

Zwischen Mehrsprachigkeit und Ressourcenlücke: “Kleine Fächer” in den deutschsprachigen DH

Vorschlag 2: Zwischen Mehrsprachigkeit und Ressourcenlücke: Quo Vadis “Kleine Fächer” in den deutschsprachigen DH?

Kurzzusammenfassung (100–150 Wörtern)

Sogenannte ‘Kleine Fächer’, wie auch auf Regionen des Globalen Südens bzw. andere Länder der Welt spezialisierte Forschende in allen Disziplinen, sind im Rahmen ihrer digitalen Forschung immer wieder mit Herausforderungen bei Ressourcen, Tools und Infrastrukturen konfrontiert. Diese beziehen sich z.B. auf die Verfügbarkeit von Daten und Metadaten in verschiedenen Sprachen, insbesondere in nichtlateinischen Schriften, die Verwendbarkeit von Tools wie auch die allgemeine Förderung zum Datenkorpora-Aufbau. Das Panel nimmt sich dieser Themen an und diskutiert anhand von User Stories der Panelist:innen aus der Arabistik, Computerlinguistik, Religionswissenschaft, Romanistik, Sinologie und Ukrainistik sowohl allgemeine Fragen zu Ungleichheiten in Infrastrukturen zu verschiedenen Schriften und Sprachen, als auch zu konkreten Beispielen die Auswirkungen dieser Ressourcenlücken in den DH. Abschließend sollen Zukunftsperspektiven für die Stärkung multilingual ausgerichteter DH in den sogenannten Kleinen Fächern und darüber hinaus formuliert werden.

Stichworte (5)

Mehrsprachigkeit, Kleine Fächer, Infrastruktur, Ressourcenlücke

Abstract (1200-1500 Wörter)

Sogenannte “Kleine Fächer” in den Geistes-, Sozial- und Kulturwissenschaften mit regionalen Schwerpunkten – von der Arabistik bis zur Ukrainistik – sind in besonderem Maße darauf angewiesen, Ressourcen der Digital Humanities für und in verschiedenen Sprachen zu verwenden, sei es als multilinguale Korpora, Metadaten in nichtlateinischen Schriften oder als Regelwerke für die Codierung wie z.B. TEI. Innerhalb der deutschen DHd-Community sind “Kleine Fächer”, aber auch die sogenannten “Regionalstudien”, wie auch sonst in der Wissenschaftsstruktur, noch unterrepräsentiert.

Infolgedessen gibt es einen "ressourcedness gap" (Nicholas, Bhatia 2023) bei der digitalen Präsenz dieser Disziplinen, sowie in Bezug auf Daten, Werkzeuge und Lösungen, die für die Forschung und Lehre im Bereich der DH verwendet / entwickelt werden.

Andererseits wird im Rahmen von Initiativen wie der NFDI in Deutschland von den "Kleinen Fächern" erwartet, dass sie ihre "Bedürfnisse" und "User Stories" einbringen:

"Kleine Fächer stellten aufgrund ihrer spezifischen Gegenstände besondere Herausforderungen an die Digitalisierung. Gleichzeitig profitierten sie von der Digitalisierung in besonderer Weise. Hier wurde insbesondere die Schaffung digitaler Strukturen in der Lehre kleiner Fächer als eine große Chance gesehen. Ihre spezifischen Bedarfe sollten kleine Fächer insgesamt stärker in den Aushandlungsprozess und in politische Diskurse einbringen." (Arbeitsstelle Kleine Fächer 2020)

Das Panel greift diese Herausforderung auf und will bezüglich der CfP-Frage nach der Entwicklung der Digital Humanities im DACH-Raum theoretische, methodische und technische Zukunftsperspektiven für eine mehrsprachige DH diskutieren und Umsetzungsstrategien vorschlagen. Dabei stehen nicht nur die bestehenden Infrastrukturen auf dem Prüfstand für die nachhaltige Einbindung "Kleiner Fächer", es soll weiterhin aus der Perspektive dieser Disziplinen diskutiert werden, welche Bedarfe noch unzureichend oder gar nicht erfüllt sind oder werden können.

Zentral ist weiterhin die Frage, wie "Kleine Fächer" sich vernetzen können, um so gemeinsam als Akteur:in auftreten zu können. Da die betroffenen Disziplinen für sich genommen selten über eine kritische Masse an Personal, Infrastruktur und Geldern verfügen ("ressourcedness gap"), um nachhaltig und über die eigenen Fachgrenzen hinaus Änderungen zu erwirken, ist ein gemeinsames Auftreten auch im Sinne des interdisziplinären Charakters der DH notwendig. Das geschieht im Einzelnen dort, wo fachliche Zusammenarbeit auch schon vor der digitalen Wende stattgefunden hat. Für eine nachhaltige Besserung der Situation aller betroffenen Fächer reicht das aber nicht aus, da Fachinitiativen sich selten über diese Fachzusammenarbeit hinaus vernetzen (siehe Grallert et al 2023).

Abschließend muss die Berücksichtigung von Mehrsprachigkeit in der Forschungsförderung und die Ent- und Weiterentwicklung von neuen Wissenschaftsinfrastrukturen thematisiert werden. Bislang ist es nicht nötig, bei der Entwicklung neuer Systeme oder der Aufarbeitung bestehender Infrastrukturen Mehrsprachigkeit als inhärent jede Fachrichtung einer globalisierten Wissenschaft betreffendes Phänomen mitzudenken.

Leitfragen

Die folgenden Punkte stehen im Zentrum der Input-Referate und sollen in der Diskussion im Panel mit den Teilnehmer:innen, welche die Perspektiven verschiedener Statusgruppen von Masterstudierenden bis Professor:innen repräsentieren, vertieft werden:

- Welche Herausforderungen zeigen sich in konkreten Forschungsprojekten, die auf fehlende Ressourcen, einsprachige Standards zurückzuführen sind?
- Welche Lösungen gibt es dafür?
- Was wäre eine mehrsprachige best practices für Forschungsprojekte in Kleinen Fächern?

- Welche Forderungen gibt es aus den Kleinen Fächern an die DH-Community, um ein Gleichgewicht herzustellen?
- Welche Infrastruktur braucht eine mehrsprachige DH?
- Wie sollte ein "Pflichtenheft" für "multilingually enhanced knowledge infrastructures" (Spence 2021) aussehen, das multilinguale und -scripturale DH stärkt?

Sprachen

In Anerkennung der Vielsprachigkeit der DH im deutschsprachigen Raum und der Inklusion möglichst vieler Kolleg:innen sind die Panelsprachen Deutsch und Englisch.

Statements der Panelist:innen

Der lange Weg zur Mehrsprachigkeit: Humanities Computing / Digital Humanities Forschung, internationale Konferenzen und Fachverbände

Prof. Dr. Elisabeth Burr, Universität Leipzig

Mehrsprachigkeit war nicht von Anfang an ein zentrales Thema im Humanities Computing / in den Digital Humanities. Das lag zum einen sicher am ASCII Code und an fehlenden Ressourcen, zum anderen aber auch an der Entstehung des Humanities Computing / Computing in the Humanities im angelsächsischen und US-amerikanischen Raum und damit verbunden der Konzentration auf das Englische.

In meinem kurzen Beitrag möchte ich einen Blick zurück wagen und anhand von u. a. (veröffentlichten) Konferenzbeiträgen (cf. Weingart et al. 2020), den CLiP Seminars (cf. Fiormonte / Usher 2001, Fiormonte 2003, Nicolás / Moneglia 2005), dem ADHO Standing Committee on Multi-lingualism & Multi-culturalism (cf. Burr 2006) und Global Outlook::Digital Humanities (GO::DH) das „Eindringen“ des Themas „Mehrsprachigkeit“ in die DH nachzeichnen. Dabei werden Erfolge und Misserfolge eine Rolle spielen. Zudem soll nach dem Verhältnis von institutioneller Ebene und Bewusstwerdung der Zentralität dieses Themas sowie nach den Wissenschaftssprachen gefragt werden.

Multilingual Challenges from a Religious Studies perspective

Prof. Dr. Frederik Elwert, Ruhr Universität Bonn

From the perspective of Religious Studies, I see three challenges with regard to multilingual DH. First, in addition to a resource gap regarding non-Latin scripts, we encounter an infrastructural gap in working with the *content* of non-Western material. Controlled vocabularies for art history are heavily biased towards Western and Christian art. Similarly, there are still no good vocabularies for classifying data with regards to religious traditions.

Second, solving the problem of working with multilingual historical material in Western institutions is not necessarily the same as including non-Western (or even non-English) research traditions.

Finally, national initiatives such as the NFDI are an important step, but they also run the risk of provincialisation, which adds to these challenges. For example, the use of GND for cataloging improves interoperability, but mainly within *German* institutions. This is a particular challenge for small disciplines that don't have the critical mass in a national context and rely on international collaboration.

Multilingual requirements for software and infrastructure

Dr. Cristina Vertan, Berlin Brandenburgische Akademie der Wissenschaften

My contribution will address the challenges of digital textual scholarship in “small research fields” with a particular focus on the investigation of historical languages. This is a field lacking dramatically of resources. Due to the specificity of such languages general digital solutions are problematic and involve in many cases a flat processing. Many researchers in area studies DH lack awareness about the restrictions already developed tools impose. These tools are often advertised as “easy to be adapted”, “work for any language”. However, they were only tested for English or the “rich resourced languages” incl. Chinese and Modern Arabic. As a consequence, research projects relying on “not yet multilingually adapted” tools often have to drop or adjust their research goals.

The situation can be regarded similar to the 2000s when many EU Languages were very low resourced. Maybe the actions which were taken for these languages could be followed also for strengthening software and infrastructures for “Kleine Fächer” DH.

Finally, my contribution will also tackle the impact on “Kleine Fächer” by new developments in AI with methods relying especially on large amounts of training data.

Keine ‘Under-resourced languages’ und ‘Kleine Fächer’: Perspektiven der Arabistik und Islamwissenschaft

Natalie Kraneiß, M.A. Arabistik, Islamwissenschaft, Universität Münster

Arabistik und Islamwissenschaft stehen als “Kleine Fächer” vor spezifischen inhaltlichen, multidisziplinären und sprachlichen Herausforderungen. Während ein institutionalisierter Austausch mit anderen “Kleinen Fächern” unerlässlich ist, können fachspezifische Probleme nicht allein durch Zusammenschlüsse und Konsortien gelöst werden. Stattdessen müssen sie innerhalb des Fachs in Zusammenarbeit mit verwandten Disziplinen wie Turkologie, Iranistik und Semitistik diskutiert werden. Derzeit fehlt im DACH-Raum ein Raum für diese wichtigen Diskussionen.

Der Zugang zu den Digital Humanities gestaltet sich gerade in den „Kleinen Fächern“ als schwierig, da er zu großen Teilen von persönlichen Kontakten und Engagement abhängt (z. B. in der Lehre oder durch kollegiale Unterstützung). Die Verwendung von Labels wie „Kleine Fächer“ oder „under-resourced languages“ erschwert zusätzlich die Sichtbarkeit und Auffindbarkeit von Ressourcen zu digitalen Methoden in den entsprechenden Zielsprachen, insbesondere für Anfänger, was eine Zusammenstellung entsprechender Ressourcen nötig

macht. Viele in den „großen Fächern“ erprobte Tools und Methoden sind nämlich nicht ohne Probleme auf fachspezifisches Material anzuwenden, u.a. wegen mangelnder Unterstützung der Kernsprachen. Daher bedarf es institutionalisierter, universitätsübergreifender und fachgebundener Kooperationen, die allen Interessierten den Zugang ermöglichen.

KI-basierte Transliterationsmodelle: Herausforderungen und Zukunftsperspektiven am Beispiel des Ukrainischen und weiterer Sprachen

Dr. Aleksej Tikhonov, Universität Zürich, Slavistik

Die Ukraine ist ein mehr- und vielsprachiger Kulturraum *par excellence*. Im Rahmen des MultiHTR-Projekts (Multilingual Handwritten Text Recognition) haben wir ein auf maschinellem Lernen basierendes Transliterationsmodell für das handgeschriebene Ukrainische des 19.-20. Jahrhunderts veröffentlicht. MultiHTR widmet sich dabei neben dem Ukrainischen auch anderen Sprachen und Schriften. Neben dem Umgang mit Ressourcenlücken und schwer zugänglichen Textdaten ist Mehrsprachig- und -schriftlichkeit eine zentrale Herausforderung unseres Projekts., So zielen wir beispielsweise auf die Erkennung nicht-lateinischer Alphabete (Serbisch, Osmanisch, Jiddisch), unterschiedlicher Schreibrichtungen (Osmanisch, Jiddisch) und komplexer Schreibsysteme (deutsche Stenografie). Dies erfordert innovative und skalierbare Lösungen. Anhand ausgewählter Beispiele werden sowohl die Herausforderungen als auch die Besonderheiten einzelner Sprachen/Schriften mit den entsprechenden Lösungsansätzen skizziert.

Parallelentwicklungen zwischen deutscher und taiwanesischer DH

Damir Padieu, Sinologie, DH, Trier Center for Digital Humanities

Für die Digitalisierung und Bearbeitung chinesischsprachiger Dokumente sind die Standard-Tools der westlichen DH oft unzureichend oder unpassend. Die chinesischsprachige DH-Community hat deshalb eigene Werkzeuge entwickelt; ein Beispiel dafür ist das taiwanesisches Markup-Format DocuXML, welches eine Alternative zum TEI-Format darstellt und in Taiwan standardmäßig benutzt wird. Hier kann weniger von einer Ressourcenlücke die Rede sein als von einer Parallelentwicklung. Statt internationaler Zusammenarbeit haben sich voneinander isolierte Systeme entwickelt. Eine Herausforderung besteht darin, über Sprachgrenzen hinweg bessere Kooperationsstrukturen aufzubauen, um Ressourcen zu teilen, Interoperabilität zu fördern und den Mehraufwand der Parallelentwicklung zu vermeiden.

Bibliographie

Arbeitsstelle kleine Fächer. 2020. *Dokumentation des Informations- und Vernetzungsworkshops „Digitalisierung in Lehre und Forschung kleiner Fächer“*.
<https://www.kleinefaecher.de/beitraege/blogbeitrag/dokumentation-des-informations-und-vernetzungsworkshops-digitalisierung-in-lehre-und-forschung-kle> [17.07.2023].

Asef, Esther, and Cosima Wagner. 2018. "Workshop-Bericht 'Nicht-lateinische Schriften in multilingualen Umgebungen: Forschungsdaten und Digital Humanities in den Regionalstudien'." In *DHd Blog*. <https://dhd-blog.org/?p=10669>.

BMBF. "Kleine Fächer – Große Potenziale - BMBF." Bundesministerium für Bildung und Forschung - BMBF. [17.07.2023].
https://www.bmbf.de/bmbf/de/forschung/geistes-und-sozialwissenschaften/kleine-faecher/kleine-faecher_node.html.

Burr, Elisabeth. 2006. "Discussion Paper: Internationalization, Multi-lingual & Multi-cultural agenda". In *Alliance of Digital Humanities Organizations (ADHO): Leadership*.
https://adho.org/wp-content/uploads/2022/07/MLMC_DiscussionPaper_2006.pdf
[17.07.2023].

Dombrowski, Quinn. 2020. *What's a 'Word': Multilingual DH and the English Default*.
<https://quinndombrowski.com/blog/2020/10/15/whats-word-multilingual-dh-and-english-default> [17.07.2023].

Fiormonte, Domenico und Jonathan Usher (Hrsg.). 2001. *New Media and the Humanities: Research and Applications*. Proceedings of the first seminar Computers, Literature and Philology, Edinburgh 7-9 September 1998. Oxford: Humanities Computing Unit, University of Oxford.

Fiormonte, Domenico (Hrsg.). 2003. *Informatica Umanistica*. Dalla ricerca all'insegnamento. Atti dei convegni Computers, Literature and Philology, Università di Roma "La Sapienza" (1999) e Università di Alicante, Spagna (2000). Roma: Bulzoni.

Fiormonte, Domenico. 2021. "Taxation against Overrepresentation? The Consequences of Monolingualism for Digital Humanities." In *Alternative Historiographies of the Digital Humanities*, hg. von Dorothy Kim und Adeline Koh, 333–76. Earth: punctum books.

Gil, Alex, und Élika Ortega. 2016. "Global Outlooks in Digital Humanities: Multilingual Practices and Minimal Computing." In *Doing Digital Humanities: Practice, Training, Research*, 22–34. Abingdon: Routledge.

ADHO Special Interest Group. *Global Outlook::Digital Humanities(GO::DH)*.
<http://www.globaloutlookdh.org>. [17.07.2023].

Grallert, Till, Xenia Monika Kudela, Eliese-Sophia Lincke, Colinda Lindermann, Jana-Katharina Mende, Jonas Müller-Laackman, und Larissa Schmid. 2023. *Umgang mit Multilingualität im DACH und DHd Verband (v1.0.0)*. Zenodo.
<https://doi.org/10.5281/zenodo.7957187>.

Horváth, Alíz. 2023. "DH in Japanese Studies, Japanese Studies in DH: Recent Trends, Tools, and Concepts." *International Journal of Digital Humanities* 4, no. 1: 213–23.
<https://doi.org/10.1007/s42803-022-00063-6>.

Kim, Dorothy, und Jesse Stommel (Hrsg.). 2018. *Disrupting the Digital Humanities*. punctum books. <https://doi.org/10.21983/P3.0230.1.00>.

Nicholas, Gabriel, und Aliya Bathia. 2023. *Lost in Translation. Large Language Models in Non-English Content Analysis*. Center for Democracy & Technology.
<https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/> [17.07.2023]

Nicolás, Carlota / Moneglia, Massimo (Hrsg.). 2005. *La gestione unitaria dell'eredità culturale multilingue europea e la sua diffusione in rete*. Atti della conferenza internazionale CLiP 2003 Computers, Literature and Philology, Firenze 04.-06.12.2003. Firenze: Firenze University Press.

Spence, Paul. 2021. *Disrupting Digital Monolingualism: A Report on Multilingualism in Digital Theory and Practice*. London: Language Acts & Worldmaking project.
<https://doi.org/10.5281/zenodo.5743283>.

Weingart, S.B., N. Eichmann-Kalwara und M. Lincoln, et al. 2020. *The Index of Digital Humanities Conferences*. Carnegie Mellon University. <https://doi.org/10.34666/k1de-j489>.

Notizen

Gelöscht:

Format

Das Panel ist als interdisziplinäre Diskussion gedacht. Nach der Einführung durch die Organisator:innen (5 Minuten) wird jede:r Panelist:in ein kurzes Statement zu ihrem Themenfokus abgeben (je 5 Minuten), bevor wir das Panel für eine erste Diskussion unter den Panelist:innen entlang der skizzierten Leitfragen öffnen (30 Minuten). Eine zweite Runde wird dann mit dem Publikum durchgeführt (30 Minuten).

Herausgeschnitten bei Cristina:

In the conflict between the necessity to develop multilingually enhanced tools versus the production of (often unsustainable) “island solutions”, there is also a general resourcedness gap of funding possibilities.

Orcid:

Jana: <https://orcid.org/0000-0001-7433-4351>

ca. 100-150 Wörter/Panelist:in (bis 30.6. ja/nein + ca 100 Wörter Exposé wenn Zusage bis zum 7.7.)

- ORCID
 - orcid.org/0000-0002-5739-8094
- Tikhonov, Aleksej (PostDoc), Slavistik
- Damir Padieu (Stud.)
- Christina Vertan (BBAW / ab 10/2023: Herder Institut), Osteuropa
- Natalie Kraneiß, Doktorandin, Arabistik, Münster
- Elisabeth Burr, Prof., Romanistik, Leipzig, Präsidentin der EADH
- Frederik Elwert, Prof., Ruhr-Universität Bochum

Reserve:

eine Infrastrukturperson? Marie Effinger (Amalia S. Levi???)

- Diego (Bochum, Leiter DH Center)

Fragen an Panelist:innen:

- Mit welchen Sprachen/Schriften arbeiten Sie?
- Was stellt in Ihrem Fach eine Ressourcedness Gap dar?
- Wie lösen Sie die Probleme?
- Aus welcher Perspektive betrachten Sie die Frage?
- Welche Wünsche haben Sie?