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
<table>
<thead>
<tr>
<th>Item</th>
<th>Title</th>
<th>Contributors</th>
<th>Issue Date</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>My Dear Selina (Richmond, Monday 12th Oct. 1829)</td>
<td>Campbell, Alexander</td>
<td>1829-10-12</td>
<td>More Info</td>
</tr>
<tr>
<td>2</td>
<td>My dear Selina (Richmond Oct 5, 1829)</td>
<td>Campbell, Alexander</td>
<td>1829-10-05</td>
<td>More Info</td>
</tr>
<tr>
<td>3</td>
<td>My dear Selina (Richmond Dec 23rd 1829)</td>
<td>Campbell, Alexander</td>
<td>1829-12-23</td>
<td>More Info</td>
</tr>
<tr>
<td>4</td>
<td>My dear Selina (Richmond Va. Nov. 4 1829)</td>
<td>Campbell, Alexander</td>
<td>1829-11-04</td>
<td>More Info</td>
</tr>
</tbody>
</table>
Arbeidsmarktperspectieven voor schoolverlaters tot 2008

Golsteyn, B.
Golsteyn, B.
Corvers, F.
Corvers, F.
Dupuy, A.
Dupuy, A


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

<table>
<thead>
<tr>
<th>Title</th>
<th>Issue Date</th>
<th>Submitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item2 The effects of computer training and internet usage on the use of everyday technology by older adults a randomized controlled study</td>
<td>2007</td>
<td>Lieven Droogmans</td>
</tr>
<tr>
<td>Item3 The effects of computer training and internet usage on the use of everyday technology by older adults: a randomized controlled study</td>
<td>2007</td>
<td>Lieven Droogmans</td>
</tr>
</tbody>
</table>

Actions:

Select all metadata from: Item2 Select
Select all bitstreams from: Item3 Select

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:

*The effects of computer training and internet usage on the use of everyday technology by older adults a randomized controlled study*

*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
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)
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)
- Ablene Christian University Brown Library
- Ablene 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?
## Successfully modified items

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

<table>
<thead>
<tr>
<th>Title</th>
<th>Restoration Serials Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Collection(s)</strong></td>
<td>Restoration Sites</td>
</tr>
<tr>
<td>contributor.editor</td>
<td>John Doe</td>
</tr>
</tbody>
</table>

**Stone-Campbell Restoration Movement Resources; McMillan, Jim**

<table>
<thead>
<tr>
<th>Title</th>
<th>Stone-Campbell Restoration Movement Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Collection(s)</strong></td>
<td>Restoration Sites</td>
</tr>
<tr>
<td>contributor.editor</td>
<td>John Doe</td>
</tr>
</tbody>
</table>

Show full item record
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