

# Das machen wir selbst

Der Aufbau eines eigenen Artikelindex als Alternative  
zu proprietären Angeboten

Jens Lazarus

Universitätsbibliothek Leipzig

## *Begrifflichkeiten*

**Artikelindex**, Unified Index, Aggregierter Index, Central Index

Aus unterschiedlichen Quellen aggregierter Index von (bibliographischen) Daten, die mit möglichst großer Erschließungstiefe Informationsressourcen nachweisen.

Zeitschriftenaufsätze, Buchkapitel, Sammelwerkbeiträge, Abstracts, Volltexte

## *Rahmenbedingungen*

**Discovery Systeme** als Voraussetzung für einen Artikel-Index  
(ermöglicht die Einbeziehung großer Datenmengen)

Anbieter von Discovery Systemen & aggregierten Indices

- EBSCO Discovery Service
- Ex Libris Primo
- OCLC WorldCat Discovery Services
- ProQuest Summon
- ...

## *Rahmenbedingungen UB Leipzig & Sachsen*

Open Source Discovery-System (VuFind) unter dem Namen *finc*  
seit 2012 im produktiven Einsatz

betrieben von einer Nutzergemeinschaft sächsischer Universitäts-  
und Hochschulbibliotheken [folgender Vortrag von Leander Seige]

bis 2015 in Kombination mit dem aggregierten  
Index Primo Central (Ex Libris)



## *Rahmenbedingungen UB Leipzig & Sachsen*

Auslaufen des Lizenzvertrages für Primo Central führte zu der Frage:

Ist es möglich, für die *finc* Nutzergemeinschaft **mit vertretbarem Aufwand** und **ohne qualitative und quantitative Einschränkung** einen eigenen Index zu bauen?

Prüfung von Alternativen vor dem Hintergrund **gewachsener Erfahrungen und Kompetenz** im Metadatenmanagement und IT

## *Motivation*

### Transparenz und Handlungsoptionen in eigenen Nachweissystemen

- Quellen
- Qualität
- Erschließung
- Prozesse
- Abdeckung, Aktualität
- Abgleich eigener Lizenzen

## *Motivation*

Daraus resultieren

- **mehr Möglichkeiten, Selbstbestimmung**
  - eigenes Pre-Processing ermöglicht Analyse und Profilierung der Daten nach eigenen Parametern
  
- **mehr Ressourcen**
  - bedarfsspezifischer Informationsressourcen
  - mehr Anforderungen an Personal und Infrastruktur
  - Zuwachs an Kompetenzen und Entwicklungspotential

## *Zeitlicher Rahmen*

Q3 / Q4 2014	Projektvorlauf: Evaluierung verfügbarer Datenquellen, Vorgespräche mit Lizenzgebern
Jan 2015	Entscheidung für den Aufbau eines eigenen Artikelindex für die finc Nutzergemeinschaft
Q1 / Q2 2015	technische und inhaltliche Realisierung
seit Juli 2015	produktiv in Anwendung



## *Freie Datenquellen*

### CrossRef

nicht-kommerzielles Netzwerk wissenschaftlicher Verlage & DOI  
Registration Agency der internationalen DOI-Foundation

Verlage liefern Metadaten, die DOI-referenziert werden

„Nebenprodukt“ ist umfangreicher Datenpool von Metadaten  
wissenschaftlicher Publikationen, für *Members* frei verfügbar

## *Freie Datenquellen*

Directory of Open Access Journals (DOAJ)

Social Science Open Access Repository (SSOAR)

am Leibniz-Institut für Sozialwissenschaften gepflegte Volltextserver,  
sozialwissenschaftlich relevante Literatur im Open Access

Open Access als Bestandteil der Erwerbung & Lizenzierung

## *Lizenzierte Datenquellen*

DeGruyter

JSTOR

GBI-Genios - Entwicklungspartnerschaft

Wiley - bislang nicht realisiert, Untersetzung im Lizenzvertrag  
notwendig wie jetzt bei Elsevier und Springer

sehr **unterschiedlicher Entwicklungsstand in Bezug auf  
Metadatenkompetenz und Bereitstellung von Metadaten**

## *Lizenzierte Datenquellen*

### Anforderungen an Anbieter

- **Metadaten**
  - konsistente Feldbelegung
  - stabile Datenformate, Standards
  - persistente Identifier
  
- **Datenbereitstellung**
  - strukturierte nachvollziehbare Datenlieferung
  - Option (semi)automatisierter Prozesse
  - transparenter update Zyklus

## *Metadatenverwaltung*

erfolgt über **amsl**, ein auf Linked Data Technologie basiertes, an der UB Leipzig entwickeltes ERM.

Mit **amsl** werden gesteuert und verwaltet

- Anbieterseitige Kollektions-Listen (KBART)
- Datenquellen und darin enthaltene Kollektionen
- Kollektions-Auswahl für beteiligte Bibliotheken
- Zugriff der beteiligten Bibliotheken (Rechtemanagement)
- EZB-Holding-Files (KBART) – automatisierter Abruf wünschenswert (angefragt)
- Technische, Bearbeitungs- und Bezugs-Informationen
- Bereitstellung über Web-API für Indexierung

## *Technische Basis*

Aufbau einer besonders **leichtgewichtigen und stark parallelisierten Verarbeitungskette** für Metadaten. Weitgehender Verzicht auf die Speicherung in Datenbanken.

**Maßgeschneiderte Softwarekomponenten und Datenformate** ermöglichen die performante Prozessierung großer Datenmengen

Alle Komponenten sind in Form von **Open Source Software** öffentlich zugänglich (Dokumentation folgt)

## *Auswirkungen im Katalog*

Qualitätssicherung, stärkere Profilierung durch die Auswahl von Teildaten im Pre-Processing möglich

transparenter und selbstbestimmter Ranking-Algorithmus, gleicher Algorithmus über alle Quellen (zuvor unterschiedliche Algorithmen)

Schnelle und eigenverantwortliche Reaktion auf Problemfälle, zum Beispiel bei fehlerhaften Updates

Bessere funktionale Integration der Daten (Export, Favoritenlisten)

In der Perspektive: Kataloganreicherung über Linked Open Data

## *Lessons Learned*

aus Nutzungssicht **problemloser Umstieg** von Primo Central zum eigenen Artikelindex

enge **Abstimmung zwischen IT und Erwerbung & Lizenzierung** notwendig

**unterschiedliche Metadatenkompetenz auf Anbieterseite**, daraus resultierend höherer Aufwand in der Datenaufbereitung;  
wünschenswert: abgestimmte Verhandlungen von Anforderungsprofilen in Konsortien



## *Fazit*

Erwartungen weitgehend erfüllt; die Erfahrungen im Projekt bestärken uns in der Entscheidung, eigene Kompetenzen und Kapazitäten zu nutzen und auszubauen.

*Vielen Dank.*

## Kontakt

Jens Lazarus

Universitätsbibliothek Leipzig

Bereichsleiter Medienbearbeitung

Beethovenstr. 6

D-04107 Leipzig

Telefon: +49 (0)341 97 30 511

E-Mail : [lazarus@ub.uni-leipzig.de](mailto:lazarus@ub.uni-leipzig.de)

[www.ub.uni-leipzig.de](http://www.ub.uni-leipzig.de)

## Abstract

Das machen wir selbst - der Aufbau eines eigenen Artikelindex als Alternative zu proprietären Angeboten

Jens Lazarus

Universitätsbibliothek Leipzig, Leipzig, Deutschland

Die Universitätsbibliothek Leipzig hat einen eigenen aggregierten Artikelindex aufgebaut und im Juli 2015 in Betrieb genommen. Mit dem Index werden Zeitschrifteninhalte auf Artikelebene im Katalog nachgewiesen. Anders als kommerzielle, cloudbasierte Lösungen, ermöglicht es dieser Artikelindex die Daten individuell zu prozessieren und anzureichern. Dadurch bestehen weitreichende Möglichkeiten zur Umsetzung innovativer Funktionen.

Der Aufbau und der Betrieb des Artikelindex erfolgt in Kooperation mit anderen sächsischen Hochschulbibliotheken. Eingesetzt wird der Index momentan auch von der SLUB Dresden, den Universitätsbibliotheken in Chemnitz und Freiberg, sowie der Hochschulbibliothek Zittau/Görlitz.

Vorangegangen war eine Evaluation zur Verfügbarkeit von Metadaten, die Prüfung von Lizenzbedingungen und Gespräche mit verschiedenen Anbietern. Innerhalb relativ kurzer Zeit konnten konkrete Anwendungsfälle entwickelt werden. In der Folge wurde der Artikelindex für den produktiven Betrieb implementiert und weiter ausgebaut.

Der Beitrag stellt die Entwicklungsstufen und erste Erfahrungen im Kontext von *finc* (Open Source Discovery System) und *amsi* (Open Source ERM) dar. Schwerpunkte liegen hierbei auf der Integration verschiedener lizenzierter und offener Datenquellen und der Entwicklungsperspektive des Artikelindex. Technische Aspekte werden in Grundzügen dargestellt.