



info@datatools4heart.eu



DataTools4Heart



DataTools4Heart is funded by the European Union's Horizon Europe Framework Under Grant Agreement No. 101057849

Vision

Cardiovascular disease is the main cause of mortality worldwide, accounting for about a third of annual deaths. Medical care produces a large amount of data every year which represent an opportunity to manage patients' health and for research.

DataTools4Heart will design solutions to allow the reuse of such data fostering research in the cardiovascular field.

The DataTools4Heart toolbox will be designed, tested, and deployed in several countries while ensuring the compliance of all legal (e.g., privacy) constraints.

Mission

DataTools4Heart aims to improve the research on patients suffering from cardiovascular diseases by developing a comprehensive, federated, privacy-preserving toolbox for data reuse in cardiology.

The tools include a platform that will securely access data from different hospitals and transform them into their digital counterparts (synthetic data) to be used by researchers and clinicians across the world.

The access into large-scale multi-source cardiology data will be facilitated by the creation of virtual assistants.

Objectives



Data ingestion and harmonisation

DataTools4Heart will develop a common data extraction tool to improve data interoperability while addressing their heterogeneity across European regions and cardiology units. This tool will be developed and validated through a modular and flexible Data Ingestion Suite which will be deployed in 7 European sites and guaranteed with at least 4 standard-based data models.



Natural Language Processing

DataTools4Heart will introduce a multilingual Natural Language Processing (NLP) suite to standardise the structuring of cardiology reports across Europe, including specific entity recognition and machine translation. Such suite will include adaptation of 7 language models to the cardiology domain in English, Spanish, Italian, Romanian, Czech, Swedish, and Dutch, using EHR data from clinical site partners.



Federated machine learning and data synthesis

By developing innovative methods for synthesising data, DataTools4Heart will build a privacy-preserving cardiology data toolbox to improve data reusability, while adhering to ethical/legal standards. Differentially private synthetic data will allow to handle data representative of a target population, scalable, shareable for research purposes, and able to reduce bias in algorithms. An open-source, synthetic dataset, CardioSynth, will be published and shared with the community.