DataTools4Heart: A European Health Data Toolbox for Enhancing Cardiology Data Interoperability, Reusability and Privacy

Description

Cardiovascular disease (CVD) remains the leading cause of mortality worldwide, accounting for about a third of annual deaths. Refusing structured and unstructured data has the potential for significant health benefits for the population suffering from CVD. Healthcare data reuse in Europe faces privacy and fragmentation issues, high data formats and language diversity, and a lack of technical and clinical interoperability.

DataTools4Heart (DT4H) will tackle such challenges and develop a comprehensive, federated, privacy-preserving cardiology data toolbox. This will include, in an integrated platform, standardised data ingestion and harmonisation tools providing a standard data model, multilingual natural language processing, federated machine learning, differentially private data synthesis generation, and seven language models adapted to the cardiology domain. DT4H virtual assistants will help scientists and clinicians navigate large-scale multi-source cardiology data. These tools will be:

i) implemented, ensuring privacy-by-design and thorough compliance with European regulations and data standards;

ii) optimised based on multi-stakeholder user-centred requirements, and

iii) validated in 7 clinical sites across Europe.

DT4H will unlock inaccessible health data in unstructured data and allow multi-site federated data use. Together with its toolbox, DT4H will leave the legacy of a federated learning platform with an embedded metadata catalogue, AI virtual assistants, and the CardioSynth open database of synthetic data remaining available for further research and AI experimentation. Effective use of the federated learning platform will improve and enable improved AI diagnostic and treatment tools. Deployment of regulated solutions will extend existing healthcare management paradigms to reduce the disease burden.

Finally, DT4H tools, systems and methodology are highly generalised and will translate well to other clinical and research areas in medicine. Computer sciences, information science and bioinformatics, Cardiovascular disease, AI diagnostic and treatment tools, Cardiology Data Interoperability, Reusability and Privacy, health care, heart, AI virtual assistants, medicine, Life Sciences

Barcelona Supercomputing Center - Centro Nacional de Supercomputación