

Prozessautomatisierung für institutionelle Repositorien unter Nutzung offener Datenquellen

Nicolai Hauf (ZHAW) und Selina Sturzenegger (FHNW)

DSpace Praxistreffen, Bamberg, 27. bis 28. März 2025



Diese Präsentation steht unter der Lizenz [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/). (Stand 21.03.2025)

Die Marken von ZHAW, FHNW und swissuniversities sind von der vorliegenden Lizenz CC BY 4.0 unberührt. Gemäss Abschnitt 2.b.2 der Lizenz werden Patent- und Kennzeichenrechte durch die vorliegende Public License nicht lizenziert.

Projekt AURORA

Projektpartner: ZHAW Zürcher Hochschule für Angewandte Wissenschaften, Fachhochschule Nordwestschweiz FHNW

Laufzeit: Dezember 2023 – März 2025

Projektvolumen: CHF 395'000

Drittmittelgeber: Bund, PgB Open Science I (kofinanziert von swissuniversities)

Was?

Entwicklung eines automatisierten Prozesses für institutionelle Repositorien, um

- Hochschulpublikationen aus offenen bibliografischen Datenbanken zu sammeln,
- diese mit Informationen zur Zweitveröffentlichung anzureichern,
- den Metadatensatz im Repository zu erstellen, und
- Publizierende über ihre Möglichkeiten zum freien Teilen ihres Beitrags zu informieren.

Warum?

- Arbeitsaufwand für Publizierende bei der Einreichung in Repositorien verringern
- Mehr Klarheit und Sicherheit bei der Zweitveröffentlichung
- Mehr frei zugängliche Volltexte
- Höherer Abdeckungsgrad von Hochschulpublikationen in institutionellen Repositorien

Quellenauswahl

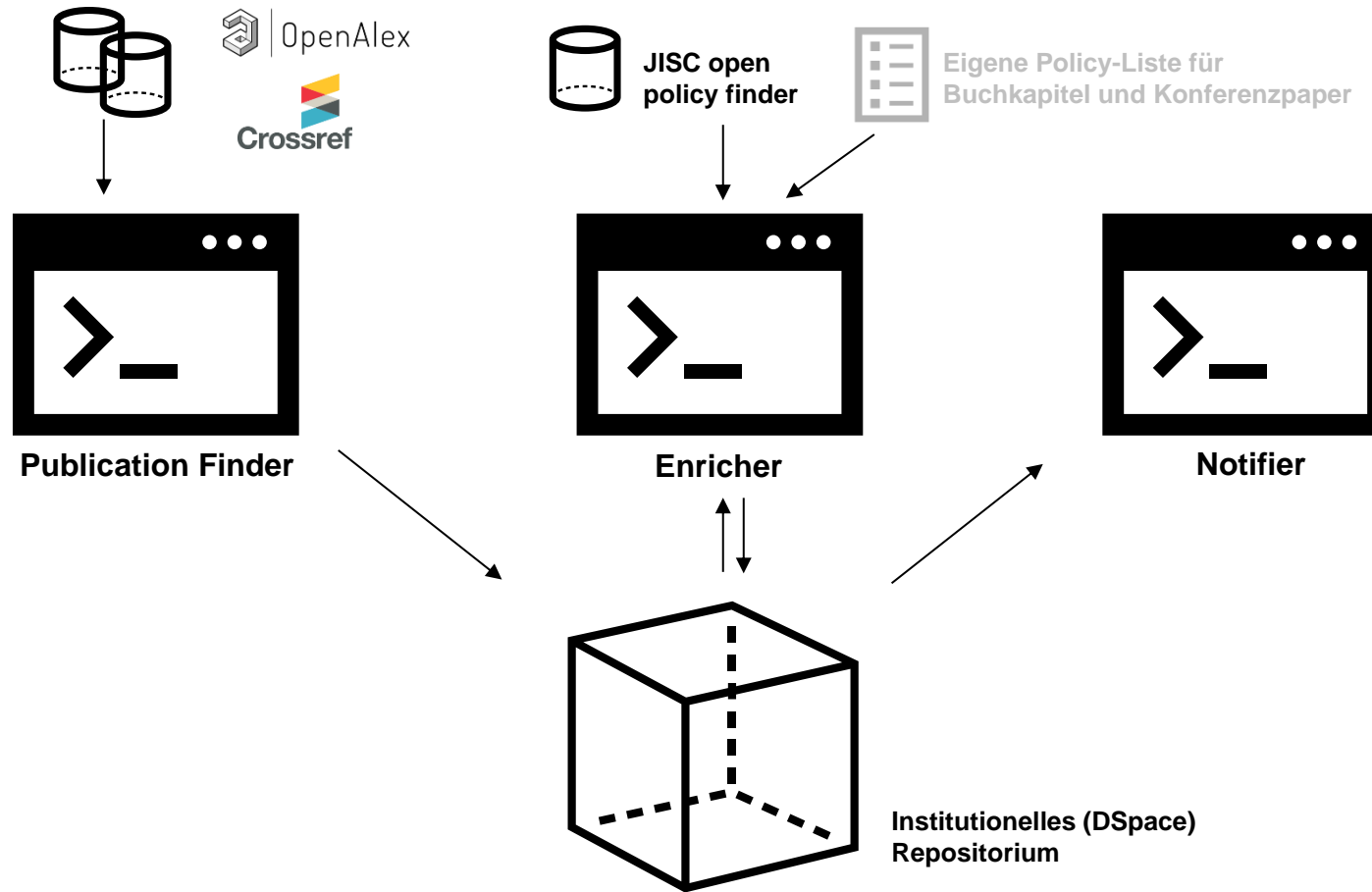
Evaluation von etablierten bibliografischen Datenbanken

(Semantic Scholar, Scopus, Web of Science, Dimensions, Crossref, DataCite, OpenAlex, BASE, OpenAIRE)

Wichtigste Kriterien:

- Datengrundlage und Abdeckung (möglichst viele Publikationen, verschiedene Publikationstypen)
 - Umfang und Qualität der Metadaten
 - Suche nach Affiliation über eine (offene) technische Schnittstelle
- Initiale Umsetzung mit **OpenAlex** und **Crossref**
(mit Potential zur Erweiterung)

Übersicht



Publication Finder I – Datenabfrage

- Abfrage der OpenAlex und Crossref APIs nach Publikationen der Hochschule
- Suchparameter: Affiliation (Namen und ROR ID) und Datum
- Kombination von verschiedenen Suchabfragen für eine genauere Abdeckung
- Ergebnisse werden als JSONs zurückgeliefert

Beispiel

https://api.openalex.org/works?filter=raw_affiliation_strings.search:ZHAW,from_publication_date:2023-01-01&per_page=200&cursor=*

https://api.openalex.org/works?filter=raw_affiliation_strings.search:Zürcher Hochschule für Angewandte Wissenschaften,from_publication_date:2023-01-01&per_page=200&cursor=*

...

https://api.openalex.org/works?filter=institutions.ror:https://ror.org/05pmsvm27,from_publication_date:2023-01-01&per_page=200&cursor=*



Publication Finder II – Verarbeitung der Metadaten

- Löschen von Dubletten innerhalb einer Datenquelle
- Transformation der relevanten Metadaten in ein internes Format und anschliessend in die hochschulspezifischen Formate
- Identifizierung gleicher Publikationen aus den verschiedenen Quellen (Match) und Zusammenführung der Metadaten nach einer Vorzugsreihenfolge (Merge)

Beispiel

Liste gematchter Publikationen in OpenAlex

ID	dc.title	dc.contributor.author	dc.identifier.doi	dc.identifier.isbn
A	10.1016/j.ece.2022.10.001	0926-5805 1872-7891
...



Liste gematchter Publikationen in Crossref

ID	dc.title	dc.contributor.author	dc.identifier.doi	dc.identifier.isbn
3	10.1016/j.ece.2022.10.001	0926-5805
...



Zusammenführung von Metadaten über eine Vorzugsreihenfolge (Merge)

Ergebnisliste für Publikationen in beiden Quellen

ID	dc.title	dc.contributor.author	dc.identifier.doi	dc.identifier.isbn
A (OA)	... (CR)	... (CR)	10.1016/j.ece.2022.10.001 (CR)	0926-5805 1872-7891 (OA)
...

Publication Finder III – Datenimport in DSpace

- Dublettenkontrolle mit bestehenden Items im Repository
- Import neuer Publikationen über die DSpace API in einer spezifischen Sammlung im Bearbeitungsbereich

Beispiel

Vorschau bild nicht verfügbar

wartet auf Bearbeitung (Waiting for Controller)
Publikation

Dissecting Structural Requirements of Leucino... and Mammalian Toxicity
(American Chemical Society (ACS), 2025) Rimle, Lukas; Pliatsika, Dimanthi; Arnold, Noëlle; Kurth, Sandra; Kai Lefleuganan, a clinical stage drug candidate for the treatment of cutaneous leishmaniasis, is a synthetic nona...
▼ Zeige mehr

Einreichender: **Import User** ←

Sammlung: aurora ←

Übernehmen Anzeige

Vorschau bild nicht verfügbar

wartet auf Bearbeitung (Waiting for Controller)
Publikation

The neurobench framework for benchmarking neuromorphic computing algorithms and systems
(Nature Portfolio, 2025) Yik, Jason; Berghe, Korneel Van den; Blanken, Douwe den; Bouhadjar, Younes; Fabre...
Keine Zusammenfassung
▼ Zeige mehr

Einreichender: **Import User**









Sammlung: aurora

Enricher

- Ein Job in DSpace verwendet Metadaten aus einem Item, um Bedingungen für eine Zweitveröffentlichung abzufragen
- Bei Erfolg werden die Informationen zu Artikelversion, Sperrfrist, Lizenz und Quelle in einem zusätzlichen Metadatenfeld im Item gespeichert
- Zeitschriftenartikel → JISC open policy finder API (über ISSN)

Beispiel

https://v2.sherpa.ac.uk/cgi/retrieve_by_id/cgi/retrieve_by_id?item-type=publication&api-key=XXX&format=Json&identifier=1872-7891

 Published	Option with associated OA fees, no embargo & CC BY licence	
 Published	Option with funder pre-requisites, associated OA fees, no embargo & CC BY licence	
 Accepted	Option with no embargo & CC BY-NC-ND licence	
 Accepted	Option with 24 months embargo & CC BY-NC-ND licence	

→ Version: Accepted *** Embargo: 24 Months *** Licence: CC BY-NC-ND *** URL: <https://v2.sherpa.ac.uk/id/publication/16733>

Notifier

- Erweiterung in DSpace, um ausgewählte Autor:innen mit einer vorgefertigten Mail auf ihre Möglichkeiten einer Zweitveröffentlichung hinzuweisen

Beispiel

1

Autor:in *

<input checked="" type="radio"/> Künzli	<input type="radio"/> Hansjörg	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="radio"/> Burtscher	<input type="radio"/> Michael Josef	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="radio"/> Martins	<input type="radio"/> Michael	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="radio"/> Delamuerte	<input type="radio"/> Francesca	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2

Sende Email an Autor:in

Wollen Sie wirklich eine Email an die verfassende Person senden?

3

Guten Tag

Über einen automatisierten Prozess wird in ausgewählten Datenbanken nach neuen Publikationen von FHNW-Angehörigen gesucht und geprüft, ob eine Zweitveröffentlichung im IRF möglich ist. Wir haben folgenden Artikel von Ihnen gefunden:

Apéry-type series and colored multiple zeta values (2024)
<https://doi.org/10.1016/j.aam.2023.102610>

Es ist erlaubt, folgende Version des Artikels im IRF zu veröffentlichen:

- Akzeptiertes Autorenmanuskript (AAM)
- Sperrfrist: 24 Monate
- Lizenz: CC BY-NC-ND 4.0

Wenn Sie von der Möglichkeit der Zweitveröffentlichung Gebrauch machen möchten, dürfen Sie uns diese Version des Dokuments gerne zustellen. Wir werden es im IRF hochladen und den Eintrag für Sie fertigstellen und im IRF veröffentlichen.

Vielen Dank für Ihre Rückmeldung innert 14 Tagen.
Bei Fragen zum Vorgehen oder den rechtlichen Grundlagen stehen wir Ihnen gerne zur Verfügung.

Freundliche Grüsse
Selina

Fachhochschule Nordwestschweiz FHNW
Vizepräsidium Hochschulentwicklung
Bibliothek FHNW / Digitale Dienste

Institutional Repository FHNW (IRF)

Quantitativer Eindruck

Gefundene Werke mit einem Publikationsdatum von 2020 bis 2024, ZHAW

	2020	2021	2022	2023	2024	total
Gefundene eindeutige Publikationen in OpenAlex und Crossref	757	837	848	964	703	4109
Zahl bestehender Publikationen im Repository (Dubletten)	468	544	496	533	391	2432
Zahl fehlender Publikationen im Repository	289	293	352	431	312	1677

Kontext: von vergleichbaren Publikationstypen bearbeiten wir jährlich ca. 1250 Einträge (+27%)

- Verdeutlicht für uns den Wert, zusätzlich zu den Eingaben der Publizierenden weitere Datenquellen nach Hochschulpublikationen zu durchsuchen.
- Verdeutlicht den erwarteten Mehraufwand für die Bearbeiter:innen

Dokumentation und Code

Unsere Ausführungen sind verfügbar auf GitHub:

<https://github.com/zhaw-hsb>

- aurora-publication-finder
- aurora-enricher
- aurora-notifier



Enthält jeweils den Java-Code und ein Readme

Angestrebt war eine einfache Erweiterung um weitere Datenquellen und die Nutzung durch weitere Institutionen.

Vielen Dank für Ihre Aufmerksamkeit. Feedback? Fragen?

Nicolai Hauf

ZHAW Zürcher Hochschule für Angewandte Wissenschaften
nicolai.hauf@zhaw.ch / digitalcollection@zhaw.ch

Selina Sturzenegger

Fachhochschule Nordwestschweiz FHNW
selina.sturzenegger@fhnw.ch / irf@fhnw.ch

