



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



**Universität
Zürich** UZH

HBZ / Zentrale Informatik



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



JDB Journals		SHERPA/RoMEO Journals		Vergleich	=	(~)	≠	leer	
					>	<	leer		
eprintid	9982	outcome	singleJournal						
title	Special Care in Dentistry	jtitle	Special Care in Dentistry	match exact	61.0	5.5	12.1	21.4	
iso_abbreviation	Spec Care Dentist	issn	0275-1879	print					
issn_print	0275-1879								
issn_electronic	1754-4505								
issn_link	0275-1879								
accepted_manuscript	not_specified	3	postprints_postarchiving	restricted	5	conflict	21.3	57.3	21.4
embargo_accepted		16	postprints_postrestrictions	<num>12</num> <period	55	conflict	2.1	9.2	88.7
published_manuscript	not_specified	3	pdfversion_pdfarchiving	cannot	5	conflict	23.3	55.3	21.4
embargo_published		13	pdfversion_pdfrestrictions		35		0.5	1.4	98.1
lastmod	22/10/2012 16:02:24		lastmod	11/08/2014 14:58:29		JDB older	50.7	27.9	21.4
JDB_URL	http://www.jdb.uzh.ch/9982/								
the_publisher	9694								
publisher	Wiley-Blackwell Publishing, Inc.		publisher	Wiley					
			publisher_alias	12 months					
			publisher_homeurl	http://eu.wiley.com/WileyCDA/					
			preprints_prearchiving	can					
			preprints_prerestrictions						
			conditions	Some journals have separate					
			paidaccess_name	OnlineOpen					
			paidaccess_notes	A paid open access option is					
			paidaccess_url	http://olabout.wiley.					
			copyrightlink_texts	Self-archiving	2				
			copyrightlink_urls	http://olabout.wiley.	2				
sherpa_romeo_id (J)	580	conflict	publisher_id	2050		J	conflict	P	conflict
sherpa_romeo_id (P)	23		message						



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%)