Technical Specification

ISO/TS 9321

First edition
2024-06

Health informatics — General requirements of multi-centre medical data collaborative analysis

Informatique de santé — Exigences générales des analyses collaboratives multicentriques de données médicales

Reference number
ISO/TS 9321:2024(en)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Symbols and abbreviate terms **2**
5 General requirements **2**
5.1 General 2
5.2 Data isolation 2
5.3 Terminology standardization 2
5.4 Data standardization 2
5.5 Data incremental expansion 2
5.6 Distributed network framework 3
5.7 Network scalability 3
5.8 System modularization 3
5.9 Security 3
6 Architecture and workflow **3**
6.1 Architecture 3
6.2 Workflow 5
6.2.1 Multi-centre collaborative research initiation workflow 5
6.2.2 Multi-centre collaborative research implementation workflow 6
7 Functional framework and requirements **7**
7.1 Functional framework 7
7.1.1 General 7
7.1.2 User layer 8
7.1.3 Service layer 8
7.1.4 Resource layer 8
7.1.5 System security 8
7.2 Functional requirements 9
7.2.1 General 9
7.2.2 User layer functional requirements 9
7.2.3 Service layer functional requirements 10
7.2.4 Resource layer functional requirements 12
7.2.5 System security functional requirements 13
Annex A (informative) **Examples of multi-centre medical data collaborative analysis networks** **14**
Annex B (informative) **Reference implementations** **15**
Bibliography **18**

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 215, *Health informatics*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Medical data collaborative analysis across multiple medical centres plays an important role in enabling extensive and universally applicable outcomes in medical research. The establishment of multi-centre medical data collaborative analysis systems aims to enable researchers to securely and efficiently utilize medical data among medical centres.

This document introduces an architecture for multi-centre medical data collaborative analysis, which differs from the conventional centralized data analysis approach. Its purpose is to address various challenges encountered in current practices. These challenges include:

— ensuring that data analysis is performed under robust safety and privacy measures;

— handling data heterogeneity;

— maintaining consistency of research findings;

— implementing effective authority controls;

— meeting general service requirements.

The ultimate objective of this document is to foster trust among researchers and medical centres by implementing regulated data protections and standardized research processes. It aims to expedite the results obtained from collaborative analysis efforts of large-scale medical data.

ISO 29585[10] provides a framework for healthcare and data reporting, addressing both the opportunities and the responsibilities of the handling of the data, emphasizing the framework for data governance, privacy, security, acquisition, processing, loading and reporting. This document, on the other hand, places greater emphasis on the collaborative analysis of healthcare data and other requirements in multi-centre scenarios.

Specifically, this document presents a detailed scope, elucidates key concepts, outlines the resulting architecture, and provides comprehensive and standardized instructions to assist medical centres in establishing or participating in a robust and cohesive multi-centre medical data collaborative analysis system.

This document holds various potential applications, including:

— guiding developers to establish new medical data collaborative analysis systems;

— aiding technicians to seamlessly and securely integrate local medical resources into collaborative analysis systems;

— supporting supervisors to effectively manage the research processes;

— enabling physicians and medical researchers to conduct multi-centre medical data collaborative analysis;

— providing a fundamental set of functional requirements to ensure the essential functionality and security, while allowing for gradual enrichment of system features.

Health informatics — General requirements of multi-centre medical data collaborative analysis

1 Scope

This document outlines the general requirements for conducting a multi-centre medical data collaborative analysis, covering various aspects such as system architecture, data storage, data standardization, collaborative research management and security. The data considered in this standard primarily encompasses electronic health record data for multi-centre collaborative researches, including structured data, medical text data, image data, etc.

This standard is applicable to a wide range of individuals and institutions, including developers, maintainers, management personnel, researchers, and data-owning organizations.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 27799, *Health informatics — Information security management in health using ISO/IEC 27002*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
multi-centre
consortium of at least two medical centres

3.2
terminology base
collection of formally structured concepts and relationships serving as standardized expressions for specific entities with the capability to manage and maintain this collection for collaborative research

3.3
multi-centre medical data standardization
protocols and capabilities implemented across multiple centres to ensure the integrity and consistency of medical data

3.4
interoperability
ability of two or more systems or components to exchange information and use the information that has been exchanged

[SOURCE: ISO/TS 27790:2009, 3.39]

3.5
cohort
in observational studies, a group of individuals who share a common set of characteristics, such as age, sex, ethnicity

3.6
common data model
CDM
standardized, systematic approach to structuring and organizing data, ensuring that data from different sources or systems can be easily integrated, compared and analysed in a consistent manner

4 Symbols and abbreviate terms

CDM common data model

ETL extract-transform-load

API application programming interface

5 General requirements

5.1 General

Participants can conduct medical data collaborative analysis across multiple medical centres while preserving data privacy. Secure multi-party computing and federated learning methods can be employed to acquire statistical and meta-analysis results across multiple medical centres, as well as for model developments and applications.

The multi-centre medical data collaborative analysis system shall incorporate a collaboration network, function modules, and system security modules to facilitate diverse multi-centre collaborative researches. The original medical data shall remain within medical centres, adhering to ethical and legal considerations. The system shall provide robust security, strict confidentiality and exceptional reliability to researchers. It shall also aim to reduce the cost of medical data collaborative analysis and enhance the quality of research outcomes. Some existing systems are listed in Annex A. The general requirements are as detailed in 5.2 to 5.9.

5.2 Data isolation

The original medical data of each medical centre shall be stored securely within its internal database, ensuring it remains within local medical centre and shall not be transferred outside of local centres. Data isolation shall adhere to ISO 27799. Country-specific legal requirements can apply.

5.3 Terminology standardization

The medical centres shall use standard health terminology for local medical data. Detailed requirements regarding terminology are listed in 7.2.3.1. The medical centres shall have the capability to standardize their health terminology to ensure terminology consistency between medical centres.

5.4 Data standardization

The transformed data shall meet the requirements of format consistency, structural consistency and semantic consistency and be stored in a CDM.

5.5 Data incremental expansion

The transformed database of medical centres shall have the capability to automatically and incrementally import and store clinical data. The newly imported data shall not impact or conflict with the existing data.

5.6 Distributed network framework

All service modules associated with the original medical data shall be built within the local networks of each centre. The local medical centre can communicate with the coordination centre and other centres.

5.7 Network scalability

The network framework shall support the enrolment and withdrawal of new medical centres. The access of additional medical centres shall not impact or conflict with existing network connections.

5.8 System modularization

The system should be decomposed into multiple independent logical entities.

5.9 Security

The system shall implement robust security controls to ensure the reliability of confidentiality and integrity for medical data, encryption keys, intermediate and final results. The system's security requirements should align with ISO/IEC 27001[9], while the data security protection requirements should adhere to ISO 22857[6]. The utilization of anonymized data follows the classification outlined in ISO/TS 14265[2] for effective data consistency management.

6 Architecture and workflow

6.1 Architecture

Figure 1 presents a detailed architecture of multi-centre medical data collaborative analysis, specifically focusing on scenarios where the original medical data cannot leave the respective medical centres. The key components depicted in the figure are as follows.

— ETL: this module is responsible for transforming original medical data into the transformed database utilizing a CDM.

— Trusted logging tools: these tools record all operations conducted within the system to ensure the integrity of the records and prevent tampering.

— Health terminology base: the health terminology base manages terminology standardization of the local medical centres. Functionally, it standardizes the locally-used medical concepts to the global standard health terminology and stores them to ensure consistency of the terminology across multiple centres.

— Analysis tools: these tools are utilized for conducting multi-centre medical data collaborative analysis and managing the process of multi-centre collaborative researches.

— Coordination centre: the coordination centre serves as a central hub to facilitate collaborations among multiple centres. Collaborations between the medical centres and the coordination centre enable the secure transmission of essential non-original data for collaborative analysis services and applications.

— Terminology management: the terminology management provides protocols and methods for medical centres and the coordination centre to manage the terminology and maintain terminology consistency, including standardization, extension and quality control of health terminology.

— Multi-centre medical data standardization: the data standardization provides protocols and methods for medical centres to perform data cleaning, data standardization, and quality control to ensure integrity and consistency of medical data for collaborative analysis.

These components collectively enable efficient medical data collaborative analysis while preserving privacy and data integrity of the original medical data within the respective medical centres.

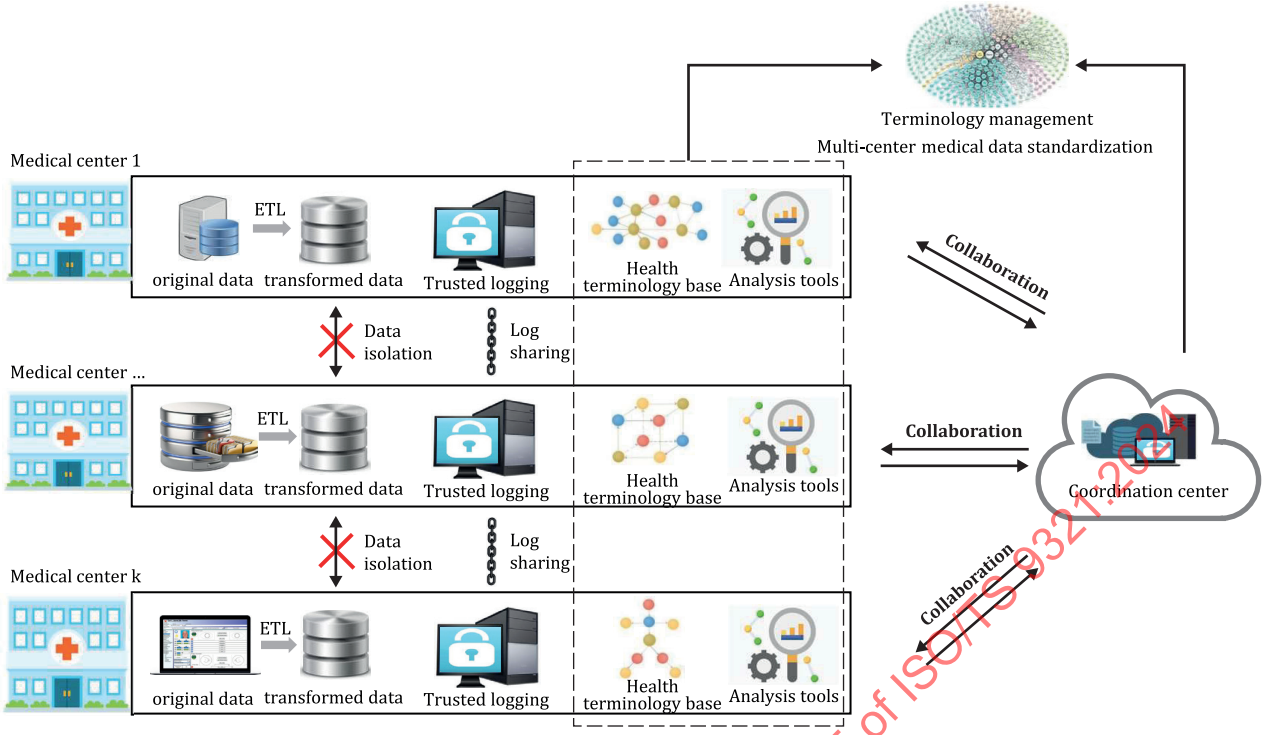

Figure 1 — Architecture of multi-centre medical data collaborative analysis

With stringent requirements for data isolation, security, and privacy in the use of medical data, collaborative analysis procedures shall adhere to a general standardized process, as illustrated in Figure 2. “Member grouping” involves collaborative research centres forming a research team and defining their respective research tasks. “Review & adjusting” entails conducting research ethics reviews and assigning data access rights by the medical centre. While “Member grouping”, “Result reporting”, and “Review & adjusting” are best discussed and determined within a centralized environment, “Data gathering” and analysis based on original data shall only be conducted in a distributed environment.

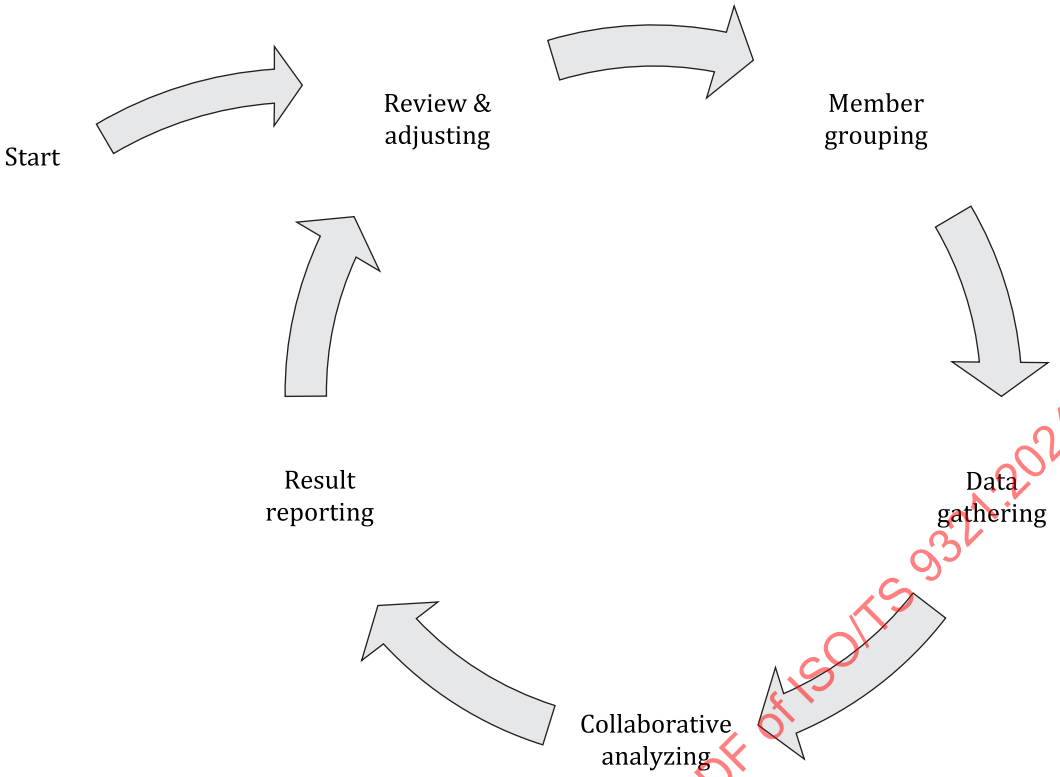

Figure 2 — General process for medical data collaborative analysis

6.2 Workflow

6.2.1 Multi-centre collaborative research initiation workflow

The initiation forms the foundation for multi-centre medical data collaborative analysis. The system offers users collaborative research functions based on the available data. Users initiate the multi-centre collaborative research based on the research interests. Figure 3 illustrates the multi-centre collaborative research initiation workflow. Someone who initiates the research is called the initiator, while the participants are those who participate in the research.

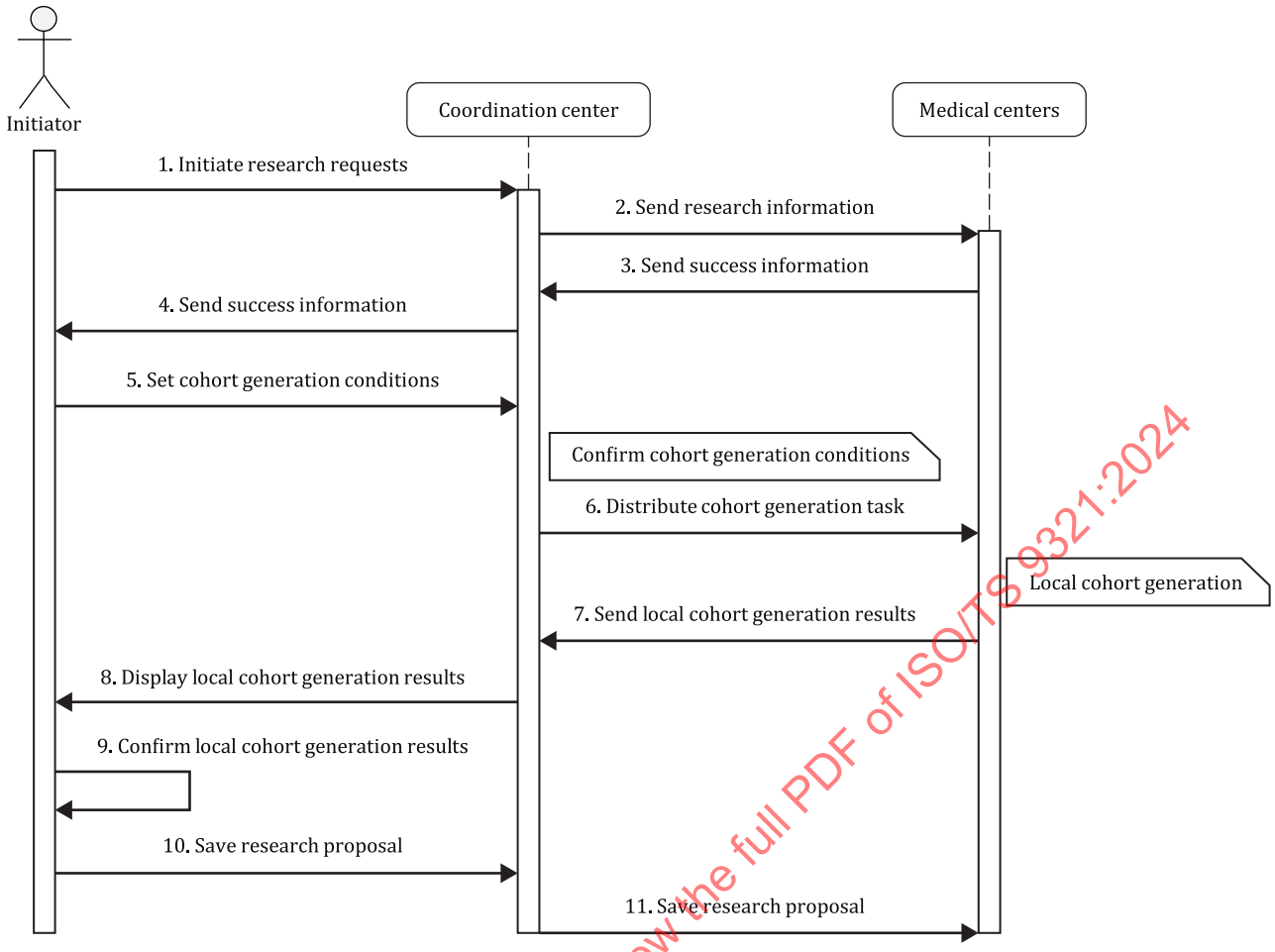

Figure 3 — Multi-centre collaborative research initiation workflow

6.2.2 Multi-centre collaborative research implementation workflow

The system aims to facilitate multi-centre collaborative research by utilizing multi-centre medical data. Users are categorized into two roles: research initiators and research participants. Research initiators initiate the research within the system and invite other users to participate. Users who agree to participate become the corresponding participants of the research. Once the initiator initiates the analysis, each medical centre completes the cohort construction locally, initializes the model, and begins sharing and exchanging model parameters until all models are trained. Research participants have access to individual training results and collaborative analysis results of the multi-centre medical data. The multi-centre collaborative research implementation workflow is shown in Figure 4. The initiator centre refers to the medical centre where the initiator of the research is located. The participant centres refer to the medical centres where the participants of the research are located.

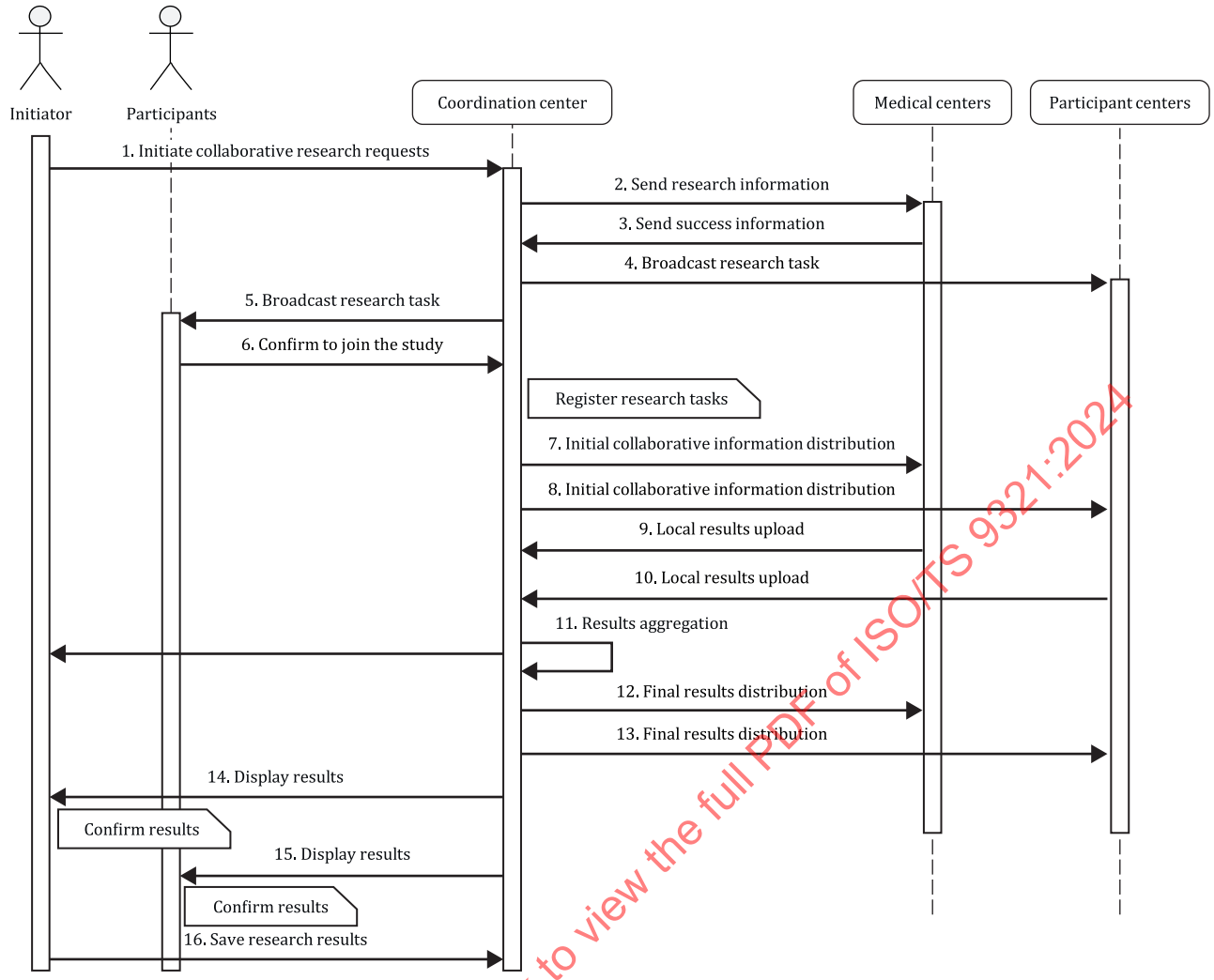

Figure 4 — Multi-centre collaborative research implementation workflow

7 Functional framework and requirements

7.1 Functional framework

7.1.1 General

In accordance with the standard process of medical data collaborative analysis, the multi-centre medical data collaborative analysis system shall encompass four functional layers: the user layer, the service layer, the resource layer, and a comprehensive set of multi-layer system security functions. Figure 5 illustrates these functional layers. To ensure seamless interoperability between the coordination centre and the medical centres, conformity to ISO 23903[8] should be attained.

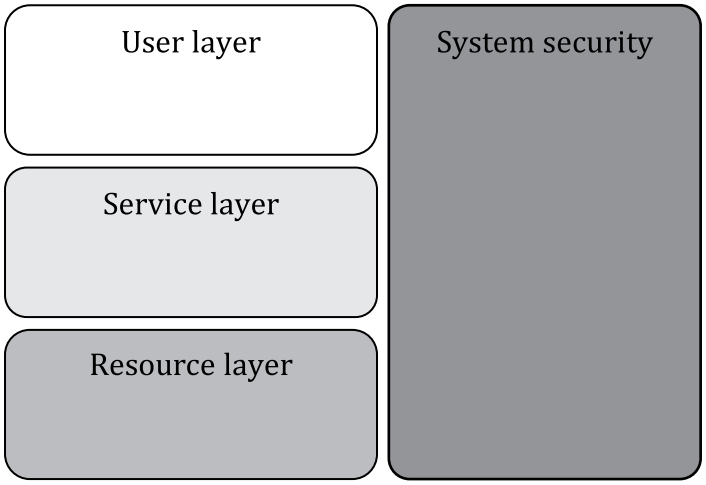

Figure 5 — Functional framework of multi-centre medical data collaborative analysis system

7.1.2 User layer

The user layer functions as an interface, enabling user interaction with the system through activities such as data entry and display. It communicates and responds to users' requirements, enhancing the overall user experience. The integration rules and capabilities of the user layer differ across scenarios. The user roles and activities of the system are shown in Clause B.3, which explains in detail how different roles cooperate to complete the multi-centre medical data collaborative analysis.

7.1.3 Service layer

The service layer comprises the implementations of services provided by the multi-centre medical data collaborative analysis system. It interacts with the user layer to receive and address user service requirements. It utilizes the necessary resources through the resource layer and executes processes through its included services.

The service layer shall include implementations for utilizing, integrating, controlling, and arranging the services. It shall provide unified and controlled interface methods for users to access services through the user layer.

The service layer shall provide unified and controlled resource connection methods for accessing various types of resources in the resource layer. The individual services shall exclusively utilize the unified resource connection methods to access and utilize the required resources.

7.1.4 Resource layer

The resource layer contains multiple types of essential resources for multi-centre medical data collaborative analysis. These include network connections, health terminology databases and specialized computing devices, allocated across various medical centres. Additionally, some resources, such as the coordination structure, and authentication files, are stored or generated at the coordination centre.

7.1.5 System security

The multi-layer system security functions provide secure environments for the entire multi-centre medical data collaborative analysis process. In secure environments, protected objects include: patient data, intermediate and final results, and intellectual property belonging to researchers or multi-centre medical data collaborative analysis system.

7.2 Functional requirements

7.2.1 General

7.2 describes the architecture of multi-centre medical data collaborative analysis system in terms of the common set of functions. A function is a functional element of the multi-centre medical data collaborative analysis system which is used to perform an activity or some part of an activity and has an implementation artefact in a concrete realization of the architecture. Figure 6 presents a high-level overview of the multi-centre medical data collaborative analysis system functions organized by means of the layering framework.

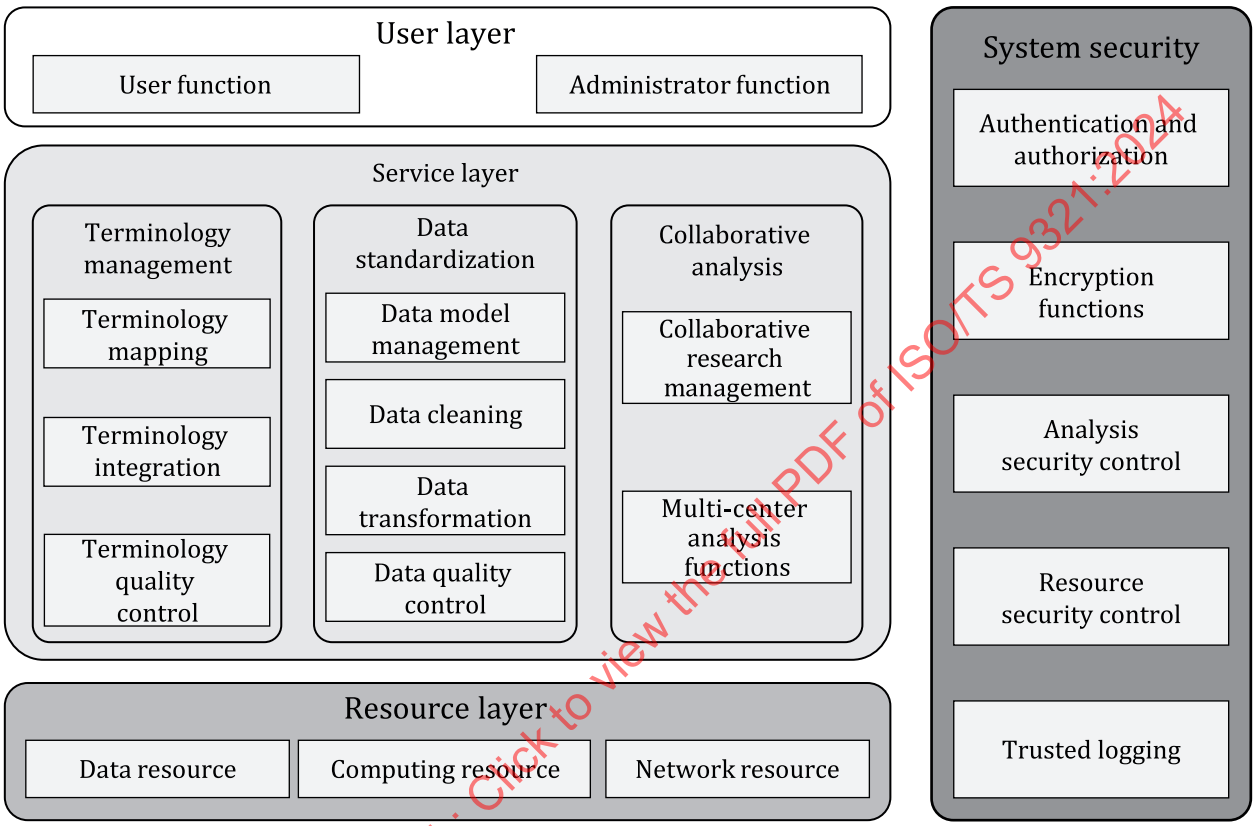

Figure 6 — Functions of multi-centre medical data collaborative analysis system

7.2.2 User layer functional requirements

7.2.2.1 General

The user layer is responsible for providing interaction interfaces that encompass the interactive functions of multi-centre medical data collaborative analysis. The functions of the user layer are under the security protection system. Users shall pass the trusted identity authentication before performing corresponding operations.

The user layer functions include:

— user function;

— administrator function.

7.2.2.2 User function

The user function supports the user to access and use services related to multi-centre medical data collaborative analysis. It is recommended to include user-interface functions of terminology management, data standardization, collaborative analysis, and user management to user function.

7.2.2.3 Administrator function

The administrator function supports administrational user functions, such as granting user authority, operation control, operation recording, operation querying, and user qualification review. The specific administrator function requirements are found in the ISO 22600 series[5].

7.2.3 Service layer functional requirements

7.2.3.1 Terminology management

7.2.3.1.1 General

The terminology management functions are responsible for maintaining standardization and consistency of medical concepts between medical centres. The terminology management functions use a global standard terminology as reference and provide functions for medical centres to map, update and manage medical concepts. The terminology management functions should provide ability to include new medical concepts for medical data collaborative analysis. An example of terminology management process is given in Clause B.1 (Annex B).

The terminology management functions include:

— terminology mapping;

— terminology integration;

— terminology quality control.

7.2.3.1.2 Terminology mapping

The terminology mapping functions provide capabilities to create mappings between medical concepts with same semantics among medical centres and the global standard terminology. Medical centres employ their respective terminology system and utilize the terminology mapping functions to map local concept to the global standard concepts for consistency.

7.2.3.1.3 Terminology integration

The terminology integration functions provide capabilities to extract new medical concepts from medical centres and add to the global standard terminology system. The terminology integration function is essential to extend the global standard terminology system to meet the specific medical concept requirements of medical centres.

7.2.3.1.4 Terminology quality control

The terminology quality control function is responsible for checking the accuracy and semantic consistency of the terminology mapping, ensuring that the medical concepts in medical centres are mapped to the correct global standard medical concept. The terminology quality control function also checks the format and quality of the terminology integration.

7.2.3.2 Data standardization

7.2.3.2.1 General

The data standardization functions are responsible for standardizing the heterogeneous medical data in medical centres and providing data quality control ability to support multi-centre medical data collaborative analysis. The data standardization functions provide functions and protocols to enable medical centres to maintain data integrity, ensure semantic consistency and control data quality.

The data standardization functions include:

— data model management;

— data cleaning;

— data transformation;

— data quality control.

7.2.3.2.2 Data model management

The data model management function provides the capability of CDM management and version control. The data model management function maintains a CDM as a reference for data standardization.

7.2.3.2.3 Data cleaning

The data cleaning function provides the capability of cleaning the medical data in medical centres to meet the data integrity requirements. The data cleaning function deals with duplicate data, incorrect data, missing data, and format errors.

7.2.3.2.4 Data transformation

The data transformation function provides the capability to transform the medical data into standard structure of the CDM. The data transformation function analyses the data structure in medical centres and maps the data elements into the CDM structure, ensuring the consistency of the data structure across multiple centres. The data transformation function also transforms the concept coding of the medical data into standard terminology codings, ensuring the data semantic consistency across multiple centres.

7.2.3.2.5 Data quality control

The data quality control function provides the capability to evaluate the accuracy, completeness and consistency of the data. The data quality control function should ensure that the data cleaning results meet the quality requirements and the standardized medical data are consistent across multiple centres.

7.2.3.3 Collaborative analysis

7.2.3.3.1 General

The collaborative analysis functions are responsible for supporting multi-centre medical data collaborative analysis, such as condition-based cohort generation functions, statistical analysis methods, and machine learning model constructions. The collaborative analysis functions are responsible for protecting all data utilized in medical data collaborative analysis by privacy computing techniques (e.g. homomorphic encryption, secure multi-party computation, federated learning) to protect patients' privacy and the data security. The collaborative analysis functions should also support research process management.

The collaborative analysis functions include:

— collaborative research management;

— multi-centre analysis functions.

7.2.3.3.2 Collaborative research management

The collaborative research management function provides the capabilities to record and review the key information in collaborative research, such as the establishment of collaborative teams, team information, research goals, scope of data used, qualification documents, achievement distribution plans, etc.

The collaborative research management function should also support medical centres' supervisors in organizing corresponding legal and ethical review processes. Besides, it shall construct an effective review

and confirmation mechanism to recognize the conformity and rationality of medical researches and final outcomes.

7.2.3.3.3 Multi-centre analysis functions

The multi-centre analysis functions provide visualized and interactive services that support the research workflows of cohort construction, feature construction, inter-hospital cohort matching, data preprocessing, analysis model construction and result analysis.

The multi-centre analysis functions provide sufficient data visualization services with user-customized statistical calculations.

The multi-centre analysis functions support the distributed computation on data statistics, preprocessing, cohort matching, and distributed model analysis, under a data isolation condition. The requirements for distributed computing are found in ISO/IEC TR 30102[11].

The multi-centre analysis functions support ciphertext equivalent analysis, a crucial technology in medical data collaborative analysis using ciphertext data. It should control the ciphertext equivalent analysis processes in either the medical centres or the coordination centre, and perform machine learning and iterative utilization based on encrypted data. Besides, it enables the decryption and output of the calculation results under the permission of all participating medical centres.

7.2.4 Resource layer functional requirements

7.2.4.1 General

The resource layer functions are responsible for the storage, organization and utilization of varied types of resources. During multi-centre medical data collaborative analysis, the resource layer functions provide unified managements for the dispatch of resources.

The resource layer functions include:

— data resource;

— computing resource;

— network resource.

7.2.4.2 Data resource

The data resource function is responsible for the independently, non-task-oriented, and mostly local storage of medical data among medical centres. The data resource function also allows the contained data resources to be utilized by the service layer functions and system security functions.

The data resources function is recommended to maintain management abilities on the common types of data resources, including medical records (structured and unstructured medical records), medical meta (database structure, health terminology base, knowledge graph, prediction models, etc.), and administration data (user data, coordination data, analysis results, trusted log, etc.) to fulfil the general requirements of this document.

7.2.4.3 Computing resource

The computing resource function is responsible for collection, management, and dispatch of all computational forces that the multi-centre system provides, including CPU, memory, GPU, other computing elements, and their combined high-level environments such as local servers, NaaS, IaaS, PaaS or SaaS.

7.2.4.4 Network resource

The network resource function is responsible for the network communication to support the coordination, collaborative analysis, computing resource distribution, trusted logging and service application between the coordination centre and all local medical centres.

The network resource function shall include a feasible networking protocol, which allows a flexible centre joining and quitting, and sub-network construction.

7.2.5 System security functional requirements

7.2.5.1 General

The system security functions are responsible for applying security-related controls to mitigate the security threats in medical data collaborative analysis. The system security functions encompass all the security facilities required to support multi-centre medical data collaborative analysis.

The security systems functions include:

— authentication and authorization;

— encryption functions;

— analysis security control;

— resource security control;

— trusted logging.

7.2.5.2 Authentication and authorization

The authentication and authorization function provides the capability of user management and access control. It can determine the role of any involved user, and specify the roles that can perform specific actions and get access to certain information.

7.2.5.3 Encryption functions

The encryption functions provide capabilities for selecting encryption schemes, managing keys and decryption, and encrypting sensitive information such as patient data, intermediate results, model parameters, and other data that requires protection.

7.2.5.4 Analysis security control

The analysis security control functions prevent any unauthorized disclosure of patient's privacy throughout the entire analysis process. They should also identify malicious behaviours by both internal and external adversaries.

7.2.5.5 Resource security control

The resource security control function provides the capability to avoid theft of code, knowledge graphs, models, and other contents about intellectual property. It should also be able to monitor the condition of computing and network resources, and prevent them from being illegally occupied.

7.2.5.6 Trusted logging

The trusted logging function should support recording, storing, querying and verifying functions of the operation logging message occurred on the medical centres for further tracking. It should also warn users who operate improperly. The stored logging should be trustworthy and secure, thus some information security techniques should be applied, such as detailed in Clause B.2.

Annex A
(informative)

Examples of multi-centre medical data collaborative analysis networks

A.1 General

Multi-centre collaborative research networks have been established by various organizations and institutions for collaborative analysis of multi-centre medical data. They have conducted research in several aspects, including common data models, data standardization, terminology management, collaborative research processes and collaborative analysis methods. There are some existing systems in this area.

A.2 Observational Health Data Science and Informatics (OHDSI)

A collaborative system for clinical research, OHDSI,[13] led by Columbia University, provides a solution to medical data analysis, with emphasis on pharmaceutical research. OHDSI uses a CDM to standardize clinical observational data such as electronic medical records to support analysis of clinical data from multiple sources. The OHDSI research network has access to more than 80 medical databases from 74 countries, pooling data from 1,2 billion patients (duplicates).

A.3 Multi-centre intelligent medical information system

The multi-centre intelligent medical information system, led by Zhejiang Lab in China, provides solutions for large-scale, trans-regional, and multi-centre medical data analysis, supporting medical research and serving clinical patients. The system has completed the access work of 5 large hospitals, covering 20,1 million patient data, and can conduct retrospective analysis and research on clinical data of nearly 20 years.

A.4 FeederNet

FeederNet,[12] a nationwide collaborative research network, has been established in Korea. Seventy percent of tertiary teaching hospitals in the Republic of Korea have joined the research network. The network has been connected to 47 hospitals and collected 30 billion real world data on 57 million patients.