsheet of records before proceeding with the batch revise process.

- **Date format**: A record will fail in the batch revise process if the record’s date field value is not formatted as a date (YYYY-MM-DD HH:MM) in your spreadsheet. When batch revising dates, make sure the dates in your spreadsheet are formatted as dates (YYYY-MM-DD HH:MM), instead of plain text.

- **Number of records**: If your collection is large (10,000+ submission), the batch export will be broken up into multiple spreadsheets (in increments of 5,000).

Batch export

To export a Digital Commons collection’s records as an Excel spreadsheet, follow the steps below.

1. Login to your Digital Commons repository.
2. Navigate to the collection/publication to export.
3. Click “Manage Submissions” at the top of the page, and you will see the Manage Submissions page.
4. Follow the steps “To batch export” in the “Batch Export and Revise” section of Digital Commons’ Batch Upload, Export, and Revise documentation.

After downloading a batch revise spreadsheet, please do not make any changes to records in Digital Commons. If you do, the Batch Revise process will fail during upload.
Cleanup exported data

Please consider the steps below as you clean up the data in your exported Excel spreadsheet:

1. Import your Excel file into Excel or OpenRefine for cleanup. Your spreadsheet includes a column for each metadata field in your collection, and a row for each record.
   - **Excel**: The fields highlighted in red are required fields, and the right-hand, gray columns should be left intact for system use. If you hover the cursor over a column header, additional information about each field will display.
   - **OpenRefine**: Refer to “OpenRefine Metadata Cleanup” in Alliance Digital Collections Documentation for information about data cleanup in OpenRefine.

2. Edit your data to be compliant with the Alliance Dublin Core Best Practices Guidelines. Refer to the “Cleanup issues unique to Digital Commons” section below for more information.

3. When you’re finished cleaning up your data, save the spreadsheet in Excel 97-2003 Workbook (.xls) format.

Cleanup issues unique to Digital Commons

*Embargo date*

For batch revisions, enter the date of expiration in YYYY-MM-DD format; if embargo date is a required field but the record is not under embargo, enter today’s date to proceed.

*Required fields*

The presence of the following fields is required in the batch export/batch revise spreadsheet.

- Title
- Author1_fname
- Author1_lname
- Embargo_date
- publication_date
**Controlled Lists**

Some Digital Commons fields, like “Document Type” and “Disciplines,” rely on controlled vocabulary lists. For these fields, specific values are required. If any fields in your spreadsheet use controlled lists, make sure to only use controlled values when you clean up these fields.

**Date format**

The Digital Commons publication date field requires the YYYY-MM-DD HH:MM format. When cleaning up this date field, enter new dates as YYYY-MM-DD HH:MM and make sure the dates in your spreadsheet are formatted as dates (YYYY-MM-DD HH:MM), instead of plain text.

To use uncertain/approximate dates or date ranges, see the “Adding additional fields” section below.

**Adding additional fields**

One way to conform to the Alliance’s Dublin Core Best Practices Guidelines, while retain your existing metadata, is by adding additional metadata fields that can be mapped to Dublin Core properties in the Digital Commons OAI output. These additional fields will be visible on the back-end, but they can be hidden so they do not display to your users. To add and configure additional fields, contact your bepress rep.

Additional fields may be added to accommodate the following:

- **Uncertain/approximate dates or date ranges**
  The Digital Commons “Publication Date” field requires the YYYY-MM-DD format. This works well for exact dates. But the data type of this field does not accommodate uncertain/approximate dates, which may include “?”, “~”, or “u”. Nor does it accommodate date ranges. However, you can work around this limitation by adding a text field and populating it with date ranges and/or uncertain/approximate dates. Then you can ask your bepress rep to map this new field to dc:date.

- **DCMI Type terms**
  The Digital Commons “Document Type” field uses a controlled vocabulary list. Perhaps you use this field for type values other than DCMI Type terms, and you are wondering how to include the required DCMI Type terms in your records. You can contact your bepress rep, request an additional field, and ask your rep to map this new field to dc:type. Then populate this field with the required DCMI Type terms.

**Batch revise**

Use your cleaned up spreadsheet to batch revise fields in Digital Commons records by following the steps below:

1. Login to your Digital Commons repository.
2. Navigate to the collection/publication you would like to revise.
3. Click “Manage Submissions” at the top of the page, and you will see the Manage Submissions page.

4. Click the “Batch revise Excel” sidebar link. **Do not click the “Batch upload Excel” link** because this function will create new Digital Commons records instead of editing values in existing Digital Commons records.

5. Follow the steps “To batch revise” in the “Batch Export and Revise” section of Digital Commons’ [Batch Upload, Export, and Revise documentation](#).

6. If a record fails to batch revise, you will receive an email explaining the source of the error. Make corrections to your spreadsheet and then re-upload. If the error relates to recency of changes, you will need to export a new spreadsheet, clean up, and then upload again.

**Check the status of a batch revision**

The batch status tool stores information about your batch revisions for 30 days. You can use the batch status tool at any time to check the status of your batch revision. To check the status of a batch revision, follow the steps below.

1. Login to your Digital Commons repository.

2. Navigate to the collection/publication you revised.

3. Click “Manage Submissions” at the top of the page, and you will see the Manage Submissions page.

4. Click the “Batch status” sidebar link.