

Daten als Brücke: Interdisziplinäre Zusammenarbeit zwischen Digital Humanities und Data Science im Projekt DataNord

Fuchs, Nele

n.fuchs@uni-bremen.de
Universität Bremen, Deutschland
ORCID: 0009-0004-6837-6968

Steinmann, Lena

lena.steinmann@uni-bremen.de
Universität Bremen, Deutschland
ORCID: 0000-0001-5443-0581

Drechsler, Rolf

drechsler@uni-bremen.de
Universität Bremen, Deutschland
ORCID: 0000-0002-9872-1740

Der gegenwärtige Data Turn und die fortschreitende Datafication in den Geisteswissenschaften ermöglichen einen kritischen Selbstreflexionsprozess innerhalb der Fachdisziplin. Diese Entwicklung zeigt sich im Diskurs um einen geisteswissenschaftlichen Datenbegriff (z. B. Drucker, 2011; Lavin, 2021; Geiger, 2023; Koch, 2024). In einer Zeit zunehmender Datafizierung von Kultur und Lebenswelt steht die geisteswissenschaftliche Forschung vor der Herausforderung, einen adäquaten Datenbegriff zu entwickeln. Diese Diskussion sollte interdisziplinär geführt werden, um die Zusammenarbeit mit anderen Fachdisziplinen zu stärken. Das interdisziplinäre Datenkompetenzzentrum DataNord¹ adressiert diese Entwicklungen, indem es unter anderem gezielte Unterstützung für Forschende in den Digital Humanities bietet.

Dieses Poster präsentiert die Rolle von DataNord als Katalysator für die Integration von Data Science in den Humanities und die Entwicklung eines geisteswissenschaftlichen Datenbegriffs. Es gibt einen Einblick in die Praxis der interdisziplinären Zusammenarbeit an der Schnittstelle von Data Science, Forschungsdatenmanagement (FDM) und Digital Humanities (DH) und lädt zur kritischen Reflexion des geisteswissenschaftlichen Datenbegriffs ein.

DataNord ist ein interdisziplinäres Datenkompetenzzentrum (iDKZ) für die Region Bremen, das als Teil der U Bremen Research Alliance (UBRA) – dem Kooperations-

netzwerk der Universität Bremen und zwölf außeruniversitärer Forschungsinstitute – aufgebaut wird. Die Hauptziele des Projekts sind die Stärkung der datenwissenschaftlichen Kompetenzen bei Forschenden aller Karrierestufen und die Schaffung eines Ortes für interinstitutionelles datenwissenschaftliches Lernen, Forschen und Vernetzen. In dem Projekt arbeiten fünf Data Scientists am Data Science Center (DSC) der Universität Bremen (Steinmann et al., 2023), die die Profilbereiche Umwelt- und Meereswissenschaften, Sozialwissenschaften, Material- und Ingenieurwissenschaften, Gesundheitswissenschaften und Geisteswissenschaften unterstützen. Die Datenkompetenzen sollen gezielt auf Disziplinen ausgeweitet werden, die in ihrer Fachtradition eher fern zu den Datenwissenschaften stehen, wie es in den Geisteswissenschaften der Fall ist, um damit die DH in der Region zu stärken. Data Science Methoden, wie Text-Mining- und Topic-Modelling-Verfahren oder Word-Embedding-Modellen, bieten ein großes Potential für die geisteswissenschaftliche Forschung, denn so können neue Forschungsfragen generiert und große Datenmengen analysiert werden.

Ein weiterer Aspekt der Arbeit von DataNord ist die Unterstützung von Forschenden im Bereich des FDM. Im aktuellen Diskurs über FDM wird häufig auf die Besonderheiten geisteswissenschaftlicher Forschung verwiesen. Allzu oft scheinen die Richtlinien der Open Science Initiativen und die FAIR-Prinzipien mit den Forschungsdaten nicht so recht zusammenzupassen. Beispielsweise ist es mit einem erheblichen Mehraufwand verbunden, die Wiederverwendbarkeit qualitativer sensibler Daten sicherzustellen. Es könnte jedoch argumentiert werden, dass der Eindruck der Besonderheit der Forschungsdaten entsteht, da die Open Science Initiativen und deren Normsetzung für FAIRe wissenschaftliche Praxis ursprünglich nicht aus den datenferneren Disziplinen entstanden sind. Die FAIR-Prinzipien entwickelten sich aus den Life Sciences, einer eher quantitativ arbeitenden Disziplin, um Maschinenlesbarkeit und Wiederverwendbarkeit der Forschungsdaten sicherzustellen (Wilkinson et al., 2016). Braukmann et al. (2023, 5 und 20) stellen fest, dass qualitative Daten in diesen Entwicklungen häufig vernachlässigt werden. Die FAIR-Prinzipien wurden ursprünglich entwickelt, um den Umgang mit quantitativen Daten zu optimieren, was zu einer unzureichenden Berücksichtigung der spezifischen Anforderungen qualitativer und hermeneutischer Forschungsansätze führte. Dieser Fragestellung wurde in einigen Publikationen bereits begegnet (z. B. Lemaire und Minn, 2017; Rizzolli et al., 2024; Kallius, 2023). Im iDKZ DataNord werden Forschende aus den geisteswissenschaftlichen Disziplinen vor diesem Hintergrund in der praktischen Umsetzung von FDM-Strategien und den FAIR-Prinzipien unterstützt.

Die Integration von Data Science Methoden und Forschungsdatenmanagement in die Geisteswissenschaften erfordert die Entwicklung eines adäquaten Datenbegriffs sowie die Förderung des interdisziplinären Dialogs zwischen den Humanities und datenintensiven Disziplinen, um ein gemeinsames Verständnis und eine effektive Zusammenarbeit zu ermöglichen. Eine einheitliche Definition des Be-

griffs gibt es in den Geisteswissenschaften bisher nicht und auch die wissenschaftstheoretische Reflexion steht noch am Anfang. In der Posterpräsentation wird die These Geigers verfolgt, dass eine einheitliche, geisteswissenschaftliche Definition des Begriffs nicht unbedingt zielführend ist (2023, 1). Der Begriff könnte viel mehr als Reflexionsinstrument genutzt werden. Johanna Drucker schlägt beispielsweise vor den Begriff *Capta* dem Datenbegriff vorzuziehen, da damit der konstitutive und situierte Charakter der Wissensproduktion betont würde (2011, 2). Dagegen konstatiert Matthew Lavin, dass es nicht notwendig sei den Datenbegriff zu ersetzen, sondern es ausreichen würde von *situated data* auszugehen (2021). Damit rekurriert er auf Donna Haraways Begriff des *situierten Wissens* (1988), der die epistemologische Abhängigkeit der Forschenden von den gesellschaftlichen Kontextbedingungen beschreibt. Dieser kleine Ausschnitt des Diskurses über den Datenbegriff innerhalb der (digitalen) Geisteswissenschaften zeigt bereits das Potential, welches dieser Zugang zur Wirklichkeit, die forschend beschrieben wird, haben kann. Dabei bauen wir auf eine reiche Fachtradition auf, in welcher die epistemologischen Bedingungen der Wissensproduktion im Forschungsprozess reflektiert werden. Diese epistemologischen Fragestellungen müssen nun durch die Übersetzung der beforschten Wirklichkeit in ein Datum erweitert werden: Was lässt sich in Daten darstellen und welcher Teil unserer qualitativen Daten lässt sich nicht übersetzen? Eine Reflexion, die auch für die datenfernen Disziplinen von Interesse sein kann.

Um jedoch die interdisziplinäre Kommunikation mit den datennahen Wissenschaften, der Informatik und der Data Science weiterhin zu ermöglichen, sollte die enge technische Definition des Datenbegriffs auch im Fokus bleiben: Daten als verarbeitbare Informationseinheiten, die von Computern auf binäre Werte (0/1) reduziert werden (Koch, 2024, 5). Um eine adäquate Zusammenarbeit zu ermöglichen, ist es zudem notwendig, ein Verständnis für die Methoden und Fachkulturen zu entwickeln, um einander verstehen zu können. Häufig arbeiten wir an ähnlichen Fragestellungen, nutzen jedoch andere Fachtermini und Prämissen. In der interdisziplinären Zusammenarbeit kommt der Übersetzungsarbeit daher eine entscheidende Rolle zu. Welchen Zugang und Grundsätze die wissenschaftlichen Disziplinen zur Welt haben, wird im interdisziplinären Forschungsprozess selten explizit adressiert. Um einen geeigneten Reflexionsrahmen zu schaffen, in welchem sich Forschende über solche Fragen austauschen können nutzen wir am DSC das COLLAB Game (COLLAB, 2024), welches eigenständig durch ein Set von datenbezogenen Fragen erweitert werden kann. Damit kann ein Verständnis für die Daten anderer Disziplinen entwickelt werden, ein positiver Effekt, der im interdisziplinären Team im Projekt DataNord für die Zusammenarbeit unabdingbar ist.

Das Poster lädt zur Diskussion über die Zukunft der Digital Humanities, den geisteswissenschaftlichen Datenbegriff und FDM im Kontext der FAIR-Prinzipien ein und gibt Inspirationen, wie andere Institutionen ähnliche Strukturen zur Förderung der Datenkompetenzen in den Geisteswis-

senschaften aufbauen können. Wir werden erste Einblicke in die Tätigkeiten von DataNord in den Geisteswissenschaften geben und deren Wirksamkeit diskutieren.

Fußnoten

1. Das Projekt ist BMBF-gefördert und läuft von 2023 bis 2026 (U Bremen Research Alliance, 2024).

Bibliographie

COLLAB. 2024. "Interdisciplinary Games. Spiele zur Förderung interdisziplinärer Zusammenarbeit." URL: <https://de.interdisciplinarygames.net>

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." In: *Digital Humanities Quarterly*, Vol 5, Nr. 1. URL: <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>

Geiger, Jonathan. 2023. "Daten / Forschungsdaten." In: AG Digital Humanities Theorie des Verbandes Digital Humanities im deutschsprachigen Raum e. V. (Hg.): Begriffe der Digital Humanities. Ein diskursives Glossar (= Zeitschrift für digitale Geisteswissenschaften / Working Papers, 2). Wolfenbüttel. 25.05.2023. Version 2.0 vom 14.03.2024. HTML / XML / PDF. DOI: 10.17175/wp_2023_003_v2

Haraway, Donna. 1988. "Situated knowledges. The science question in feminism and the privilege of partial perspective." *Feminist Studies* 14: 575–599.

Kailus, Angela. 2023. "Handreichung für ein FAIRes Management kulturwissenschaftlicher Forschungsdaten" (Version 1.0.1). Zenodo. <https://doi.org/10.5281/ZENODO.7716941>

Koch, Gertraud. 2024. "Digitale Remediation und Datafizierung der ethnografischen Forschung." In: Ernst, C., Krtilova, K., Schröter, J., Sudmann, A. (Hg.): Handbuch Medientheorien im 21. Jahrhundert. Springer VS, Wiesbaden. DOI: 10.1007/978-3-658-38128-8_31-1

Lavin, Matthew. 2021. "Why Digital Humanists Should Emphasize Situated Data over Capta." _Digital Humanities Quarterly_, 15 (2). <https://www.digitalhumanities.org/dhq/vol/15/2/000556/000556.html>

Lemaire, Marina, Gisela Minn. 2017. "Forschungsdatenmanagement in den Geisteswissenschaften. Eine Planungshilfe für die Erarbeitung eines digitalen Forschungskonzepts und die Erstellung eines Datenmanagementplans." *Universität Trier eScience Working Papers*, Nr. 3, Trier. URN: <https://nbn-resolving.org/urn:nbn:de:hbz:385-10715>

Rizzolli, Michaela, Sabine Imeri, Elisabeth Huber. 2024. "Ethnografische Forschungsmaterialien zur Archivierung und Nachnutzung vorbereiten und dokumentieren – ein Überblick für Forschende." Universität Bremen. <https://doi.org/10.26092/ELIB/2723>

Steinmann, Lena, Heike Thöricht, Sandra Zänkert, Rolf Drechsler. 2023. "Das Data Science Center an der

Universität Bremen: Interdisziplinärer Knotenpunkt und Service-Infrastruktur für die datenintensive Forschung." In: Heuveline, Vincent und Bisheh, Nina (Hrsg.): E-Science-Tage 2023: Empower Your Research – Preserve Your Data, Heidelberg: heiBOOKS. DOI: 10.11588/heibooks.1288.c18083

U Bremen Research Alliance, 2024 "Interdisziplinäres Datenkompetenzzentrum für die Region Bremen." URL: <https://www.bremen-research.de/datanord>

Wilkinson, Mark et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship." **Scientific Data**, 3 (1), DOI: <http://dx.doi.org/10.1038/sdata.2016.18>