

Data Science Services an der UB Mannheim

Daten neu denken in den digitalen Geisteswissenschaften

Will, Larissa

larissa.will[at]uni-mannheim.de
Universitätsbibliothek Mannheim, Deutschland
ORCID: 0009-0004-6220-8939

Shigapov, Renat

renat.shigapov[at]uni-mannheim.de
Universitätsbibliothek Mannheim, Deutschland
ORCID: 0000-0002-0331-2558

Schmidt, Thomas

thomas.schmidt[at]uni-mannheim.de
Universitätsbibliothek Mannheim, Deutschland
ORCID: 0000-0003-3620-3355

Zusammenfassung. Das Forschungsdatenzentrum der Universitätsbibliothek Mannheim bietet ein breit gefächertes Angebot an digitalen Services, Plattformen und Unterstützungsformaten für die digitalen Geisteswissenschaften und im Bereich der Data Literacy. Dazu zählen spezialisierte Services wie MAObjects zur Präsentation digitaler Objektsammlungen mit Omeka sowie die automatisierte Spracherkennung mit whisper zur Unterstützung von Forschungsprojekten. Als Teil des Kompetenzzentrums OCR fördert die UB Mannheim gemeinsam mit der UB Tübingen die automatisierte Texterkennung – sowohl innerhalb der Universität als auch für externe Vorhaben. Ergänzt wird das Portfolio durch digitale Lernangebote wie MaDaLi² und durch projektbasierte Lehrinitiativen zur Förderung datenbezogener Kompetenzen. Ziel ist es, sowohl Forschende als auch Studierende im Umgang mit digitalen Methoden zu qualifizieren und die nachhaltige Verankerung digitaler Infrastrukturen in der wissenschaftlichen Praxis zu unterstützen. Der Beitrag gibt einen Überblick über das vernetzte Service- und Projektangebot der UB Mannheim, stellt ausgewählte Anwendungsszenarien vor und diskutiert zentrale Herausforderungen an der Schnittstelle von Infrastruktur, Qualifizierung und methodischer Innovation.

1 Data Science Services der UB allgemein

Das Forschungsdatenzentrum (FDZ) der Universitätsbibliothek Mannheim hat institutionelle Data Science Services aufgebaut, um Forschende an der Universität Mannheim umfassend bei datenbasierten Projekten zu unterstützen (Shigapov et al. 2025a und Shigapov et al. 2025b). Ziel ist es, Forschungsdatenmanagement (FDM) zu stärken, Open Science zu fördern und zur Reproduzierbarkeit wissenschaftlicher Arbeiten beizutragen – unabhängig vom technischen Vorwissen der Forschenden (UB Mannheim 2025b, Forschungsdatenzentrum).

Das FDZ unterstützt entlang der Data-Science-Pipeline – von der Datenerhebung (z. B. durch Web Scraping, OCR oder Audio-Transkription) über Analyse, Visualisierung und Modellierung bis hin zur Veröffentlichung. Es berät bei Förderanträgen, begleitet datenwissenschaftliche Vorhaben und fördert die Veröffentlichung von Forschungsdaten – etwa in interaktiven Webanwendungen oder Knowledge Graphs (Universität Mannheim 2025d, KGI). Besonderes Augenmerk liegt auf offenem Code, Reproduzierbarkeit und den FAIR-Prinzipien.

Universitätsinterne Vernetzung wird durch Veranstaltungen, Workshops und Kooperationen, etwa mit dem Mannheim Center for Data Science und der Graduate School of Economic and Social Sciences, intensiv gefördert (Universität Mannheim 2025b, MCDS und Universität Mannheim 2025c, GESS). Zudem bündeln Veranstaltungsreihen wie die „Research Skills“ (UB Mannheim 2025f, Research Skills) sowie dezidierte Data-Literacy-Projekte wie MaDaLi² fachübergreifendem Wissenstransfer im Bereich von Data Literacy, FDM und Open Science für Studierende und Forschende der Universität (Universität Mannheim 2025a, Data Literacy und Sidiropoulos 2024 und Bröcher et al. 2025).

2 Digitale Geisteswissenschaften im Fokus

FDM-Angebote für die digitalen Geisteswissenschaften sind fester Bestandteil des FDZ. Bereits seit 2013 beschäftigt sich die UB intensiv mit Optical Character Recognition (OCR). Neben eigenen Digitalisierungsprojekten war die UB an drei Phasen der OCR-D-Initiative (DFKI 2025 und UB Mannheim 2025e, Projekte der UB) beteiligt und hat gemeinsam mit der Universitätsbibliothek Tübingen das Kompetenzzentrum OCR aufgebaut (Will und Huff 2023).

Ein zentrales Angebot ist die monatliche OCR-Sprechstunde, die auf großes überregionales Interesse bei Forschenden, GLAM-Einrichtungen und Privatpersonen stößt. Das FDZ unterstützt sowohl uniintern als auch vereinzelt extern bei OCR-Vorhaben durch individuelle Beratung, Support und Projektmitarbeit. Auch hierbei wird gezielt auf Open-Source-Lösungen gesetzt, um nachhaltige und anschlussfähige Ergebnisse zu schaffen (UB Mannheim 2025d, OCR-BW).

Ein weiterer zentraler Service ist die automatisierte Transkription multimedialer Inhalte (UB Mannheim 2025a, Automatisierte Spracherkennung). Das FDZ begleitet den gesamten Prozess – von der Erstberatung über die Einrichtung des Workflows bis zur Anpassung der Ausgabeformate. So können wertvolle neue Forschungsdaten aus Quellen wie Zeitzeug*innen-Interviews, Rundfunk- und Fernsehaufnahmen effizient generiert werden. Die Transkription erfolgt intern auf den Servern der UB mithilfe von *whisply* (Schmidt 2025).

Der Service *MAObjects* unterstützt die Konzeption und Umsetzung digitaler Ausstellungen und Datenbanken mit der Software *Omeka* (Omeka 2025.). Diese lassen sich langfristig auf den Servern der UB hosten – eine benutzerfreundliche und attraktive Lösung für die Geisteswissenschaften (Schumm und Will 2024 und UB Mannheim 2025c, *MAObjects*.).

Neben diesen Infrastrukturservices unterstützt das FDZ auch aktiv Geisteswissenschaftler*innen bei datengetriebenen Forschungsprojekten und entwickelt teils maßgeschneiderte Data-Science-Pipelines. Als Beispiel hierfür kann ein Projekt aus der Romanistik dienen, in dem annotierte Textkorpora (18. bis 20. Jahrhundert) zur figurativen Verwendung des Paralyse-Begriffes erstellt und, auf diesen aufbauend, ein Multilabel-Classifer trainiert wurde (Kuschel und Shigapov 2025).

Ziel der FDZ-Services für Geisteswissenschaftler*innen ist es, digitale Methoden in den Humanities zugänglich zu machen, auch für Forschende mit wenig oder keiner Programmiererfahrung. Wichtig dafür sind Low- und No-Code-Tools und die Beratung zu passenden Workflows und Technologien. Das Poster bietet einen kompakten Einblick in das Serviceangebot und stellt aktuelle Projekte des Forschungsdatenzentrums der UB Mannheim vor, die den praktischen Einsatz digitaler Methoden in den Geisteswissenschaften veranschaulichen.

Bibliografie

- Bröcher, L., T. Schmidt, I. Hartig, J. Schwenzer, I. Schumm, und N. Baldus. 2025. Optimierung eines E-Learning-Angebots durch Studierenden-Interviews und KI-Tools. In *Neues Handbuch Hochschul-lehre: Lehren und Lernen effizient gestalten*, Band 119, herausgegeben von B. Berendt, H.-P. Voss und J. Wildt, 83–104. Stuttgart: Franz Steiner Verlag.
- Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI). 2025. „OCR-D – Deutsches Projekt zur optischen Zeichenerkennung.“ Zugriff am 29. April 2025. <https://ocr-d.de/de/>.
- Kuschel, Daniela, and Renat Shigapov. 2025. “Paraly: An (annotated) dataset for exploring the concept of paralysis (fr. ‘paralysie’) in a digital corpus of French Literature.” *Data in Brief* (June 2025). <https://doi.org/10.1016/j.dib.2025.111577>.
- Omeka. 2025. „Omeka – Web-Publishing-Plattform für digitale Sammlungen und Ausstellungen.“ Zugriff am 29. April 2025. <https://omeka.org/>.
- Schmidt, Thomas. 2025. *whisply*. GitHub-Projekt. Zugriff am 29. April 2025. <https://github.com/tsmdt/whisply>.
- Shigapov, R., Schumm, I., Schmidt, T., Kamlah, J., & Will, L. (2025a, März 17). Institutional Data Science Services at FDZ UB Mannheim: Enhancing Research Data Management. E-Science-Tage 2025, Heidelberg.
- Shigapov, R., Schumm, I., Kamlah, J., & Will, L., Schmidt, T. (2025). FDZ UB Mannheim: Three challenges in establishing sustainable research data management, data science, and AI services. Data-Science-Community Workshop: Aktuelle Entwicklungen und Perspektiven (an Hochschulen) im Bereich Data Science (DSCW), Potsdam.
- Schumm, Irene, und Larissa Will. 2024. „OMEKA als Software für digitale Objektsammlungen und virtuelle Ausstellungen – Ein Erfahrungsbericht.“ *AKMB-News* 30 (1): 31–35. Düsseldorf: AKMB. <https://journals.ub.uni-heidelberg.de/index.php/akmbnews/article/view/82883>.

- Sidiropoulos, Christos. 2024. „MaDaLi — Data Literacy E-Learning Plattform an der Uni Mannheim.“ Posterpräsentation bei der Biblio-Con 2024, Hamburg, 4. Juni 2024. Zugriff am 29. April 2025. <https://nbn-resolving.org/urn:nbn:de:0290-opus4-189325>.
- Universität Mannheim. 2025a. „Data Literacy – Lehrprojekte.“ Zugriff am 29. April 2025. <https://www.uni-mannheim.de/universitaet/lehre/lehrprojekte/data-literacy/>.
- Universität Mannheim. 2025b. „Mannheim Center for Data Science (MCDS).“ Zugriff am 29. April 2025. <https://www.uni-mannheim.de/datascience/>.
- Universität Mannheim. 2025c. „GESS – Graduate School of Economic and Social Sciences.“ Zugriff am 29. April 2025. <https://www.uni-mannheim.de/gess/>.
- Universität Mannheim. 2025d. „Knowledge Graph Infrastructure (KGI).“ Zugriff am 29. April 2025. <https://kgi.uni-mannheim.de/>.
- Universitätsbibliothek Mannheim. 2025a. „Automatisierte Spracherkennung.“ Zugriff am 29. April 2025. <https://www.bib.uni-mannheim.de/lehren-und-forschen/forschungsdatenzentrum/fdz-services/automatisierte-spracherkennung/>.
- Universitätsbibliothek Mannheim. 2025b. „Forschungsdatenzentrum (FDZ).“ Zugriff am 29. April 2025. <https://www.bib.uni-mannheim.de/lehren-und-forschen/forschungsdatenzentrum/>.
- Universitätsbibliothek Mannheim. 2025c. „MAObjects – Digitale Objektsammlungen und virtuelle Ausstellungen.“ Zugriff am 30. April 2025. <https://www.bib.uni-mannheim.de/lehren-und-forschen/forschungsdatenzentrum/fdz-services/maobjects/>.
- Universitätsbibliothek Mannheim. 2025d. „OCR-BW – Kompetenzzentrum OCR der Universitätsbibliotheken Mannheim und Tübingen.“ Zugriff am 29. April 2025. <https://ocr-bw.bib.uni-mannheim.de/>.
- Universitätsbibliothek Mannheim. 2025e. „Projekte der UB.“ Zugriff am 29. April 2025. <https://www.bib.uni-mannheim.de/ihre-ub/projekte-der-ub/#c207994>.

- Universitätsbibliothek Mannheim. 2025f. „Research Skills.“ Zugriff am 29. April 2025. <https://www.bib.uni-mannheim.de/lehren-und-forschen/research-skills/>.
- Will, Larissa, und Dorothee Huff. 2023. „Kompetenzzentrum OCR – Automatische Texterkennung als Serviceangebot.“ Präsentation auf der FORGE 2023, Tübingen, 4.–6. Oktober 2023. <https://doi.org/10.5281/zenodo.8392499>.