

# Lost in Transcription

Rolle und Aufgaben von Bibliotheken  
für die Multilingual Digital Humanities

Cosima Wagner & Jonas Müller-Laackman  
BiblioCon 2024

# Struktur

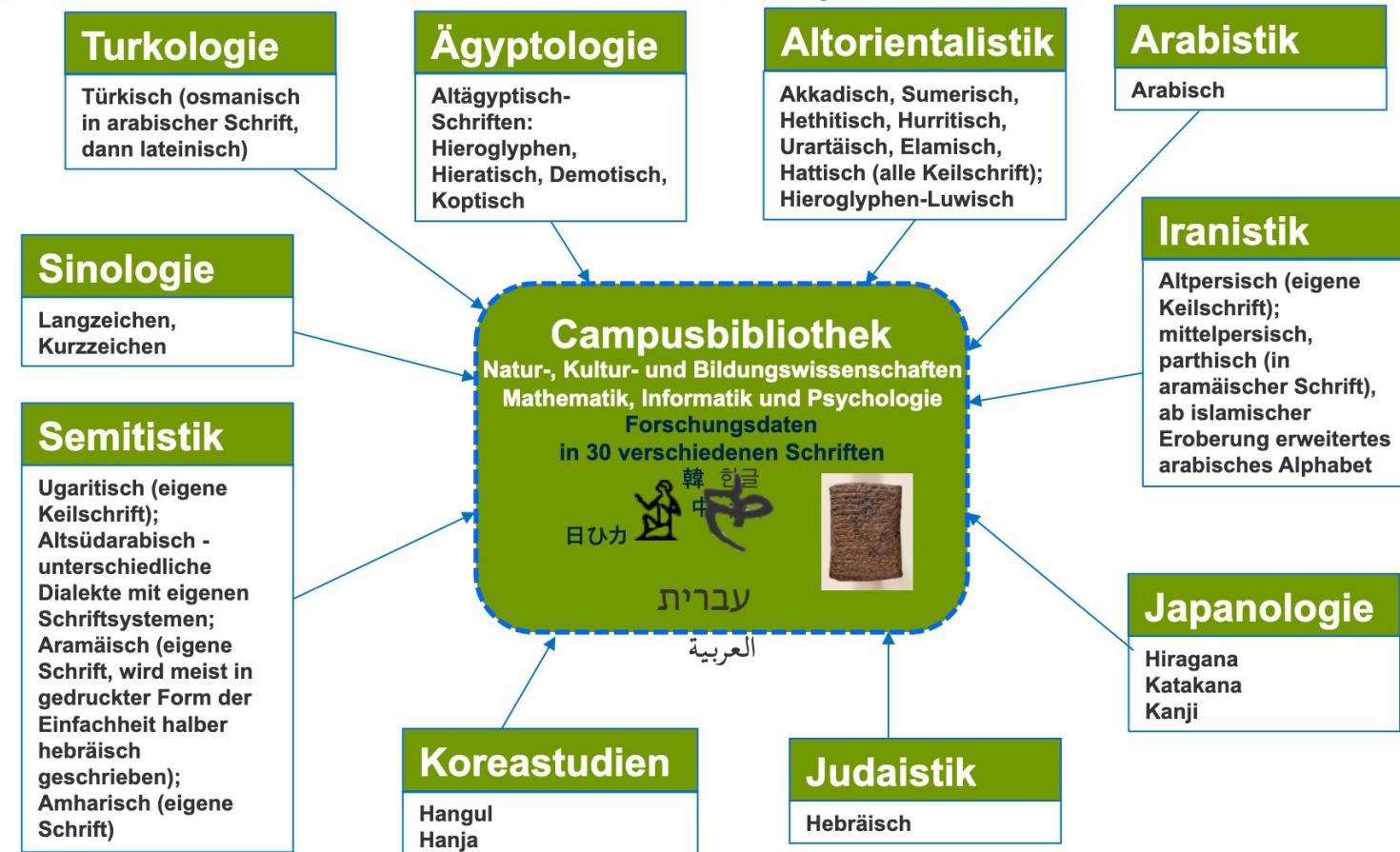
- Einführung
- Problematisierung: Mehrsprachigkeit und –schriftlichkeit im digitalen Raum
- Multilingual Digital Humanities als problemorientiertes Schnittstellenfeld
- Bibliotheken als Akteure in den Multilingual Digital Humanities
- Perspektiven, Chancen, Ausblick

# Einführung: Wer sind wir?

- DHd AG Multilingual DH
  - Digital Humanists
  - Historiker:innen
  - Literaturwissenschaftler:innen
  - Liaison Librarians (Cosima Wagner, FU Berlin)
  - Digitale Forschungsdienste (Jonas Müller-Laackman, SUB Hamburg)
- Fachsprachen in der AG
  - Arabisch, Hebräisch, Japanisch, Chinesisch, Russisch, Ukrainisch

# Beispiel: Campusbibliothek der FU Berlin

Multilinguale Forschungsdaten aus den Regionalwissenschaften bzw. den sogenannten „Kleinen Fächern“ in der Freien Universität / Campusbibliothek



# Problematisierung: Mehrschriftlichkeit

- Warum eigentlich relevant?
  - Unicode unvollständig
  - OCR Qualität
  - Keine oder konkurrierende Transliterations/Transkriptions-Standards
    - Beispiel: Makka, Mekka, Mekkah, Makkah, ... Für مكة
  - Retrieval: Unterschiede in Syntax Morphologie und Semantik
  - Zeitangaben und Kalendersysteme
  - Darstellung
  - Mangelnde Unterstützung in NLP-Prozessen

# Quod erat demonstrandum:

members of the DH community (O'Donnell et al., 2015).

The second major issue confronting multilingual DH is technical. As the name implies, DH relies on pre-existing digital technologies that themselves perpetuate English as a dominant language. As Ramsey Nasser notes in the overview of his programming language بلق, these are almost exclusively based on the ASCII character set, which encodes Latin characters; this bias underlies most of our digital platforms (Nasser, N/D). Moreover, both programming syntax and vocabulary usually use the English language as its base, and the provided documentation is

Hier müsste es قلب heißen...

Btw.: auch OnlyOffice kann das nicht... بلق

# Problematisierung: Bibliotheken

- Probleme mit Mehrsprachigkeit in:
  - Beschaffung
  - Digitalisierung
  - Erschließung und Auffindbarkeit
  - Fachpersonal und Beratungskapazitäten
  - Software und Systeminfrastruktur
  - Provinienzforschung und Dekolonialisierung

# Problematisierung I: Beschaffung

- Was wird für wen (nicht) beschafft?
  - „Privilege to Select“ (Schmidt 2020: *The Privilege to Select. Global Research System, European Academic Library Collections, and Decolonisation*)
  - Mangelnde Repräsentation der sogenannten „Kleinen Fächer“ bei Budgetierungs- und Beschaffungsentscheidungen
- Große Unterschiede in Anschaffungskosten und –aufwänden (z.B. Zoll, Auflagenzahl, Verfügbarkeit)
  - Mangelnde Personalkapazitäten

# Problematisierung II: Fachpersonal

- Wenig und oft nur zufällig Personal mit Fachexpertise an Entscheider:innenstellen
- Inkompatible oder unattraktive Positionen für wissenschaftliches Personal mit Fachkenntnis
- Keine Beratungs- und Einschätzungsmöglichkeiten zum Bestand in NLS, wenn kein Fachreferat oder Fachbibliothekar:in für Area Studies vorhanden
- Kaum Fachexpertise im GND-Kontext

# Problematisierung III: Dekolonialisierung

- Ignoranz gegenüber Mehrschriftlichkeit blendet Kolonialgut inhaltlich oft aus
- Keine bibliotheksinterne Möglichkeit für die Erforschung von Provenienz und Überlieferungsgeschichte
- Provoziertes Unwissen über große Teile bibliothekarischer Bestände aus dem globalen Süden

# Problematisierung IV: Auffindbarkeit

- Recherchesysteme unterstützen von Haus aus oft nur Sprachen in lateinischer Schrift
  - Kommerzielle Lösungen mit hermetischen Feature Request-Workflows
  - Fehlende Auffindbarkeit je nach Index, Erschließung, Schreibweise
- » Bestände werden unwillentlich dem wissenschaftlichen Zugriff vorenthalten

# Problematisierung: Fazit

# Ressourcenlücke

# Schnittstellenfeld: Multilingual DH

- Digital Humanities als Schnittstellendisziplin für Wissenschaft und wissenschaftliche Bibliotheken
- Mehrsprachige Digital Humanities mit Fokus auf nicht-lateinische Schriften (NLS, z.B. Arabisch, Japanisch, Kyrillisch)
- Beispiele:
  - Open Arabic Periodical Editions
  - Bibliotheca Arabica
  - Closing the Gap in Non-Latin Script Data

# Herausforderungen in den MLDH

- Fehlende Infrastrukturen, die NLS unterstützen
- Fehlende Unterstützung in Bibliotheken (...!)
- Vokabulare mit Bias des Globalen Nordens
- Herausforderungen in der Vernetzung mit Akteuren des Globalen Südens
  - Digital Humanities – wie so vieles – wird eindeutig von Akteuren aus dem Globalen Norden dominiert

# Bibliotheken als Akteure

- Synergetische Problemfelder
  - Bibliotheken sind mangels Kapazitäten auf die Wissenschaft angewiesen
  - Wissenschaftler:innen sind mangels Kapazitäten auf Bibliotheken angewiesen
- Bibliotheken bieten stabilere Infrastruktur für die Problematisierung von Mehrsprachigkeit: Weniger Projektierung und mehr Daueraufgaben mit Bezug zu Mehrschriftlichkeit
- Skillsharing und IK im Bereich der (Multilingual) Data Literacy

# Bibliotheken als Akteure: Chancen

- Nachhaltigkeit als Paradigma verpflichtet
- „Linguistic Equity and Diversity“ („Enhancement request“ der Harvard University Library an ExLibris)
- Dekolonisierung nicht nur als Lippenbekenntnis
- Vernetztes Denken verlangt Auffindbarkeit
- Ko-Design für neue bibliothekarische Services in den Digital Humanities: Library-Faculty Collaboration auch im Bereich Mehrschriftlichkeit

Quelle enhancement request der Harvard University Library:

<https://wiki.harvard.edu/confluence/display/LibraryStaffDoc/HOLLIS+Known+Issues+and+Enhancement+Requests#HOLLISKnownIssuesandEnhancementRequests-PleaservotefortheseenhancementrequestsinIdeasExchange>

# Bibliotheken als Akteure: Diversität

“....the meaning of language and discourse in sustaining cultural heritage and the world’s languages, and cultural diversity. It has been acknowledged that culture constitutes a part of social sustainability, and covers such aspects as social justice and equity, social participation, economic needs and work, and, above all, awareness of the range of sustainability ideas.”

Kamińska, Anna M., Łukasz Opaliński, and Łukasz Wycislik. 2022. "The Landscapes of Sustainability in the Library and Information Science: Systematic Literature Review" *Sustainability* 14, no. 1: 441, p.18-19, <https://doi.org/10.3390/su14010441>

# Perspektiven und Chancen

- Reduzierung von Data Bias
- Stärkung der Bibliotheksinfrastruktur für die Forschung
- Nachhaltigkeit nicht nur für lateinische Schriften
- Förderung von Inklusion und Diversität
- Synergetische Unterstützung von Dekolonisierungsbemühungen

# Ausblick: Was geht schon?

- Arbeitsgruppen: DHBib, DHd AG Multilingual DH, Decolonize the Library, ADHO-SIG MLDH, DARIAH WG MLDH, Cambridge University Library decolonisation working group
- NFDI als Aktivitätsfeld wird erschlossen
- Thema rückt in wissenschaftlichen Bibliotheken zunehmend in den Fokus

... Zumindest reden wir schon mal drüber.

# Ausblick: Was muss noch? (1/2)

- Nutzung von Open Source Systemen statt kommerzieller Locked-In Angebote von Data Brokern
- Katalog- und Recherchesysteme kontinuierlich auf alle bestandsrelevanten Schriften ausbauen
- Personalressourcen und –expertise schaffen durch Weiter- und Fortbildungsangebote
- Konsistente Erschließungsrichtlinien für Umschriftsysteme

# Ausblick: Was muss noch? (2/2)

- Mehrschriftlichkeit muss eine größere Rolle in Entscheidungsgremien spielen
- Einbettung von Mehrschriftlichkeit als Thema in die bibliothekarische Ausbildung
- Attraktivere Möglichkeiten für Quereinsteiger:innen
- Mehrschriftlichkeit in der Projekt- und Infrastrukturentwicklung von Beginn an mitdenken
- ...

# Vielen Dank.

# Etwas Eigenwerbung muss auch noch...



DHd AG Multilingual  
DH auf GitHub



Discord-Channel der  
AG mit Themenkanälen  
und Community Hour

# Initiativen / Arbeitsgruppen

## Multilingual DH

- DHd AG Multilingual DH:  
<https://m-l-d-h.github.io/DHd-AG>
- DARIAH WG Multilingual DH:  
<https://multilingual.hypotheses.org/about>
- ADHO Special Interest Group Multilingual DH:  
<https://adho.org/sigs/#multilingual-DH>

## Decolonizing the library

- Netzwerk Dekolonialisierung von Bibliotheken im DACH-Raum  
<https://decolonizethelibrary.mirah-eze.org/wiki/Hauptseite>
- Cambridge University Library decolonisation working group  
<https://www.lib.cam.ac.uk/about-library/diversifying-collections-and-practices/cambridge-university-libraries-decolonisation>

# Literaturauswahl

Bloc, Sharon (2020). “Erasure. Misrepresentation and Confusion: Shortcomings of JSTOR Topics on Women’s and Race Histories,” *Digital Humanities Quarterly* 14:1.

Coleman, Catherine Nicole (2020). “Managing Bias When Library Collections Become Data.” *International Journal of Librarianship* 5.1: 8–19.

Fioronte, D. (2021). “Taxation against Overrepresentation? The Consequences of Monolingualism for Digital Humanities”. Kim D. and Koh A. (eds), *Alternative Historiographies of the Digital Humanities*. Santa Barbara: Punctum Books, 333–376.

Grallert, T.; Kudela X. M.; Lincke, E.-S.; Lindermann, C.; Mende, J.-K., Müller-Laackman, J.; Schmid, L. (2023c). “Umgang mit Multilingualität im DACH und DHd Verband.” Zenodo, May 22, 2023.  
<https://doi.org/10.5281/zenodo.7957187>.

Malínek, V., Umerle, T., Gray, E., Heibi, I., Király, P., Klaes, C., Korytkowski, P., Lindemann, D., Moretti, A., Panušková, C., Péter, R., Tolonen, M., Tomczyńska, A., Vimr, O. (2024). “Open Bibliographical Data Workflows and the Multilinguality Challenge”, *Journal of Open Humanities Data*, 10, 27, 1-14.  
10.5334/johd.190

Schmidt, Nora. 2020. *The Privilege to Select. Global Research System, European Academic Library Collections, and Decolonisation*. Lund: Lund University: Faculties of Humanities and Theology. DOI: 10.5281/zenodo.4011296