

LOC-DB: Linked Open Citation Database

Entwicklung von Werkzeugen und Prozessen zur verteilten Erschließung von Zitationen auf Basis der Linked-Data-Technologie

Anne Lauscher und Prof. Dr. Kai Eckert
Web-based Information Systems and Services
Hochschule der Medien, Stuttgart

Motivation

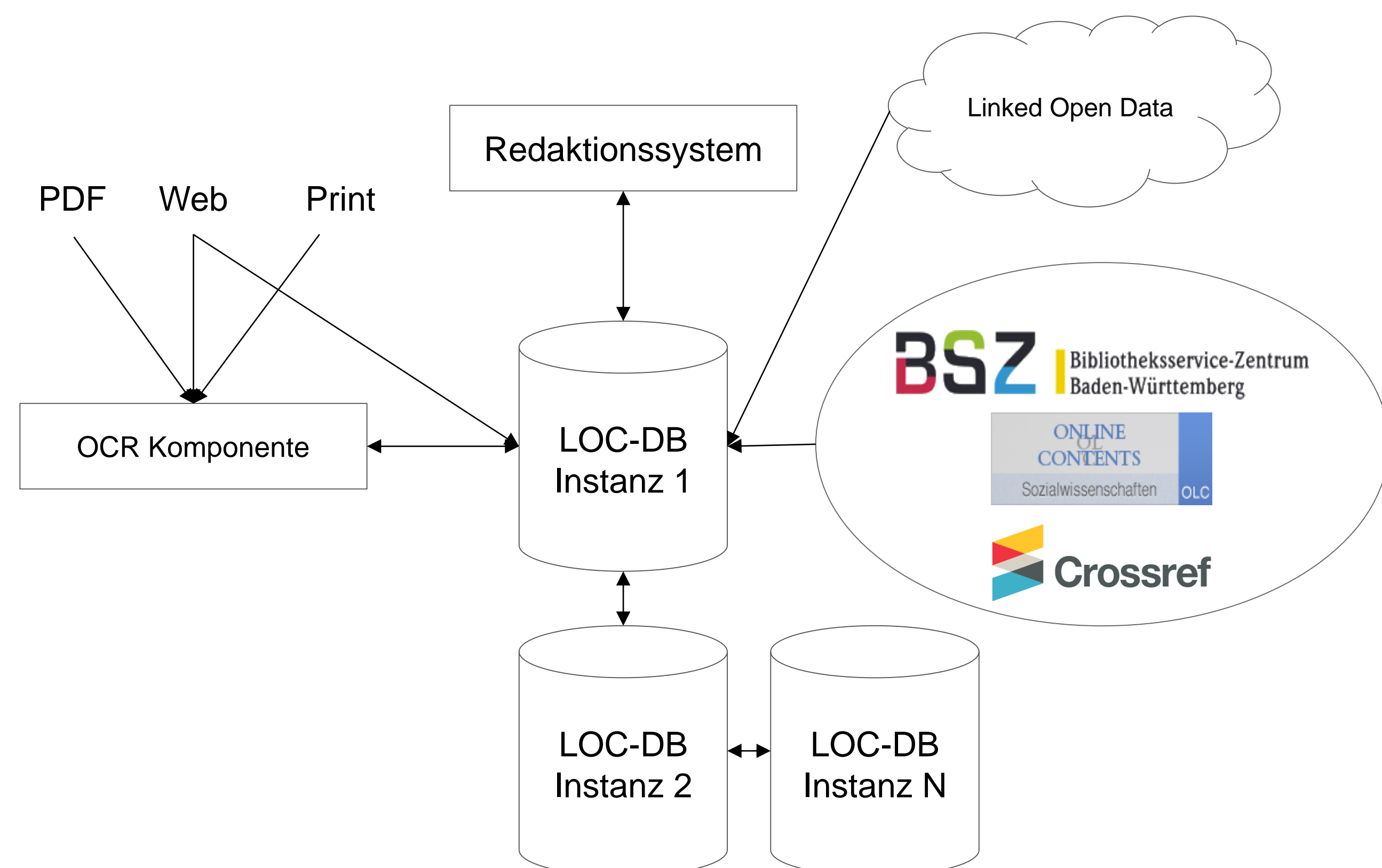
Zitationen sind eine wertvolle Ressource:

- Bewertung von Wissenschaftlern über Zitationsindizes.
- Zitationen als Recherchemittel.
- Zitationsanalyse zur Unterstützung vieler Aufgaben im Rahmen der automatischen Auswertung wissenschaftlicher Publikationen, z.B. zur automatischen Generierung einer Zusammenfassung eines Forschungsfelds.

Problem: Zitationen werden während der Katalogisierung in Bibliotheken nicht erfasst.

Projektziel: Erbringen des Nachweises, dass eine effiziente und nachhaltige Erschließung von Zitationen in Bibliotheken möglich ist.

Infrastrukturübersicht



Datengrundlage

- Fachbestand Soziologie
- Phase 1 (Jahresmenge): Alle 2011 erworbenen Medien der UB Mannheim
- Phase 2 (Workflowintegration): Ausweitung auf den laufenden Jahrgang 2017
- Art der Medien:
 - Monographien
 - Sammelwerke
 - Zeitschriften
 - Elektronisch und gedruckt

LOC-DB Hauptkomponenten

LOC-DB-OCR-Komponente

Verantwortlich für das automatische Erkennen und Extrahieren sowie Segmentieren von Referenzen aus einer Vielzahl von Quellen, z.B. aus Scans oder PDF-Dokumenten, unter Einsatz von maschinellem Lernen.

LOC-DB-Zentralkomponente

Zum zentralen Management und zum Speichern der erzeugten Dateien und Daten sowie zur Generierung von Vorschlägen aus externen Systemen, wie z.B. Crossref, SWB-BSZ oder Google Scholar.

LOC-DB-Redaktionssystem

Als Benutzerschnittstelle für Bibliothekare zum effizienten Verlinken der erkannten Referenzen mit externen und internen bibliographischen Ressourcen auf Basis aktueller UI-Technologien.

Redaktionssystem

Beispiel: Verlinkung der Referenzen auf einer gescannten Buchseite.

Datenmodell

Gewährleistung optimaler Nutzbarkeit und Interoperabilität der erzeugten Daten durch

- Adaption des OpenCitations Metadaten-Modells.
- Veröffentlichung der Daten im RDF-Format unter Verwendung der Semantic Publishing and Referencing (SPAR) Ontologies.

