

MAILE

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"



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.



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



Manual Comparison

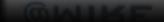
- Potential duplicates overview:
 - Compact overview of potential overview
 - More details available
- Marking records as different:
 - Ensures similar but different records are no longer proposed as duplicates
- Viewing differences:
 - Clear overview of exactly what differs in the metadata



Item	Title	Contributors	Issue Date	Actions	
▼ 1	My Dear Selina (Richmond, Monday 12th Oct. 1829)	Campbell, Alexander	1829-10-12	More Info	
□ 2	My dear Selina (Richmond Oct 5, 1829)	Campbell, Alexander	1829-10-05	More Info	
⊟ 3	My dear Selina (Richmond. Dec 23rd 1829)	Campbell, Alexander	1829-12-23	More Info	
□ 4	My dear Selina (Richmond Va. Nov. 4 1829)	Campbell, Alexander	1829-11-04	More Info	
Merge Compare Delete					
All unrelated					

All unrelated





Title (title):

item1 item2 Arbeidsmarktperspectieven voor schoolverlaters tot 2008

Author (contributor.author):

itemi Golsteyn, B.

item2 Golsteyn, B-

item2 Corvers, F-

item1 Corvers, F.

item1 Dupuy, A.

item2 Dupuy, A

Description (description):

Item1 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur pretium, ligula vel facilisis accumsan, urna est ornare nisl, a consectetur eros tortor vitae magna. Maecenas accumsan lectus vitae nisl. In hac habitasse platea dictumst. Etiam luctus risus nec dui. Aenean iaculis lobortis lectus.

Item2 Lorem ipsum dolor sit amet_ consectetur adipiscing elit. Curabitur pretium_ ligula vel facilisis accumsan_ urna est ornare nisl_ a consectetur eros tortor vitae magna. Maecenas accumsan lectus vitae nisl. In hac habitasse platea dictumst. Etiam luctus risus nec dui. Aenean iaculis lobortis lectus.

lobortis lectus.

pretium- ligula vel facilisis accumsan- urna est ornare nisl- a consectetur eros tortor vitae magna. Maecenas accumsan lectus vitae nisl. In hac habitasse platea dictumst. Etiam luctus risus nec dui. Aenean iaculis

Lorem ipsum dolor sit amet- consectetur adipiscing elit. Curabitum





- 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: Item2 Select

Select all bitstreams from: Item3 Select

Select metadata values:

Γitle	e (title):			
Keep:				
ullet	None			
0	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:				
0				
	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			



- 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



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



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

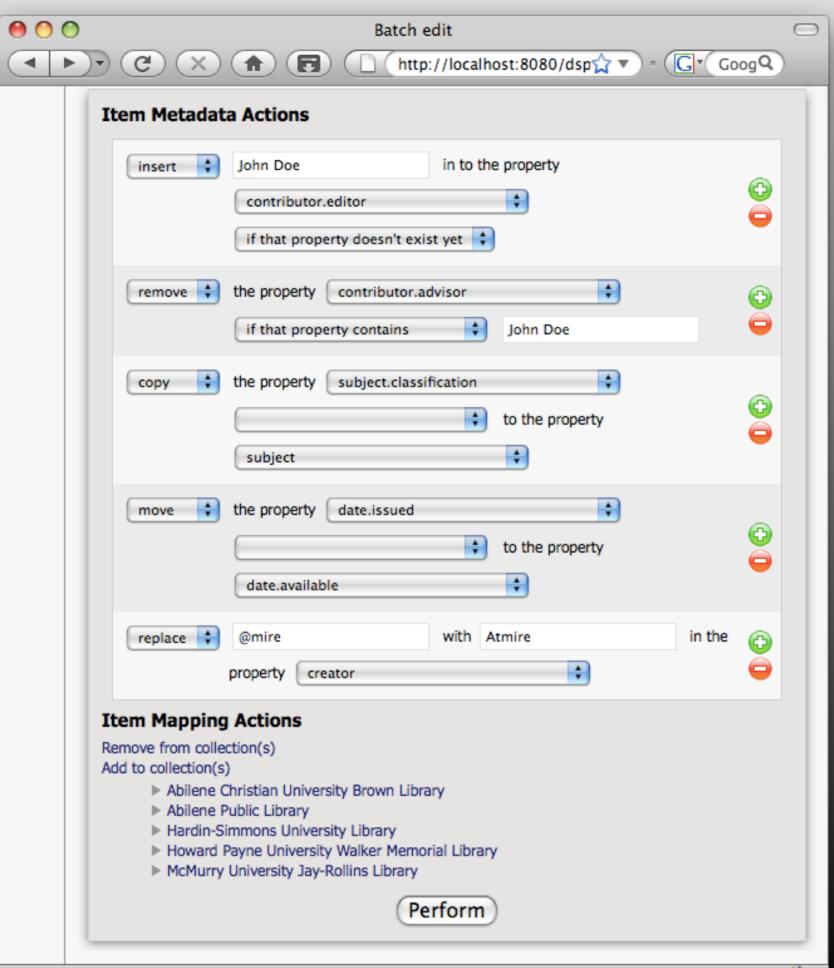
- View compact description of resulting items
 - Displayed metadata depends on previous search terms
- Link to full item record



■ 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)

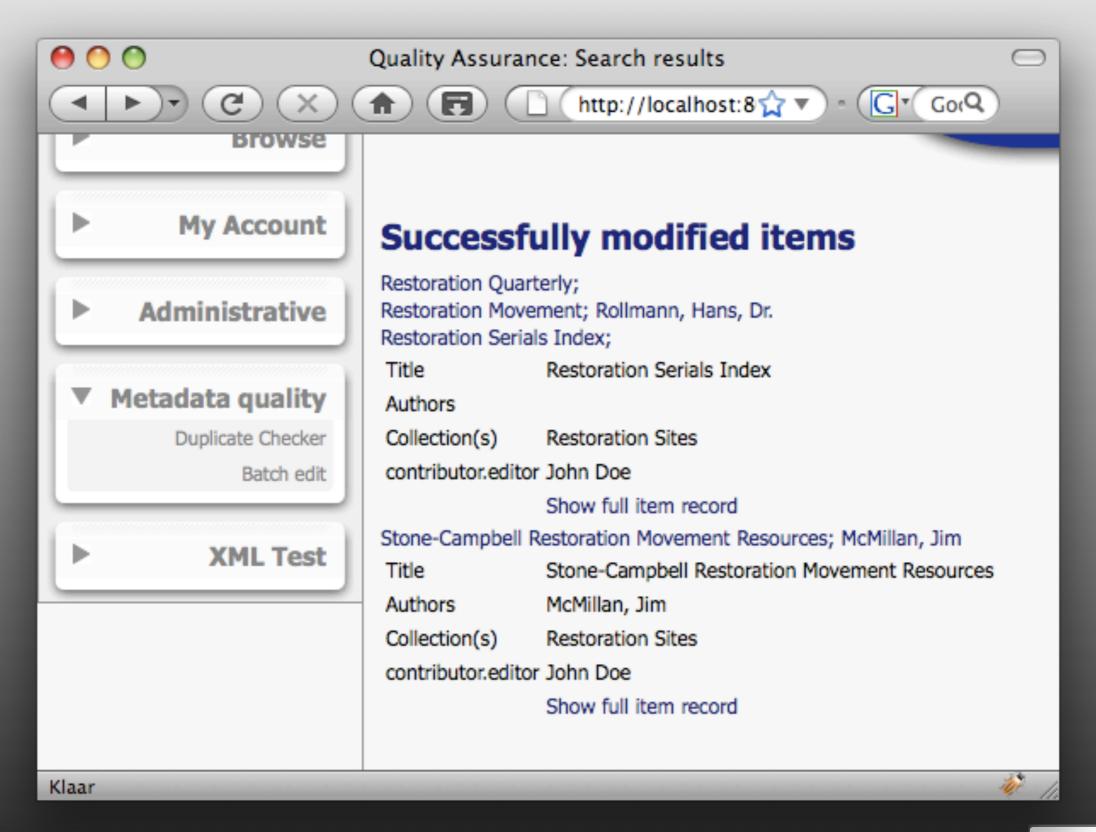




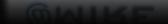


- Checking results
 - Overview of results:
 - Overview of all changed items displaying updated fields
 - What happened when some of the items were changed in the meantime?











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
 - **x** ...



THANK YOU!



COLVINE

Questions?

ben@atmire.com