



# **Das Projekt LOC-DB und die Vision der offenen Zitationen**

Dr. Annette Klein  
Erster LOC-DB Workshop  
Mannheim, 07.11.2017

## Bibliotheken und Metadaten

Bibliotheken katalogisieren  
selbständige Werke

Bibliotheken lizensieren  
elektronische Medien (und  
deren Metadaten)

Bibliotheken nutzen Fremddaten

Bibliotheken erstellen  
Satzdatenbanken

Bibliotheken  
Resource Dis  
systeme

... und jetzt auch noch  
Literaturverzeichnisse?

Bibliotheken scannen  
Inhaltsverzeichnisse

Bibliotheken lizensieren  
bibliographische Datenbanken

Bibliotheken betreuen  
Hochschulbibliographien

Bibliotheken erstellen  
bibliometrische Analysen

## Das Linked Open Data-Versprechen

„Die Ideen der kollaborativen Katalogisierung und Erschließung [...] werden auch im Semantic Web unterstützt und weitergeführt. Die Wissensdatenbanken der Bibliotheken werden damit mit der gesamten Wissensbasis des Semantic Webs verknüpft.“ (Zumstein 2012)

→ mehr Information, mehr Effizienz, neue Services

Warum ist die Revolution bisher ausgeblieben?





Fehlen

- die richtigen Instrumente?
- die Ressourcen?
- der Wille, neue Wege zu gehen?

## Das Projekt LOC-DB

- **Fragestellung:** Was ist notwendig (Personal, Infrastruktur, Material), um die Materialien, die Bibliotheken für ihre Nutzer bereitstellen, vollständig mit allen ihren Beziehungen zu erschließen, d.h.
  - Selbständige Werke (Bücher, Zeitschriften; gedruckt und elektronisch)
  - Enthaltene unselbständige Werke (Kapitel in Sammelwerken, Zeitschriftenartikel)
  - Zitierte Werke (selbständig und unselbständig)
- **These:** Wenn möglichst viele **vorhandenen Daten nachgenutzt** werden und die **kooperative Erschließung** durch **Automatisierung** optimal unterstützt wird, ist eine neue Dimension der Erschließung realisierbar.

## Welche Daten lassen sich nachnutzen?

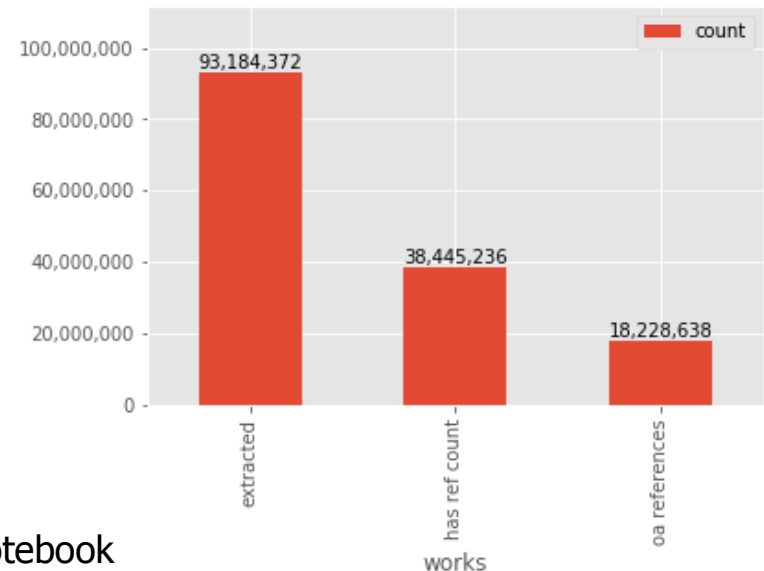
- Bücher, Zeitschriften  $\Rightarrow$  Verbundkataloge, ZDB 
- Zeitschriftenartikel  $\Rightarrow$  Crossref, Online Contents 
- Kapitel in Sammelwerken  $\Rightarrow$  Crossref 
- Zitationsbeziehungen  $\Rightarrow$  Crossref, OpenCitations 

Außerdem: Google Scholar, DataCite, Wikidata... 

## Welche Daten lassen sich nachnutzen?

**Crossref:** „not-for-profit membership organization for scholarly publishing“

- Metadaten-Registrierung und DOI-Vergabe für wiss. Verlage
- Aktuell über 93 Mio Datensätze
  - Davon 41% mit Zitationen
  - Davon 47% frei nachnutzbar
- Trend zur Freigabe von Zitationsdaten durch [I4OC](https://www.i4oc.org/)



## Welche Daten lassen sich nachnutzen?

**OpenCitations:** Repositorium offener, verlinkter Zitationsdaten

- Leitung: David Shotton (Oxford), Silvio Peroni (Bologna)
  - Umfang: Ca. 11,5 Mio. Zitationen aus 270.000 Artikeln
  - Ursprünglich Fokus auf Lebenswissenschaften, Ausweitung angestrebt → möglicher "Hub" für Linked Open Citation Data weltweit
- Vortrag von David Shotton um 11:50 Uhr (per Webkonferenz)

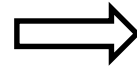
## Was fehlt?

- Metadaten v.a. für (nur) gedruckte Publikationen
- Zitationsbeziehungen, die von diesen Publikationen ausgehen (und einige mehr)
- Einheitliche Strukturierung und Datenqualität der vorhandenen Metadaten und Zitationen
- Verknüpfung zwischen bibliothekarischen Nachweissystemen und anderen Datenquellen



## Was lässt sich automatisieren?

- Nachnutzung vorhandener Daten, Verknüpfung mit bibliothekarischen Nachweissystemen



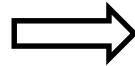
LOC-DB Redaktionssystem mit integrierten Schnittstellen

→ Vorträge von

- Philipp Zumstein, Laura Erhard, Sylvia Zander (9:30 Uhr)
- Kai Eckert, Anne Lauscher, Lukas Galke (9:50 Uhr)

## Was lässt sich automatisieren?

- Erschließung gedruckter Publikationen

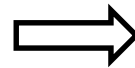


OCR, automatische Informationsextraktion und Segmentierung

→ Vortrag von Sheraz Ahmed und Syed Tahseen Raza Rizvi (10:10 Uhr)

## Was lässt sich automatisieren?

- Verbesserung der Datenqualität, Standardisierung



Verknüpfung mit Bibliotheksdaten, automat. Segmentierung

→ Vorträge von

- Behnam Ghavimi (10:50 Uhr)
- Timotheus Chang-whae Kim (11:10 Uhr)

## Wie können wir kooperieren?

- Datenproduktion: Netz mit verteilten LOC-DB-Instanzen als Vision
  - Nachnutzung der Daten: Bibliothekskataloge, Discovery Services, bibliometrische Anwendungen ...
  - Aufgabe für Verbünde, FIDs, Zentrale Fachbibliotheken, einzelne wissenschaftliche Bibliotheken?
- Vortrag von Tamara Pianos (11:30 Uhr) und Podiumsdiskussion (15:20 Uhr)

## Organisatorisches

- Mittagspause mit Imbiss von ca. 12:10-13:00 Uhr, Kaffeepausen 10:30-10:50 Uhr und 15:00-15:20 Uhr
- Praxis-Sessions ab 13:00 Uhr
  - Session 1 und 2 (Workflows testen): Fuchs-Petrolub-Saal
  - Session 3 (Expertengespräch): Raum SW 277
  - Ergebnisse und Diskussion 14:30-15:00 Uhr: Fuchs-Petrolub Saal

Vielen Dank für Ihre Aufmerksamkeit!