UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Research data integration for AI development: process and learnings [Abstract]

Dominik Müller, Frank Kramer

Angaben zur Veröffentlichung / Publication details:

Müller, Dominik, and Frank Kramer. 2026. "Research data integration for AI development: process and learnings [Abstract]." *Biomedical Engineering / Biomedizinische Technik* 71 (s1). https://doi.org/10.1515/bmt-2026-1001.

Nutzungsbedingungen / Terms of use:

CC BY 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
CC-BY 4.0: Creative Commons: Namensnennung
Weitere Informationen finden Sie unter: / For more information see:
https://creativecommons.org/licenses/by/4.0/deed.de

Research Data Integration for AI Development: Process and Learnings

Dominik Müller, IT-Infrastructure for Translational Medical Research, University of Augsburg, Augsburg, Germany, dominik.mueller@uni-a.de
Frank Kramer, IT-Infrastructure for Translational Medical Research, University of Augsburg, Augsburg, Germany, frank.kramer@uni-a.de

Introduction

With AI models, the field of healthcare data analytics is advancing rapidly. However, these models often rely on limited public datasets. The organization of clinical studies including the acquisition as well as annotation of clinical data represents a time-consuming challenge, consuming significant efforts before the actual model development can begin. Therefore, we present insights into the current practical process of research data integration for medical AI at a German university hospital.

Methods

Based on multiple clinical AI projects conducted in collaboration with clinical partners, we identified and systematized a recurring ten-step process as a reference framework (order and overlap may vary by project): (0) Stakeholder finding and task definition, (1) ethics committee, (2) data protection compliance, (3) data access and query strategy, (4) data export from clinical systems (5) further data selection and quality control, (6) de-identification and pseudonymization, (7) data transfer to a secure analytics environment, (8) annotation service deployment and process, (9) final data structuring for processing, and (10) model development and application. Each step requires coordinated input from clinical experts and data scientists, covering responsibilities such as patient cohort identification, filtering, annotation, and local clinical infrastructure or workflows.

Results

National efforts like the Medical Informatics Initiative aim to standardize and streamline this process through Medical Data Integration Centers, which extract, harmonize, and provide clinical data upon approval by local Use and Access Committees. Still, such service-based research data integration remains largely unfeasible for complex modalities like high-dimensional clinical data, biosignals, and imaging.

Conclusion

There is a clear need for standardized protocols for research data integration at university hospitals. Particularly for complex clinical data, there is currently no alternative to the presented multi-step reference framework. Establishing robust institutional processes would substantially reduce the overhead required before medical AI development, enabling faster and more reproducible healthcare analytics research.

Open Access. © 2026 The Author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.