

Nicht-lateinische Schriften in multilingualen Umgebungen

Beispiele für die Verbesserung von Kuratierung in kooperativen Netzwerkstrukturen

Bibliothekskongress, 18.3.2019, Leipzig
Esther Asef, Martin Lee

Campusbibliothek der Freien Universität
Berlin

Agenda Kontext

Projekt FDM_OAS-Orient

Technische Herausforderungen

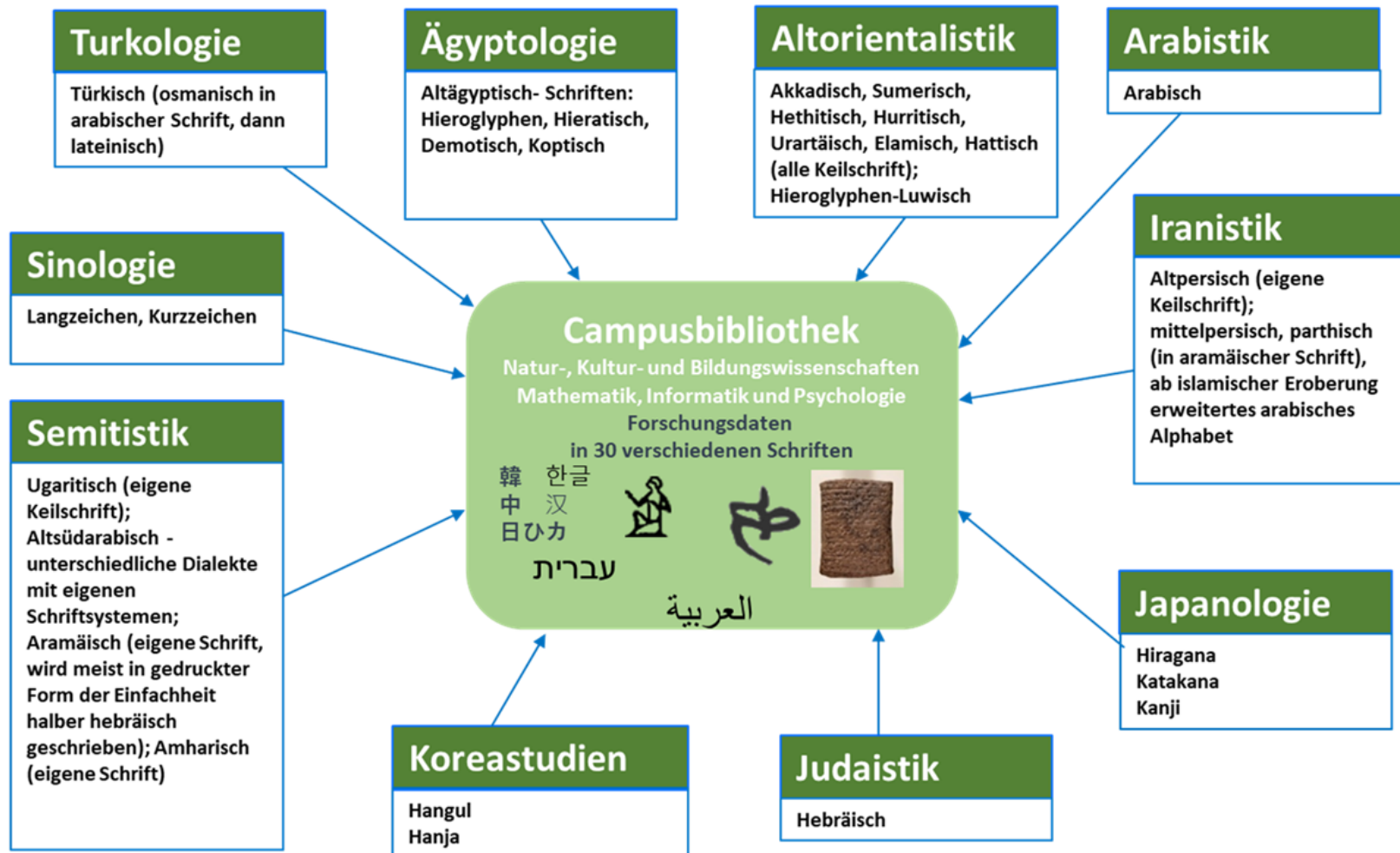
Sichtbarkeit

Vernetzung

Zusammenarbeit

Ausblick

Multilinguale Forschungsdaten: Beispiele aus der Campusbibliothek



BMBF-Forschungsprojekt „Aufbau und Erprobung von Strategien zum Forschungsdatenmanagement“ (FDM_OAS-Orient) 2017-2018

Projektphasen

Bedarfserhebung
Experten-
interviews

Daten-
kurations-
konzepte

Workshop
Vernetzung

Begleitung von 5 Pilotprojekten

- Fächer: Altorientalistik/Archäologie, Japanologie, Koreastudien, Sinologie
- Diverse methodische Ansätze
- Verschiedene Datentypen
- Projektstadien von Antragsstellung bis Abschlussphase

Ziele:

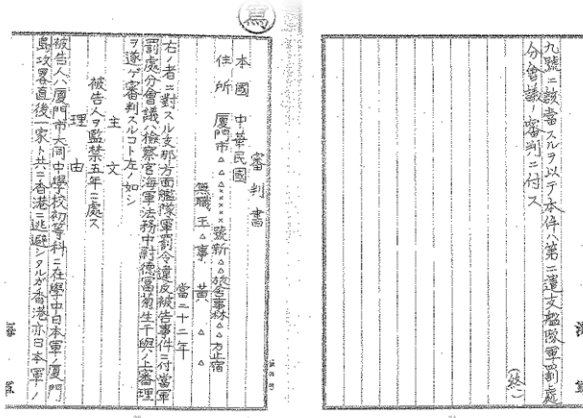
Fachspezifische
Bedarfe identifizieren

Sensibilisierung/
“Raising Awareness”

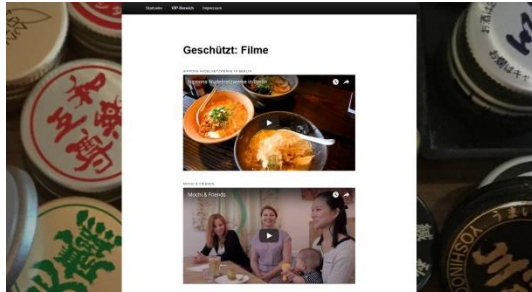
Konzeption von FDM-
Services für Forschung
und Lehre

Datenbeispiele Projekt FDM_OAS_Orient

Quellenmaterial (Prozessakten etc.) aus dem Archiv des japanischen Verteidigungsministeriums zur Rechtsprechung während des Zweiten Japanisch-Chinesischen Krieges 1939-1945 (Japan als Besatzungsmacht in China)



Lehrforschungsprojekt „Forschungswerkstatt Japanologie: Praxisnahe Vermittlung sozialwissenschaftlicher Methoden zum Thema ‚Japanische Küche in Berlin‘“
 Forschungsdaten hier Videos



Beispiel Forschungsthema „Food Safety“ in Japan: pdf des „White Book of Food Education 2015“ 平成27年度 食育白書


第2部 第8章 調査、研究その他の施策の推進

第1節 調査、研究等の実施

1 「日本人の食事摂取基準」の作成・公表、活用促進

国民の健康の維持・増進、生活習慣病の予防を目的として国民が健全な食生活を営むことができるように、「日本人の食事摂取基準」を策定し、5年ごとに改定しています。日本人の食事摂取基準(2015年版)は、平成27(2015)年度から平成31(2019)年度まで使用する予定です。食事摂取基準(2015年版)では、高齢化の進展や糖尿病有病者数の増加等を踏まえ、生活習慣病の発症予防に加えて、重症化予防も視野に入れて策定を行い、エネルギーの指標として目標とするBMI (Body Mass Index)の範囲を定めているほか、炭水化物、たんぱ

Zählscheibe (7383.004.003)


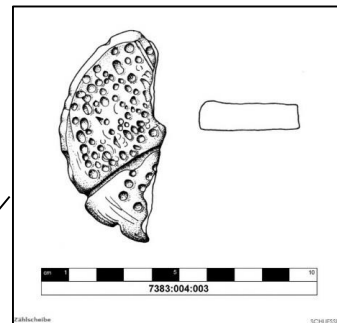


3 von 135

Fundort	Oymaağaç (Oym S 91)
Fundstadium	16.08.2007
Anlage	20.08.2007 18.06
Anmeldung	13.09.2017 17:07
Kennzahl	2005.004 (Oymaağaç)
Objekt	036
Höhe	201,47 bis 201,52
Feststoff	Metallfund
Material	Zinn
Fundstadium	Verarbeitung
Datierung	Spätbronzezeit

Rundliche Zählscheibe von (sekundär?) gelatertem, hellbraunem Ton, etwa 40% erhalten, der Rand weist in unregelmäßigen Abständen von ca. 1,3-1,5 cm vertikale Einbrüche auf, die handgemachte Scheibe ist unterschiedlich dick. Während die Unterseite glatt ist, weist die Oberseite 4 Quadranten auf, die durch Ritzlinien voneinander getrennt sind (vgl. Munbaga), die Quadranten tragen zahlreiche runde Einbrüche von 0,3 cm Durchmesser, die sich stellenweise stark überschneiden. Man gewinnt das Gefühl, als hätte man innerhalb eines rechtlichen kurzen Zeitraumes eine Vielzahl von Fingerdrücken anbringen sollen (Zählung einer großen Zahl von Objekten?). Durch die Facettiertheit mit Munbaga sind die Zählscheiben mit großer Wahrscheinlichkeit in die Spätbronzezeit (20. u. 19. Jh. v. Chr. zu datieren. Durchmesser: 7,5 cm (unvollständig), D: 1,1 cm.

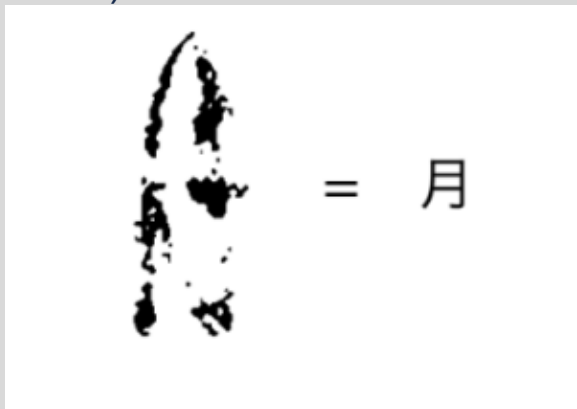
Mit Wasser und Zahnbürste gereinigt. (Forster)

Ausgrabungsfund aus dem Projekt Oymaağaç-Nerik: Fotos, Skizzen, Geo- und Messdaten
 Quelle:
<http://www.nerik.de/ausgrabung/funde/fund.php?fundid=346>

Spezielle Herausforderungen Texterkennung

Beispiel schlechte Scanqualität von historischen Texten (hier von einem Mikrofilm):



Aus den Folien Shoji YAMADA von [Historical Character Recognition: An Overview](#); Vortrag bei [Digital Approaches: Japanese Studies \(25./26.05.\)](#)

Beispiel verschiedene Schriften und Zeichen, die sich berühren:

公園の片隅に通りがかりの人を相手に演説をしている者がある。向うから来た釜形の尖った帽子を被ずいて古ぼけた外套を猫背に着た爺さんがその傍みを佇めて演説者を見る。演説者はひたりと演説をやめてかつかとこの村夫子のたたずぬる前に出て来る。二人の視線がひたと行き

Fig.2 Single- and multiple-touching characters underlined by single lines and double lines, respectively.

Nguyen, Kha Cong; Nguyen, Cuong Tuan; NAKAGAWA, Masaki (2017): A Segmentation Method of Single- and Multiple-Touching Characters in Offline Handwritten Japanese Text Recognition. In: IEICE Trans. Inf. & Syst. E100.D (12), S. 2962–2972. DOI: 10.1587/transinf.2017EDP7225, S. 2962 und 2963, Fig. 2

Weitere Herausforderungen: Schriftrichtungen, verschiedene Schriften in einem Dokument, Variationen in Zeichengrößen, unterschiedliche Zeichenabstände, Segmentierung und Tokenisierung

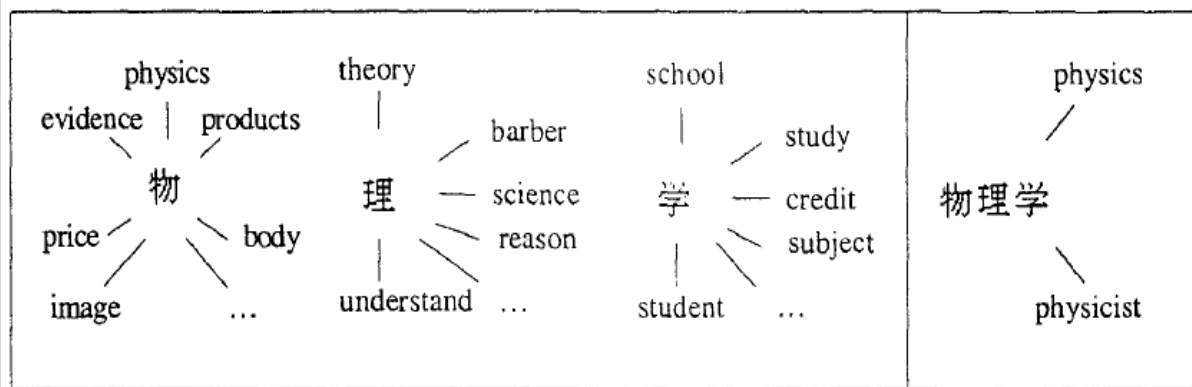
Technische Herausforderungen für NLS-Daten in multilingualen Discovery Systemen

Segmentierung von Wörtern

“It's as if the text always looks like this but the software expects it to look like this.”

Tokenisierung und Index-Bildung

A sentence in Chinese	我喜欢新西兰花	
Interpretation 1	我 喜欢 新西兰 花	= Ich mag neuseeländische Blumen
Interpretation 2	我 喜欢 新 西兰花	= Ich mag frischen Brokkoli



Beispiele aus: Teahan, W. J., McNab, R., Wen, Y. 6 Witten, I. H. (2000) A compression-based algorithm for chinese word segmentation. Comput.Linguist., 26(3), 375-393, hier S.376.

Herausforderung des Nebeneinanders von Transkriptionskonventionen bei CJK-Schriften

Transkriptionsvarianten 中国经济 (chinesische Wirtschaft):

- **"Zhong guo jing ji"** (Hanyu pinyin in Einzelsilben)
- **"Zhongguo jing ji"** (Hanyu pinyin, mehrsilbige Namen zusammengeschrieben)
- **"Zhongguo jingji"** (Hanyu pinyin mit Wortbildung)
- **"Zhongguo-jingji"** (Hanyu pinyin mit Wortbildung, Wörter mit Bindestrichen durchgekoppelt)
- **"Chung-kuo ching-chi"** (Wade- Giles-Umschrift)

Sichtbarkeit

disziplinspezifische Hilfen

- Umgang mit Daten
- Recherche
- Werkzeuge

CAMPUSBIBLIOTHEK NATUR-, KULTUR- UND BILDUNGSWISSENSCHAFTEN, MATHEMATIK, INFORM

ANSPRECHPARTNER/INNEN BEREICHE BIBLIOTHEK FAQ KONTAKT LITERA

Startseite > Sites > Campusbibliothek der Freien Universität Berlin > Bibliothek > Forschungsdatenmanagement > I

Digitale Tools und Datenbanken für die Sinologie

Diese erste Sammlung entstand im Rahmen eines Rechercheprojekts von Lea Schneider innerhalb von FDM_OAS-Orient. Die Sammlung befindet sich im Aufbau und wird fortlaufend ergänzt.

(Stand Juni 2018)

Themenübergreifend

DIRT Digital Research Tools

- umfassende Übersicht über so gut wie alle aktuell verfügbaren Tools für digitale Forschung
- sortiert nach Aufgabenbereichen wie Schreiben, Daten analysieren und visualisieren, Annotieren, Publizieren, Kollaborieren, ...
- sehr viele frei zugängliche Open-Source-Programme
- interessante hier aufgeführte Tools sind zum Beispiel:
 - [Editors' Notes](#): ein webbasiertes Tool zum Aufzeichnen, Organisieren, Teilen und Archivieren von Research Notes, besonders nützlich für Editionsprojekte oder die Arbeit mit Archivmaterial
 - [Palladio](#): ein webbasiertes Tool zur Visualisierung komplexer historischer Daten entlang von Raum- und Zeitachsen oder in Netzwerken
 - [Overview](#): ein Tool zur Analyse, Auswertung und Visualisierung sehr großer Dokumentenmengen (Text Mining), sehr nützlich für Inhalts- und Diskursanalysen
 - [ImagePlot](#): ein Tool zur Analyse, Auswertung und Visualisierung von Bilder- und Videosammlungen
 - [Neatline](#): ein intuitives Karten-Tool zum Mapping von Daten in Zeit und Raum u.v.m.

CrossAsia

- die zentrale Plattform für Fachinformation in den Asienwissenschaften
- über CrossAsia kann zentral in fast allen nationalen und internationalen Bibliothekskatalogen sowie in einer Vielzahl von Datenbanken gesucht werden
→ ein einziger Sucheinstieg für nahezu alle Quellen!
- Print-Materialien und digitale Materialien können gleichermaßen recherchiert werden
- Zugriff auf digitalisierte historische Materialien, z.B. historische Bestände der Staatsbibliothek Berlin
- mit dem Modul CrossAsia Campus kann ein eigener virtueller Arbeitsraum eingerichtet werden, in dem auch kollaborativ mit Wissenschaftler*innen an deren Institutionen zusammengearbeitet werden und eigene Dokumente hochgeladen und bearbeitet werden können
- über 200 [Datenbanken](#) zu China und Taiwan bieten Informationen und Materialien aus den Bereichen Politik, Gesellschaft, Kultur, Literatur, Philosophie, Religion, etc., darunter z.B.
 - [全国人民代表大会资料库](#): Database of the National People's Congress
 - [中国政府资料库](#): Database of the Chinese Government
 - [中国年鉴网络出版总库](#): China Yearbooks Full-text Database (CNKI)
 - [漢學文庫](#): Chinese ancient texts
 - [古今圖書集成](#): The complete classics collection of ancient China

<https://www.fu-berlin.de/sites/campusbib/bibliothek/Forschungsdatenmanagement/oasorient.html>

Sichtbarkeit

Sichtbarkeit von Projekten, Kompetenzen und eingesetzten Technologien

ÜBERSICHT PROJEKTE LEHRE ARBEITSGEMEINSCHAFT KONTAKT

Startseite > Fachbereich > Digital Humanities > Projekte

DH-Projekte am Fachbereich Geschichts- und Kulturwissenschaften

Auf dieser Seite finden Sie eine Liste von Projekten aus dem Bereich *Digital Humanities*, die mindestens einen Projektpartner am Fachbereich Geschichts- und Kulturwissenschaften aufweisen. Diese Liste wird fortlaufend aktualisiert. Ergänzungen zu laufenden oder bereits abgeschlossenen Projekten sind jederzeit herzlich willkommen.

Hier können Sie die aktuelle Liste der DH-Projekte am Fachbereich GeschKult herunterladen.

Um ein Forschungsprojekt im Bereich *Digital Humanities* zu melden, nutzen Sie bitte unser [Online-Formular](#).

SPOTLIGHT! Neue DH-Projekte in der Projektliste



https://www.geschkult.fu-berlin.de/fachbereich/digital_humanities/Projekte_v2/index.html

(Auf der Seite befindet sich eine herunterladbare Tabelle mit allen Angaben)

Vernetzung

- AG Digital Humanities am Fachbereich Geschichts- und Kulturwissenschaften (Austausch, Webseite, Positionspapier,...)
- Workshop-Bericht „Nicht-lateinische Schriften in multilingualen Umgebungen: Forschungsdaten und Digital Humanities in den Regionalstudien“: <https://dhd-blog.org/?p=10669>
- NLS-Mailingliste zum Austausch von Erfahrungen mit Software und selbst entwickeltem Code oder fachspezifische Kurationsleitfäden: <https://lists.fu-berlin.de/listinfo/nicht-lateinische-schriften>


Workshop

Nicht-lateinische Schriften in multilingualen Umgebungen mit dem Fokus auf Forschungsdaten und Digital Humanities in den Regionalstudien: ein Beitrag zur NFDI

3. Juli 2018

Campusbibliothek, Freie Universität Berlin



Campus 
bibliothek

GEFÖRDERT VON
 Bundesministerium
für Bildung
und Forschung

Zusammenarbeit

Digitale Kuratierung deutsch-koreanischer Archivmaterialien zur Wiedervereinigung: Semantische Aufbereitung und visueller Zugriff




Semantische Analyse KT_FUB

- Named Entity Recognition und Entity Linking deutsch und koreanisch
- Semantische Suchen über Ontologien
- Clustern von Dokumenten
- Automatische Extraktion von Schlüsselbegriffen (“2+4-Gespräche“)
- Topic Detection

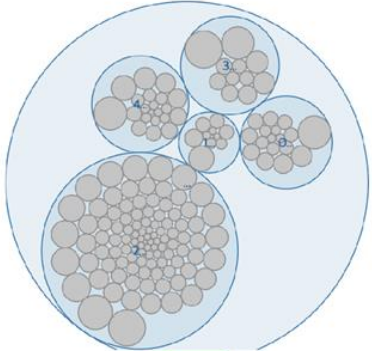
Curation Dashboard KT_FUB

🏠
🔌
⏴
📖
🔍
📄
📄
⚙️



Dashboard

CLUSTERS



[View more!](#)

HIGHLIGHT

...

1990년 2월 9일 Bundesminister für Arbeit und Sozialordnung (BMA), Norbert Blum; Bundeskanzler Helmut Kohl


출처

Bundesarchiv, BArch, B 136 /21660

2 사회현장을 위한 기본원칙과 입장 - 동독 원탁회의의 토론 1990년 3월 5일 Zentraler Runder Tisch der DDR;
 Unabhängiger Frauenverband (UFV);
 Ministerium für Arbeit und Löhne; Initiative Frieden und Menschenrechte (IFM);
 Gewerkschaften; Ministerium für Gesundheitswesen (MfG); Sektion Rechtswissenschaft der Humboldt-Universität; Institut für Soziologie und Sozialpolitik der Akademie der Wissenschaften

[View more!](#)

ENTITIES



[View more!](#)

© 2018 DFKI. All Rights Reserved | Design by w3layouts.

Zusammenarbeit „embedded librarianship“

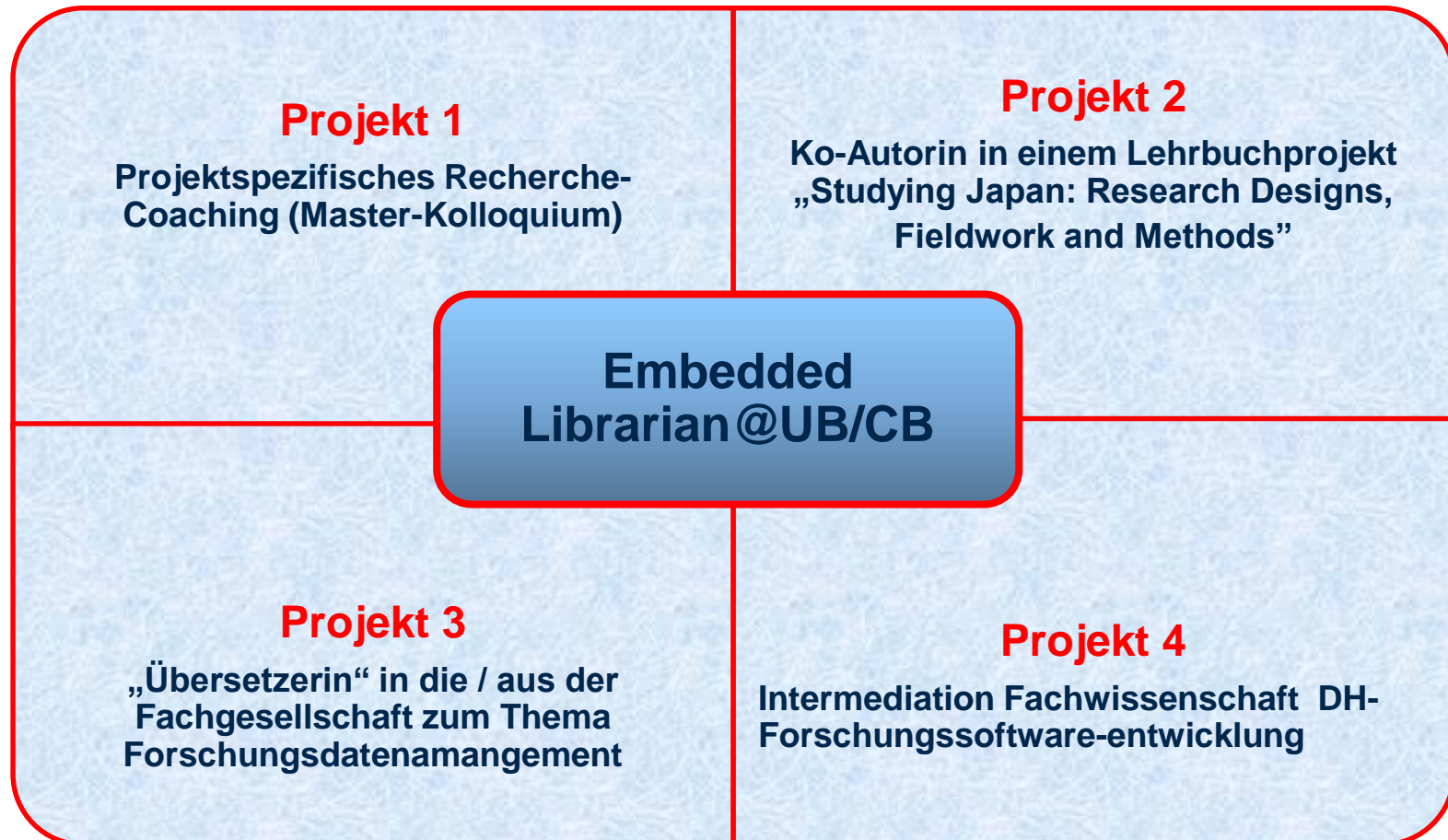
ATTRIBUTES OF EMBEDDED LIBRARIANS

TRADITIONAL	EMBEDDED
Responsive	Anticipatory
Individual customers	Team of collaborators
Standardised	Customised
Single transaction	Ongoing projects
Service	Partnership

Differences between Traditional & Embedded Librarianship (Schumaker, 2009)

Quelle: [Shahril Effendi](https://www.slideshare.net/ShahrilEffendi/embedded-librarian-in-higher-education) <https://www.slideshare.net/ShahrilEffendi/embedded-librarian-in-higher-education>

Pilotprojekte embedded librarian



Ausblick

- **Workshop DH Conference Utrecht 2019:** Towards multilingualism in Digital Humanities: Achievements, failures and good practices in DH Projects with non-latin scripts (Bibliothek, Forschung, Rechenzentren)
- **Themenheft NLS** in Open Access Zeitschrift 027.7 Zeitschrift für Bibliothekskultur, multidisziplinär, gemeinsam von Bibliotheken und Forschung
- **AI4NFDI:** Bibliothek, Forschung (DH & Informatik), Rechenzentren

Ende

спасибо 谢谢
GRACIAS 谢谢
THANK YOU
ありがとうございました MERCI
DANKE धन्यवाद
شُكراً **OBRIGADO**

Basierend auf dem BMBF-Projekt: Aufbau und Erprobung von Strategien zum Forschungsdatenmanagement mit dem Schwerpunkt von Forschungsdaten in nicht-westlichen Sprachen, insbesondere aus dem Ostasiatischen Raum und dem Vorderen Orient („FDM_OAS-Orient“)



Mitwirkende:

Asef, Esther: esther.asef@bbaw.de

Gräff, Andreas Dr.: A.Graeff@smb.spk-berlin.de

Lee, Martin: m.lee@fu-berlin.de

Wagner, Cosima Dr.: cosima.wagner@fu-berlin.de