



Research on Multi-agency Data Fusion Mode Under Regional Medical Integration

Yixin Wang¹, Weiqing Fang¹, Wei Zhu¹(✉), and Jinshun Ding²

¹ Changshu No.2 People's Hospital, Changshu 215500, Jiangsu, China

² Changshu Meili Hospital, Changshu 215500, Jiangsu, China

Abstract. 2020 is not only the stage of intensive implementation of medical informatization related policies, but also a key year for the further development of regionalization of medical informatization projects. The medical community data sharing technology using multi-source heterogeneous data fusion solves the problem of different hospitals, different procedures, different database structures, and information islands in each hospital. Through ETL technology, using the SSIS tool in Microsoft SQL Server, a relatively standard data system is built for the original information system of each hospital in the medical community group to centrally convert, clean and transfer to a standardized data model to form a data set: Patient Master Index (EMPI), Master Data Management (MDM), etc., to solve the problem of reducing repeated statistics and discrepancies in various hospitals, improve data quality, complete interconnection and data sharing.

Keywords: Data sharing · Data fusion · ETL technology

1 Introduction

2020 is not only the stage of intensive implementation of policies related to medical informatization, but also a key year for the further development of regionalization of medical informatization projects [1]. In accordance with the overall planning of the supporting implementation measures for the pilot program for further deepening the comprehensive reform of public medical institutions in Changshu City (Changbanfa [2018] No. 124), the city's second hospital regional medical community was established to promote the sharing of regional medical resources and provide the overall capabilities and capabilities of the medical service system. Performance. The medical community data sharing technology using multi-source heterogeneous data fusion solves the problem of different hospitals, different procedures, different database structures, and information islands in each hospital [2–4]. In order to carry forward the advantages of medical regional integration of the Second People's Hospital of Changshu City, fully tap and play the value of data, integrate the clinical data accumulated for many years into the clinical data center (CDR), and plan to carry out a number of community service centers and central health centers [4–6]. The data resource integration of the hospital realizes the data resource sharing between the general hospital and community hospitals.

2 Research Status

The ultimate goal of regional medical-oriented clinical data exchange is to solve the two problems of information integration and data standardization. In order to solve the various challenges faced by information system integration, promote the information sharing and coordination of medical information systems, and provide integration solutions between different subsystems, the North American Association of Radiological Medicine and the American Association of Medical and Health Information and Management Systems established Medical in 1988 Institutional integration [7]. In North America, organizations such as RSNA, HIMSS, and the American Heart Association were responsible [8, 9]. The first five-year plan was formulated in 1999. The main participants include RSNA, HIMSS, California healthcare institute (CHI), etc. In Europe, IHE is under the responsibility of the European Society of Radiology and the association of the medical imaging industry [10]. The participants mainly include the Hospital Information System Exhibition Organization, the French Radiological Society, the German-Italian Radiological Society, and the Italian Radiological Society. In Asia, countries such as Japan, South Korea, and Taiwan have participated in the IHE project. Taiwan has conducted several IHE connection tests before this [11].

In China, there is still very little research and application of IHE, but it has gradually entered this field. IHE China was established in Shanghai in 2007 [12, 13]. In order to promote the popularization and promotion of IHE in China, the University of Shanghai for Science and Technology undertakes the technical preparation and organization of the test, and the first test was conducted in 2007. In 2008, the China Medical Equipment Association also started testing work and conducted the first test in Beijing Tiantan Hospital, including Mindray, General Medical, Siemens and other domestic and foreign manufacturers [14].

In terms of clinical data exchange standardization, in the process of medical informationization, countries such as the United Kingdom, the United States, and France have invested a lot of manpower and material resources in standardization work [15, 16]. Many standards have been widely recognized and applied. For example, the international code of diseases (ICD) standard coding is used in the classification of diseases and treatments; ICD-9, ICD-10 and other standard codes are used in diseases and surgical operations; Adopt systemized nomenclature of human and veterinary medicine (SNOMED) coding system in the electronic medical record system [17–19]. In addition, there are the US medical and health information transmission and exchange standards (health level seven, HL7), digital imaging and communications in medicine (digital imaging and communications in medicine, DICOM), and laboratory information system and inspection equipment interface standards [20].

3 Research Purpose

3.1 Purpose of Research and Application

- (1) Solve the problem that the clinical information data cannot be effectively shared due to the different hospital systems in the medical community; it can optimize the continuity of the clinical data of the referred patients in the region and quickly view the historical case data for efficiency, Quickly trace the patient's entire treatment process and disease development; integrate the clinical basic data unit in the medical community, so as to more efficiently and conveniently trace the patient's entire treatment process and disease changes in the future [21, 22].
- (2) Using the data center experience that has been initially established in Changshu Second People's Hospital, combining the actual situation of each hospital system in the medical community, using the sharing technology of multi-source heterogeneous data fusion, they have accumulated a large number of years. Standardized storage and classification of clinical data and operational data are used to improve the medical quality and performance during the daily operation of the hospital, and to make full use of existing data to improve the level of diagnosis and treatment in the hospital [23].
- (3) Use shared data technology to further strengthen the quality of operational indicators and data between hospitals in the medical community, integrate data, and provide a basis for decision-making.

3.2 Characteristics of Multi-agency Data Fusion Model

- (1) Overall characteristics
System construction considers the protection and utilization of existing IT investment, based on the existing business system, without affecting the original system, to achieve seamless data and business connection between functions and the existing business system. The specific performance is: ease of use: implement the principle of end-user-oriented, establish a friendly user interface, make user operation simple and intuitive; security: security is a necessary condition for effectiveness and usability, and data cannot be lost due to various reasons. Or it may be illegally modified to affect or even stop the development of hospital medical services. In addition to hardware configuration measures, the security of the medical information system should also be addressed in many aspects such as system operation mode, software settings, rules and regulations; flexibility: the system must change accordingly with changes in health policies and regulations, and full consideration must be given to the system's Flexibility, reasonable arrangement of system structure.
- (2) Data characteristics
Data input: It must be formatted data, and use selective input as much as possible to input it quickly and efficiently; Data statistics: Data statistics functions for all selected items must be provided, so that users can freely choose and count freely; Data backup: With data backup function, it can be divided into automatic timing backup, automatic remote backup and manual operation backup; data authority: data use authority function must be provided; data security: the system should

provide a complete user authority data access system, and database backup and different Provide scientific and reasonable security measures for machine backup and recovery. See Fig. 1.

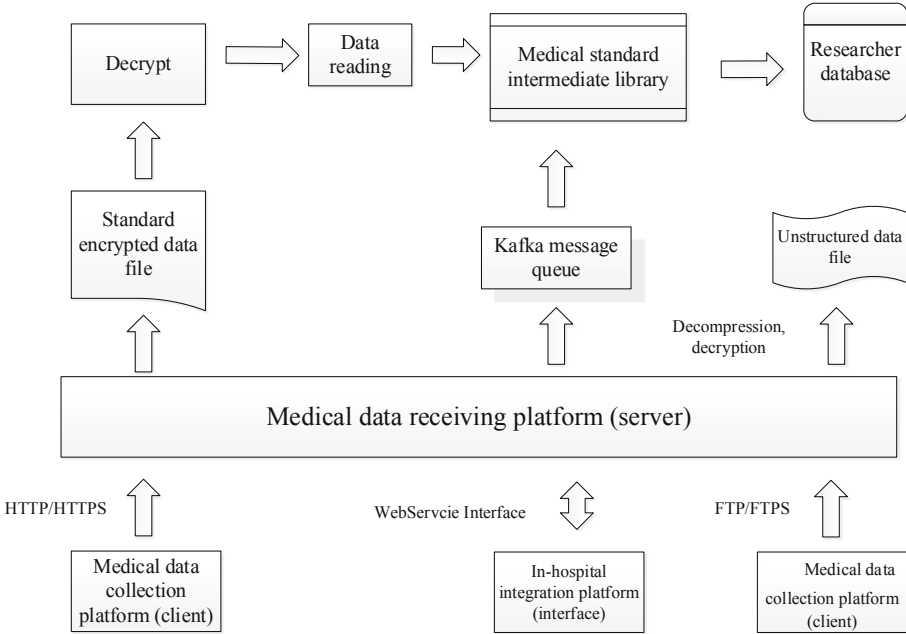


Fig. 1. Data processing diagram of multi-institution data platform

(3) Function parameters

Clinical Data Center (CDR): the data accessed by the clinical data center, and the organization of the access data is distinguished; hospital operation management: the system needs to increase the organization selection function, which is convenient for querying the indicators of the headquarters and community hospitals; switching institutions The function requires the authority control function, which can be conveniently and quickly set up the authority; the main patient index EMPI: to identify patients from multiple institutions, through the integration of EMPI numbers, it is convenient for other systems to identify and display patients with different card numbers; patient 360 view: Yes The medical information of multiple institutions is integrated to facilitate better diagnosis and treatment by medical workers.

4 Architecture Design

4.1 The Overall System Architecture is Divided into Five Levels

- (1) Basic business layer: including HIS, LIS, PACS and other business systems that have been built by the member units of the medical community and various community service centers. See Fig. 2.
- (2) Platform data layer: build a hospital-wide integrated data center, including business replication database and ODS, clinical data center (CDR), master data (MDM), data warehouse (DW), the main focus of this layer construction includes data The scientific organization of the domain model, the storage of high-quality data after cleaning and conversion, the theme design of the data warehouse, and the design of the data mart (including the data cube) model, etc.
- (3) Platform service layer: Build the core service components of the hospital data center. These services provide services for hospital data integration, data statistical analysis, and platform construction for application layer data applications, including UI engine, patient main index EMPI, etc.
- (4) Platform application layer: The platform application layer is to provide hospital platform users with key application systems based on the hospital data center and provide application support for other application systems. The project mainly includes patient views, hospital operation management, scientific research applications, etc.
- (5) Platform portal layer: establish single sign-on services to break the physical location constraints of the platform, including manager portals, doctor portals, patient portals, researchers portals, and public portals.

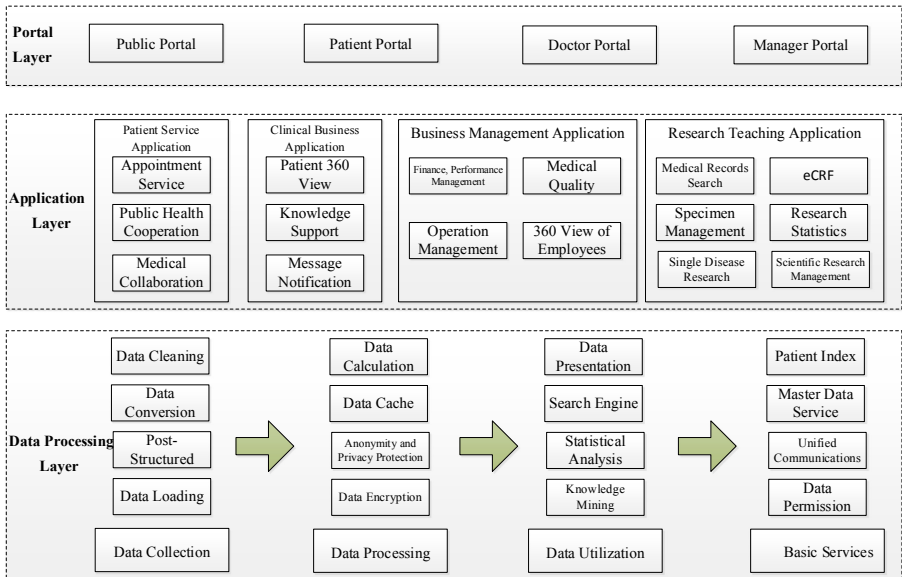


Fig. 2. Layered schematic diagram of the overall system architecture

4.2 Network Data Structure

Take Changshu Second People’s Hospital as the center, as shown in Fig. 3, each in the region as branch points, through ETL technology to effectively centralize the clinical data sets of each hospital, and establish a relatively standard data system for centralized conversion, cleaning and transfer A data set is formed in a standardized data model to effectively integrate regional internal data to serve the clinic.

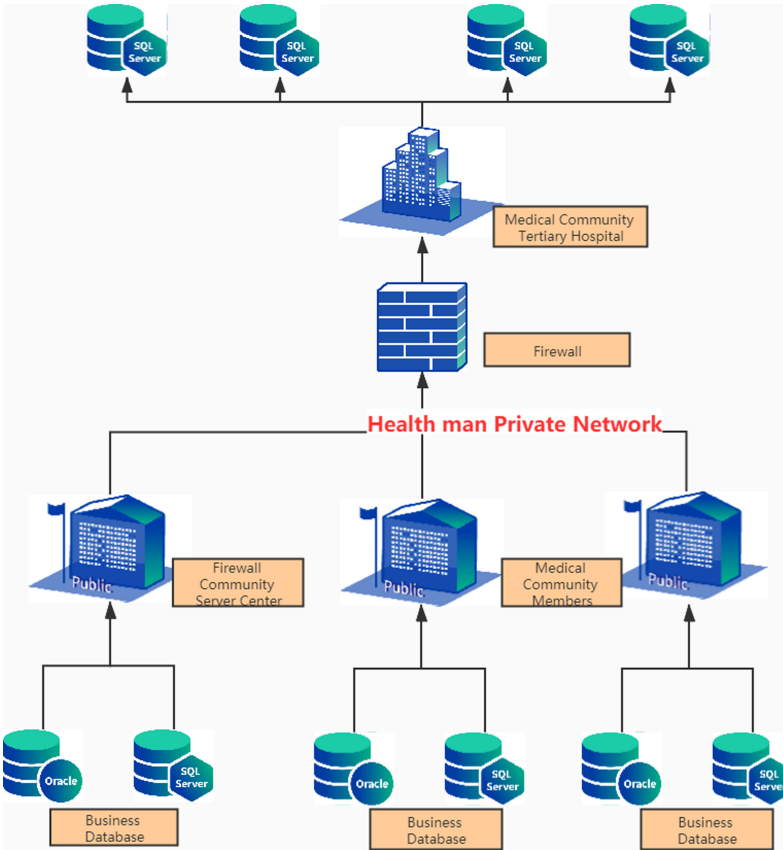


Fig. 3. Schematic diagram of regional center network structure

5 Key Technologies

Using the current experience of establishing a small clinical data center in Changshu Second People’s Hospital, a research method for data sharing based on multi-source heterogeneous data fusion under the medical community has been established to solve the different database structures of different procedures in different hospitals within the

medical community. The problem of information isolation in various hospital systems. Through ETL technology, using the SSIS tool in Microsoft SQL Server, the original information system of each hospital in the medical community group, including hospital management information system (HIS), electronic medical record information system (EMR), nursing information system (NIS), Laboratory information system (LIS), medical image cataloging (PACS), etc. for data extraction, combined with the his table structure to establish a relatively standard data system, centralized conversion, cleaning, and transfer to a standardized data model to form a data set: Patient main index (EMPI), Master Data Management (MDM), etc., to solve the problem of reducing repeated statistics and disagreements in various hospitals, and propose to target the five-level application level of hospital electronic medical records, standardize integration to build the big data of the medical community, and unify statistics Caliber, improve data quality, complete interconnection and data sharing.

5.1 Data Model

- (1) The data model should include the main business areas of the hospital and support at least the following data queries. See Table 1, Table 2, and Table 3.

Table 1. Data model business areas

Field	Patient Management Domain	Cost	Laboratory Domain
Structure	Basic patient information Registered information Admission and admission registration information Diagnostic information Admission information	Outpatient settlement master record Outpatient settlement details Inpatient settlement master record Inpatient settlement details	Application registration information Specimen information Clinical examination and biochemical report Microbiology report Pathology report

- (2) Data consolidation, including at least the following key dictionaries, hospital identification, update time, import time, and time stamp. See Table 4.

5.2 Main Patient Index

The Patient Master Index Service (EMPI) uses unique algorithms and technologies to create, search, and maintain the patient’s basic information index throughout the hospital, which can intelligently assist medical staff in effectively searching for patients. Medical institutions use the main index (EMPI) to identify, match, merge, and cancel duplicate data, purify patient records, and use the main index to obtain a complete and single patient view. EMPI will create a unique identifier for each patient and establish a mapping between the identifiers of the medical records in the relevant system.

Table 2. Data model business areas

Field	Observation Domain	Medical Record Domain	Care Domain	Surgical Domain
Structure	Observation report Vital signs observation information Allergy information observation information	Medical Record Home Medical record diagnosis Medical record surgery	Medical order execution record Care delivery record	Surgical registration Operation record Surgical diagnosis Surgical anesthesia information Postoperative recovery information Surgery participants

Table 3. Data model business areas

Field	Medical Order Domain		Medical Record Field
Structure	Outpatient drug prescription Outpatient examination prescription Outpatient inspection prescription Outpatient treatment prescription Outpatient surgery prescription	Inpatient Drug Order Hospital checkup Inpatient examination order Hospital surgery doctor's order Hospital care order Hospital blood transfusion order Hospitalization Hospital Meal Order	Medical record master data Medical record segment data Medical record style data Full-text index of medical records Unstructured medical record data Hospital outpatient medical records

5.3 Master Data Management

Master data management MDM (Master Data Management) mainly stores dictionary data stored in the hospital business system, as well as national standards, industry standards, landmarks, hospital departments, service units, service personnel, service items and other custom data elements used in the hospital business.

Table 4. Key dictionaries for data consolidation

Serial number	Column name	Data type	length primary	key	Column description
1	HospitalNo	nvarchar(30)	30	√	Medical institution code
2	HospitalName	nvarchar(60)	60		Medical institution name
3	PatientID	nvarchar(30)	30	√	Patient identification
4	MedicalRecordNo	nvarchar(30)	30		Medical record number
5	PatientName	nvarchar(60)	60		Patient name
6	PY	nvarchar(60)	60		pinyin
7	Wb	nvarchar(60)	60		Wubi

6 Summary

- (1) The medical community data sharing technology using multi-source heterogeneous data fusion solves the problem of different hospitals, different procedures, different database structures, and each hospital system information island.
- (2) Medical community data collection uses ETL technology to collect, and the scalability of the system can be realized through the database engine. Keep all data in the database at all times, avoid data loading and export, thereby ensuring efficiency and improving system monitorability. It is also possible to optimize the distribution of data in parallel, and use the inherent functions of the database to optimize disk I/O.
- (3) It is proposed to organize the clinical data of patients with patients as the main line, and store the clinical data of patients in a comprehensive, standard and unified manner, which provides a unified platform support for the sharing of clinical data of the medical community, and provides some applications of clinical data Also laid a solid foundation.
- (4) Through data sharing technology based on multi-source heterogeneous data fusion, extract clinical quality-related data, improve the medical quality and performance during the daily operation of the hospital, and improve the level of diagnosis and treatment in the hospital.
- (5) Through medical community research and practical application, the mature medical community data sharing technology and information big data company will be promoted to other regional medical communities to better serve clinical and hospital management.

Funding. 2019 Changshu City Science and Technology Development Plan (Social Development) Project, Research and application of sharing technology based on multi-source heterogeneous data fusion under the medical community applied to clinical-related data quality (No.CS201913).

References

1. Wang, G., Liu, X., Wu, G., Guo, Y., Ma, S.: Research on data fusion method based on rough set theory and BP neural network. In: ICCEA 2020, pp. 269–272, March 2020
2. Gao, J., Li, P., Chen, Z., Zhang, J.: A survey on deep learning for multimodal data fusion. *Neural Comput.* **32**(5), 829–864 (2020)
3. Lin, D.: Research on key technologies of regional synergy emergency system based on medical data center. *Proc. Comput. Sci.* **154**, 732–737 (2019)
4. Jiemin, Z.: Analyzing the models of medical data center on cloud computing. In: 2015 10th International Conference on Computer Science & Education (ICCSE), pp. 76–9 (2015)
5. Apao, N.J., Feliscuzo, L.S., Romana, C.L.C.S.: Developing a patient information and descriptive analytics system for data actors of university of bohol medical and rehabilitation center: Towards policy making. In: ACM International Conference Proceeding Series, DSIT 2019, pp. 42–48, 19 July 2019
6. Yang, Y., et al.: A new medical imaging sharing service network based on professional medical imaging center. *Progress in Biomedical Optics and Imaging*, vol. 10954, p. 109540U (2019)
7. Jiemin, Z., Jinsheng, L.: The model of district medical data center. In: 2010 International Conference on Computer Application and System Modeling (ICCA SM 2010), pp. 471–474 (2010)
8. Biswas, N., Sarkar, A., Mondal, K.C.: Efficient incremental loading in ETL processing for real-time data integration. *Innovations in Systems and Software Engineering*, vol. 16, no. 1, pp. 53–61, 1 March 2020
9. Oliveira, B., Oliveira, Ó., Santos, V., Belo, O.: ETL development using patterns: a service-oriented approach. In: The 21st International Conference on Enterprise Information Systems, ICEIS 2019, vol. 1, pp. 204–210 (2019)
10. Wojciechowski, A., Wrembel, R.: On case-based reasoning for ETL process repairs: Making cases fine-grained. In: *Communications in Computer and Information Science*, CCIS, vol. 1243, pp. 235–249 (2020)
11. Muddasir, N.M., Raghuvver, K.: A novel approach to handle huge data for refreshment anomalies in near real-time ETL applications. In: *Soft Computing: Theories and Applications. SoCTA 2019. Advances in Intelligent Systems and Computing* (1154), pp. 545–54 (2020)
12. Mandal, S., Jha, R.R.: Exploring the importance of collaborative assets to hospital-supplier integration in healthcare supply chains. *Int. J. Prod. Res.* **56**(7), 2666–2683 (2018)
13. Greenroyd, F.L., Price, A., Demian, P., Hayward, R., Sharma, S.: Modeling and simulating hospital operations in a 3D environment. In: *Proceedings - Winter Simulation Conference, WSC 2017*, pp. 2952–2963, 28 June 2017
14. Mandal, S., Jha, R.R.: Exploring the importance of collaborative assets to hospital-supplier integration in healthcare supply chains. *Int. J. Prod. Res.* **56**(7), 2666–2683 (2018)
15. Tsumoto, S., Hirano, S., Kimura, T., Iwata, H.: From hospital big data to clinical process: a granular computing approach. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 2669–78 (2018)
16. Usama, M., et al.: Deep feature learning for disease risk assessment based on convolutional neural network with intra-layer recurrent connection by using hospital big data. *IEEE Access* **6**, 67927–67939 (2018)

17. Kazancigil, M.A.: Innovations in medical apps and the integration of their data into the big data repositories of hospital information systems for improved diagnosis and treatment in healthcare. *Smart Innovation, Systems and Technologies*, vol. 189, pp. 183–192. *Human Centred Intelligent Systems - Proceedings of KES-HCIS 2020 Conference* (2021). https://doi.org/10.1007/978-981-15-5784-2_15
18. Tao, J.: Application of the big data processing technology in the hospital informatization construction. *Lecture Notes in Electrical Engineering*, vol. 551 LNEE, pp. 1589–1595, 2020, *Frontier Computing - Theory, Technologies and Applications, FC* (2019)
19. Liu, Z., Pu, J.: Analysis and research on intelligent manufacturing medical product design and intelligent hospital system dynamics based on machine learning under big data, *Enterprise Information Systems* (2019)
20. Sirisawat, P., Hasachoo, N., Kaewket, T.: Investigation and prioritization of performance indicators for inventory management in the university hospital. In: *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, p. 691–695 (2019)
21. da Silva Etges, A.P.B., et al.: Proposition of a shared and value-oriented work structure for hospital-based health technology assessment and enterprise risk management processes. *Int. J. Tech. Assessment Health Care* **35**(3), 195–203 (2019)
22. Canha, M., Loureiro, R., Marques, C.G.: The impact of the introduction of logistics management systems in an organization: a case study in a hospital center. In: *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, p. 4 (2018)
23. Xinlei, C., Xiaogang, R., Yue, W., Jiufeng, Y.: Design and realization of a comprehensive management system for severe mental disorders based on FLUX mode. *J. Med. Imaging Health Inform. ASP* **10**(2), 522–527 (2020)