

# Batch Metadata Editing

- 1 [Introduction](#)
- 2 [User Interface Functionality](#)
  - 2.1 [XML User Interface](#)
  - 2.2 [JSP User Interface](#)
- 3 [Command Line Functionality](#)
  - 3.1 [Export parameters](#)
  - 3.2 [Import parameters](#)
- 4 [The CSV Files](#)
- 5 [Editing Collection Membership](#)
- 6 [Adding Metadata-Only Items](#)
- 7 [Deleting Metadata](#)
- 8 [Performing 'actions' on items](#)
- 9 [Migrating Data or Exchanging data](#)
  - 9.1 [Common Issues](#)

## Introduction

DSpace provides a batch metadata editing tool. The batch editing tool is able to produce a comma delimited file in the CSV format. The batch editing tool facilitates the user to perform the following:

- Batch editing of metadata (e.g. perform an external spell check)
- Batch additions of metadata (e.g. add an abstract to a set of items, add controlled vocabulary such as LCSH)
- Batch find and replace of metadata values (e.g. correct misspelled surname across several records)
- Mass move items between collections
- Mass deletion, withdrawal, or re-instatement of items
- Enable the batch addition of new items (without bitstreams) via a CSV file
- Re-order the values in a list (e.g. authors)

For information about configuration options for the Batch Metadata Editing tool, see [Batch Metadata Editing Configuration](#)

### DSpace Versioning Compatibility



Out of the box, the batch metadata editing features do not support the DSpace versioning system. Changes are applied straight on the item metadata and no versions of these items are being generated and stored as part of these edit operations. Be careful when using these features.

## User Interface Functionality

Both the XML and JSP user interfaces allow administrator users to:

- **Export** metadata into a CSV file.
- **Import** metadata in a CSV, with the goal of modifying existing items or creating new ones.

## XML User Interface

In a few easy steps you can export metadata from a collection, change the metadata and import the changed metadata back into the repository. These screenshots were taken on the demo DSpace.

### 1) Exporting metadata for a particular collection

**DSpace Repository** Profile: Demo Administrator | Logout  
English Deutsch Français


DSpace Home → Sample Community → **Collection of Sample Items**

## Collection of Sample Items

**Browse by**

- By Issue Date
- Authors
- Titles
- Subjects

Search within this collection:  **Go**

  
open source initiative

This is a DSpace Collection which contains sample DSpace Items.

**Collections in DSpace may only contain Items.**

This particular Collection has its own logo (the [Open Source Initiative](#) logo).

This introductory text is editable by System Administrators, Community Administrators (of a parent Community) or Collection Administrators (of this Collection).

**News**

This is the **news** section for this Collection. System Administrators, Community Administrators (of a parent Community) or Collection Administrators (of this Collection) can edit this News field.

*If this collection had a specific copyright statement, it would be placed here.*

[Submit a new item to this collection](#)

**Recent Submissions**

1  
a, a (2012)  
a

**Search DSpace**

**Go**

☒ Search DSpace  
☐ This Collection

[Advanced Search](#)

**Browse**

All of DSpace  
[Communities & Collections](#)  
[By Issue Date](#)  
[Authors](#)  
[Titles](#)  
[Subjects](#)

This Collection  
[By Issue Date](#)  
[Authors](#)  
[Titles](#)  
[Subjects](#)

**My Account**

[My Exports](#)  
[Logout](#)  
[Profile](#)  
[Submissions](#)

**Context**

[Edit Collection](#)  
[Item Mapping](#)  
[Export Collection](#)  
[Export Metadata](#)

**Administrative**

[Access Control](#)  
[People](#)  
[Groups](#)  
[Authorizations](#)  
[Registries](#)

Whenever you are on a collection, you will have the possibility to export the metadata of that specific collection. You just have to click "Export Metadata" in the Context menu

## 2) Go to the Import Metadata page

The screenshot shows the DSpace Demo Repository homepage. A purple box highlights the 'DSpace Home' link in the top navigation bar. A purple arrow points from the text 'On the homepage you select "Import metadata" in the Administrative menu.' to the 'Import Metadata' link in the 'Administrative' menu on the right side of the page. The 'Import Metadata' link is circled in purple. The homepage content includes a welcome message, a list of enabled features, demo user information, a warning about automatic resets, and a list of communities.

**DSpace Repository** Profile: Demo Administrator | Logout  
English Deutsch français

DSpace Home

## DSpace Demo Repository

Welcome to the DSpace.org demonstration/sample repository (Manakin / XMLUI interface)!

This repository is currently running on **DSpace version 4.1**, with the following extra features enabled:

- Authority Control (on Authors and Publishers) in Submission process
- Community/Collection/Item Statistics are publicly visible
- OpenSearch
- Google Scholar metadata
- DSpace Discovery (faceted search/browse)
- SWORD Client
- Mobile Theme
- Elastic Search Statistics (login as admin and browse to any community/collection/item, then choose the last 'View Usage Statistics' link from the menu)

The following Demo Users are set up in the system (all users have password equal to the lowercase name of this software):

- Demo Site Administrator = dspacedemo+admin@gmail.com
- Demo Community Administrator = dspacedemo+commadmin@gmail.com
- Demo Collection Administrator = dspacedemo+colladmin@gmail.com
- Demo Submitter = dspacedemo+submit@gmail.com

**WARNING:** All content on this site resets automatically **every Saturday at 23:59 UTC**, at which point any changes are lost.

### Communities in DSpace

Select a community to browse its collections.

- [OCC](#)
- [Prueba Rio Negro](#)
- [Sample Community](#)
- [Test Community](#)
- [Test Community](#)

### Recently Added

Yoga  
Hitesh, P (2000-04-01)

### Search DSpace

  
[Advanced Search](#)

After you have altered the metadata, you can import it back into the repository quite simply. You just need to go to the homepage.

### 3) Choose file and upload CSV

The screenshot shows the 'Import Metadata' page in the DSpace Demo Repository. The page has a header with the DSpace logo and navigation links. The main content area includes a 'Choose File' button, a 'No file chosen' status, and an 'Upload CSV' button. The right sidebar contains search and browse options.

**DSpace Repository** Profile: Demo Administrator | Logout  
English Deutsch français

DSpace Home → Import Metadata

## Import Metadata

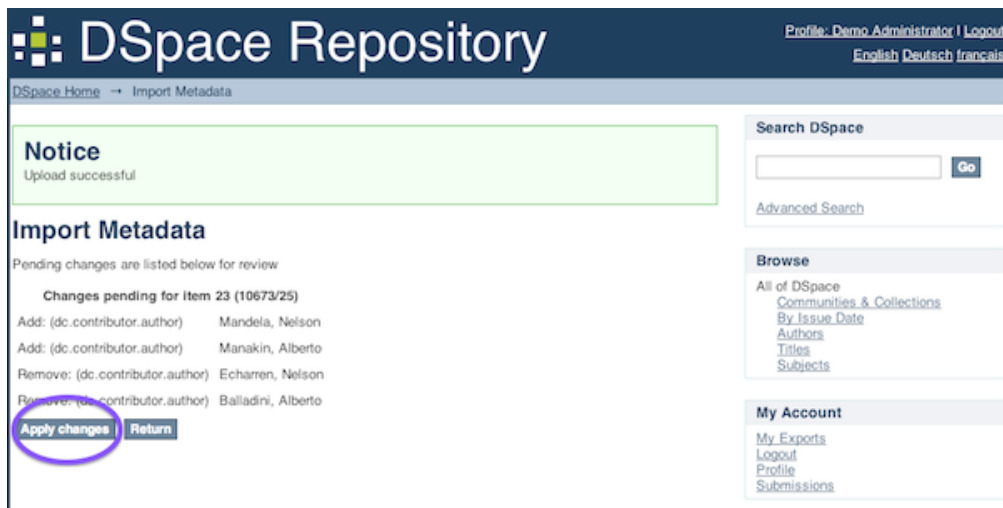
No file chosen

### Search DSpace

  
[Advanced Search](#)

Choose a file from your computer and click on "Upload CSV"

### 4) Review and approval of changes

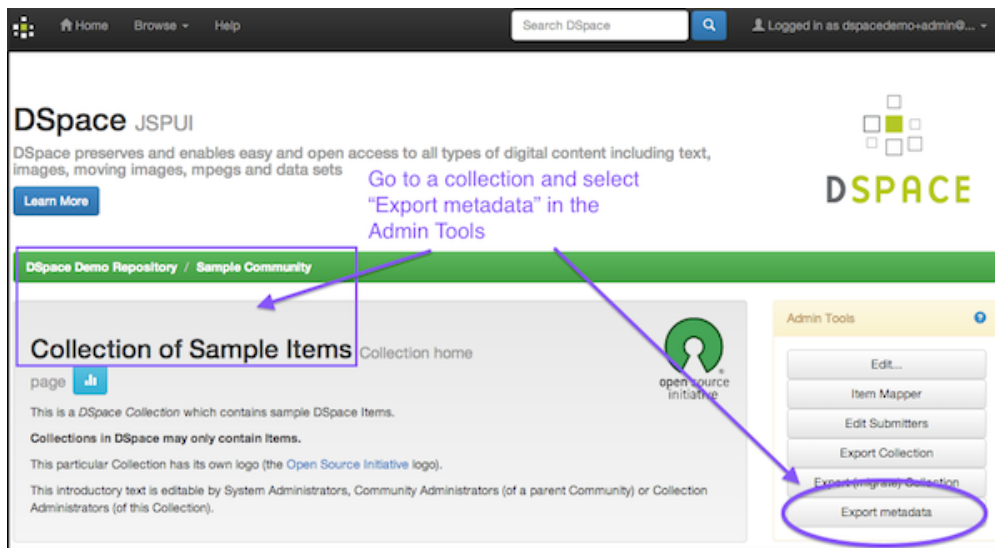


On this page you can see which changes you have made within the CSV-file. You can now either accept these changes and click "Apply changes" or not, in that case click "Return".

## JSP User Interface

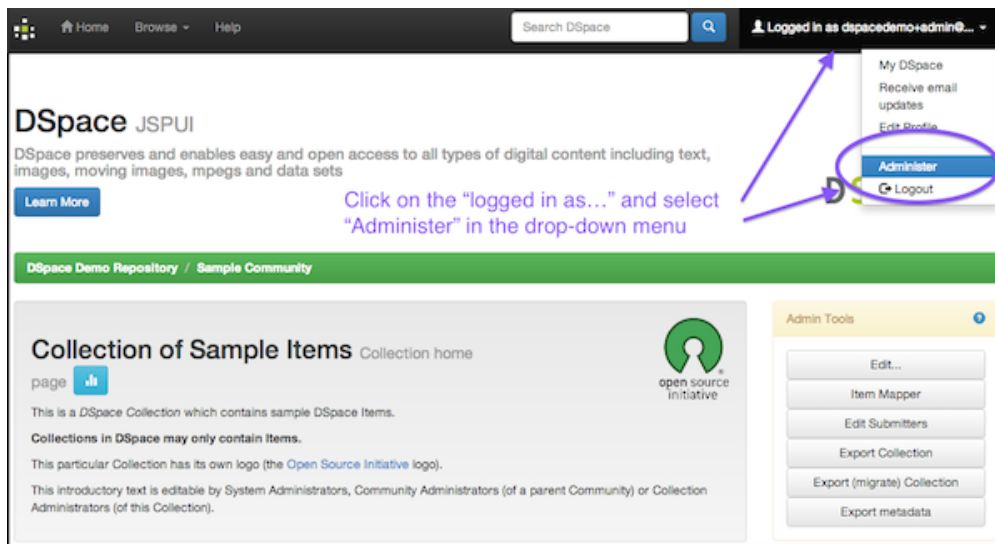
In a few easy steps you can export metadata from a collection, change the metadata and import the changed metadata back into the repository. These screenshots were taken on the demo DSpace.

### 1) Exporting metadata for a particular collection



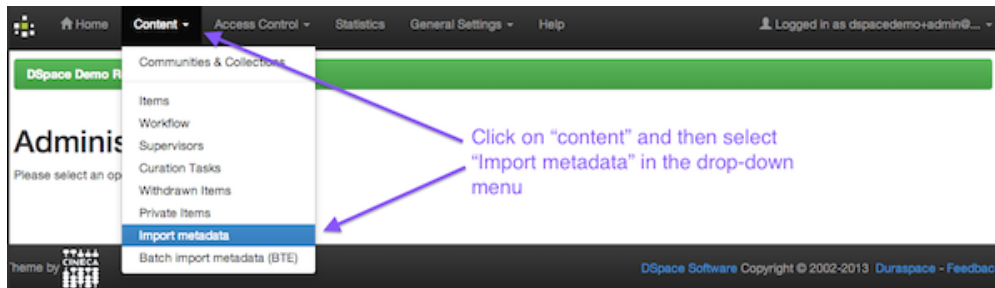
Whenever you are on a collection, you will have the possibility to export the metadata of that specific collection. You just have to click "Export Metadata" in the Admin tools.

### 2) Go to the Administrator Tools



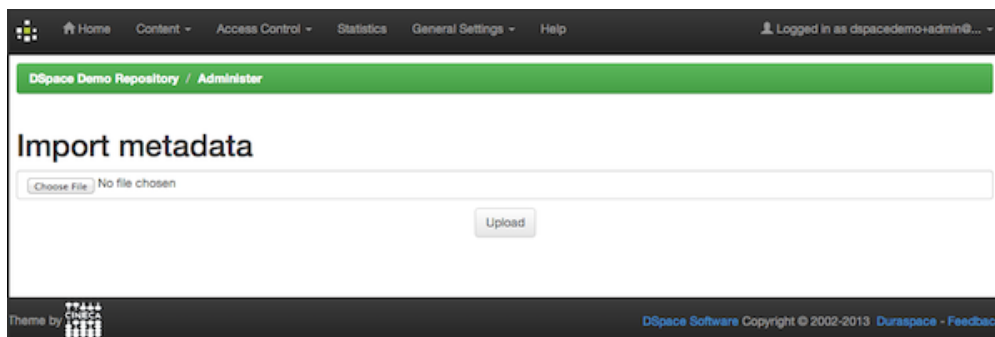
After you have altered the metadata, you can import it back into the repository quite simply. On (almost) every page of the repository you can access the administrator tools.

### 3) Go to the Import Metadata page



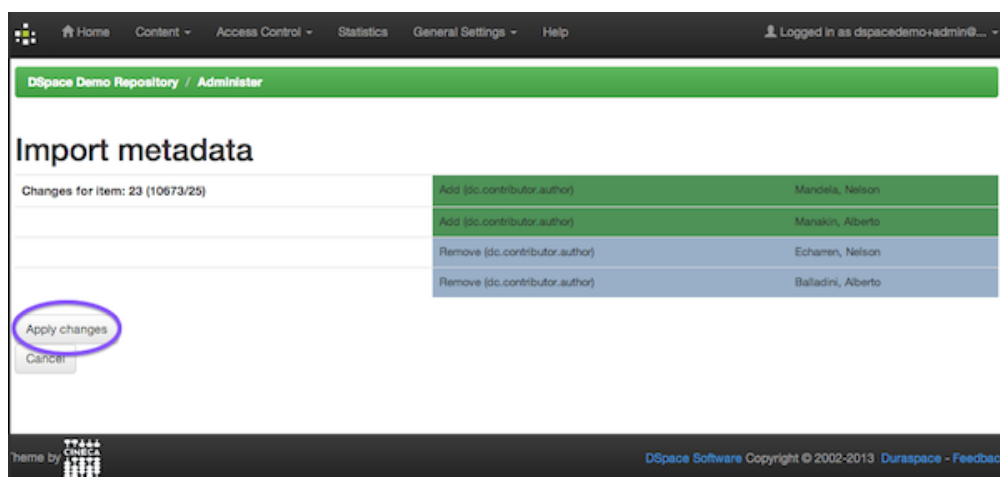
Once you are in the administrator tools, just click on "content" and then you only have to select "import metadata" from the list that drops down.

### 4) Choose file and upload CSV



Choose a file from your computer and click on "Upload CSV"

### 5) Review and approval of changes



On this page you can see which changes you have made within the CSV-file. You can now either accept these changes and click "Apply changes" or not, in that case click "Return".

## Command Line Functionality

### Export parameters

The following table summarizes the basics.

Command used:	<code>[dspace]/bin/dspace metadata-export</code>
Java class:	<code>org.dspace.app.bulkedit.MetadataExport</code>
Arguments short and (long) forms):	Description
<code>-f</code> or <code>--file</code>	Required. The filename of the resulting CSV.
<code>-i</code> or <code>--id</code>	The Item, Collection, or Community handle or Database ID to export. If not specified, <b>all</b> items will be exported.
<code>-a</code> or <code>--all</code>	Include all the metadata fields that are not normally changed (e.g. provenance) or those fields you configured in the <code>[dspace]/config/modules/bulkedit.cfg</code> to be ignored on export.
<code>-h</code> or <code>--help</code>	Display the help page.

#### Example commands

To run the batch editing exporter, at the command line:

```
[dspace]/bin/dspace metadata-export -f name_of_file.csv -i 1023/24
```

Example:

```
[dspace]/bin/dspace metadata-export -f /batch_export/col_14.csv -i /1989.1/24
```

In the above example we have requested that a collection, assigned handle `'1989.1/24'` export the entire collection to the file `'col_14.csv'` found in the `'/batch_export'` directory.

### Import parameters

The following table summarizes the basics.

Command used:	<code>[dspace]/bin/dspace metadata-import</code>
Java class:	<code>org.dspace.app.bulkedit.MetadataImport</code>
Arguments short and (long) forms:	Description
<code>-f</code> or <code>--file</code>	Required. The filename of the CSV file to load.

-s or --silent	Silent mode. The import function does not prompt you to make sure you wish to make the changes.
-e or --email	The email address of the user. This is only required when adding new items.
-w or --workflow	When adding new items, the program will queue the items up to use the Collection Workflow processes.
-n or --notify	when adding new items using a workflow, send notification emails.
-t or --template	When adding new items, use the Collection template, if it exists.
-h or --help	Display the brief help page.

Silent Mode should be used carefully. It is possible (and probable) that you can overlay the wrong data and cause irreparable damage to the database.

### Example commands

To run the batch importer, at the command line:

```
[dspace]/bin/dspace metadata-import -f name_of_file.csv
```

### Example

```
[dspace]/bin/dspace metadata-import -f /dImport/col_14.csv
```

If you are wishing to upload new metadata **without** bitstreams, at the command line:

```
[dspace]/bin/dspace metadata-import -f /dImport/new_file.csv -e joe@user.com -w -n -t
```

In the above example we threw in all the arguments. This would add the metadata and engage the workflow, notification, and templates to all be applied to the items that are being added.

### Importing large CSV files



It is not recommended to import CSV files of more than 1,000 lines. When importing files larger than this, it is hard to accurately verify the changes that the import tool states it will make, and large files may cause 'Out Of Memory' errors part way through the process.

## The CSV Files

The csv files that this tool can import and export abide by the [RFC4180](#) CSV format. This means that new lines, and embedded commas can be included by wrapping elements in double quotes. Double quotes can be included by using two double quotes. The code does all this for you, and any good csv editor such as Excel or OpenOffice will comply with this convention.

**File Structure.** The first row of the csv must define the metadata values that the rest of the csv represents. The first column must always be "id" which refers to the item's id. All other columns are optional. The other columns contain the dublin core metadata fields that the data is to reside.

A typical heading row looks like:

```
id,collection,dc.title,dc.contributor,dc.date.issued,etc,etc,etc.
```

Subsequent rows in the csv file relate to items. A typical row might look like:

```
350,2292,Item title,"Smith, John",2008
```

If you want to store multiple values for a given metadata element, they can be separated with the double-pipe '||' (or another character that you defined in your modules/bulkedit.cfg file. For example:

```
Horses || Dogs || Cats
```

Elements are stored in the database in the order that they appear in the csv file. You can use this to order elements where order may matter, such as authors, or controlled vocabulary such as Library of Congress Subject Headings.

When importing a csv file, the importer will *overlay* the data onto what is already in the repository to determine the differences. It only acts on the contents of the csv file, rather than on the complete item metadata. This means that the CSV file that is exported can be manipulated quite substantially before being re-imported. Rows (items) or Columns (metadata elements) can be removed and will be ignored. For example, if you only want to edit item abstracts, you can remove all of the other columns and just leave the abstract column. (You do need to leave the ID column intact. This is mandatory).

## Editing Collection Membership

Items can be moved between collections by editing the collection handles in the 'collection' column. Multiple collections can be included. The first collection is the 'owning collection'. The owning collection is the primary collection that the item appears in. Subsequent collections (separated by the field separator) are treated as mapped collections. These are the same as using the map item functionality in the DSpace user interface. To move items between collections, or to edit which other collections they are mapped to, change the data in the collection column.

## Adding Metadata-Only Items

New metadata-only items can be added to DSpace using the batch metadata importer. To do this, enter a plus sign '+' in the first 'id' column. The importer will then treat this as a new item. If you are using the command line importer, you will need to use the -e flag to specify the user email address or id of the user that is registered as submitting the items.

If you are using the web user interface for adding metadata-only items, any activated collection workflow steps are effectively bypassed. As a result, these added items are immediately archived into the repository.

## Deleting Metadata

It is possible to perform metadata deletes across the board of certain metadata fields from an exported file. For example, let's say you have used keywords (dc.subject) that need to be removed *en masse*. You would leave the column (dc.subject) intact, but remove the data in the corresponding rows.

## Performing 'actions' on items

It is possible to perform certain 'actions' on items. This is achieved by adding an 'action' column to the CSV file (after the id, and collection columns). There are three possible actions:

1. **'expunge'** This permanently deletes an item. Use with care! This action must be enabled by setting 'allowexpunge = true' in `modules/bulkedit.cfg`
2. **'withdraw'** This withdraws an item from the archive, but does not delete it.
3. **'reinstate'** This reinstates an item that has previously been withdrawn.

If an action makes no change (for example, asking to withdraw an item that is already withdrawn) then, just like metadata that has not changed, this will be ignored.

## Migrating Data or Exchanging data

It is possible that you have data in one Dublin Core (DC) element and you wish to really have it in another. An example would be that your staff have input Library of Congress Subject Headings in the Subject field (dc.subject) instead of the LCSH field (dc.subject.lcsh). Follow these steps and your data is migrated upon import:

1. Insert a new column. The first row should be the new metadata element. (We will refer to it as the TARGET)
2. Select the column/rows of the data you wish to change. (We will refer to it as the SOURCE)
3. Cut and paste this data into the new column (TARGET) you created in Step 1.
4. Leave the column (SOURCE) you just cut and pasted from empty. Do not delete it.

## Common Issues

[Metadata values in CSV export seem to have duplicate columns](#)