Demonstrating Study Discovery and Exploration Through a Knowledge Graph

Lea Gütebier[1,*], Dagmar Waltemath[1,2] and Ron Henkel[1]

[1]Medical Informatics Laboratory, Institute for Community Medicine, University Medicine Greifswald, Walther-Rathenau-Straße 48, D-17475 Greifswald, Germany
[2]Data Integration Center, University Medicine Greifswald, Walther-Rathenau-Straße 49a 17475 Greifswald

Abstract

Efficiently finding and contextualizing clinical studies is a critical challenge for clinician scientists and also impacts patients. To improve the access to information about clinical studies, we developed a knowledge graph that integrates diverse resources, including information from ClinicalTrials.gov and the a German Portal for Medical Data Models. We enrich the knowledge graph with entries from biomedical terminologies such as the Unified Medical Language System (UMLS). The knowledge graph facilitates enhanced interoperability and complex query capabilities, enables targeted searches and identification of study similarities, and allows scientists to efficiently explore study eligibility criteria. Through demonstration queries, we illustrate these abilities.

Keywords

Knowledge Graph, Clinical Studies, Unified Medical Language System, FAIR Principles, Study Discovery and Exploration

1. Introduction

Locating and accessing relevant studies is an essential process that benefits a wide range of stakeholders, including patients, scientists, and clinicians [1]. Patients are enabled to locate clinical studies with matching inclusion criteria [2]. Scientists and clinicians benefit from more successful recruitment, a comprehensive design of study protocols, a coherent gathering of evidence for interventions, and the contextualization of research findings. The efficiency of these processes plays a key role in advancing medical knowledge and improving resource utilization.

Despite its importance, finding relevant studies is often a tedious and time-consuming endeavour [3]. The complexity of medical databases, the volume of published research, and the variability in metadata quality contribute to the laborious nature of this task [4], underscoring the need for more efficient and accessible solutions as required by the FAIR principles [5].

2. The knowledge graph

To improve the findability of clinical studies and support the exploration of study information, we developed a knowledge graph. It is based on data from ClinicalTrials.gov [6], the MDM Portal [7], and standardized terminologies such as the Unified Medical Language System (UMLS) [8].

Integration of these datasets into a single knowledge graph results in a unified representation of study information. The semantic enrichment provided by UMLS and related terminologies allows for complex queries and gives additional insights, transforming the way studies are searched and contextualized.

In this demonstration, we showcase the potential of the knowledge graph through example queries that highlight its capability to improve study findability and accessibility.

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025
*Corresponding author.
lea.guetebier@uni-greifswald.de (L. Gütebier); dagmar.waltemath@uni-greifswald.de (D. Waltemath); ron.henkel@uni-greifswald.de (R. Henkel)
0000-0001-5504-5108 (L. Gütebier); 0000-0002-5886-5563 (D. Waltemath); 0000-0001-6211-2719 (R. Henkel)

3. Querying the knowledge graph

The demonstration highlights the practical applications of the knowledge graph through several use cases that underscore its utility in exploring study data. The hierarchical structure and cross-references of ontologies enable researchers to identify studies of interest efficiently.

Example 1: The structured eligibility criteria in the graph allow for targeted exploration of linked data, such as locating studies with specific inclusion criteria that exclude particular patient groups.

Example 2: The knowledge graph leverages ontologies to uncover studies with similar features, including shared conditions, comparable outcome measures, and analogous study designs. This capability supports nuanced queries that go beyond simple keyword searches, providing deeper insights into the relationships between studies.

Example 3: Another valuable use case involves identifying studies relevant to a patient's disease or pathology. By incorporating data about recruitment centers and their geographic proximity to the patient, the graph facilitates the discovery of studies that are still actively recruiting and conveniently located.

Graph algorithms further enhance the graph's functionality by enabling complex queries. These algorithms allow for sophisticated analyses, such as identifying hidden relationships within the data, evaluating study networks, and uncovering patterns that would be challenging to discern through traditional methods.

4. Conclusion

The presented knowledge graph significantly contributes to enhancing the findability and accessibility of clinical and research studies. By integrating diverse resources, the graph simplifies the exploration of studies, enabling researchers and clinicians to perform more targeted and efficient searches.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] K. A. Robinson, S. N. Goodman, A systematic examination of the citation of prior research in reports of randomized, controlled trials, Annals of internal medicine 154 (2011) 50–55.

[2] Q. Su, G. Cheng, J. Huang, A review of research on eligibility criteria for clinical trials, Clinical and experimental medicine 23 (2023) 1867–1879.

[3] I. Pigeot, W. Ahrens, J. Darms, J. Fluck, M. Golebiewski, H. K. Hahn, X. Hu, T. Intemann, E. Kasbohm, T. Kirsten, et al., Making Epidemiological and Clinical Studies FAIR Using the Example of COVID-19, Datenbank-Spektrum (2024) 1–12.

[4] C. Chen, Z. Li, H. Huang, B. E. Suzek, C. H. Wu, U. Consortium, A fast peptide match service for UniProt knowledgebase, Bioinformatics 29 (2013) 2808–2809.

[5] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[6] D. A. Zarin, T. Tse, R. J. Williams, R. M. Califf, N. C. Ide, The ClinicalTrials. gov results database—update and key issues, New England Journal of Medicine 364 (2011) 852–860.

[7] S. Riepenhausen, M. Blumenstock, C. Niklas, S. Hegselmann, P. Neuhaus, A. Meidt, C. Püttmann, M. Storck, M. Ganzinger, J. Varghese, et al., Europe's Largest Research Infrastructure for Curated Medical Data Models with Semantic Annotations, Methods of Information in Medicine (2024).

[8] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, Nucleic acids research 32 (2004) D267–D270.