Freie Universität Berlin

Towards Linked Library Data @ Freie Universität Berlin

Michael Franke-Maier, Dominik Hagel, Johannes Hercher, Cosima Wagner

Universitäts Bibliothek CeDiS

Wer wir sind:

Freie Universität Berlin

Michael Franke-Maier, M.A., MA LIS

Universitätsbibliothek
Stellvertretender Leiter der Zugangsabteilung
- Inhaltserschließung,
- Dokumentationszentrum Vereinte Nationen - Europäische Union,
- Fachreferate Sozialwissenschaften und Recht

franke@ub.fu-berlin.de
Tel: 030 / 838 56392

Johannes Hercher, Dipl.-Dokumentar

Universitätsbibliothek
Systembibliothekar
- Administration Discovery-System PRIMO
- Linked Open Library Data
- Leitung der AG Digitalisierung

hercher@ub.fu-berlin.de
Tel: 030 / 838 57619

Dr. Cosima Wagner

Universitätsbibliothek / Campusbibliothek
Projekt „Bibliotheken als Partner der Internationalisierungsstrategie der FU“
- Fachlicher Hintergrund: Japanologie, Science & Technology Studies
- Multilingualität und Forschungsinfrastrukturen
- Forschungsdatenmanagement und Digital Humanities

cosima.wagner@fu-berlin.de
Tel: 030 / 838 60307

Dr. Michael Dominik Hagel

Universitätsbibliothek
Projekt „Evaluierung und Entwicklungspotentiale der Universitätsbibliographie“
- Fachlicher Hintergrund: Germanistik
- Forschungsnahe Bibliotheksservices
- Daten, Metadaten und geisteswissenschaftliche Forschung

dominik.hagel@fu-berlin.de
Tel: 030 / 838 64049

Universitäts Bibliothek CeDiS

Agenda

1. Daten der UB, des CeDiS und des FU-Archivs an der Freien Universität Berlin (Auswahl)

2. Strukturierte Erfassung der Daten
   - inhaltliche und formale Erschließung
   - Normdaten

3. Möglichkeiten der Datenextraktion

4. Linked Open Data in der Forschung

1. Daten der UB + CeDiS + Archiv (Auswahl)

Katalogdaten

~ 5,4 Mio Titel im Katalog (physische Exemplare + eRessourcen)
http://primo.fu-berlin.de

~ Bestand in Originalschrift, z.B.: CJK, Arabisch, Hebräisch = 90.000 Ex. physisch, 128.393 mit eRessourcen)

> 3 Mrd Metadaten zu wissenschaftlichen Publikationen via Primo Central / Central Discovery Index

Datenbanken und Informationssysteme

~ 6.600 lizenzierte Datenbanken (via DBIS http://dbis.ur.de), Textmining fallweise zu prüfen

1. Daten der UB + CeDiS + Archiv (Auswahl)

Repositorium (refubium)

~ 14.000 Dissertationen (Charité und FU)

~ 20.000 Artikel

~ 200 gescannte Bücher (Service Digitalisierungsauftrag)

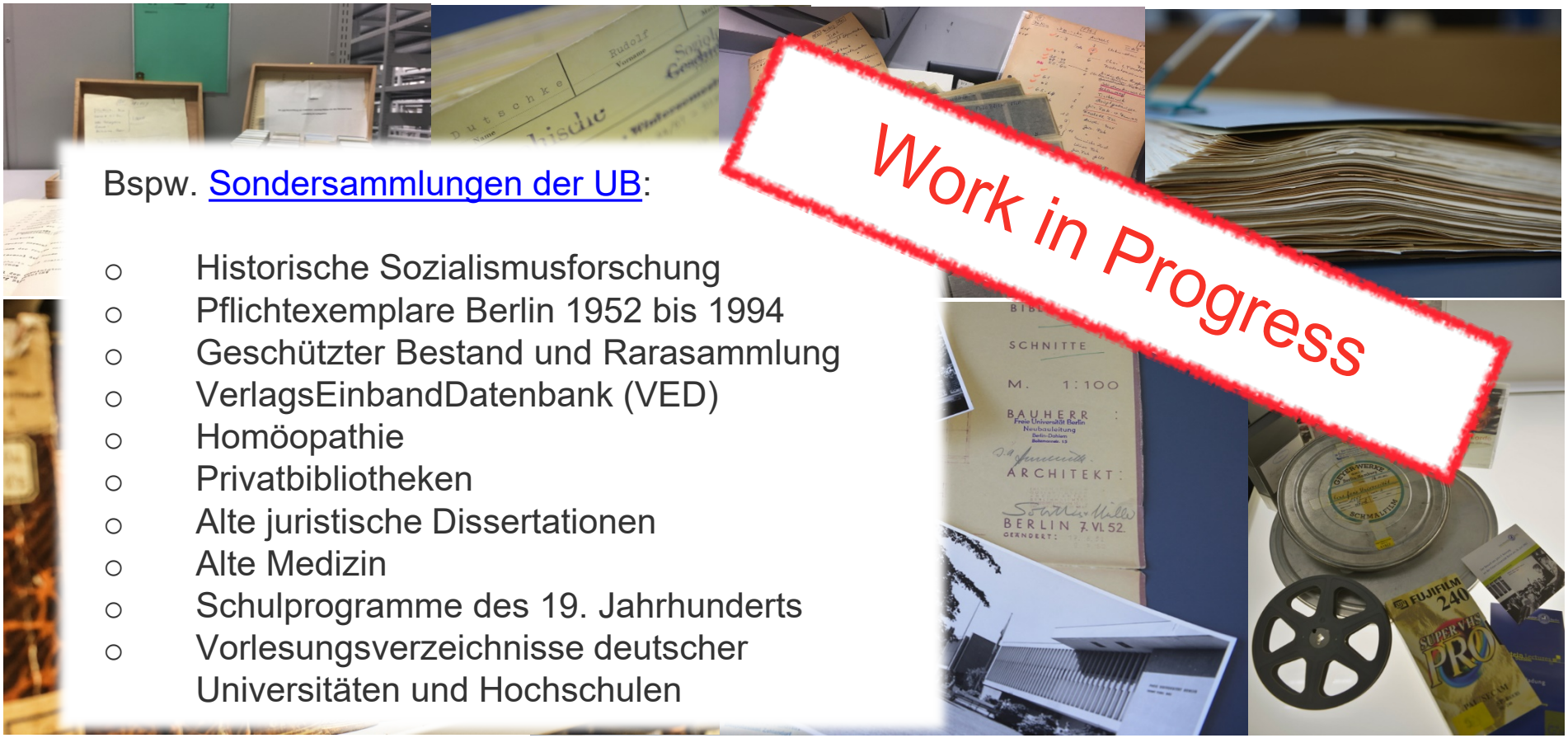
FU-Archiv

~ 4.200 Datensätze zu Hochschullehrer*innen seit Gründung

CeDiS

~ 2.000 Digitale Interviewsammlungen (Videos mit Zeitzeugen, Schwerpunkt Nationalsozialismus und Holocaust) (/50.000 mit externen Interviews) (AV Daten, Textdaten, perspektivisch mit TEI erschlossen)

Sammlungen



Bspw. Sondersammlungen der UB:

- Historische Sozialismusforschung
- Pflichtexemplare Berlin 1952 bis 1994
- Geschützter Bestand und Rarasammlung
- VerlagsEinbandDatenbank (VED)
- Homöopathie
- Privatbibliotheken
- Alte juristische Dissertationen
- Alte Medizin
- Schulprogramme des 19. Jahrhunderts
- Vorlesungsverzeichnisse deutscher Universitäten und Hochschulen

2. Strukturierte Erfassung der Daten

Katalogdaten der UB

- **Formale** Erschließung nach RAK (Regel für die alphabetische Katalogisierung), RDA (Resource Description and Access) seit 2015

- **Inhaltliche** Erschließung mit
  - Regeln für die Schlagwortkatalogisierung (RSWK)
  - Erschließung mit der Dewey Decimal Classification (DDC)
  - Erschließung mit der Regensburger Verbundklassifikation (RVK)
  - Medical Subject Headings (MeSH) - nur Charité, Veterinärmedizin

2. Strukturierte Erfassung der Daten

Katalogdaten der UB (Fortsetzung)

- **Erschließung mit GND-ID** (seit 2014)
  - Werktitel
  - Geografika
  - Personen, Körperschaften, Konferenzen
  - Sachbegriffe
- **Weitere Normdaten:** z.B. LCSH, STW über den Bayerischen Verbund Katalog (B3Kat), Übersicht der Classification Source Codes (in vielen Fällen unnormiert)

Archiv und CeDiS

- CeDiS: Einsatz von GND, Kooperation mit der Bayerischen Staatsbibliothek
- Archiv: bislang keine Nutzung von Normdaten

Recherche mit Identifiern (Beispiele)

Art	FU-Katalog via Primo	ALMA via SRU	Refubium via OAI	CDI via Primo
BVB-Nummer	x	x	x	-
OCLC-id	-	x	-	-
ORCiD	-	-	-	(x)
GND	x	x	-	-
VIAF	-	x	-	-
RVK	x	x	-	-
DDC	x	x	x	-
Wikidata	-	-	-	-
ISBN/ISSN	x	x	x	x

3. Datenextraktion: Überblick Schnittstellen

ALMA (bibliografische Daten) :

- SRU serialisiert als: marcxml | dc | mods … (max. 10.000 Datensätze!!)
- OAI-PMH Konfiguration auf Nachfrage
- Linked Data Schnittstellen (JSON-LD, Bibframe, RDF)
- manueller Export von .mrc via FTP (auf Nachfrage)

Refubium (bibliografische Daten / Volltexte / Digitalisate inkl. OCR):

- OAI-PMH
- Beispiel FU-Dissertationen

FU-Archiv:

- (manueller) CSV-Export möglich

CeDis Interviewsammlungen

- OAI-PMH (im Aufbau)

3. Datenextraktion via SRU

Auswahl von Titeln via SRU "4006804-3" (Biografie) AND "4017604-6" (Flüchtling)

Export als CSV via Catmandu

```
catmandu convert SRU
        --version 1.2
        --parser marcxml
        --base http://fu-berlin.alma.exlibrisgroup.com/view/sru/49KOBV_FUB
        --recordSchema marcxml
        --query 'authority_id=4006804-3 AND authority_id=4017604-6'
to CSV > ~/Desktop/MMSiDs.csv
```

Beispiele und Dokumentation:
http://bit.ly/FUB-LLD-SRU

Weitere Indizes:

- title, creator, language
- dewey_decimal_class_number
- digital_object_identifier
- u.v.a.m. siehe: SRU-explain

3. Datenextraktion: Weitere Schnittstellen

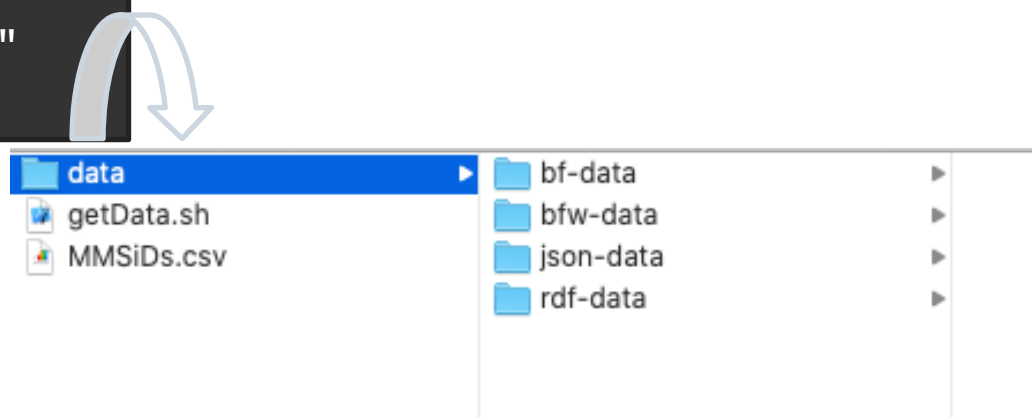
ALMA Linked Data Schnittstellen **nur über Systemnummern** (**MMSiD**) nutzbar!

- **JSON-LD**: https://open-na.hosted/.../ bibs/**9958929138202883**.jsonld
- **RDF**: https://open-na.hosted/.../rda/entity/manifestation/**9958929138202883**.rdf
- **Bibframe**: https://open-na.hosted.../bf/entity/instance/**99004943554040288 3**

Download mehrerer Datensätze mit **MMSiDs.csv**

=> Script: http://bit.ly/FUB-getData

```
./getData.sh -c MMSiDs.csv -f "bf"
./getData.sh -c MMSiDs.csv -f "json"
./getData.sh -c MMSiDs.csv -f "rdf"
```



Hilfe und Dokumentation:

http://bit.ly/FUB-LLD-LD

4. Linked Data für die Forschung

- Exploration der Potentiale von linked-data-services für die Wissenschaft, auch im Rahmen von Forschungsaktivitäten in der Bibliothek

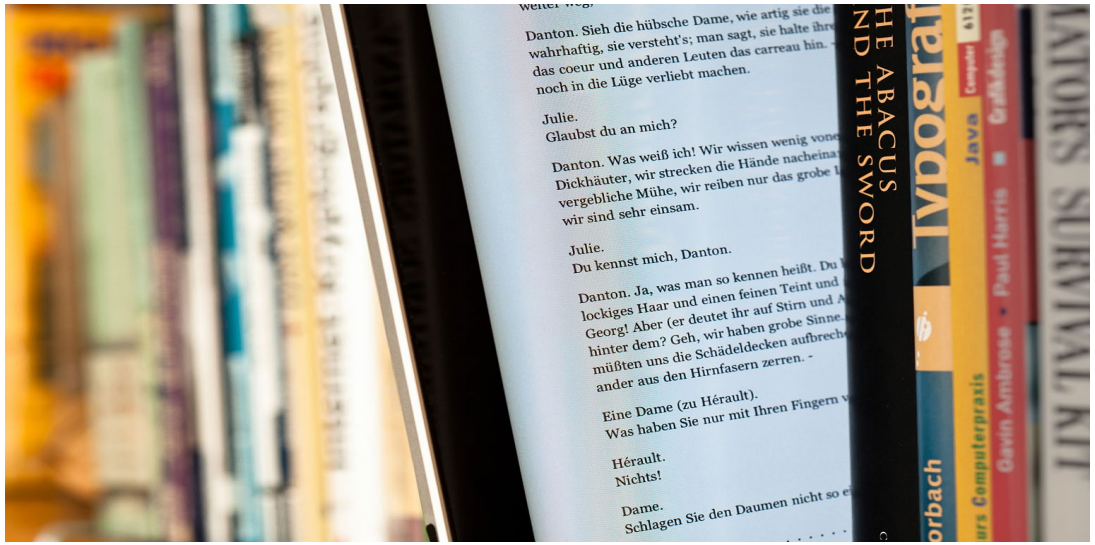
- Bereitstellung von Linked (Open) Library Data und Weiterentwicklung von Services unter Berücksichtigung der Anforderungen des Clusters

- Ideen, Anregungen erwünscht
  => http://bit.ly/FUB-LLD-Ideen

Literaturauswahl: Linked (Open) Library Data als Forschungsthema

- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012, May 24). Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models. 78th IFLA General Conference and Assembly.
- Altenhöner, Reinhard. „*Data Curation* als Leitbegriff für eine neue Facette im bibliothekarischen Berufsbild.“ BIT online: Bibliothek, Information, Technologie. 21, no. 3 (2018): 197–204.
- Bubenhofer, Noah; Klaus Rothenhäusler. „*Korporatheken*: Die digitale und verdatete Bibliothek.“ 027.7 Zeitschrift für Bibliothekskultur / Journal for Library Culture 4, no. 2 (November 11, 2016): 60–71. https://doi.org/10.12685/027.7-4-2-154.
- DLF Linked Open Data Zotero Group: https://www.diglib.org/groups/linkeddata/ .
- Hoolandvan, Seth Van; Ruben Verborgh. Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata. US ed. Chicago: Neal-Schuman, 2014.
- Jäschke, Robert: „Uncovering Hidden Treasures: Libraries and Data Science“ held at KNVI Smart Humanity Conference, November 15, 2019, Amsterdam, The Netherlands. https://slides.igada.de/2019-11-14-knvi_smart_humanity/
- Ders.; Fischer, Frank: „Liebe und Tod in der Deutschen Nationalbibliothek. Der DNB-Katalog als Forschungsobjekt der digitalen Literaturwissenschaft“. DHd 2018 Konferenz, March 1, 2018. lehkost.github.io/slides/2018-koeln/
- Maier, Petra. „Digital Humanities und Bibliothek als Kooperationspartner. Faktoren einer aktiven Unterstützung im Bereich der Metadaten“ (2016), DARIAH-DE Working Paper Nr. 19, Göttingen. http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2016-19.pdf
- WC3 Library Linked Data Incubator Group: https://www.w3.org/2005/Incubator/lld/
- Wilson Emma Annette; Alexander, Mary (2016): “When Metadata Becomes Outreach: Indexing, Describing, and Encoding for DH.” dh+lib. http://acrl.ala.org/dh/2016/07/29/when-metadata-becomes-outreach

Vielen Dank!



Source.: Maximilian Schönherr: https://de.wikipedia.org/wiki/E-Book#/media/File:EBook_between_paper_books.jpg

Kontakt:

- Michael Franke-Maier franke@ub.fu-berlin.de
- Dominik Hagel dominik.hagel@fu-berlin.de
- Johannes Hercher hercher@ub.fu-berlin.de
- Cosima Wagner cosima.wagner@fu-berlin.de