MEDICAL INFORMATICS INITIATIVE GERMANY

# GeMTeX

## GeMTeX – German Medical Text Corpus

### AUTOMATIC INDEXING OF MEDICAL TEXTS FOR RESEARCH

Medical texts from routine care contain large amounts of complex and unstructured data, such as disease progression, diagnoses, and treatments. This data can be very useful for research and patient care. However, the structure and content of clinical documentation texts often vary widely between institutions. This makes them difficult to use for automatic natural language processing (NLP), which is the basis for all automation processes and analysis. Due to the lack of standardization of medical free texts, the potential of this treasure trove of data for research and healthcare cannot be fully exploited. This is where the GeMTeX methodology platform comes in.

### A LARGE COLLECTION OF GERMAN MEDICAL TEXTS

The GeMTeX methodology platform is a cross-consortium project of the Medical Informatics Initiative (MII) with the goal of making medical texts from patient care, such as doctors' letters or discharge summaries, accessible for research projects. The collaboration of the MII consortia DIFUTURE, HiGHmed, MIRACUM and SMITH aims to create the largest medical text corpus in the German language. The project is coordinated by the SMITH Consortium office.

With the consent of the patients, documents are collected from the electronic health records (EHRs) of the six university medical centers in Munich, Leipzig, Essen, Berlin, Dresden and Erlangen.

The documents are then processed using natural language processing methods and made available in anonymized form for research purposes. The resulting database can be used, for example, to train AI models and test them in everyday clinical practice.

© istockphoto.com/wutwhanfoto

### GOALS

- → Creating a broad database for medical research projects and AI models with the goal of clinical application.
- → Creating the largest corpus of medical texts in the German language.
- → Preparing texts from routine patient care in a machine-readable format and making them available for research.
- → Establishing technical and organizational standards for the representation and structuring of medical texts.
- → Expanding the core data set of the Medical Informatics Initiative (MII).

The GeMTeX Use Case started on 1 June 2023 and is funded by the German Federal Ministry of Education and Research (BMBF) with around seven million euros until 31 August 2026.

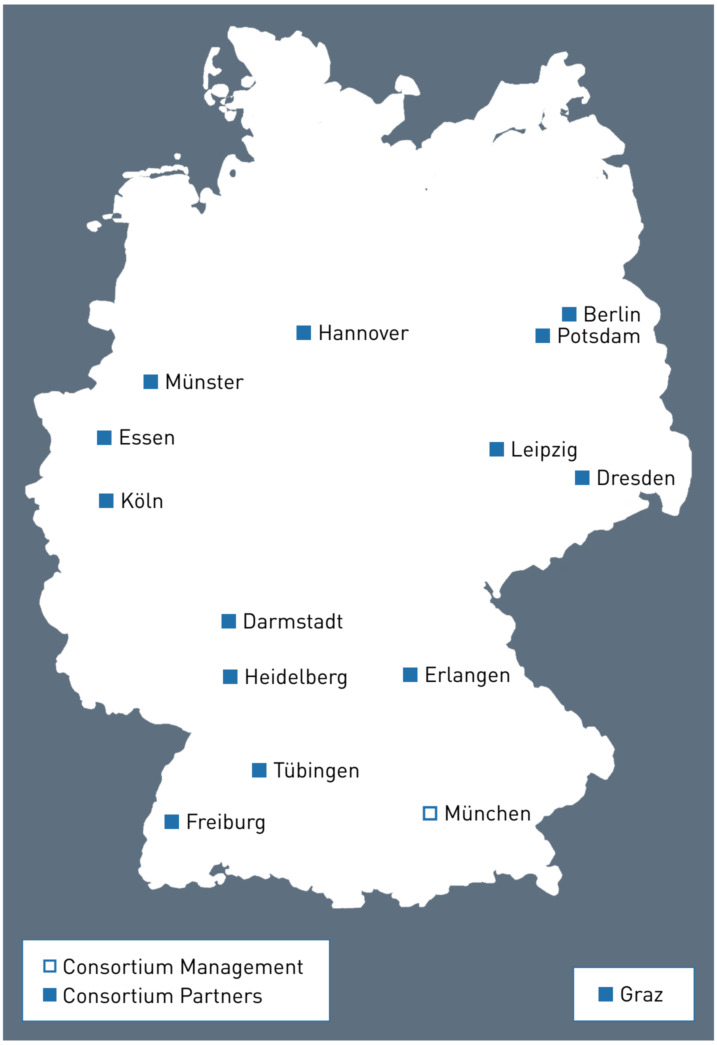
MEDICAL INFORMATICS INITIATIVE GERMANY

SPONSORED BY THE
Federal Ministry
of Education
and Research

→ **Further informations: https://www.smith.care/en/gemtex_mii/about-gemtex/**

MEDICAL
INFORMATICS
INITIATIVE
GERMANY

# GeMTeX



## Project Partners

### CONSORTIUM MANAGEMENT

**Munich**
Technical University of Munich

### CONSORTIUM PARTNERS

**Berlin**
- Charité – University Hospital Berlin
- ID GmbH & Co. KGaA

**Darmstadt**
- Technical University of Darmstadt

**Dresden**
- Dresden University of Technology

**Erlangen**
- University Hospital Erlangen

**Essen**
- University Hospital Essen

**Freiburg**
- Averbis GmbH

**Hannover**
- Hannover Medical School

**Heidelberg**
- Heidelberg University Hospital

**Cologne**
- German National Library of Medicine (ZB MED)

**Leipzig**
Leipzig University / University of Leipzig Medical Center

**Munich**
- Ludwig Maximilian University of Munich

**Münster**
- University of Münster

**Potsdam**
- Hasso Plattner Institute for Digital Engineering

**Tübingen**
- Tübingen University Hospital

### ASSOCIATED PARTNER

**Graz**
- Medical University of Graz

### COORDINATION OFFICE

**Berlin**
- TMF e. V. Head Office

DIFUTURE
Data Integration for Future Medicine

HiGHmed
Medical Informatics

miracum

SMITH
Smart Medical Information Technology for Healthcare

## Contact

### NETWORK COORDINATION

Prof. Dr. Martin Boeker
Network coordinator
Head of the DIFUTURE Consortium

Technical University of Munich /
University Hospital rechts der Isar

### PROJECT COORDINATION

Leipzig University | Faculty of Medicine
SMITH Office
Philipp-Rosenthal-Straße 27
04103 Leipzig

Phone: +49 341 97-16720
E-Mail: info@smith.care
Web: www.smith.care

Status: February 2024