Building OHDSI with Privacy Computing in Shanghai Medical College, Fudan University

Changran Wang[12], Lei Liu[12], Feizhen Wu[1], Lin Li[1]
[1]Intelligent Medicine Institute, Shanghai Medical School, Fudan University
[2]OHDSI CHINA Coordinate Center

Background

The continuous advancement of global medical informatization has resulted in vast amounts of health data, reaching exabyte and zettabyte scales. However, these data are scattered across various institutions, hindering their orderly circulation. The integration of artificial intelligence (AI) with healthcare is emerging as a key driver in transforming medical technology. Health data are essential for clinical research, and multi-center research institutions that can securely aggregate data from various sources demonstrate superior efficiency. This approach facilitates the collection of extensive datasets, enabling deeper and more comprehensive data utilization.

Shanghai Medical College is positioned to seize this historic opportunity by participating in the pilot construction of high-level local universities in Shanghai. The institution aims to leverage AI in healthcare to enhance its strengths and address its weaknesses. By fostering innovation in clinical research and promoting multi-center research collaborations, Shanghai Medical College seeks to establish a robust multi-party data collaboration model, advancing clinical medical research. This integration will promote interdisciplinary cooperation and scientific innovation, accelerating the overall development of medical disciplines. Committed to becoming a top-tier domestic and world-class medical school, Shanghai Medical College will significantly contribute to public health improvements and the advancement of a healthy China and a healthy Shanghai.

Methods

Determine the quality management content of multi-center clinical research medical data, and confirm the multi-center clinical research medical data collection process and quality control content through literature review and expert consultation. On the basis of the investigation of the medical data quality management system, a standard framework for the construction of a multi-center medical data platform was constructed, and a multi-center clinical research data quality management system was established from multiple dimensions such as operating procedures, information collection, and quality control.

Based on the core technologies of privacy computing (federated learning, secure sandbox, multi-party secure computing, etc.), the privacy computing engine (PCP) is used to ensure the safe flow of data. Referring to the experience of multi-center clinical research projects and guided

by the OHDSI-OMOP model, a multi-center clinical research approach is proposed, which closely combines the project lead and participating units, bringing together multi-party research data on the platform for joint application in research can increase the dimension and breadth of clinical research data, and relies on the project to build a data platform Cooperation in the cultivation of medical talents and the research and promotion of new technologies.

This project builds a clinical data collaborate platform (CDCP), improves data collection and governance capabilities, system design is shown as figure 1. According to the cooperation mode of the OHDSI model, create a unified medical terminology system by using web Protégé, accelerate the process of medical data circulation and application, and empower clinical research cooperation and medical data sharing with advanced technology. We also create a data security sharing mechanism, and improves the integration of data resources. Provide relevant platform support for "building multiple high-level multi-center clinical medical research projects".
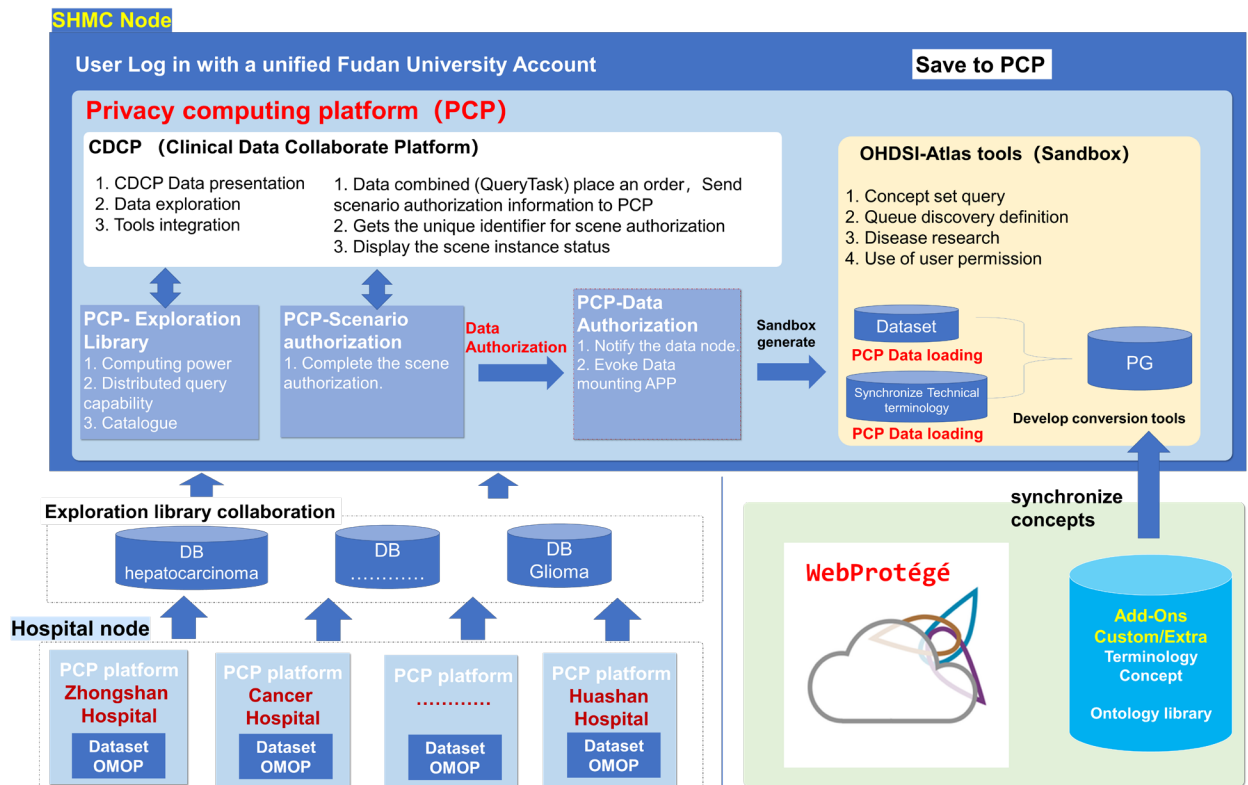

Figure1. CDCP system design

Results

As a crucial support platform for multi-center clinical research, aligned with the actual needs of Shanghai Medical College at Fudan University and in accordance with relevant laws and

regulations, this initiative provides robust medical data management support for multi-center research projects that adhere to medical ethics and pertinent regulations. A basic privacy computing platform is established within each hospital to host distributed databases, which undergo ETL processes to conform to the OHDSI format. The dataflow is shown as Figure 2.
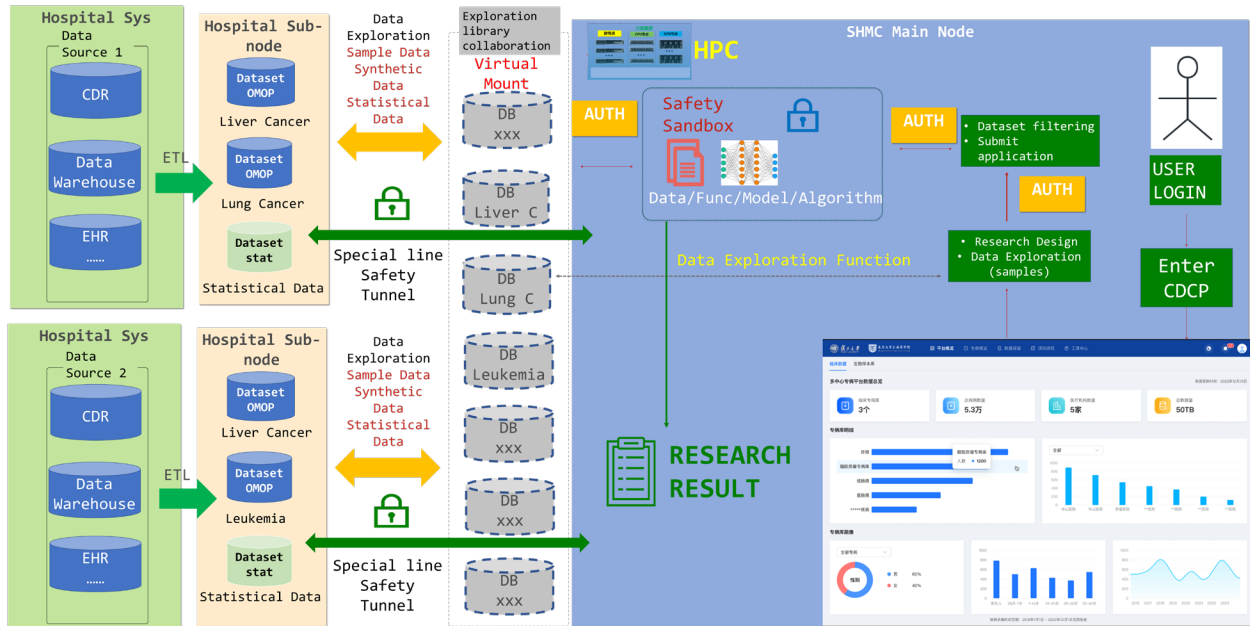

Figure 2. Dataflow

This platform empowers multicenter clinical research at Fudan University. The fusion of multi-center data offers several key advantages to the medical research efforts at Shanghai Medical College. In recent years, multi-center clinical research has been increasingly conducted across various disease fields, with multiple research units and researchers collaboratively executing work based on the same design and objectives. These studies encompass clinical drug trials and more generalized clinical investigations, including prospective and retrospective studies.

The sandbox environment created on the main node HPC cluster allows researchers to access necessary data without viewing actual patient details, thus maintaining data privacy. This invisible data access, coupled with the ability to utilize OHDSI analytic tools, provides significant advantages. Researchers can perform complex analyses and derive insights without compromising patient confidentiality. This capability not only enhances the efficiency and scope of clinical research but also fosters collaboration across institutions, driving forward medical innovation and improving patient outcomes.

Conclusion

In response to the needs of Shanghai Medical College at Fudan University and in compliance with relevant laws and medical ethics, a robust support platform for multi-center clinical research has been established. This platform features a basic privacy computing infrastructure within hospitals to host distributed databases conforming to the OHDSI format, facilitating the secure management of multi-center research data. Currently, the system comprises three sub nodes, two disease categories, and four distributed databases.

The fusion of multi-center data within this framework offers significant advantages for clinical research, including enhanced data volume and diversity, which are crucial for comprehensive medical studies. Researchers can access necessary data through a secure sandbox environment on the main node HPC cluster, ensuring patient privacy while allowing the use of advanced OHDSI analytic tools. This approach has already proven beneficial in recent studies, which encompass clinical drug trials and broader clinical investigations. The ability to perform complex analyses without compromising patient confidentiality enhances the efficiency and scope of clinical research, fostering collaboration across institutions and driving medical innovation.

Currently, the platform is focused on three sub nodes and two disease categories, utilizing four distributed databases. However, the vision for the future is expansive. Plans are underway to increase the number of sub nodes to six, significantly broadening the data and research capabilities of the platform. Additionally, there is a strategic initiative to pilot collaborations with overseas institutions and other research entities, leveraging main nodes to conduct multi-center clinical research on a global scale.

This planned expansion will further enhance the platform's capacity for high-level research, fostering greater collaboration and innovation in the field of clinical medicine. By integrating advanced privacy computing technologies and adhering to stringent ethical standards, the platform aims to set a new benchmark in multi-center clinical research, ultimately contributing to improved patient outcomes and the advancement of medical science. This comprehensive approach ensures that the platform will remain at the forefront of clinical research, continually evolving to meet the growing needs of the medical community.