Journal Database of UZH: Quality assessment and syncing with SHERPA/RoMEO

Dr. Martin Brändle

martin.braendle@id.uzh.ch
Journal Database (www.jdb.uzh.ch)

Aim

• Bibliographic data of scientific journals and series (~22K)
• Publishers (~11K)
• Copyright policies (author rights) and Refereed status
• Made by using various sources
• Used in publication submission process
  • Submitters obtain information about publisher policy
  • New journal and publisher entries can be submitted

Implementation

• EPrints repository – flat table
• Interface with ZORA | Ulrichs Periodicals
Short Demo
Current Problems

Technical

• Current ZORA-JDB interface leads to duplicates and slight alternations of submitted titles in buffer (>2K)
• ZORA and JDB only loosely coupled
• Review team overburdened in curation process
• Script for duplicate detection

➤ ZORA-JDB interface and script must be improved (Peter West)

Organisational

• Copyright policies changing
• Review team overburdened in curation process (not systematic)

➤ Script for syncing policies with SHERPA/RoMEO (me)
Syncing with SHERPA/RoMEO

- One-way: SHERPA/RoMEO → JDB (using SHERPA/RoMEO API)
- Implementation
  - Journals (via ISSN)
  - Publishers (via SHERPA/RoMEO ID, publisher name)
  - full sync/partial sync in phases
  - XML download
  - Mapping rules (under construction)
    - Overwrite/Skip
Pre-Evaluation: Quality and quantity structure

- Overlap between JDB and SHERPA/RoMEO?
- Timeliness of JDB and SHERPA/RoMEO records (newer/older)?
- Consistency and quality of JDB field contents?

Procedure
- full sync, journals & publishers, XML download
- Import of JDB and SHERPA/RoMEO records into Comparison database (FileMaker, rapid prototyping, calculation fields)
- Assessment by quality figures
- Initial revisions using SQL updates
<table>
<thead>
<tr>
<th>JDB Journals</th>
<th>SHERPA/RoMEO Journals</th>
<th>Vergleich</th>
</tr>
</thead>
<tbody>
<tr>
<td>eprintid</td>
<td>9982</td>
<td></td>
</tr>
<tr>
<td>title</td>
<td>Special Care in Dentistry</td>
<td>singleJournal</td>
</tr>
<tr>
<td>is_abbreviation</td>
<td>Spec Care Dentist</td>
<td></td>
</tr>
<tr>
<td>issn_print</td>
<td>0275-1879</td>
<td></td>
</tr>
<tr>
<td>issn_electronic</td>
<td>1754-4505</td>
<td></td>
</tr>
<tr>
<td>issn_link</td>
<td>0275-1879</td>
<td></td>
</tr>
<tr>
<td>accepted_manuscript</td>
<td>not_specified</td>
<td></td>
</tr>
<tr>
<td>embargo_accepted</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>published_manuscript</td>
<td>not_specified</td>
<td></td>
</tr>
<tr>
<td>embargo_published</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>jdb_url</td>
<td>[link]</td>
<td></td>
</tr>
<tr>
<td>the_publisher</td>
<td>Wiley-Blackwell Publishing, Inc.</td>
<td></td>
</tr>
<tr>
<td>publisher</td>
<td>Wiley</td>
<td></td>
</tr>
<tr>
<td>publisher_aka</td>
<td>12 months</td>
<td></td>
</tr>
<tr>
<td>publisher_home_url</td>
<td>[link]</td>
<td></td>
</tr>
<tr>
<td>preprints_prearchiving</td>
<td>can</td>
<td></td>
</tr>
<tr>
<td>preprints_prerestrictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conditions</td>
<td>Some journals have separate</td>
<td></td>
</tr>
<tr>
<td>paidaccess_name</td>
<td>OnlineOpen</td>
<td></td>
</tr>
<tr>
<td>paidaccess_notes</td>
<td>A paid open access option is</td>
<td></td>
</tr>
<tr>
<td>paidaccess_url</td>
<td>[link]</td>
<td></td>
</tr>
<tr>
<td>copyrightlink_texts</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>copyrightlink_urls</td>
<td>[link]</td>
<td></td>
</tr>
<tr>
<td>sherpa_romeo_id (J)</td>
<td>580</td>
<td></td>
</tr>
<tr>
<td>sherpa_romeo_id (P)</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>publisher_id</td>
<td>2050</td>
<td></td>
</tr>
<tr>
<td>message</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Vergleich:
- match exact: 61.0, 5.5, 12.1, 21.4
- print: 0.5, 27.9, 21.4
- conflict: 21.3, 57.3, 21.4
- conflict: 2.1, 9.2, 88.7
- conflict: 23.3, 55.3, 21.4
- conflict: 0.5, 1.4, 98.1
- JDB older: 50.7
Some comparison figures

Global
Journals: 79% (17K) of JDB match
Publishers: 16% (1.7K) of JDB match

Mapping Rules
Embargo periods
JDB 13-16 distinct entries, consistent
SHERPA/RoMEO >50 different entries, very inconsistent
Discussion

• Use of journal/publisher/policy data?
• Interest in JDB?
• Collaborative curation?
Initial revisions using SQL updates

sherpa_romeo_id
Journals: +4612 (+89%)
Publishers: +1235 (+111%)