UNIVERSITY OF MANNHEIM

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM



The **University Library** is offering a full-time position as

Data Scientist (all genders)

starting as soon as possible.

The position is offered as a fixed-term appointment for 24 months within the project "**FAIR Corpus: A curated dataset for training and evaluating AI-Assistants for Research Data Management and Open Science**".

The FAIR Corpus project is aimed to build the first openly licensed, high-quality dataset for training and evaluating AI-Assistants specifically focused on Research Data Management (RDM) and Open Science (OS). This corpus will contain thousands of curated documents, including data management plans, policies, training materials, standards, and open-access publications. It will be enriched with metadata and formatted for use in AI training, fine-tuning, evaluation, and retrieval-augmented generation. The corpus will serve as a foundational resource for developing trustworthy, domain-specific AI assistants for RDM and OS across universities, research data centers, national and international research infrastructures.
The project's working languages are English and German.

Your role:
As part of the project, you will contribute to the development of a high-quality, openly reusable dataset for AI applications in Research Data Management and Open Science. Your work will focus on the automated processing and quality assurance of documents, as well as the development and evaluation of AI-ready datasets for Retrieval-Augmented Generation (RAG) and Question Answering (QA).

Your tasks include:

- Design and implementation of the FAIR Corpus, including the selection of suitable data sources, the development of data and metadata models, and the application of FAIR and CARE principles while taking legal, data protection, and ethical requirements into account
- Development of automated methods for acquiring, processing, cleaning, and quality-assuring documents and metadata
- Creation of a RAG subcorpus, including the development and testing of suitable text segmentation and retrieval methods
- Development of a QA subcorpus using LLM-assisted generation of question–answer pairs and appropriate quality assurance procedures

The **University of Mannheim** is one of the leading universities in Germany with approximately 12,000 students in five schools. Particularly in business and economics as well as in the social sciences the university ranks among the top institutions both at national and international level. More than 2,600 employees profit from an exciting work environment with numerous benefits.

The University Library is the central information infrastructure unit of the University of Mannheim and supports student learning, teaching, and research with a wide range of services.

Key information

Start: at the earliest possible date

Term: 24 Month

Remuneration bracket: E 13 TV-L BW

Weekly hours: 39,5 hours,
the position is generally suitable for job sharing

Location:
Mannheim, City Center, Schloss

Application deadline:
29 November 2026

Reference number: UB-512

- Design and conduction of benchmarks to evaluate retrieval and question-answering performance
- Documentation and provision of reproducible data processing and evaluation of workflows, as well as presentation of project results

Your profile:

Required qualifications:

- A completed university degree (Master's degree or equivalent) in Computer Science, Data Science, Computational Linguistics or a related field
- Good Python programming skills and experience in processing text and metadata
- Knowledge of Natural Language Processing (NLP) and experience working with Large Language Models (LLMs)
- Interest in developing and evaluating AI-based methods, as well as in Research Data Management and Open Science
- Ability to work independently, systematically and collaboratively as part of a team
- Very good command of both German and English

Desirable qualifications:

- Experience with Retrieval-Augmented Generation (RAG), Information Retrieval, or Question Answering (QA)
- Knowledge of automated data pipeline development, data validation and quality assurance
- Experience with LLM-assisted data generation, prompt engineering or the evaluation of AI systems
- Knowledge of Research Data Management, Open Science and the FAIR and CARE principles
- Experience with Git, reproducible workflows and software development
- Experience in scientific communication and presenting research findings

What we offer:

- A challenging and varied role in a university environment with national and international partners
- A workplace in the representative Baroque Palace in the middle of the university campus; downtown location, immediate access to public transport, 5 minutes to the main train station
- Flexible working hours model, the possibility of teleworking (up to 50%), 30 days of annual leave
- Opportunity to purchase the subsidized Deutschlandticket Job
- An annual salary according to TV-L E13 and the usual social benefits and advantages of public service, like a company pension plan (VBL)

Data protection

Please find detailed information on the collection of personal data from the data subject according to Article 13 GDPR on the university' s homepage: www.uni-mannheim.de/datenschutz-bei-bewerbungen.

Submitted application documents will only be returned if you enclose a self-addressed stamped envelope. Otherwise, they will be destroyed in accordance with current data protection law after the application procedure has ended. Electronic applications will be deleted accordingly..

If you apply by e-mail, please note that protection of confidential data cannot be guaranteed as unauthorized third parties might gain access to unencrypted e-mails during transmission.



Innovative academic working environment



Flexible work time and workplace models (flexitime, telework)



Family-friendly higher education institution



International orientation



Deutschlandticket



Sports program for all staff members

Applications from persons with a disability are given preferential consideration provided that they are appropriately qualified. The University of Mannheim is committed to increasing the quota of women in areas where they are underrepresented and thus encourages women with appropriate qualifications to apply.

We are looking forward to your application using our University of Mannheim's job portal by 29 November 2026 at the latest.

If you have any questions about the position, please contact Dr. Irene Schumm (Head of the Research Data Center Department) at

irene.schumm@uni-mannheim.de



