

# Nicht-lateinische Schriften in multilingualen Umgebungen

## Handlungsmöglichkeiten für Bibliotheken am Beispiel der Campusbibliothek der FU Berlin

Esther Asef, *Campusbibliothek der FU Berlin*; Martin Lee, *Campusbibliothek der FU Berlin*

### Zusammenfassung

Im Einklang mit dem Diskurs über die digitale Transformation der akademischen Forschung und Lehre steigt der Bedarf an stärkerer Zusammenarbeit. Insbesondere bei drittmittelfinanzierten Forschungsprojekten können spezifische Kenntnisse über Digital Humanities und nicht-lateinische Schriften selten innerhalb der Organisation vermittelt werden und Projektwerkzeuge und -plattformen können oft nicht länger als die Dauer des Projekts gepflegt werden. Eine Lösung könnte in der Einbeziehung von Informationsinfrastruktureinrichtungen wie Bibliotheken oder Rechenzentren sein. Diese Institutionen sind jedoch in der Regel nicht in der Lage, alle Sprachen und Disziplinen zu unterstützen. Der Artikel zeigt praxisbezogene Beispiele für institutionsübergreifende Kollaboration und die Ansätze eines Expertennetzwerkes zu nicht-lateinischen Schriften in multilingualen Umgebungen auf.

### Summary

In line with the discourse on the digital transformation of academic research and teaching, the need for stronger cooperation is increasing. Especially in third-party funded research projects, specific knowledge of digital humanities and non-Latin scripts can rarely be shared within the organisation, and project tools and platforms can often not be maintained longer than for the duration of the project. One solution could be to involve information infrastructure facilities such as libraries or data centres. However, these institutions usually are not able to support all languages and disciplines. The article shows practical examples of cross-institutional collaboration and the approach of an expert network on non-Latin scripts in multilingual environments.

**Zitierfähiger Link (DOI):** <https://doi.org/10.5282/o-bib/2019H4S136-150>

**Autorenidentifikation:** Asef, Esther: ORCID: 0000-0003-2411-4953;

**Lee, Martin:** ORCID: 0000-0001-7550-7365

**Schlagwörter:** Forschungsdaten; Digital Humanities; Datenkuration; CJK; Regionalwissenschaften; Ostasienwissenschaften; Forschungsdatenmanagement; Research Data; Data Curation; Area Studies; East Asian Studies; Research Data Management

## 1. Ausgangssituation und Kontext

Die Freie Universität Berlin zeichnet sich durch ein breites Fächerspektrum mit einer besonderen Vielfalt auch an „kleinen“ Fächern in den Geisteswissenschaften wie archäologische Wissenschaften, Orientalwissenschaften und Ostasienwissenschaften aus. Die Campusbibliothek ist aus 24 heterogenen Bibliotheken aus fünf Fachbereichen entstanden.<sup>1</sup> Sie betreut die Fächergruppe Altertumswis-

1 Für mehr Informationen vgl. Lee, Martin; Hilliger, Kirsten; Gräff, Andreas: 24 in 1 – Der Umzug der Campusbibliothek der Freien Universität Berlin, in: ABI Technik 35 (2), 01.01.2015. Online: <<https://doi.org/10.1515/abitech-2015-0020>>.

senschaften (Cluster Altertum: Ägyptologie, Altorientalistik, Klassische Archäologie, Prähistorische Archäologie, Vorderasiatische Archäologie), Fächergruppe Orientwissenschaften (Cluster Orient: Arabistik, Iranistik, Islamwissenschaft, Turkologie, Semitistik), Fächergruppe Ostasienwissenschaften (Cluster Ostasien: Japanologie, Koreastudien, Sinologie) und Fächergruppe Religionen (Cluster Religionen: Judaistik, Religionswissenschaft, katholische Theologie) und bietet daher große Chancen für multidisziplinäre Synergieeffekte.

Die Campusbibliothek versteht sich als Dienstleister für die Forschung. Als wissenschaftliche Bibliothek stellt sie das „zentrale Labor“ für Wissenschaftlerinnen und Wissenschaftler (in den Geisteswissenschaften) dar. Neben der Zurverfügungstellung der Literaturliste werden dabei auch Services im Bereich Digitalisierung, Digitale Werkzeuge und Methoden sowie Management von Forschungsdaten immer wichtiger.

Um diesen neuen Anforderungen aus der Wissenschaft, insbesondere in Bezug auf den an anderen Stellen bisher wenig berücksichtigten Aspekt „nicht-lateinische Schriften“ (NLS) optimal begegnen zu können, wurde im Forschungsprojekt FDM\_OAS-Orient an der Campusbibliothek in enger Zusammenarbeit mit Forschungsprojekten Fragestellungen zu Forschungsdaten in nicht-lateinischen Schriften untersucht.<sup>2</sup> Dabei wurden Bedarfe gesammelt, Herausforderungen herausgearbeitet und erste Lösungsansätze skizziert.

Sprachen in nicht-lateinischer Schrift, insbesondere aus dem ostasiatischen Raum, haben bei bisherigen Entwicklungen zum Forschungsdatenmanagement in Deutschland kaum Berücksichtigung gefunden und die Aktivitäten im Bereich Forschungsdatenmanagement in ostasiatischen Ländern selbst sind noch nicht sehr ausgeprägt.<sup>3</sup>

Das Projekt FDM\_OAS-Orient wurde zwischen 2017 und 2018 für 18 Monate vom BMBF gefördert und war in drei Projektabschnitte eingeteilt: Bedarfserhebung und modellartige Untersuchung von Forschungsprozessen im Zusammenhang mit NLS-Materialien, eine Konzeptionsphase für erste Services und in den letzten Monaten des Projekts stand der Austausch mit den Fachcommunities sowie die Vernetzung durch einen Workshop im Fokus. Als Grundlage diente die Zusammenarbeit mit fünf Pilotprojekten aus den Fächern Altorientalistik/Archäologie, Japanologie, Koreastudien und

- 2 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, Förderkennzeichen:16FDM022; Laufzeit 2017-2018; Förderrichtlinie zur Erforschung des Managements von Forschungsdaten in ihrem Lebenszyklus an Hochschulen und außeruniversitären Forschungseinrichtungen. Bundesanzeiger vom 19.08.2016, Bundesministerium für Bildung und Forschung - BMBF, <<https://www.bmbf.de/foerderung/bekanntmachung-1233.html>>, Stand: 28.06.2019. Projektwebseite FDM\_OAS-Orient, 05.06.2018, <<https://www.fu-berlin.de/sites/campusbib/bibliothek/Forschungsdatenmanagement/16fdm022.html>>, Stand: 28.06.2019.
- 3 Vgl. Kim, Jihyun: Data Sharing from the Perspective of Faculty in Korea, in: Libri 67 (3), 28.01.2017. Online: <<https://doi.org/10.1515/libri-2016-0116>>; Liu, Xia; Ding, Ning: Research data management in universities of central China, in: The Electronic Library 34 (5), 2016, S. 808-822. Online: <<https://doi.org/10.1108/EL-04-2015-0063>>; Chu, Wenjing; Chu, Zhaohui; Pang, Hongsheng u. a.: Research on the Standardized Process of Research Data Management in CHINA, in: Proceedings of the 2017 2nd International Seminar on Education Innovation and Economic Management (SEIEM 2017), Penang, Malaysia 2018. Online: <<https://doi.org/10.2991/seiem-17.2018.60>>; Cooper, Danielle; Daniel, Katherine; Albuero, Jade u. a.: Supporting the Changing Research Practices of Asian Studies Scholars, Ithaca S+R, 21.06.2018. Online: <<https://doi.org/10.18665/sr.307642>>.

Sinologie. Die Projekte befanden sich in verschiedenen Phasen des Forschungsprozesses, sodass die Bedarfe von Antragstellung bis Abschlussphase abgefragt und begleitet werden konnten. Mit diesen Pilotprojekten wurden zu Beginn teilstrukturierte Experteninterviews durchgeführt und in der folgenden Zeit wurden die Projekte in allen Fragen im Bereich Forschungsdaten (FD) und Digital Humanities (DH) begleitet. So konnten nicht nur fachspezifische Bedarfe herausgearbeitet werden, sondern auch erste Services/Beratungsangebote auf Basis tatsächlicher Anfragen entworfen werden. Der Fokus der Untersuchung lag zwar immer auf multilingualen FD in nicht-lateinischen Schriften, dennoch waren viele der erhobenen Bedarfe grundlegender Natur, da viele Forschende mit dem Thema Forschungsdaten bisher kaum Berührung hatten und es noch zu wenige (Beratungs-) Angebote in diesem Bereich gab und gibt.

## 2. Spezielle Herausforderung NLS

In der Auseinandersetzung mit der Thematik NLS und durch den Austausch mit Forschenden aus den verschiedenen Disziplinen wurden im Projekt FDM\_OAS-Orient Bedarfe von Forschungsprojekten mit NLS-Daten zusammengetragen. Viele Anforderungen beschäftigen auch andere Disziplinen, wie die Klärung von rechtlichen Fragen, Langzeitarchivierung und Nachhaltigkeit. Diese Punkte sollten dennoch im fachinternen Diskurs der verschiedenen Regionalwissenschaften behandelt werden, denn zum einen gibt es spezielle Unterfragen, die sich in anderen Disziplinen bisher nicht stellen (z.B. die rechtliche Lage für Daten, die im Ausland gesammelt wurden und wie diese Rechte mit deutschen bzw. europäischen Urheber- und Persönlichkeitsrechten vereinbar sind). Zum anderen können die Regionalwissenschaften ihre Perspektive in die allgemeine Diskussion einbringen, da sie Erfahrungen mit vielen disziplinübergreifenden Methoden (und somit auch Daten) vorweisen können.

Neben diesen allgemeinen Ansprüchen an Forschungsdaten stehen verschiedene Problemstellungen, die spezifisch im Zusammenhang mit Daten in nicht-lateinischen Schriften stehen und bisher selten im interdisziplinären Diskurs berücksichtigt werden. Im Folgenden werden exemplarisch typische Daten und drei technische Problemstellungen behandelt, die sich im Umgang mit Daten in nicht-lateinischen Schriften während des Forschungsprozesses oder bei der Speicherung, Langzeitverfügbarkeit und der Auffindbarkeit zeigen.

### 2.1. Beispiel-Datensets aus dem Projekt

In Regionalstudien kommt eine Vielfalt an verschiedenen Methoden (z.B. historische, ethnologische, philologische, politikwissenschaftliche, sozialwissenschaftliche etc.) zum Einsatz, was sich in den untersuchten Materialien und erhobenen Daten widerspiegelt. In den fünf Projekten kamen neben Textmaterialien Bild-, Audio- und Videodateien vor sowie aufbereitete Daten in Form von Tabellen, 3D-Modellen und komplexen Datenbanken. (Digitalisierte) Archivmaterialien und (ebenfalls digitalisierte) graue Literatur waren in den meisten Projekten zumindest teilweise Gegenstand der Untersuchung. Die Projektmitarbeiter/innen digitalisierten meistens selbst und der fehlende Zugang zu geeigneter Hard- und Software sowie fehlende Absprachen und Koordination des Digitalisierungsprozesses führte zu erheblichen Unterschieden in der Datenqualität, Benennungen und Anordnung (Ordnerstrukturen). Des Weiteren wurden häufig nicht konsistent (standardisierte) Metadaten

vergeben und es kam keine Software wie z.B. Layoutanalyse, Optical Character Recognition (OCR) und Named Entity Recognition (NER) zum Einsatz.

Hier zeigt sich ein wichtiger Ansatzpunkt für Bibliotheken und Archive, die meist langjährige Expertise zur Digitalisierung von Printmaterialien aufgebaut haben, zumeist über geeignete Hard- und Software verfügen und umfangreiche Metadatenschemata zur Beschreibung solcher Ressourcen entwickelt und etabliert haben. Sofern es keine rechtlichen Einschränkung gibt, wäre ein Digitalisierungsservice für die Geisteswissenschaften außerdem eine Möglichkeit diese „Roh“-Daten zusammen mit den dann erhobenen strukturierten Metadaten gleich (oder nach einer Embargozeit) zu veröffentlichen, anstatt dass sie über einen Zeitraum von Jahren ungeordnet und ohne Metadaten auf Projektfestplatten und Netzlaufwerken liegen bleiben würden und nicht oder nur mit viel Aufwand veröffentlicht werden könnten.

Da ein solch umfassender Digitalisierungsservice jedoch mit viel Aufwand verbunden ist, wäre es zunächst ein möglicher Schritt, eine Beratung zum Digitalisierungsprozess anzubieten, bei der Expertise zu Qualitätsstandards, Metadaten und geeigneter Software weitergegeben wird.<sup>4</sup> Insbesondere bei Materialien mit nicht-lateinischen-Schriften sollte an dieser Stelle bereits geprüft werden, ob und welche Bearbeitungs-Software für die jeweiligen Schriftsysteme geeignet sind und in welcher Sprache bzw. Schrift die Metadaten erfasst werden sollen. Außerdem sollte überprüft werden, ob besondere rechtliche Einschränkungen bestehen, wenn die Materialien aus dem Ausland stammen.

## 2.2. Texterkennung

Bei der Arbeit mit Daten in NLS zeigt sich, dass Software, Informationssysteme und Infrastrukturen häufig nicht oder nur bedingt für die Verwendung von NLS ausgelegt sind.<sup>5</sup> Auch viele semantische Kuratierungstechnologien, die z.B. auf Mustererkennung, Deep Learning, Natural Language Processing (NLP), OCR und Handwritten Text Recognition (HTR) basieren, sind in vielen Bereichen für NLS noch nicht bzw. nicht gleich weit entwickelt wie für lateinische Schriften. Zahlreiche digitale Werkzeuge und Methoden können daher bisher nicht für Forschung mit nicht-lateinischen Schriften verwendet werden.<sup>6</sup>

4 z.B. Deutsche Forschungsgemeinschaft: DFG Praxisregeln „Digitalisierung“, DFG-Vordruck 12.151 – 12/1, 2013, S. 81. Online: <[https://www.dfg.de/formulare/12\\_151/12\\_151\\_de.pdf](https://www.dfg.de/formulare/12_151/12_151_de.pdf)>, Stand: 28.06.2019.

5 Call for papers, DH Asia conference 2018, Stanford University: "(...) when we look at DH in Western Europe and the Americas, we find a vibrant intellectual environment in which even college and university undergraduates – let alone more advanced researchers – can download off-the-shelf analytical platforms and data corpora, and venture into new and cutting-edge research questions; while, in the context of Asian Studies, we find an environment in which many of the most basic elements of DH research remain underdeveloped or non-existent". Sophie, Hélène et: Call for proposals Digital humanities Asia : Harnessing Digital Technologies to Advance the Study of the Non-Western World, 26-29 April 2018, Stanford University, 2017, Online: <<https://carnetcase.hypotheses.org/3165>>, Stand: 28.06.2019.

6 Für die Keilschrift vgl. Rothacker, Leonard; Fisseler, Denis; Müller, Gerfrid G. W. u. a.: Retrieving Cuneiform Structures in a Segmentation-free Word Spotting Framework, in: Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing, New York, NY, USA 2015 (HIP '15), S. 129–136. Online: <<https://doi.org/10.1145/2809544.2809562>>. : "Hitherto these pictures are in the best case linked to some descriptive data or a transliteration and translation of the text and may be consulted with it for collation. The digital availability of the documents opens up new paths for computer assisted analysis, especially in the field of pattern recognition. Yet, promising solutions for optical character recognition (OCR) on cuneiform manuscripts are still unavailable."

Nicht-lateinische Schriften zeigen spezielle Charakteristika, die spezielle Herausforderungen für OCR-/HTR-Verfahren darstellen. Die lateinische Schrift ist eine waagerechte rechtsläufige Schrift, die in der Regel von links nach rechts geschrieben wird. Auf diesen „Standard“ ist die meiste Software ausgerichtet. Semitische Schriften wie Hebräisch und Arabisch werden primär von rechts nach links geschrieben, aber sind ebenfalls waagrecht ausgerichtet.<sup>7</sup> CJK-Sprachen (Chinesisch, Japanisch, Koreanisch) werden traditionell in Spalten von oben nach unten geschrieben. Heutzutage werden viele Texte jedoch, beeinflusst von der westlichen waagerechten Schreibweise, horizontal in Zeilen geschrieben. Dennoch existiert die traditionelle Schriftrichtung nebenher und es werden auch heute noch Publikationen mit vertikaler Schreibrichtung gedruckt. So kommen bei CJK-Textsammlungen nicht selten Dokumente mit unterschiedlichen Schreibrichtungen vor, die dann nicht automatisch von der OCR-Software verarbeitet werden können. Solche Textsammlungen müssen z.B. vorsortiert und für die unterschiedlichen Variationen getrennte Trainingsmaterialien vorbereitet werden.

Ein wichtiger Prozess bei der Texterkennung ist die Segmentierung, bei der i. d. R. in einem mehrstufigen Verfahren das Layout, die Textzeilen (bzw. -spalten) und insbesondere einzelne Worte eingeteilt werden. Im Japanischen wie auch im Chinesischen werden Wortgrenzen nicht mit einem Leerzeichen markiert, sondern alle Zeichen sind mit einem gleichen Abstand aneinandergereiht.<sup>8</sup> Die Texterkennungssoftware kann hier nicht wie bei der lateinischen Schrift die Wortzwischenräume als Indikator für den Beginn und das Ende eines Wortes verwenden (White-Space-Tokenisierung). Für die Segmentierung in semantische Texteinheiten müssen für diese Schriften alternative Tokenisierungsverfahren verwendet werden.

Dies sind nur zwei der vielen speziellen technischen Herausforderungen bei der Texterkennung für verschiedene NLS-Sprachen.<sup>9</sup> Für einige Schriften wird zwar fortlaufend an einer Verbesserung der Verfahren gearbeitet, die Verbreitung und Umsetzung der gewonnenen Erkenntnisse bleiben jedoch ein wichtiges Anliegen, um Wissenschaften mit NLS-Daten und -Quellen bei der Quellenerschließung zu unterstützen.<sup>10</sup>

### 2.3. Metadaten und Retrieval für NLS-Daten in multilingualen Discovery-Systemen

Die meisten Informationsinfrastrukturen sind auf westliche Sprachen ausgelegt. Im Projekt wurden verschiedenen Repositorien im deutschsprachigen Raum<sup>11</sup>, entweder von Institutionen mit Regionalwissenschaften oder solche, die von Forschenden mit NLS-Daten genutzt werden, stichprobenartig

7 Beispielhaft genannt seien hier die Erfahrungen des Projekts Europeana mit der OCR-Erkennung osmanischer Zeitungen, die im für das osmanische Türkisch adaptierten arabischen Alphabet Trefferquoten von 20 % hatten, vgl. Neudecker, C.; Antonacopoulos, A.: Making Europe's Historical Newspapers Searchable, in 2016. Online: [https://www.primaresearch.org/www/assets/papers/DAS2016\\_Neudecker\\_HistoricalNewspapers.pdf](https://www.primaresearch.org/www/assets/papers/DAS2016_Neudecker_HistoricalNewspapers.pdf), Stand: 28.06.2019.

8 Bericht der National Diet Library (NDL) zum Stand OCR <<http://www.ndl.go.jp/jp/aboutus/digitization/fulltextreport.html>> (japanisch) 4/2010: Von November 2010 bis Januar 2011 testete die NDL OCR-Erkennungssoftware mit ihrem Bestand aus der Meiji- (1868–1912), Taishō- (1912–1926) und Shōwa-Zeit (1926–1989), insgesamt 20.000 Bücher mit ca. 5 Millionen Einzelseiten. Die durchschnittliche Erkennungsrate betrug 70 %, je neuer die Texte desto besser.

9 Weitere Herausforderungen z.B. für Japanisch vgl. Das, Soumendu; Banerjee, Sreeparna: Survey of Pattern Recognition Approaches in Japanese Character Recognition, in: International Journal of Computer Science and Information Technologies (IJCSIT) 5 (1), 2014, S. 93–99.

10 Vgl. z.B. Dai, Ruwei; Liu, Chenglin; Xiao, Baihua: Chinese character recognition: history, status and prospects, in: Frontiers of Computer Science in China 1 (2), 2007, S. 126–136. Online: <<https://doi.org/10.1007/s11704-007-0012-5>>.

11 Unter den geprüften Repositorien waren Zenodo, DARIAH, IANUS u.a. Die Liste kann bei den Autoren angefragt werden.

geprüft sowie Anfragen über den Umgang mit NLS-Daten gestellt. Bei keinem Repository wurden Hinweise oder Empfehlungen gefunden wie speziell mit (Meta-) Daten in nicht-lateinischen Schriften umgegangen werden soll. Metadaten können zwar meistens in der gewünschten Sprache eingegeben und angezeigt werden, die implementierten Suchalgorithmen können diese Daten jedoch nicht zufriedenstellend durchsuchen.

Tests im institutionellen FD-Repository der FU<sup>12</sup> und auch die Rücksprache mit anderen Repositorien und Forschenden zeigen Beispiele der mangelnden multilingualen Anpassung von Suchalgorithmen in gängigen Suchmaschinen/Discovery-Systemen:

- Suchalgorithmen erkennen nicht, in welcher Sprache/Schrift gesucht wird.
- Für viele Schriften sind keine Stoppwort- oder Homonym-/Synonym-Listen implementiert.
- Es findet kein Mapping zwischen verschiedenen Zeichen (z.B. chinesischen Kurz- und Langzeichen oder Schriftzeichen in japanischer und chinesischer Schreibung), Zeichen und Umschrift (z.B. Kurzzeichen/Pinyin im Chinesischen), verschiedenen Umschriften (z.B. Hepburn-Umschrift/Kunrei-Umschrift für die Transkription des Japanischen) und verschiedenen Sprachen (Englisch/Chinesisch) statt. Das bedeutet, dass nur ein geringer Anteil der vorhandenen Literatur gefunden wird bzw. dass der/die Nutzer/in in fünf bis sechs unterschiedlichen Schreibweisen nach einem Titel suchen muss.
- In CJK-Schriften gibt es üblicherweise keine Leerzeichen, daher müssen die Texte mit computerlinguistischen Verfahren in Wörter unterteilt werden. Zum jetzigen Zeitpunkt werden Suchstrings und Metadatenfelder im Repository der FU als ein Wort interpretiert. Beispiel: Steht im Titelfeld „清末中国における日本観と西洋観“ („Chinese views of Japan and the West in the late Ch'ing period“) bleibt eine Suche nach den letzten drei Zeichen „西洋観“ („views of the West“, ein wichtiger Fachbegriff) ohne Treffer, weil die Technologie nicht erkennen kann, dass diese Zeichen einen Teil des Titels ausmachen.
- Treffermengen werden nicht logisch sortiert.

Diese technischen Voraussetzungen vermindern die Auffindbarkeit und damit Sichtbarkeit von NLS-Forschungsdaten, auch wenn die Daten noch so gut kuratiert und beschrieben sind. Solche Rahmenbedingungen schaffen für die Wissenschaft wenig Anreize, regionalwissenschaftliche FD aufzubereiten und zu publizieren. Aufgrund fehlender Kenntnis der technischen Herausforderungen von NLS werden bei viele Repositorien und Infrastrukturen keine Lösungen für solche Daten entwickelt. Dadurch ist es nicht-lateinischen Daten selten möglich, dem Anspruch „Findable“ der FAIR-Prinzipien gerecht zu werden.<sup>13</sup>

Die Nachvollziehbarkeit von Forschungsergebnissen sowie die nationale und internationale Nutzbarkeit von Forschungsdaten ließen sich durch Berücksichtigung der NLS-Sprachen bei der Entwicklung und Standards für multilinguale Metadaten und Beschreibungen wesentlich verbessern.

---

12 Refubium, basiert auf den Technologien DSpace und Apache Solr. Online: <<https://refubium.fu-berlin.de/>>, Stand: 28.06.2019.

13 FORCE11: The FAIR Data Principles, <<https://www.force11.org/group/fairgroup/fairprinciples>>, Stand: 28.06.2019.

## 2.4. Herausforderung des Nebeneinanders von Transkriptionskonventionen bei den CJK-Schriften

Ein Ansatz, die vielen technischen Herausforderungen von nicht-lateinischen Schriften in Informationssystemen zu umgehen und eine bessere ‚Lesbarkeit‘ für Nutzerinnen und Nutzer mit keinen NLS-Sprachkenntnissen zu ermöglichen, ist die Verwendung von Transkriptionen für Metadaten. Durch die Transkription werden die ursprünglichen Zeichen auf Basis der Aussprache/Phonetik in ein anderes Schriftsystem übertragen, z.B. japanische Hiragana in lateinische Buchstaben. Transkriptionen sind zeitaufwändig und benötigen fundierte Fach- und Sprachkenntnisse. Wenn ausschließlich Metadaten in Umschrift in einem System verwendet werden, wird darüber hinaus die große Gruppe von Muttersprachlerinnen und Muttersprachler der jeweiligen Sprache ausgeschlossen, denn die meisten Transkriptionssysteme werden in den betroffenen Regionen selbst nicht verwendet. Verwendung von Transkriptionen oder Transliterationen sind dennoch in vielen Informationssystemen wie z.B. in Bibliothekskatalogen Praxis.<sup>14</sup>

Mit der Transkription fällt zwar die in 2.3 erläuterte komplexe Segmentierung und Tokenisierung weg, jedoch tritt an deren Stelle eine andere Herausforderung: Für viele Sprachen gibt es etablierte standardisierte Transkriptionsregeln wie beispielsweise Hanyu Pinyin für Chinesisch oder die Hepburn-Umschrift (DIN 32708:2014-08) für Japanisch. Für andere Sprachen gibt es jedoch (noch) keine Standards bzw. verschiedene Transkriptionsregeln werden parallel verwendet. So wird in deutschsprachigen Bibliotheken zumeist die ROK 2000-Umschrift für Koreanisch angewandt<sup>15</sup>, während in der Wissenschaft und in nordamerikanischen Bibliotheken der Standard McCune-Reischauer etabliert ist.<sup>16</sup>

Folgendes Beispiel mit Transkriptionsvarianten für 中国经济 (chinesische Wirtschaft) macht die Problematik deutlich:

- „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)

Damit Nutzerinnen und Nutzer zuverlässig alle Inhalte in einem Informationssystem finden können, müsste also dokumentiert werden, welche Transkriptionsregeln verwendet wurden. Dadurch könnten dann dieselben Regeln für die Suchanfrage verwendet werden. Eine andere Lösung wäre, die verschiedenen Schreibweisen im Index oder bei Retrievalanfragen zu mappen.

---

14 Vgl. Panel „Herausforderungen der CJK-Erschließung“ auf dem 107. Bibliothekartag in Berlin 2018.; Vgl. auch „Geschäftgangsregel originalschriftliche Katalogisierung in der ZDB“. Online: <[https://www.zeitschriftendatenbank.de/fileadmin/user\\_upload/ZDB/pdf/arbeitshilfen/OSK\\_Geschaefstgang\\_Version2-2014.pdf](https://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/pdf/arbeitshilfen/OSK_Geschaefstgang_Version2-2014.pdf)>, Stand: 28.06.2019.

15 Praxisregeln zur CJK-Erfassung (Aktualisierte Ausgabe 2017). Online: <<https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Standardisierung/AGV/agVerbundPraxisregelnCjk2017.pdf>>, Stand: 28.06.2019.

16 „The Library of Congress will continue to follow the McCune-Reischauer system (...)“, <<https://www.loc.gov/catdir/cpso/romanization/korean.pdf>>, Stand: 28.06.2019.

### 3. Sichtbarkeit schaffen

Ein übergeordnetes Ziel des Projekts und der Bearbeitung der Thematik „nicht-lateinische Schriften“ war die Sensibilisierung der Forschenden in Disziplinen, die mit nicht-lateinischen Materialien in Berührung kommen sowie die Sensibilisierung von Informationsinfrastrukturen für die besonderen Herausforderungen, die im Zusammenhang mit Daten in nicht-lateinischen Schriften in Informationssystemen bestehen.

Zu Beginn des Projekts im Frühjahr 2017 stand die Bearbeitung der Thematik Forschungsdaten in den meisten Regionalwissenschaften selbst noch ganz am Anfang. Zu diesem Zeitpunkt gab es bereits von verschiedenen Seiten der Wissenschaftspolitik die Aufforderung, fachspezifische Perspektiven in die Entwicklung neuer Services und Infrastrukturen einzubinden, die später in den Entschluss einer von Fachdisziplinen getriebenen Nationalen Forschungsdateninfrastruktur (NFDI) mündete.<sup>17</sup>

Während des Projekts wurde auf verschiedene Weisen auf mehr Sichtbarkeit für die Thematik NLS generell und DH sowie FDM in den Regionalwissenschaften hingearbeitet. An den folgenden zwei Beispielen soll gezeigt werden, wie lokal an der eigenen Institution Sensibilisierung stattfindet und verschiedene Akteurinnen und Akteure in den Prozess eingebunden werden können.

#### 3.1. Fachspezifische Informationsangebote

Eine Möglichkeit, schnell Sichtbarkeit herzustellen, ist die gezielte Erstellung von fachspezifischen Informationsmaterialien. Es gibt bereits zahlreiche Informationsangebote zu allgemeinen Fragen des Forschungsdatenmanagements<sup>18</sup>, aber häufig fehlt es an fachspezifischen und insbesondere auf die Angebote einzelner Institute und Universitäten zugeschnittenen Informationsmaterialien. Dabei könnte die institutionelle Bibliothek die erste Anlaufstelle für viele Forschende bei Fragen zum Datenmanagement sein, so wurde es jedenfalls in den im Projekt geführten Experteninterviews zurückgespiegelt.<sup>19</sup>

Im Projekt gesammelte Informationen wurden daher z.B. niedrigschwellig auf der Website der Bibliothek für alle frei zur Verfügung gestellt.<sup>20</sup> Darunter eine „Frequently Asked Questions“-Seite, die an Fragen der begleiteten Pilotprojekte an das FDM\_OAS-Orient-Projekt orientiert ist. Dadurch werden

17 Vgl. RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, in, 2016, S. 160. Online: <<http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>> ; HRK: Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien. Empfehlung der 19. Mitgliederversammlung der HRK am 10. November 2015 in Kiel, S. 6. Online: <[https://www.hrk.de/fileadmin/\\_migrated/content\\_uploads/Empfehlung\\_Forschungsdatenmanagement\\_final\\_Stand\\_11.11.2015.pdf](https://www.hrk.de/fileadmin/_migrated/content_uploads/Empfehlung_Forschungsdatenmanagement_final_Stand_11.11.2015.pdf)>, Stand: 28.06.2019.

18 Beispielhaft ist die Informationswebsite „forschungsdaten.info“, <<https://www.forschungsdaten.info/>>, Stand: 28.06.2019.

19 Aufbau von lokalen Informationsangeboten ist ein wichtiger Schritt im strategischen Ausbau von FDM vgl. Helbig, Kerstin; Dierkes, Jens; Neumann, Janna: Aufbau und Bekanntmachung von Informationsangeboten über Forschungsdatenmanagement für Forschende, 25.10.2018. Online: <<https://doi.org/10.17192/bfdm.2018.1.7821>> ; Hartmann, Niklas K.; Jacob, Boris; Weiß, Nadin: RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement, 25.01.2019. Online: <<https://doi.org/10.5281/zenodo.2549344>> ; Rans, Jonathan; Whyte, Angus: Using RISE the Research Infrastructure Self-Evaluation Framework v.1.1, Digital Curation Centre, Edinburgh 2017, S. 19. Online: <[www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)>, Stand: 28.06.2019.

20 Einstiegsseite Forschungsdatenmanagement, <<https://www.fu-berlin.de/sites/campusbib/bibliothek/Forschungsdatenmanagement/index.html>>, Stand: 28.06.2019.



dort viele allgemeine Fragen behandelt („Was sind Forschungsdaten“), aber auch sehr spezifische Themen abgedeckt („Sind Transkriptionsregeln für das FDM relevant?“). Eine FAQ-Seite ist eine effiziente Art, Informationen weiterzugeben und mit relativ wenig Recherche- und Erstellungsaufwand für die Informationseinrichtung verbunden, sofern man die Fragen bereits regelmäßig in Beratungssituationen beantwortet.

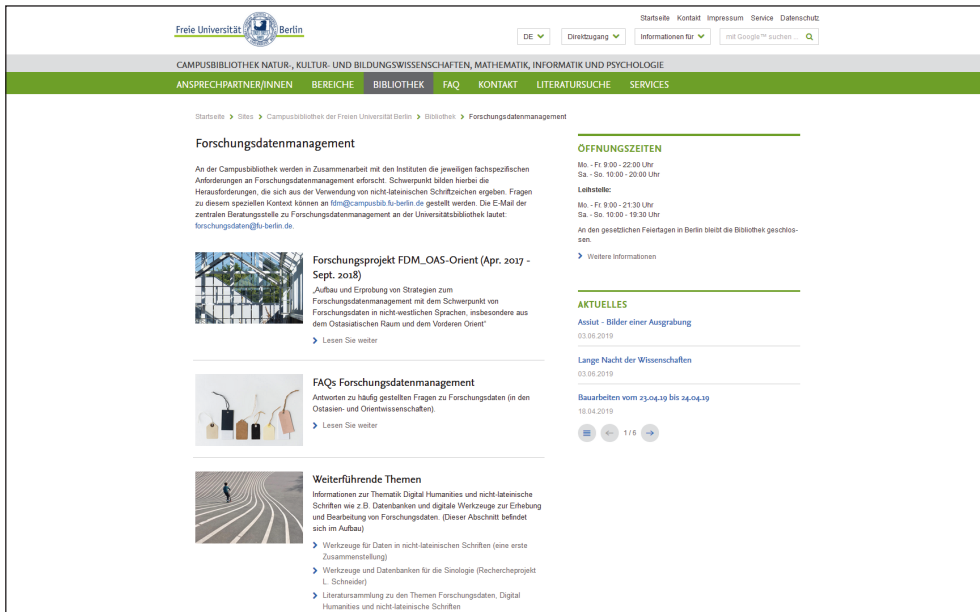


Abb 1.: Einstiegsseite zu den erarbeiteten Informationsmaterialien an der Campusbibliothek, Screenshot <<https://www.fu-berlin.de/sites/campusbib/bibliothek/Forschungsdatenmanagement/index.htm>>, Stand: 28.06.2019.

Außerdem wurde im Projekt eine Liste mit digitalen Werkzeugen kuratiert, mit denen nicht-lateinische Daten bearbeitet werden können.<sup>21</sup> Mit dem Informationsmaterial „Digitale Tools und Datenbanken für die Sinologie“ wurden fachspezifische Recherche-Hinweise sowie eine Liste mit passenden Werkzeugen zur Arbeit mit chinesischen Materialien veröffentlicht.<sup>22</sup>

Für alle diese Informationsangebote sollte eine regelmäßige Pflege der Daten eingeplant werden, da sich insbesondere das Angebot an digitalen Werkzeugen und Datenbanken stetig verändert bzw. erweitert wird. Auch wenn eine regelmäßige Überarbeitung nicht möglich ist, sollte dies nicht von einer Publikation des vorhandenen Wissenstands abhalten, denn diese Information kann dennoch ein Einstieg für Forschende darstellen, von dem aus sie weiter recherchieren können. Die Aktualität solcher Listen sollte durch den letzten Bearbeitungsstand angegeben werden.

21 DH-Werkzeuge für nicht-lateinische Daten, <<https://www.fu-berlin.de/sites/campusbib/bibliothek/Forschungsdatenmanagement/tools-os/index.html>>, Stand: 28.06.2019.

22 Digitale Tools und Datenbanken für die Sinologie, <<https://www.fu-berlin.de/sites/campusbib/bibliothek/Forschungsdatenmanagement/oasorient.html>>, Stand: 28.06.2019.

Die Erstellung fachspezifischer Informationsangebote könnte z.B. von Fachreferentinnen und Fachreferenten oder von entsprechenden „Embedded Librarians“ (s.u.) übernommen werden.

### 3.2. Vernetzung

Auch über die Initiative der Campusbibliothek entstand über Gespräche mit dem Dekan des Fachbereichs Geschichts- und Kulturwissenschaften eine AG Digital Humanities, in welche alle (digitalen) Forschungsprojekte des Fachbereichs eingeladen wurden. Dadurch wurde ein Raum geöffnet, um vorhandene Kompetenzen, Erfahrungen und Ressourcen im Fachbereich zu bündeln und auszutauschen. Die AG bietet Gelegenheit, andere digitale Projekte kennenzulernen und den gegenseitigen Austausch und somit eine Vernetzung zu fördern. So werden mögliche Synergie-Effekte zwischen den Projekten erkennbar und der Ausbau der Digital Humanities in Forschung und Lehre im Fachbereich gemeinsam vorangebracht.

Eine der ersten Aufgaben der AG war die Erstellung eines FU-internen Positionspapiers, um den Status Quo des Fachbereichs zum Thema DH festzustellen. Neben den Forderungen nach einer besseren, koordinierten Infrastruktur an der Universität wurde das Fehlen einer Übersicht der DH-Projekte des Fachbereichs mit Ansprechpartner/inne/n und verwendeter Software moniert. Deshalb wurde im Rahmen der AG innerhalb von wenigen Wochen eine Website mit einer solchen Liste erstellt. Dort werden die verschiedenen Digital-Humanities-Projekte zusammen mit technischen Angaben verzeichnet.<sup>23</sup>



Abb 2: Beispiel für Sichtbarkeit von Digital Humanities Projekten, Screenshot <[https://www.geschkult.fu-berlin.de/fachbereich/digital\\_humanities/Projekte\\_v2/index.html](https://www.geschkult.fu-berlin.de/fachbereich/digital_humanities/Projekte_v2/index.html)>, Stand: 28.06.2019.

Um den interdisziplinären Austausch und die gemeinsame Bearbeitung der Thematik NLS zu fördern, wurde im Juli 2018 an der Campusbibliothek ein Workshop mit dem Titel „Nicht-lateinische Schriften in multilingualen Umgebungen: Forschungsdaten und Digital Humanities in den Regionalstudien“ ausgerichtet. Dabei kamen erstmalig Forschende, IT-Expertinnen und -Experten sowie Bibliothekarinnen und Bibliothekare aus verschiedenen Regional- und Altertumswissenschaften<sup>24</sup> zusammen,

23 Liste mit DH-Projekten am Fachbereich Geschichts- und Kulturwissenschaften, vgl. Digital Humanities am Fachbereich Geschichts- und Kulturwissenschaften, <[https://www.geschkult.fu-berlin.de/fachbereich/digital\\_humanities/index.html](https://www.geschkult.fu-berlin.de/fachbereich/digital_humanities/index.html)>, Stand: 28.06.2019.

24 Es waren Expert/inn/en für 28 nicht-lateinischen Schriften vertreten.

um gemeinsam Herausforderungen im Zusammenhang mit Forschungsdaten zu identifizieren, Erfahrungen zu teilen und mögliche Lösungsansätze zu skizzieren.<sup>25</sup>

Da sich viele Synergien zwischen den verschiedenen Projekten zeigten und ein weiterer Wissensaustausch zu Spezifika des Managements von Daten in nicht-lateinischen Schriften von allen Beteiligten gewünscht wurde, gründete sich ein Expertennetzwerk zu nicht-lateinischen Schriften. Das Netzwerk besteht auch über das Projektende von FDM\_OAS-Projekt hinaus und fördert die Zusammenarbeit verschiedener Forschenden und Personen aus Informationsinfrastrukturen.<sup>26</sup>

### 3.3. Zusammenarbeit

#### 3.3.1. Beispiel Projekt DFKI - Campusbibliothek - Institut für Koreastudien

Gemeinsam mit dem Institut für Koreastudien und dem deutschen Forschungsinstitut für Künstliche Intelligenz (DFKI) wurde eine digitale Plattform erarbeitet, welche eine Visualisierung der eingespeisten Dokumente sowie die semantische Analyse und Anreicherung ermöglicht. Die Datengrundlage bildet das Forschungsprojekt „Sharing German Government’s Documents on Unification and Integration, and Building a data-base on German Unification“<sup>27</sup>, in dem offizielle deutsche Regierungsdokumente zur deutschen Wiedervereinigung gesammelt und intellektuell kuratiert, analysiert und interpretiert wurden.

Im gemeinsamen Projekt „Digitale Kuratierung deutsch-koreanischer Archivmaterialien zur Wiedervereinigung: Semantische Aufbereitung und visueller Zugriff“ wurden die vorliegenden PDF-Dokumente über Apache Tesseract 4.0 (für Deutsch) bzw. 3.5 (für Koreanisch)<sup>28</sup> in reinen maschinenlesbaren Text verwandelt (OCR). Die Ground Truth Daten wurden mit Transkribus<sup>29</sup> und die Auswertungen mit ocrevalUation<sup>30</sup> erstellt. In einem nächsten Schritt wurde OpenNLP<sup>31</sup> implementiert, um Personen, Orte und Organisationen automatisch zu erkennen (NER). Nach dem Erkennen der Entitäten wurde über DBPedia SPARQL Endpoint<sup>32</sup> eine automatische Verlinkung mit einem einheitlichen Bezeichner für Ressourcen (URI) vorgenommen. Die URIs wurden über WEKA geclustert.<sup>33</sup> Diese einzelnen Elemente werden in einem Dashboard und über einen Curation Workflow Manager online visualisiert.<sup>34</sup>

25 Ausführlicher Workshopbericht: Asef, Esther; Gräff, Andreas; Lee, Martin u. a.: Workshop-Bericht „Nicht-lateinische Schriften in multilingualen Umgebungen: Forschungsdaten und Digital Humanities in den Regionalstudien“, DHdBlog Digital Humanities im deutschsprachigen Raum, 24.10.2018, <<https://dhd-blog.org/?p=10669>>, Stand: 28.06.2019.

26 Mailingliste: <<https://lists.fu-berlin.de/listinfo/nicht-lateinische-schriften>>.

27 Projektwebseite: <<https://www.geschkult.fu-berlin.de/en/e/oas/korea-studien/forschung/projektTongilbu/index.html>>, Stand: 28.06.2019.

28 Tesseract Open Source OCR Engine Wiki. <<https://github.com/tesseract-ocr/tesseract/wiki>>, Stand: 28.06.2019.

29 Transkribus, <<https://transkribus.eu/Transkribus/>>, Stand: 28.06.2019.

30 The ocrevalUation tool, <<https://sites.google.com/site/textdigitisation/ocrevaluation>>, Stand: 28.06.2019.

31 Apache OpenNLP, <<http://opennlp.apache.org>>, Stand: 28.06.2019.

32 DBPedia Sparql <<https://dbpedia.org/sparql>>, Stand: 28.06.2019. Auch die Einbindung von GND ist in Planung. <[https://web.archive.org/web/20190702012604/https://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedata\\_node.html](https://web.archive.org/web/20190702012604/https://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedata_node.html)>, Stand: 28.06.2019.

33 Hall, Mark; Frank, Eibe; Holmes, Geoffrey u. a.: The WEKA data mining software: an update, in: ACM SIGKDD Explorations Newsletter 11 (1), 16.11.2009, S. 10. Online: <<https://doi.org/10.1145/1656274.1656278>>.

34 Mehr dazu in: Rehm, Georg; Lee, Martin; Moreno Schneider, Julian u.a.: Curation Technologies for Cultural Heritage Archives. Analysing and transforming a heterogeneous data set into an interactive curation workbench, In: 3rd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2019). Digital Access to Textual Cultural Heritage, May 9–10 Brussels Belgium ACM 2019.

Das Projekt ist ein gelungenes Beispiel für eine bedarfsorientierte Zusammenarbeit zwischen Forschenden, Bibliotheken und Expert/innen auf dem Gebiet von maschinellem Lernen bzw. Sprachtechnologien.

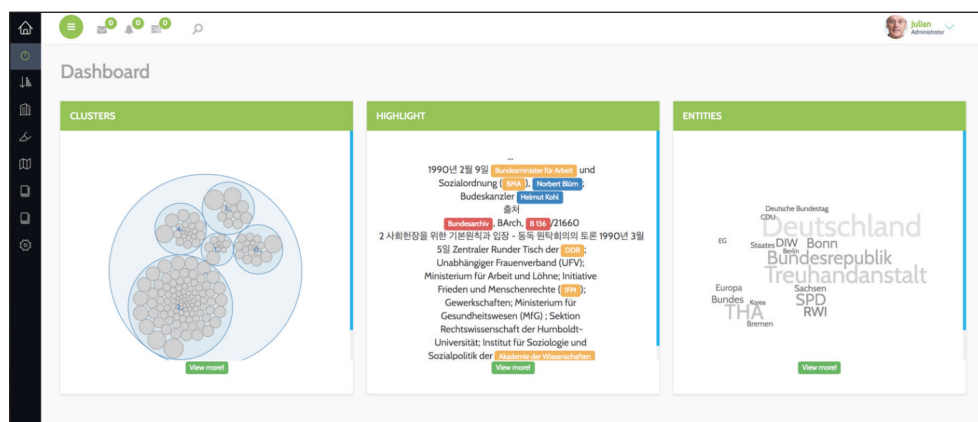


Abb 3.: Das entwickelte Curation-Dashbord des Projekts „Digitale Kuratierung deutsch-koreanischer Archivmaterialien zur Wiedervereinigung: Semantische Aufbereitung und visueller Zugriff“

### 3.3.2. Beispiel Embedded Librarian

In zahlreichen Gesprächen mit Forschenden an der FU Berlin wurde der Wunsch nach einer stärkeren Involvierung der Bibliothek in den Instituten genannt. Dieses Konzept von „Embedded Librarian“ ist nicht neu.<sup>35</sup> Auf Basis der theoretischen Grundlagen und des konkreten Bedarfs wurde an der FU Berlin eine der Referendarstellen als „Embedded Librarian“ konzipiert, um für den Bereich Ostasien als besonderer Kontakt zur Verfügung zu stehen und die Zusammenarbeit mit diesen Instituten auszubauen. Konkret wurden dann mehrere Projekte realisiert. Zwei Beispiele: (1) Für ein Master-Kolloquium wurde eine maßgeschneiderte Recherche-Beratung in einer „realen“ und virtuellen Kursumgebung durchgeführt. Grundlage war die Beobachtung der Dozentin, dass sich die Quellengrundlage in Hausarbeiten seit Einführung eines Discovery-Systems als Recherche-Tool der Bibliothek stetig verschlechterte, da dieses nicht für den Umgang mit Japanischer Originalschrift optimiert war. So hielt die Bibliothekarin zu Kursbeginn einen Vortrag über die Besonderheiten der Recherche in Originalschrift und erläuterte das Recherche-Coaching Konzept, in welchem sie als Co-Dozentin auf der E-Learning Plattform für individuelle Coaching-Sessions zur Verfügung stand.

(2) In einem zweiten Beispiel war die Bibliothekarin „embedded“ in ein Lehrbuchprojekt zu „Studying Japan: Research Designs, Fieldwork and Methods“<sup>36</sup>. Da Regionalstudienfächer wie die Japanologie keine eigene „Methode“ haben, muss diese aus dem Nebenfach, in Seminaren oder im Selbststudium projektspezifisch erarbeitet werden. Studierende haben Probleme, für ihre Hausarbeiten/

35 Shumaker, David; Talley, Mary; Miervaldis, Wendy: Models of Embedded Librarianship: Final Report, 2009. Online: <[http://www.talleypartners.com/wp-content/uploads/2013/10/Models-of-Embedded-Librarianship\\_FinalReport-Rev-copy.pdf](http://www.talleypartners.com/wp-content/uploads/2013/10/Models-of-Embedded-Librarianship_FinalReport-Rev-copy.pdf)>, Stand: 28.06.2019.

36 Fokus: Qualitative Sozialforschung. In: Reiher, Cornelia and Kottmann, Nora, (eds.), Studying Japan: Research Designs, Fieldwork and Methods. Baden Baden: Nomos Verlag. (Forthcoming)

Abschlussarbeiten „Methoden“ zu identifizieren und anzuwenden. Hier wurden die Kapitel zu Informationsrecherche und Forschungsdaten speziell für die Japanologie verfasst. Eine enhanced-publication-Website ist auf der Homepage der Bibliothek in Vorbereitung.

## 4. Ausblick

Die Resonanz auf die Aktivitäten des Projekts und das Feedback aus den verschiedenen Communities zeigt, dass die Themen nicht-lateinische Schriften und Multilingualität eine größere Bedeutung erhalten müssen. Dafür sind eine breitere Diskussion und Vernetzung notwendig. Ein weiterer Schritt in diese Richtung war der Workshop „Towards Multilingualism In Digital Humanities: Achievements, Failures And Good Practices In DH Projects With Non-latin Scripts“<sup>37</sup> im Rahmen der DH2019 in Utrecht.<sup>38</sup> Der eintägige Workshop folgte dem Aufruf zur Mehrsprachigkeit und Multikulturalität in den Digitalen Geisteswissenschaften und diskutierte Erfolge, Misserfolge und bewährte Praktiken in DH-Projekten mit nicht-lateinischen Schriften. Im Rahmen des Workshops wurden praktische Einblicke in „Do's and Dont's“ im NLS-Kontext gegeben und mögliche übertragbare Praktiken auf andere Sprachen und Disziplinen in den Sitzungen identifiziert. Dieser Workshop basierte auf den Erfahrungen des Workshops „NLS in multilingual (Software-)Umgebungen“, der 2018 an der Freien Universität Berlin stattgefunden hatte (s.o.). Dadurch konnte das internationale Netzwerk von Forschenden sowie Expertinnen und Experten im Bereich NLS weiter gestärkt werden, die – unabhängig von ihrer Arbeitszugehörigkeit (z.B. in Universitäten, Bibliotheken und Museen) – spezifisches NLS-Wissen entwickeln, pflegen und verbreiten. Ein nachnutzbares Beispiel solch zusammengetragenen Wissens ist eine Übersicht von multilingualen NLP-Werkzeugen auf Github.<sup>39</sup>

Darüber hinaus ist ein NFDI<sup>40</sup>-Konsortium namens NFDI4Language in Vorbereitung, welches auf Basis der Fächer Altertumswissenschaften, Nicht-europäische Kulturen und Jüdische Studien in Verbindung mit Informatik sowie Bibliotheken agiert.<sup>41</sup>

Ziel all der beschriebenen Aktivitäten ist es, eine größere „Community of Practice“ im Bereich der NLS und Multilingualität zu erlangen, um die Herausforderungen wie einheitliche Standards in digitalen Umgebungen und die Berücksichtigung spezifischer Bedarfe gemeinsam besser zu meistern.

---

37 Informationen und Begleitmaterialien zum Workshop <<https://hackmd.io/s/ry0yFF1oE>>, Stand: 28.06.2019.

38 DH2019 – Digital Humanities conference 2019 <<https://web.archive.org/web/20190607194758/https://dh2019.adho.org/>>, Stand: 28.06.2019.

39 Multilingual NLP: <<https://github.com/multilingual-dh/nlp-resources>>, Stand 28.06.2019.

40 DFG, Nationale Forschungsdateninfrastruktur, <<https://www.dfg.de/foerderung/programme/nfdi>>, Stand 28.06.2019.

41 Lee, Martin, Müller-Birn, Claudia, Rehm, Georg: NFDI4Language, Letter of Intent: <[https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen/2021/2021\\_nfdi\\_4lang.pdf](https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen/2021/2021_nfdi_4lang.pdf)>, Stand 28.06.2019.

## Literaturverzeichnis

- Asef, Esther; Gräff, Andreas; Lee, Martin; Wagner, Cosima. Workshop-Bericht „Nicht-lateinische Schriften in multilingualen Umgebungen: Forschungsdaten und Digital Humanities in den Regionalstudien“, DHdBlog Digital Humanities im deutschsprachigen Raum, 24.10.2018, <<https://dhd-blog.org/?p=10669>>, Stand: 28.06.2019.
- BMBF: Bekanntmachung: Förderrichtlinie zur Erforschung des Managements von Forschungsdaten in ihrem Lebenszyklus an Hochschulen und außeruniversitären Forschungseinrichtungen. Bundesanzeiger vom 19.08.2016, Bundesministerium für Bildung und Forschung - BMBF, <<https://www.bmbf.de/foerderungen/bekanntmachung-1233.html>>, Stand: 28.06.2019.
- Chu, Wenjing; Chu, Zhaohui; Pang, Hongsheng u. a.: Research on the Standardized Process of Research Data Management in CHINA, in: Proceedings of the 2017 2nd International Seminar on Education Innovation and Economic Management (SEIEM 2017), Penang, Malaysia 2018. Online: <<https://doi.org/10.2991/seiem-17.2018.60>>.
- Cooper, Danielle; Daniel, Katherine; Alburo, Jade u. a.: Supporting the Changing Research Practices of Asian Studies Scholars, Ithaka S+R, 21.06.2018. Online: <<https://doi.org/10.18665/sr.307642>>.
- Dai, Ruwei; Liu, Chenglin; Xiao, Baihua: Chinese character recognition: history, status and prospects, in: Frontiers of Computer Science in China 1 (2), 2007, S. 126–136. Online: <<https://doi.org/10.1007/s11704-007-0012-5>>.
- Das, Soumendu; Banerjee, Sreeparna: Survey of Pattern Recognition Approaches in Japanese Character Recognition, in: International Journal of Computer Science and Information Technologies (IJCSIT) 5 (1), 2014, S. 93–99. Online: <<http://www.ijcsit.com/docs/Volume%205/vol5issue01/ijcsit2014050120.pdf>>, Stand: 28.06.2019.
- Deutsche Forschungsgemeinschaft: DFG Praxisregeln „Digitalisierung“, DFG-Vordruck 12.151–12/1, 2013, S. 81. Online: <[https://www.dfg.de/formulare/12\\_151/12\\_151\\_de.pdf](https://www.dfg.de/formulare/12_151/12_151_de.pdf)>, Stand: 28.06.2019.
- Deutsche Nationalbibliothek: Praxisregeln zur CJK-Erfassung (Aktualisierte Ausgabe 2017), 2017. Online: <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Standardisierung/AGV/agVerbundPraxisregelnCjk2017.pdf>, Stand: 28.06.2019.
- Dombrowski, Quinn: Multilingual NLP: <<https://github.com/multilingual-dh/nlp-resources>>, Stand: 28.06.2019.
- FORCE11: The FAIR Data Principles, 03.09.2014, <<https://www.force11.org/group/fairgroup/fairprinciples>>, Stand: 28.06.2019.
- Hall, Mark; Frank, Eibe; Holmes, Geoffrey u. a.: The WEKA data mining software: an update, in: ACM SIGKDD Explorations Newsletter 11 (1), 16.11.2009, S. 10. Online: <<https://doi.org/10.1145/1656274.1656278>>.
- Hartmann, Niklas K.; Jacob, Boris; Weiß, Nadin: RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement, 25.01.2019. Online: <<https://doi.org/10.5281/zenodo.2549344>>.
- Helbig, Kerstin; Dierkes, Jens; Neumann, Janna: Aufbau und Bekanntmachung von Informationsangeboten über Forschungsdatenmanagement für Forschende, 25.10.2018. Online: <<https://doi.org/10.17192/bfdm.2018.1.7821>>.
- Hochschulrektorenkonferenz: Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien, 2015. Online:

- <[https://www.hrk.de/uploads/tx\\_szconvention/Empfehlung\\_Forschungsdatenmanagement\\_\\_final\\_Stand\\_11.11.2015.pdf](https://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement__final_Stand_11.11.2015.pdf)>, Stand: 28.06.2019.
- Institut für Korea-Studien der Freien Universität Berlin: Sharing German Government's Documents on Unification and Integration, and Building a data-base on German Unification. Online: <<https://www.geschkult.fu-berlin.de/en/e/oas/korea-studien/forschung/projektTongilbu/index.html>>, Stand: 28.06.2019.
  - Kim, Jihyun: Data Sharing from the Perspective of Faculty in Korea, in: Libri 67 (3), 28.01.2017. Online: <<https://doi.org/10.1515/libri-2016-0116>>.
  - Lee, Martin; Hilliger, Kirsten; Gräff, Andreas: 24 in 1 – Der Umzug der Campusbibliothek der Freien Universität Berlin, in: ABI Technik 35 (2), 01.01.2015. Online: <<https://doi.org/10.1515/abitech-2015-0020>>.
  - Lee, Martin, Müller-Birn, Claudia, Rehm, Georg: NFDI4Language, Letter of intent: <[https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen/2021/2021\\_nfdi\\_4lang.pdf](https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen/2021/2021_nfdi_4lang.pdf)>, Stand: 28.06.2019.
  - Library of Congress: Korean Romanization and Word Division. Online: <<https://www.loc.gov/catdir/cpso/romanization/korean.pdf>>, Stand: 28.06.2019.
  - Liu, Xia; Ding, Ning: Research data management in universities of central China, in: The Electronic Library 34 (5), 2016, S. 808–822. Online: <<https://doi.org/10.1108/EL-04-2015-0063>>.
  - Neudecker, C.; Antonacopoulos, A.: Making Europe's Historical Newspapers Searchable, 2016. Online: <[https://www.primaresearch.org/www/assets/papers/DAS2016\\_Neudecker\\_HistoricalNewspapers.pdf](https://www.primaresearch.org/www/assets/papers/DAS2016_Neudecker_HistoricalNewspapers.pdf)>, Stand: 28.06.2019.
  - Rans, Jonathan; Whyte, Angus: Using RISE the Research Infrastructure Self-Evaluation Framework v.1.1.1, Digital Curation Centre, Edinburgh 2017. Online: <[www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)>, Stand: 28.06.2019.
  - Rehm, Georg; Lee, Martin; Moreno Schneider, Julian u. a.: Curation Technologies for Cultural Heritage Archives: Analysing and transforming a heterogeneous data set into an interactive curation workbench, in: 3rd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2019). Digital Access to Textual Cultural Heritage (DATeCH-2019), May 9-10, Brussels, Belgium, 2019.
  - Reiher, Cornelia and Kottmann, Nora, (eds.), Studying Japan: Research Designs, Fieldwork and Methods. Baden Baden: Nomos Verlag. (Forthcoming)
  - RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, 2016, S. 160. Online: <<http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>>.
  - Rothacker, Leonard; Fisseler, Denis; Müller, Gerfrid G. W. u. a.: Retrieving Cuneiform Structures in a Segmentation-free Word Spotting Framework, in: Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing, New York, NY, USA 2015 (HIP '15), S. 129–136. Online: <<https://doi.org/10.1145/2809544.2809562>>.
  - Shumaker, David; Talley, Mary; Miervaldis, Wendy: Models of Embedded Librarianship: Final Report, 2009 (Special Libraries Association). Online: <[http://www.talleypartners.com/wp-content/uploads/2013/10/Models-of-Embedded-Librarianship\\_FinalReportRev-copy.pdf](http://www.talleypartners.com/wp-content/uploads/2013/10/Models-of-Embedded-Librarianship_FinalReportRev-copy.pdf)>, Stand: 28.06.2019.
  - Zeitschriftendatenbank: Geschäftsgangsregel originalschriftliche Katalogisierung in der ZDB, 2014. Online: <[https://www.zeitschriftendatenbank.de/fileadmin/user\\_upload/ZDB/pdf/arbeitshilfen/OSK\\_Geschaefstsgang\\_Version2-2014.pdf](https://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/pdf/arbeitshilfen/OSK_Geschaefstsgang_Version2-2014.pdf)>, Stand: 28.06.2019.