
Institutional Data Science Services at FDZ UB Mannheim: Enhancing Research Data Management

Renat Shigapov ¹, Irene Schumm ¹, Thomas Schmidt ¹, Jan Kamlah ¹,
Larissa Will ¹

¹ University Library, University of Mannheim

Data science services have rapidly emerged as essential support mechanisms for research data management (RDM) at universities worldwide. In alignment with these developments, the research data center (FDZ) at the Mannheim University Library (UB Mannheim) has established institutional data science services at the University of Mannheim. Our goal is to enhance RDM, promote open science, and contribute to research reproducibility. FDZ supports researchers across the complete data science pipeline, emphasizing advanced data collection methods (OCR, audio/video transcription, API integration, web scraping) and chatbot and large language model (LLM) applications. Our services include data-science-as-a-service, education and training, consultancy, community engagement, and integration of data science methods within internal library workflows. In this paper we introduce our services, describe the challenges we faced, and analyze the impact these services have had on RDM at the University of Mannheim.

Keywords: data science services, RDM services, AI, research support

1 Introduction

Data Science Services have rapidly emerged as essential support mechanisms for research data management (RDM) at universities worldwide. Notable examples include services

Published in: Vincent Heuveline, Philipp Kling, Florian Heuschkel, Sophie G. Habinger, and Cora F. Krömer (Hrsg.): E-Science-Tage 2025. Research Data Management: Challenges in a Changing World. Heidelberg: heiBOOKS, 2025. DOI: <https://doi.org/10.11588/heibooks.1652.c23916> (CC BY-SA 4.0).

at Harvard University¹, the University of Utah², Purdue University³, NC State University⁴, the University of Arizona⁵ (Oliver et al. 2019), and the University of Groningen⁶. In Germany, several data science initiatives have recently been established. These include the Data Science Center at the University of Bremen (Steinmann, Thöricht, et al. 2023), the Bielefeld Center for Data Science (Weiß and Decker 2023), the consortium NFDI4DataScience (Schimmler et al. 2023) within the national research data infrastructure (Hartl, Wössner, and Sure-Vetter 2021) and some others listed in Steinmann, Nowotka, et al. (2023).

The latter describes the Mannheim Center for Data Science (MCDS)⁷, established in 2020. The MCDS serves as a platform to connect and enhance research in and through data science at the University of Mannheim. It organizes courses, networking events, and lectures, contributing to science communication by regularly sharing information on new projects and developments in data science. Due to limited human resources, the MCDS does not offer data-science-as-a-service, data science consultancy, education and training on data science including research data management (RDM), or IT infrastructure. To address these gaps at least partially, the Research Data Center (FDZ) at the University Library Mannheim (UB Mannheim) created institutional data science services in 2024.

In this paper, we introduce the FDZ and its data science services, describe the challenges we have faced, and analyze the impact of these services on research data management at our institution.

2 FDZ UB Mannheim

The FDZ serves as the primary contact point and support hub for all research data management (RDM) needs at the University of Mannheim. Given the university’s leading role in economics and social sciences, the FDZ focuses particularly on addressing the spe-

1 Data Science Services at Harvard University, <https://www.iq.harvard.edu/data-science-services>; *Visited on March 30, 2025.*

2 Data Science Services at University of Utah, <https://cores.utah.edu/data-science-service>; *Visited on March 30, 2025.*

3 Data Science Consulting Service at Purdue University, <https://www.purdue.edu/data-science/dscs>; *Visited on March 30, 2025.*

4 Data Science Consulting Team at NC State University, <https://datascienceacademy.ncsu.edu/data-science-consulting/data-science-consulting-team>; *Visited on March 30, 2025.*

5 Data Science & Visualization services at University Libraries, <https://data.library.arizona.edu/data-science>; *Visited on March 30, 2025.*

6 Data Science Team in the Center for Information Technology at University of Groningen, <https://www.rug.nl/society-business/centre-for-information-technology/research/services/data-science>; *Visited on March 30, 2025.*

7 Mannheim Center for Data Science, <https://www.uni-mannheim.de/datascience>; *Visited on March 30, 2025.*

cific needs and challenges of these fields, while also supporting interdisciplinary research efforts.

The FDZ provides a comprehensive suite of core RDM services, including the Mannheim data repository (MADATA), RDM consultancy and educational programs, support for creating data management plans (MADaPlan), and a secure data-sharing environment. Additionally, the FDZ plays an active role in multiple interdisciplinary projects including data literacy (MaDaLi²)⁸, chatbots for RDM (Schmidt et al. 2025; Shigapov and Schumm 2024), optical character recognition (OCR-BW⁹), knowledge graphs (Rossenova et al. 2023; Limani et al. 2023), BERD@BW (Gehrlein, Schumm, and Shigapov 2022), and BERD@NFDI (Gehrlein et al. 2022).

Participation in these diverse initiatives has significantly boosted the FDZ’s internal expertise, enabling the center to attract highly qualified personnel specialized in data science and advanced RDM practices. This has led to the establishment and expansion of consultancy services focusing on OCR techniques, AI-driven methodologies, and advanced research data management.

3 Data Science Services

The Data Science Services at the FDZ include four core offerings: data-science-as-a-service, education and training, consultancy, and community engagement and networking.

Data-science-as-a-service. Our core offerings include automatic data collection methods such as Optical Character Recognition (OCR), audio/video transcription, web scraping, data acquisition via APIs, and transforming unstructured data into structured formats. Recently, we have expanded our scope by implementing data science pipelines, developing specialized chatbots, and collaboratively authoring data papers.

Education and training. Our education and training services include regular courses and hands-on workshops covering topics such as chatbots for research, AI tools for academic studies, open-source AI assistants for research data management (RDM), OCR applications in research, and automatic audio/video transcription. These workshops are integrated into our established training series, including “Research Skills” and specialized “RDM Seminars.”

⁸ <https://www.uni-mannheim.de/universitaet/lehre/lehrprojekte/data-literacy>; *Visited on March 30, 2025.*

⁹ <https://ocr-bw.bib.uni-mannheim.de>; *Visited on March 30, 2025.*

Consultancy. We offer consultancy through multiple channels, including official FDZ email, direct personal contact with FDZ staff, subject librarian referrals, and participation in related courses and workshops. Consultations are provided through email, virtual meetings, and in-person sessions, complemented by regular consultancy hours dedicated to topics such as OCR and knowledge graphs.

Community engagement and networking. We actively participate in existing data science networking activities, particularly those organized by the MCDS. This includes contributing to the MCDS academic speed-dating events each semester, which promote collaboration among university researchers. Additionally, we engage in university-wide events such as the GESS Research Day, presenting our services directly to the research community.

Additionally, establishing these data science services enabled us to optimize internal library workflows through data science methods. This includes developing chatbots for general library use and specialized chatbots for the FDZ addressing research data management (RDM) and open science inquiries. We also work on automatic subject indexing tasks using Dewey Decimal Classification (DDC) and Regensburger Verbundklassifikation (RVK) systems, streamline RDM processes within FDZ (such as metadata review and creating data documentation), and represent the library in the university’s strategic AI initiatives, including the development of the institutional AI strategy.

4 Challenges

Establishing institutional data science services at the FDZ involved several challenges. A key issue concerning the emerging field of AI applications within the data science services of the FDZ was the initial lack of an overarching institutional AI strategy at the University of Mannheim. We started building our data science services independently, which led to uncertainty regarding alignment with broader institutional goals. However, the FDZ has since actively contributed to the development of the university-wide AI strategy, ensuring our services align closely with institutional priorities and standards.

Another considerable hurdle is the high infrastructural costs, particularly related to the need for computational resources such as GPUs, essential for delivering advanced AI-based services frequently requested by researchers. Nevertheless, the university IT plans to expand GPU infrastructure, promising support for future projects.

Personnel costs also presented significant challenges. Operating and maintaining sophisticated data science services demand highly skilled staff, which increases operational costs in the long term. To ensure sustainability, the FDZ is actively seeking more AI, data sci-

ence, and RDM-related projects, along with internal and external funding opportunities to ensure consistent support and development of these services.

Finally, scalability represents a critical challenge due to the potentially high costs associated with expanding data-science-as-a-service. To mitigate scalability issues, the FDZ is exploring multi-tier support models and strategic project partnerships. By partnering directly in projects, the FDZ can split costs, optimize resources, and sustainably grow its data science service offerings.

5 Impact on RDM

Since our data science services were established only recently, we currently provide a qualitative analysis of their initial impact on RDM at the University of Mannheim.

The primary beneficiaries of our data-science-as-a-service offerings are researchers who lack the resources or expertise to manage complex data science tasks independently. Our data collection services, including OCR, audio/video transcription, web scraping, and data structuring, have increased the availability and accessibility of FAIR-compliant research data within the university. By actively engaging in data collection, we also directly support and advise researchers on best practices in FAIR data sharing through data repositories and data papers, enhancing overall transparency and reproducibility of research projects.

Recognizing higher participation rates in AI-focused workshops compared to traditional RDM sessions, we incorporated essential RDM concepts into our AI-oriented training programs. This integrated approach broadens our outreach, ensuring that a wider group of researchers develops competencies in both AI methods and RDM practices.

Moreover, automating RDM workflows (such as metadata validation and the creation of data documentation) has improved our internal processes at the FDZ of the UB Mannheim. This automation reduces the administrative burden, resulting in faster metadata checks and smoother data submission and publication.

6 Discussion

Currently, the University of Mannheim hosts two distinct structures supporting data science activities: the MCDS and the FDZ. While the MCDS operates as a researcher-driven initiative, the FDZ serves as an infrastructural department. Despite their separate organizational setups, these entities effectively complement one another. For instance,

data scientists from FDZ actively participate in the academic speed dating events on data science organized each semester by the MCDS.

Although Putzke et al. (2023) suggest that “data science roles should be housed in newly created, independent data science centers,” our experience at the University of Mannheim demonstrates that parallel structures, such as those embodied by the MCDS and the FDZ, can also yield substantial benefits, provided their respective services are clearly defined and mutually complementary.

The integration between data science services and RDM services occurs naturally within our framework, as our data scientists also function as data stewards (RDM consultants). This dual role is highly beneficial, ensuring that every data science consultation inherently addresses RDM principles and practices, and vice versa.

7 Conclusions

Establishing institutional data science services at the FDZ has had a positive impact on research data management (RDM) practices at the University of Mannheim. Despite initial challenges (including infrastructure costs, staffing requirements, and the lack of a comprehensive institutional AI strategy), our services have quickly demonstrated benefits. By providing practical data science solutions, RDM-included AI training, and consultancy, we have expanded the amount of available research data, improved researcher competencies in data stewardship, and streamlined internal RDM workflows.

The dual role of our data scientists as RDM consultants has ensured the integration of data science and RDM practices, enhancing researcher awareness of FAIR principles and open science standards. Going forward, continued collaboration with parallel structures such as the MCDS, strategic infrastructure investment, and sustained resource planning will be critical. Further quantitative evaluation will also be essential to fully measure long-term impact and refine our service offering.

Acknowledgements

The authors thank the whole FDZ and UB Mannheim team for their contributions and support.

Authorship Contributions

- Renat Shigapov: Writing – original draft
- Irene Schumm: Writing – review & editing
- Thomas Schmidt: Writing – review & editing
- Jan Kamlah: Writing – review & editing
- Larissa Will: Writing – review & editing

Conflict of Interest

The authors are not aware of any potential conflicts of interest.

Bibliography

- Gehrlein, Sabine, Annette Klein, Irene Schumm, and Klaus Tochtermann. 2022. “BERD@NFDI für unstrukturierte Daten in den Wirtschafts- und Sozialwissenschaften”. *Zeitschrift für Bibliothekswesen und Bibliographie* 69 (1–2): 59–67. <https://doi.org/10.3196/1864295020691280>.
- Gehrlein, Sabine, Irene Schumm, and Renat Shigapov. 2022. “BERD@BW – A Science Data Center to foster Open Science in Business, Economics and Social Sciences”. In *E-Science-Tage 2021: Share Your Research Data*, edited by Vincent Heuveline and Nina Bisheh, 314–319. heiBOOKS. ISBN: 978-3-948083-55-7. <https://doi.org/10.11588/HEIBOOKS.979.C13743>.
- Hartl, Nathalie, Elena Wössner, and York Sure-Vetter. 2021. “Nationale Forschungsdateninfrastruktur (NFDI)”. *Informatik Spektrum* 44 (5): 370–373. <https://doi.org/10.1007/s00287-021-01392-6>.
- Limani, Fidan, Renat Shigapov, Auriol Degbelo, Sonja Schimmler, Benjamin Zapilko, Lozana Rossenova, Stefan Dietze, Konrad U. Förstner, Saurav Karmakar, and Andreas Czerniak. 2023. *Who is using Knowledge Graphs in NFDI? An overview by the Working Group "Knowledge Graphs"*. Zenodo. <https://doi.org/10.5281/ZENODO.8332776>.
- Oliver, Jeffrey C., Christine Kollen, Benjamin Hickson, and Fernando Rios. 2019. “Data Science Support at the Academic Library”. *Journal of Library Administration* 59 (3): 241–257. <https://doi.org/10.1080/01930826.2019.1583015>.

- Putzke, Johannes, Thomas Schimmer, Manuela Richter, and Lucas Hamel. 2023. "Towards the Establishment of Data Science Centers at Higher Education Institutions". In *INFORMATIK 2023 – Designing Futures: Zukünfte gestalten*, 117–121. Gesellschaft für Informatik e.V. https://doi.org/10.18420/INF2023_10.
- Rossenova, Lozana, Renat Shigapov, Moritz Schubotz, and Daniel Mietchen. 2023. *KGI4-NFDI: Knowledge Graph Infrastructure for the German National Research Data Infrastructure*. Zenodo. <https://doi.org/10.5281/ZENODO.8337432>.
- Schimmmler, Sonja, Bianca Wentzel, Arnim Bleier, Stefan Dietze, Saurav Karmakar, Peter Mutschke, Angelie Kraft, et al. 2023. "NFDI4DS Infrastructure and Services". In *INFORMATIK 2023 – Designing Futures: Zukünfte gestalten*. Gesellschaft für Informatik e.V. https://doi.org/10.18420/INF2023_103.
- Schmidt, Thomas, Renat Shigapov, Irene Schumm, and Jan Kamlah. 2025. *Perplexity-inspired metasearch-based alternatives to FAIR GPT: Open-source AI consultants for RDM*. Zenodo. <https://doi.org/10.5281/ZENODO.15038486>.
- Shigapov, Renat, and Irene Schumm. 2024. "FAIR GPT: A virtual consultant for research data management in ChatGPT". *arXiv*, <https://doi.org/10.48550/ARXIV.2410.07108>.
- Steinmann, Lena, Dirk Nowotka, Lea Oberländer, Helen Pfuhl, Heiner Stuckenschmidt, and Rolf Drechsler. 2023. "Workshop: „Aktuelle Entwicklungen und Perspektiven (an Hochschulen) im Bereich Data Science“". In *INFORMATIK 2023 – Designing Futures: Zukünfte gestalten*. Gesellschaft für Informatik e.V. https://doi.org/10.18420/INF2023_05.
- Steinmann, Lena, Heike Thöricht, Sandra Zänkert, and Rolf Drechsler. 2023. "Das Data Science Center an der Universität Bremen: Interdisziplinärer Knotenpunkt und Service-Infrastruktur für die datenintensive Forschung". In *E-Science-Tage 2023: Empower Your Research – Preserve Your Data*, edited by Vincent Heuveline, Nina Bisheh, and Philipp Kling, 246–252. heiBOOKS. <https://doi.org/10.11588/HEIBOOKS.1288.C18083>.
- Weiß, Katharina, and Reinhold Decker. 2023. "Data Science und Forschungsdaten als Treiber für wissenschaftliche Interdisziplinarität". In *INFORMATIK 2023 – Designing Futures: Zukünfte gestalten*. Gesellschaft für Informatik e.V. https://doi.org/10.18420/INF2023_09.