

FDZ UB Mannheim: Three challenges in establishing sustainable research data management, data science, and AI services

Renat Shigapov ¹, Irene Schumm ², Jan Kamlah ³, Larissa Will⁴ , Thomas Schmidt⁵ 

Abstract: Academic libraries are increasingly expanding their role in research support by offering data science services. The research data center at the Mannheim University Library is keeping pace with this trend, creating integrated research data management, data science, and artificial intelligence services. These developments enhance community engagement, streamlines internal processes, and fosters innovation across the library. In this paper, we describe three challenges of developing these emerging services: defining the evolving scope of our offerings, building key partnerships, and planning for long-term sustainability.

Keywords: Data Science Services, AI Services, Research Data Management, Research Data Center

1 Introduction

In recent years, universities and academic libraries worldwide have established institutional data science services [Ol19, He21a]. In Germany, prominent initiatives include the Data Science Center at the University of Bremen [St23a], the Bielefeld Center for Data Science [WD23], the consortium NFDI4DataScience [Sc23], and a few others listed in [St23b]. The concept of integrating RDM support with data science services is not entirely new and offers numerous advantages [St21].

Inspired by 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 [Sh25]. In the process, we have inadvertently aligned with several of the “ten simple rules for starting and sustaining an academic data science initiative” [PBB21] including “Rule 1: Don’t try to own everything”, “Rule 3: Have a sustainability plan (and find funding)”, “Rule 4: Hire a team, and support them”, “Rule 5:

¹ University of Mannheim, University Library, Schloss Schneckenhof, Mannheim, 68131, renat.shigapov@uni-mannheim.de,  <https://orcid.org/0000-0002-0331-2558>

² irene.schumm@uni-mannheim.de,  <https://orcid.org/0000-0002-0167-3683>

³ jan.kamlah@uni-mannheim.de,  <https://orcid.org/0000-0002-0417-7562>

⁴ larissa.will@uni-mannheim.de,  <https://orcid.org/0009-0004-6220-8939>

⁵ thomas.schmidt@uni-mannheim.de,  <https://orcid.org/0000-0003-3620-3355>

Recognize and elevate data, software, and workflow contributions”, “Rule 7: Emphasize responsible data science”, and “Rule 9: Engage with external communities”. We also incorporated elements from a data science framework tailored for academic research libraries [Ma21] including their recommendations on communication and branding, assessment and reporting, as well as partnerships and stakeholders.

According to [He21a], academic libraries face three primary challenges when developing new data science services: defining the scope, building partnerships, and ensuring sustainability. In this paper, we define the evolving scope of FDZ’s offerings, describe key partnerships, and outline our plans for long-term sustainability.

2 Defining the evolving scope of services

The FDZ UB Mannheim offers a dynamic and evolving range of services in RDM, data science, and AI. While defining the exact scope of these services can be challenging, we take a flexible and data-driven approach by continuously assessing the needs of our researchers and adapting our offerings accordingly.

Our evolving service portfolio is shaped by a combination of user demand, institutional priorities, and feasibility. We regularly analyze feedback from consultations, training sessions, and outreach events to identify recurring needs and gaps. Promising topics are often tested through workshops or pilot consultations before being formalized as core services. In line with our “learning-by-doing” approach, team members also proactively explore emerging trends (such as generative AI) by preparing training materials and prototype solutions. This allows us to respond quickly to the needs of the research community while building internal expertise.

- **Data-Science-as-a-Service.** Our 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. We also implement data science pipelines, fine-tune large language models, and develop specialized chatbots.
- **Education and Training.** Our educational 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 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."
- **Consultancy.** We offer consultancy on RDM, data science, and AI through multiple channels, including official FDZ email, direct personal contacts, and as question/answer sessions within 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 Mannheim Center for Data Science (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.
- **Data offers.** The FDZ provides researchers with access to licensed external databases alongside a growing portfolio of in-house datasets. Our internally generated collections include printed materials digitized into machine-readable formats, as well as specialized economic and social-science data, with a particular emphasis on corporate information. The FDZ also curates datasets inherited from university research projects, preserving them for long-term scholarly use. The most prominent datasets include Aktienführer data archive, Bankruptcy database, German Reichsanzeiger and Prussian Staatsanzeiger, German Statistical Yearbooks, Official municipal directories, LinguaPix, Mechanical engineering industry knowledge graph, and Wer Leitet knowledge graph.
- **Software offers.** The employees of the FDZ developed various software for RDM, data science, and AI, which are mainly hosted at GitHub organization of UB Mannheim: <https://github.com/UB-Mannheim>. For example, there are repositories with chatbots for RDM and open science: FAIR GPT [SS24] and FAIRplexica [Sc25].
- **Informational resources about RDM.** We provide useful information about RDM at the FDZ's homepage (<https://www.bib.uni-mannheim.de/lehren-und-forschen/forschungsdatenzentrum>) and in the special GitHub repository “awesome RDM” (<https://github.com/UB-Mannheim/awesome-RDM>).
- **Secure environment.** The Data Cube at UB Mannheim provides a secure on-site environment for sharing and analyzing sensitive personal, financial, and business data, ensuring privacy and compliance.
- **Research data infrastructures.** We offer researchers access to: a) Mannheim Data Repository MADATA for depositing their data and replication packages in accordance with the FAIR principles (<https://madata.bib.uni-mannheim.de>), b) the platform MaObjects for creating and presenting digital object collections and virtual exhibitions, c) an RDMO-instance MaDataPlan for planning, organisation and implementation of RDM (<https://fdz.bib.uni-mannheim.de/madataplan>), and d) Mannheim knowledge graph infrastructure for creating and hosting knowledge graphs (<https://kgi.uni-mannheim.de>).

Our data-driven approach (based on usage statistics, consultancy documentation, and continuous feedback) guides the refinement and prioritization of services, ensuring they remain relevant. Core services have been supported through internal funding, while many

innovative offerings have been realized through project-based external funding.

The initial selection of methods for our consultancy and data-science-as-a-service offers was historically driven, as the first digitization projects at UB Mannheim required OCR algorithms. Over time, we built strong expertise in this area, which naturally expanded into further OCR-based projects and services. Recently, due to increased demand in other areas, we extended our offerings to include other data collection methods (audio/video transcription and web scraping) as well as large language models (LLMs) and chatbots.

This iterative development process also benefits from strong collaborations both within and beyond the university, which will be further detailed in the next section.

3 Building key partnerships

At the FDZ UB Mannheim, our strength lies in collaborating across multiple levels: within the library, across the University of Mannheim, with external institutions, and through the national research data infrastructure (NFDI). Each of these partnerships is crucial to our success.

We would not be able to manage our dynamically evolving and innovative services without the continuous support of the UB IT department and subject librarians, our university partners, numerous projects, and collaborations within the NFDI. These strong and ongoing partnerships enable us to remain responsive and adaptable to the emerging needs of the research community.

1. Internal Library Partners.
 - a) UB IT services support us to design and maintain the technical infrastructure.
 - b) Subject librarians promote our services to researchers.
2. University-Level Collaborations.
 - a) The FDZ is part of several university-wide committees that develop policies and strategies relating to RDM, data science, and AI. This helps to build strong interdisciplinary partnerships and align the data science and RDM services with university-wide policies.
 - b) Mannheim Center for Data Science (MCDS). As recurring presenters at MCDS's "academic speed-dating" events, we introduce FDZ services to the broader research community.
 - c) Zentrum für Lehren und Lernen (ZLL). We present FDZ services at various ZLL events, including the "KI in der Hochschullehre," where we introduced our poster "AI Services at FDZ UB Mannheim."
 - d) Researchers who used our services.

3. Project-Based Partnerships.

- a) BERD@NFDI is a NFDI consortium for business, economic, and related data with multiple partners across Germany [Ge22, Kr24]. The FDZ contributes to OCR, legal topics, education and training, and knowledge graphs. We were also involved in its pilot project BERD@BW [GSS22].
- b) MaDaLi² is a project for creating cross-disciplinary e-learning courses and certificate programmes on AI and data literacy in cooperation with ZLL and departmental partners of the University of Mannheim [Br25].
- c) KGI4NFDI is a basic NFDI service for creating knowledge graphs together with TIB, ZB MED, GESIS, FIZ Karlsruhe, and ZBW [Ro24, Sa25]. The FDZ contributes to running consultancy sessions and developing guidelines.
- d) TransforMA is a project on transferring technologies and knowledge in Mannheim region together with the Technical University of Applied Science Mannheim. The FDZ contributes with automatic subject indexing, OCR, metadata extraction, and creation of chatbots.
- e) ENGAGE.EU is a project and alliance of leading European universities. The FDZ contributes its expertise with training courses on RDM and is part of an expert group on RDM.
- f) GerHisFin is an Akademie der Wissenschaften project focusing on historical financial data, conducted in collaboration with the Leibniz Institute for Financial Research SAFE. The FDZ is responsible for digitising analogue sources, performing OCR and creating KGs from structured databases.
- g) MARCIE is a project on data literacy training. The FDZ supports the Mannheim Research Group in Culture, Research and Entrepreneurship (MARCIE) by creating a video tutorial on cultural data literacy.

4. Collaborations within NFDI.

- a) ELSA Section (Ethics, Legal & Social Aspects): An FDZ team member serves as a section speaker.
- b) Working Group “Knowledge Graphs” (WG KGs): An FDZ team member serves as a WG coordinator.

The way we integrate partner offers depends on the nature of the collaboration. When partners provide services or materials that are relevant and valuable to our communities, we actively refer to them in our information materials and consultancy sessions. In broader collaborations, such as projects or NFDI consortia, we often build on existing ideas and, where appropriate, implement similar services ourselves. This allows us to consider our institutional context and strengthen in-house competences. By aligning our efforts with those of our partners, we ensure complementarity and avoid unnecessary redundancies.

Through this multi-layered network of partnerships, the FDZ not only amplifies its capacity to serve Mannheim researchers but also benefits from the wider academic data science and RDM ecosystem.

4 Sustainability plan

Ensuring the long-term sustainability of integrated data science, AI, and RDM services at FDZ UB Mannheim is a multi-faceted challenge. Our sustainability plan focuses on four key areas.

1. **Infrastructure development.** To meet the growing demand for data science and AI services, we aim to upgrade our local GPU cluster and cooperate with the University of Mannheim's IT department (UNIT) to enable access to cloud-based GPU resources. A key challenge lies in the fact that GPU infrastructure falls primarily under the responsibility of the central IT department, and we are only one of several stakeholders competing for limited resources. As a result, our specific needs are not always addressed with the necessary speed or priority. If this situation persists, we will need to further invest in local infrastructure and secure funding to ensure our services can scale in line with demand.
2. **Staff development.** To maintain high-quality services, we encourage further education and training, staff collaboration, and mentoring for newcomers. One key challenge is the breadth of our service portfolio, which spans data science, AI, and research data management. This requires team members to develop skills across multiple domains, which can be demanding. To address this, we encourage targeted upskilling based on individual strengths and project (or service) needs. One of our strategic approaches to staff development is "learning by doing." When identifying emerging topics that may be relevant to our research communities, we often assign team members to explore these areas even without prior in-depth knowledge. They dive into the topic, develop expertise through hands-on exploration, and then translate their insights into a course, seminar, or service offering. This approach not only accelerates internal capacity-building but also ensures that our services remain responsive to current developments and grounded in real user needs.
3. **Community engagement.** To improve community engagement, we continuously adapt our outreach program, strengthen partnerships within and outside the university, gather user feedback, and adapt our scope of offers. Engaging researchers in RDM remains a significant challenge. To address this, we have expanded our portfolio to include AI and data science topics, which tend to attract broader interest. For example, while each seminar in our four-day FAIR Data Week (covering the Findable [Sh23a], Accessible [Sh23b], Interoperable [Sh23c], and Reusable [Sh23d] aspects of the FAIR principles) drew fewer than 10 participants, our seminar "ChatGPT for FAIR Research Data" attracted nearly 90 [Sh24]. The associated open educational resources published on Zenodo were viewed around

200 times for each FAIR seminar, compared to over 1000 views for the ChatGPT-based RDM session. This shift illustrates the potential of integrating AI topics to improve engagement with RDM practices.

4. **Long-term funding.** We actively pursue both external and internal funding, seek collaborative project opportunities, and explore models for monetizing advanced data science services where applicable. A key challenge is the reliance on short-term project-based funding, which can limit strategic continuity. To mitigate this, we particularly aim to secure long-term projects. For example, our involvement in BERD@NFDI ensures five years of funding, while the GerHisFin project is planned for an 18-year duration. These longer funding horizons enable us to plan strategically and build sustainable services.

By focusing on infrastructure, human capital, community engagement, and diversified funding, the FDZ UB Mannheim is positioned to sustain and expand its data science, AI, and RDM services.

5 Conclusions

In this paper, we have addressed the three primary challenges faced by academic libraries in developing sustainable RDM, data science, and AI services (defining a scope, building partnerships, and ensuring sustainability) as follows.

Scope. We defined and demonstrated the expanding scope of the FDZ's offerings by detailing our integrated service portfolio, which includes data-science-as-a-service, education and training, consultancy, data and software provision, secure on-site environments, and research data infrastructures. The scope is not fixed, but flexible and evolving.

Partnerships. We outlined our multi-layered collaborations: internally with UB IT and subject librarians; across the University of Mannheim with MCDS and faculty research teams; externally with peer institutions and NFDI consortia; and through multiple project-based partnerships.

Sustainability. We presented a sustainability plan focused on upgrading infrastructure (GPUs), investing in staff development (training and mentoring), strengthening community engagement and feedback loops, and securing diversified funding (through external grants and institutional support).

By systematically tackling these three challenges, the FDZ at UB Mannheim is integrating RDM, data science, and AI services and is prepared to adapt them to evolving scholarly needs and technological advances.

Acknowledgements

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

Bibliography

- [Br25] Bröcher, L. et al.: Optimierung eines E-Learning-Angebots durch Studierenden-Interviews und KI-Tools. In (Berend B.; Voss H.-P.; Wildt J. eds.): *Neues Handbuch Hochschullehre: Lehren und Lernen effizient gestalten*. Band 119. Franz Steiner Verlag, pp. 83–104, 2025.
- [Ge22] Gehrlein, S. et al.: BERD@NFDI für unstrukturierte Daten in den Wirtschafts- und Sozialwissenschaften. *Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB)* 69(1-2), 59–67, 2022. <http://dx.doi.org/10.3196/1864295020691280>
- [GSS22] Gehrlein, S.; Schumm, I.; Shigapov, R.: BERD@BW – A Science Data Center to foster Open Science in Business, Economics and Social Sciences. In (Heuveline, V.; Bisheh, N.; Kling, P. eds.): *E-Science-Tage 2021: Share Your Research Data*. heiBOOKS, Heidelberg, pp. 314–319, 2022. <https://doi.org/10.11588/heibooks.979.c13743>
- [He21a] Herndon, J. (ed.): *Data Science in the Library: Tools and Strategies for Supporting Data-Driven Research and Instruction*. Facet, 2021. <https://doi.org/10.29085/9781783304615>
- [He21b] Herndon, J.: Designing for Data Science: Planning for Library Data Services. In (Herndon, J. ed.): *Data Science in the Library: Tools and Strategies for Supporting Data-Driven Research and Instruction*. Facet, pp. 131–140, 2021. <https://doi.org/10.29085/9781783304615.009>
- [Ma21] Mani, N. S. et al.: Creating a Data Science Framework: A Model for Academic Research Libraries. *Journal of Library Administration* 61(3), 281–300, 2021. <https://doi.org/10.1080/01930826.2021.1883366>
- [Kr24] Krieger, U.; Shigapov, R.: Die BERD@NFDI Plattform: Eine Forschungsinfrastruktur für unstrukturierte Daten in den Sozialwissenschaften. In: 112. BiblioCon 2024, Hamburg, Germany, 2024. <https://nbn-resolving.org/urn:nbn:de:0290-opus4-191214>
- [Ol19] Oliver, J. C. et al.: Data Science Support at the Academic Library. *Journal of Library Administration* 59(3), 241–257, 2019. <https://doi.org/10.1080/01930826.2019.1583015>
- [PBB21] Parker, M. S.; Burgess, A. E.; Bourne, P. E.: Ten simple rules for starting (and sustaining) an academic data science initiative. *PLOS Computational Biology* 17(2), e1008628, 2021. <https://doi.org/10.1371/journal.pcbi.1008628>
- [Ro24] Rossenova, L. et al.: KGI4NFDI – Knowledge Graph Infrastructure for the German National Research Data Infrastructure. Zenodo. <https://doi.org/10.5281/zenodo.13118749>
- [Sa25] Sauerwein, T. et al.: KGI4NFDI – Knowledge Graph Infrastructure for the German National Research Data Infrastructure (Proposal for the Integration phase of a Basic Service in Base4NFDI). Zenodo, 2025. <https://doi.org/10.5281/zenodo.15088676>

- [Sc23] Schimmler, S. et al.: NFDI4DS Infrastructure and Services. INFORMATIK 2023 – Designing Futures: Zukünfte gestalten. Bonn: Gesellschaft für Informatik e.V., pp. 919–924, 2023. https://doi.org/10.18420/inf2023_103
- [Sc25] Schmidt, T. et al.: Perplexity-inspired metasearch-based alternatives to FAIR GPT: Open-source AI consultants for RDM. In: E-Science-Tage 2025 “Research Data Management: Challenges in a Changing World”. Heidelberg, 2025.
- [Sh23a] Shigapov, R.: FAIR Data Week: F for Findable. FAIR Data Week, Online. Zenodo, 2023. <https://doi.org/10.5281/zenodo.7984881>
- [Sh23b] Shigapov, R.: FAIR Data Week: A for Accessible. FAIR Data Week, Online. Zenodo, 2023. <https://doi.org/10.5281/zenodo.7989605>
- [Sh23c] Shigapov, R.: FAIR Data Week: I for Interoperable. FAIR Data Week, Online. Zenodo, 2023. <https://doi.org/10.5281/zenodo.7993735>
- [Sh23d] Shigapov, R.: FAIR Data Week: R for Reusable. FAIR Data Week, Online. Zenodo, 2023. <https://doi.org/10.5281/zenodo.7997250>
- [Sh24] Shigapov, R.: ChatGPT for FAIR Research Data. Research Data Management Seminars at the University of Mannheim, Online. Zenodo, 2024. <https://doi.org/10.5281/zenodo.10664554>
- [SS24] Shigapov, R.; Schumm, I.: FAIR GPT: A virtual consultant for research data management in ChatGPT. arXiv preprint arXiv:2410.07108, 2024. <https://doi.org/10.48550/arXiv.2410.07108>
- [Sh25] Shigapov, R. et al.: Institutional Data Science Services at FDZ UB Mannheim: Enhancing Research Data Management. In: E-Science-Tage 2025 “Research Data Management: Challenges in a Changing World”. Heidelberg, 2025.
- [St21] Steinmann, L.; Drechsler, R.: Verzahnung von Data Stewardship und Data Science – Wege und Perspektiven. Bausteine Forschungsdatenmanagement Nr. 3/2021, S. 82–91, 2021. <https://doi.org/10.17192/bfdm.2021.3.8342>
- [St23a] Steinmann, L. et al.: Das Data Science Center an der Universität Bremen: Interdisziplinärer Knotenpunkt und Service-Infrastruktur für die datenintensive Forschung. In (Heuveline, V.; Bisheh, N.; Kling, P. eds.): E-Science-Tage 2023: Empower Your Research – Preserve Your Data. heiBOOKS, Heidelberg, pp. 246–252, 2023. <https://doi.org/10.11588/heibooks.1288.c18083>
- [St23b] Steinmann, L. et al.: Workshop: „Aktuelle Entwicklungen und Perspektiven (an Hochschulen) im Bereich Data Science“. INFORMATIK 2023 – Designing Futures: Zukünfte gestalten. Bonn: Gesellschaft für Informatik e.V., pp. 73–80, 2023. https://doi.org/10.18420/inf2023_05
- [WD23] Weiß, K.; Decker, R.: Data Science und Forschungsdaten als Treiber für wissenschaftliche Interdisziplinarität. INFORMATIK 2023 – Designing Futures: Zukünfte gestalten. Bonn: Gesellschaft für Informatik e.V., pp. 107–115, 2023. https://doi.org/10.18420/inf2023_09