



@WIKE

Improving the quality of
your metadata

Ben Bosman - @mire

Outline

- 1 - Introduction
- 2 - Duplicate Checking
- 3 - Mass Metadata Editing
- 4 - Relation to DSpace 1.6 batch edit

1 - Introduction

✦ Problem statement

✦ Data present in your repository:

- ✦ Migrated from legacy systems
- ✦ Import from online sources
- ✦ Manually submitted items

✦ Maintenance of your data:

- ✦ Metadata editors can check newly submitted items
- ✦ What about comparison to existing items?
- ✦ How to handle changes to your application profile?
 - ✦ e.g. changing “Article” to “info:eu-repo/semantics/article”

2 - Duplicate Checking

✦ Steps

- ✦ Duplicate detection:
 - ✦ A configurable duplicate detection algorithm
- ✦ Manual Comparison:
 - ✦ Check whether the records are really duplicates
- ✦ Merging records:
 - ✦ Select metadata fields and files from the records, and compose a new record
- ✦ Keeping persistent URL's:
 - ✦ Ensure all persistent URL's refer to the merged item.

2 - Duplicate Checking

✦ Duplicate detection

✦ Specify scope:

- ✦ Search your entire repository
- ✦ Search only a specific community or collection
- ✦ Search only a small set of records

✦ Configuration:

- ✦ Administrators can configure duplicate detection algorithm

✦ Multilingual:

- ✦ Has been tested in multiple languages, including Chinese

✦ Avoid duplicates during entry:

- ✦ Submitters are offered list of potential duplicates at the end of their submission

2 - Duplicate Checking

✦ Manual Comparison

✦ Potential duplicates overview:

- ✦ Compact overview of potential overview
- ✦ More details available

✦ Marking records as different:

- ✦ Ensures similar but different records are no longer proposed as duplicates

✦ Viewing differences:

- ✦ Clear overview of exactly what differs in the metadata

Item	Title	Contributors	Issue Date	Actions
<input checked="" type="checkbox"/> 1	My Dear Selina (Richmond, Monday 12th Oct. 1829)	Campbell, Alexander	1829-10-12	More Info
<input type="checkbox"/> 2	My dear Selina (Richmond Oct 5, 1829)	Campbell, Alexander	1829-10-05	More Info
<input type="checkbox"/> 3	My dear Selina (Richmond. Dec 23rd 1829)	Campbell, Alexander	1829-12-23	More Info
<input type="checkbox"/> 4	My dear Selina (Richmond Va. Nov. 4 1829)	Campbell, Alexander	1829-11-04	More Info

[Merge](#)
[Compare](#)
[Delete](#)

[All unrelated](#)

[All unrelated](#)

[Merge](#)
[Compare](#)
[Delete](#)



@MIRE

@MIRE

2 - Duplicate Checking

✦ Merging records

✦ Merging metadata fields:

- ✦ Select one or multiple values
- ✦ Ordering selected values

✦ Merging file lists:

- ✦ Select one or multiple values
- ✦ Files grouped by bundle
 - ✦ Auto-generated files can be maintained

Compose merged item



Items overview

Title	Issue Date	Submitter
item2 The effects of computer training and internet usage on the use of everyday technology by older adults a randomized controlled study	2007	Lieven Droogmans
item3 The effects of computer training and internet usage on the use of everyday technology by older adults: a randomized controlled study	2007	Lieven Droogmans

Actions:

Select all metadata from:

Select all bitstreams from:

Select metadata values:

Title (title):

Keep:

- None
- item2** The effects of computer training and internet usage on the use of everyday technology by older adults a randomized controlled study
- item3** The effects of computer training and internet usage on the use of everyday technology by older adults: a randomized controlled study

or compose:



<input type="checkbox"/> item2 The effects of computer training and internet usage on the use of everyday technology by older adults_ a randomized controlled study
<input type="checkbox"/> item3 The effects of computer training and internet usage on the use of everyday technology by older adults; a randomized controlled study
.....



2 - Duplicate Checking

- ✦ **Keeping persistent URL's**
 - ✦ In metadata:
 - ✦ Information can be reused in other services
 - ✦ Browser redirects:
 - ✦ Ensure the user will be displayed the merged record



Duplicate checking DEMO

3 - Mass Metadata Editing

✦ Steps

- ✦ Search for records:
 - ✦ Using advanced search, community or collection browse, or repository-wide
- ✦ Selection of records:
 - ✦ Select or deselect items to be edited
- ✦ Definition of tasks:
 - ✦ List all actions to be performed
- ✦ Checking results:
 - ✦ Overview of all changed items displaying updated fields

3 - Mass Metadata Editing

- ✦ **Search for records**
 - ✦ Use advanced search functionality
 - ✦ Select a specific community or collection
 - ✦ Combine search with community or collection
 - ✦ Update entire Repository
- ✦ **Selection of records**
 - ✦ View compact description of resulting items
 - ✦ Displayed metadata depends on previous search terms
 - ✦ Link to full item record

3 - Mass Metadata Editing

- ✦ **Definition of tasks**
 - ✦ **Metadata Editing:**
 - ✦ Remove metadata values
 - ✦ Add metadata values
 - ✦ Move metadata values to other fields
 - ✦ Copy metadata values to other fields
 - ✦ Edit metadata values
 - ✦ **Collection Mapping Editing:**
 - ✦ Remove from current collection(s)
 - ✦ Add to other collection(s)

Batch edit

http://localhost:8080/dsp

Item Metadata Actions

insert John Doe in to the property
 contributor.editor
 if that property doesn't exist yet

remove the property contributor.advisor
 if that property contains John Doe

copy the property subject.classification
 to the property
 subject

move the property date.issued
 to the property
 date.available

replace @mire with Atmire in the
 property creator

Item Mapping Actions

Remove from collection(s)
 Add to collection(s)

- ▶ Abilene Christian University Brown Library
- ▶ Abilene Public Library
- ▶ Hardin-Simmons University Library
- ▶ Howard Payne University Walker Memorial Library
- ▶ McMurry University Jay-Rollins Library

Perform



3 - Mass Metadata Editing

- ✦ **Checking results**

- ✦ Overview of results:

- ✦ Overview of all changed items displaying updated fields
 - ✦ What happened when some of the items were changed in the meantime?

Quality Assurance: Search results

http://localhost:8

Browse

My Account

Administrative

Metadata quality

Duplicate Checker

Batch edit

XML Test

Successfully modified items

Restoration Quarterly;
Restoration Movement; Rollmann, Hans, Dr.
Restoration Serials Index;

Title Restoration Serials Index

Authors

Collection(s) Restoration Sites

contributor.editor John Doe

Show full item record

Stone-Campbell Restoration Movement Resources; McMillan, Jim

Title Stone-Campbell Restoration Movement Resources

Authors McMillan, Jim

Collection(s) Restoration Sites

contributor.editor John Doe

Show full item record

Klaar



@MIRE

@MIRE

Mass Metadata Editing DEMO



4 - Relation to DSpace 1.6

batch edit

- ✦ **MQM includes a batch editing feature**
 - ✦ Different approach
 - ✦ No “per item” batch editing
 - ✦ Instead series of conditional edit actions directly in Admin UI
 - ✦ Targeted at specific but common use cases:
 - ✦ Correcting consistent errors
 - ✦ Move (some) values to other fields
 - ✦ Adding new required values based on changes of the application metadata
 - ✦ ...

THANK YOU !



Questions ?

ben@atmire.com