DCAT Meeting October 2016

Date & Time

- October 11th **15:00** UTC/GMT - **11:00** ET

This call is a Community Forum call: Sharing best practices and challenges in the use of existing DSpace features

Dial-in

We will use the international conference call dial-in. Please follow directions below.

- U.S.A/Canada toll free: 866-740-1260, participant code: 2257295
- International toll free: http://www.readytalk.com/intl
    - Use the above link and input 2257295 and the country you are calling from to get your country's toll-free dial in #
    - Once on the call, enter participant code 2257295

Agenda

Community Forum Call: DSpace Importing and Bulk Metadata Editing

Sharing best practices, challenges, and questions

- DSpace Importing and Bulk Metadata Editing
    - Building simple archive format structures/folders
    - Working with the spreadsheet bulk editing tool
    - Command line imports

Preparing for the call

Bring your questions/comments you would like to discuss to the call, or add them to the comments of this meeting page.

If you can join the call, or are willing to comment on the topics submitted via the meeting page, please add your name, institution, and repository URL to the Call Attendees section below.

Meeting notes

Batch Metadata Editing

DSpace offers a default batch metadata editing feature which allows administrators to export metadata in a CSV file. This CSV file can be imported in a spreadsheet application, after which the metadata can be altered. After editing, administrators can reconvert the file to a CSV file, and import it back in DSpace.

Georgetown University created several tools as an extension of the standard DSpace batch editing functionality. These tools will become part of the DSpace 6 codebase.

Georgetown University created tools for:

- Replacing metadata for all items in a collection
- Replacing metadata by query
- Generating bulk ingest folders (metadata and media)

UTF8 encoding issue

When using the batch metadata functionality, metadata sometimes gets corrupted when the CSV file is imported in a spreadsheet application. This is caused by some characters not being imported correctly as UTF8, which automatically results in an erroneous metadata value when the metadata is exported as a CSV file. Even if the metadata value was not altered.

According to DCAT this is due to a lack of correct encoding support by (certain)spreadsheet applications.

Openrefine

Throughout the discussion participants often mentioned OpenRefine (http://openrefine.org/) as a great application for editing CSV exports. This tool could be interesting to such an extend it may be useful to organize a workshop on the application. This workshop could be an extension of the OpenRefine workshop organized by Code4lib. DCAT members having more information on the Code4lib Openrefine workshop are invited to share their knowledge, or (links to) any affiliated documents, in the comments.

DSpace Bulk ingest & export

Simple Archive Format

DSpace offers bulk ingest through Simple Archive Format. This is an archive containing a directory for each item. Each item directory consists out of a file containing the file's metadata together with all of the item's bitstreams.

Exporting search results

DSpace 6 will come with new bulk exporting functionality, being a new tool allowing to export search results.

Blank spaces

There was the concern of blank values being introduced after exporting a CSV file out of a spreadsheet application. While in the original CSV file there was no value for a certain metadata field, there may be a blank value in the CSV file exported out of the spreadsheet editor. This however should not not be a problem as it is unlikely the DSpace batch metadata editing tool will insert a value for this blank when the CSV fil is imported in DSpace.

Call Attendees

- Maureen Walsh - The Ohio State University
- Ignace Deroost - Atmire
- Irene Berry - Naval Postgraduate School
- Anna Dabrowski - Texas A&M University (http://oaktrust.library.tamu.edu)
- Terrence W Brady - Georgetown University
- Mariya Maistrovskaya - University of Toronto
- Jose Carvalho
- Filipe Furtado
- Valerie Collins - University of Minnesota
- Marianne Reed - University of Kansas
- Felicity Dykas - University of Missouri–Columbia
- Anne Lawrence - Virginia Tech
- Sarah Potvin - Texas A&M University
- Daniel Draper - Colorado State University
- Iryna Kuchma - EIFL
- Elias Tzoc - Miami University
- Monica Rivero - Rice University
- Susan Borda - Montana State University
- Bill Kelm - Willamette University