Digital Policy Leading the Governance Logic and Ecosystem Construction of Medical Data Space: A "Technology-Management-Law" Analytical Framework Postprint

Authors: Wang Hongchuan, Zhang Jianbo, Ma Wei, Zhao Sidi, Zhang Jianbo

Date: 2025-08-06T00:00:00+00:00

Abstract

Amidst the global wave of digital transformation, the data revolution in the healthcare sector is accelerating, with countries actively formulating digital policies to drive medical data development. This study focuses on the construction of digital policy-driven medical data spaces, systematically dissecting their multidimensional challenges in data sharing, privacy security, technology convergence, and ethical governance. Drawing upon the practical experiences of the EU's "European Health Data Space" and China's "14th Five-Year Plan" policies, it leverages technologies such as federated learning and blockchain notarization to break down data silos, optimizes resource allocation through layered and transparent funding mechanisms, and simultaneously establishes cross-border data mutual recognition frameworks to enhance international cooperation. The research posits that the sustainable development of medical data spaces necessitates the urgent construction of a "technology-management-law" tripartite governance framework, building a multi-stakeholder collaborative ecosystem grounded in technology and policy to resolve core issues including insufficient data standardization, storage bottlenecks, and talent shortages. Through a theoretically and practically integrated path analysis, this work provides systematic reference for the digital infrastructure construction under the Healthy China Strategy, with the potential to facilitate the transformation of medical data from resource accumulation to intelligent services, thereby significantly advancing the democratization of precision medicine and the coordinated development of global health governance.

Full Text

Preamble

· Review and Monograph · Governance Logic and Ecosystem Construction of Healthcare Data Space Led by Digital Policies: An Analytical Framework Based on "Technology-Management-Law"

WANG Hongchuan[1],[2], ZHANG Jianbo[3]*, MA Wei[3], ZHAO Sidi[2]

1. School of Public Policy and Management, Tsinghua University, Beijing 100084, China

2. Institute for Contemporary China Studies, Tsinghua University, Beijing 100084, China

3. School of Economics and Management, Xinjiang University, Urumqi 830046, China

*Corresponding author: ZHANG Jianbo, Research assistant; E-mail: 3067754029@qq.com

Abstract Under the wave of global digital transformation, the data revolution in the healthcare industry is accelerating, with countries actively enacting digital policies to drive healthcare data development. This paper focuses on the construction of healthcare data space driven by digital policies, systematically analyzing its multi-dimensional challenges in data sharing, privacy security, technology integration, and ethical governance. By combining the practical experiences of the European Union' s "European Health Data Space" and China' s "14th Five-Year Plan" policies, it breaks through data silos with technologies such as federated learning and blockchain-based certification, optimizes resource allocation through hierarchical and penetrating capital allocation, and simultaneously establishes a cross-border data mutual recognition mechanism to strengthen international cooperation. The study argues that the sustainable development of healthcare data space urgently requires constructing a trinity governance framework of "technology-management-law," relying on technology and policies to create a multi-agent collaborative ecosystem to solve core problems such as insufficient data standardization, storage bottlenecks, and talent shortages. Through path analysis that combines theory and practice, this paper provides systematic reference for the construction of the digital foundation of the Healthy China strategy, and is expected to help transform healthcare data from resource deposition to intelligent services, significantly promoting the popularization of precision medicine and the coordinated development of global health governance.

[**Key words**] Healthcare data space; Artificial intelligence; Big data; Data security; Precision medicine

[**Chinese Library Classification**] R-056 [**Document code**] A DOI:

10.12114/j.issn.1007-9572.2025.0145

Funding: Tsinghua University School of Public Policy and Management Basic Research Seed Fund (2024JCZ2002); Tsinghua University "Double First-Class" Liberal Arts Construction Project (2024TSG06402); Xinjiang Autonomous Region Universities' Basic Research Business Expenses Scientific Research Project (XJEDU2022P024); Xinjiang University Young Teachers Cultivation Project (23CPY019)

Citation: Wang HC, Zhang JB, Ma W, et al. Governance logic and ecosystem construction of healthcare data space led by digital policies: an analytical framework based on "technology-management-law" [J]. Chinese General Practice, 2025. DOI: 10.12114/j.issn.1007-9572.2025.0145. [Epub ahead of print]. [www.chinagp.net]

1 Definition and Research Background of Healthcare Data Space

Driven by academic exploration of data management models, global data competition, and industrial intelligent transformation demands, and to balance data value mining with privacy protection, the "Healthcare Data Space" has emerged as a critical innovation. Healthcare data space refers to a medical information ecosystem supported by digital technology and centered on data sharing. By integrating physical infrastructure (such as data centers and medical equipment) and software architecture (such as cloud platforms and interoperability protocols), it enables standardized storage, cross-institutional flow, and authorized use of healthcare data in a secure and trusted environment. This space is patient-centered, supports multiple stakeholders including medical institutions, researchers, government agencies, and compliant enterprises under the premise of protecting personal privacy and respecting data sovereignty, and achieves full lifecycle management and value mining of healthcare data through unified governance frameworks and technical standards. Its core objectives are to improve medical quality, enhance management levels, promote medical research, and empower public health governance, striving to create a new healthcare model of "intelligent interconnection, data sharing, and collaborative services."

Propelled by the global intelligent era, the medical field is undergoing unprecedented data transformation. Artificial intelligence is reshaping healthcare with unprecedented depth and breadth, from intelligent assisted diagnosis to accelerated new drug research and development, making data the core engine driving medical innovation. However, prominent issues such as heterogeneous data formats across medical institutions, insufficient cross-system interoperability, and the high sensitivity of patient health information (e.g., electronic medical records, genomic data) cannot be met by existing regulatory systems, technical standards, and management mechanisms. The urgency of building a key carrier that integrates multi-party data resources and enables secure and efficient

circulation is increasingly prominent. Simultaneously, scenarios such as precision medicine, real-world evidence research [1], and public health emergencies impose urgent demands for large-scale data sharing. Governments and international organizations continue to introduce relevant policies aimed at breaking data silos, reducing privacy security risks, and addressing increasingly stringent compliance pressures. The European Union has taken the lead in promoting the "European Health Data Space (EHDS)" to achieve secure cross-border data circulation through the establishment of a digital health market [2]; the Chinese government actively promotes relevant policies, such as the "14th Five-Year Plan for Digital Economy Development" [3] emphasizing the cultivation of medical resource sharing space construction, and based on relevant norms and standards including the "National Basic Public Health Service Standards (Third Edition)" [4], comprehensively promotes the digital transformation and upgrading of primary-level medical and health institutions.

This paper focuses on the construction of healthcare data space under the background of digital policies, aiming to systematically explore practical challenges in data sharing innovation, privacy protection technology, and intelligent governance models, summarize domestic and international experiences in rule-making, technology integration, and ecosystem cultivation, and provide theoretical analysis and practical pathways for building a trusted, secure, and efficient medical intelligent ecosystem [5].

2 Current Development Status and Challenges

2.1 Accelerated Development of the Healthcare Data Market

The healthcare data space market is developing with strong momentum, powerfully driving the upgrading of the global medical industry. Currently, the global digital healthcare market is in a stage of vigorous development, with rapid growth in related sectors. According to data from relevant institutions, the global digital healthcare market size reached $286.35 billion in 2023 and is expected to continue expanding at a compound annual growth rate of 26.8%, rising to $365.67 billion in 2024 and potentially exceeding $450 billion in 2025 [8]. The "14th Five-Year Plan for Bioeconomy Development" proposes [9] to focus on enhancing original innovation capabilities in directions such as advanced diagnosis and treatment technologies and equipment, and precision medical testing. Against this backdrop, China' s intelligent medical device market has grown rapidly from 2020 to 2025, with the market size expected to reach 24.23 billion yuan in 2025; and from 2026 to 2027, this market is expected to maintain high-speed growth [10]. In market competition, the corporate landscape shows high concentration, with leading enterprises occupying dominant positions in the industry. Meanwhile, the rise of internet medical platforms has deepened scenarios such as health management and remote diagnosis and treatment, with enterprises possessing multimodal processing capabilities and compliance frameworks gradually gaining advantages in the industrial landscape. However, the market-based allocation mechanism for data remains imperfect,

with most projects relying on government funding support, and the sustainability of business models needs exploration. In the future, it is necessary to focus on solving bottleneck problems such as the digital capability gap at the primary level and insufficient algorithm interpretability to fully promote the development of artificial intelligence in the medical field toward universal benefit.

2.2 Prominent Challenges in Healthcare Data Security

Data security plays an extremely critical role in healthcare data space, providing guarantees for stable operation, rational data use, and healthy development of the medical industry. The "2020 Digital Healthcare: Research Report on Cybersecurity Risks During the Epidemic Prevention and Control Period" released by the China Academy of Information and Communications Technology [11] shows that nearly 30% of surveyed medical institutions face data asset leakage risks. Verizon's cybersecurity report indicates that worldwide, the healthcare industry is the only industry where internal threats exceed external threats, with leakage of medical data by internal practitioners reaching alarming levels [12].

Security threats to healthcare data space mainly originate from three dimensions: (1) **Technical dimension**: Weak innovation capability in data governance technology [13], security vulnerabilities in medical institution information systems and equipment, functional attenuation of traditional encryption technology in protecting diagnosis and treatment data, outdated systems not updated in time easily forming high-risk attack surfaces, and medical data being vulnerable to hacker eavesdropping, tampering, and interception during transmission; (2) **Management dimension**: Incomplete management system construction in medical institutions, weak risk management awareness, chaotic internal permission management [14], insufficient safety awareness among staff who may obtain permissions beyond their work scope, increasing the risk of unnecessary data access and leakage; (3) **Legal dimension**: Ambiguous definition of medical data ownership, imperfect data rights allocation system and relevant laws and regulations [15], loopholes in the normative use and supervision methods of data security, and current penalty standards hardly covering new types of data crimes, making it difficult to effectively curb illegal behaviors.

2.3 Innovation Opportunities from Technology-Policy Synergy

Currently, the operation of healthcare data space presents an evolutionary trend of technology-driven and policy-driven interweaving. Globally, data sharing models are transforming from centralized storage to distributed collaborative healthcare data space concepts, with core technological breakthroughs focusing on privacy-preserving computation, cross-chain interoperability, and federated learning fusion applications that support this transformation [16]. Foreign healthcare data space is developing rapidly, with regions such as the United States and the European Union leading in healthcare data space standards and ecosystem construction thanks to advanced information technology and mature market environments. The United States promotes policies such as the "21st

Century Cures Act" [17] to encourage medical institutions to share data, thereby forming a broader medical data sharing ecosystem, promoting the application of aggregated medical data resources in drug research and development and clinical decision support system development, and helping early screening and precision diagnosis in healthcare. The European Union, through the "European Health Data Space (EHDS)" strategy, breaks data barriers among member states, achieves circulation of cross-border medical data under unified rules and security frameworks, and promotes multi-center clinical research and public health collaborative governance [18]. For example, the joint Alzheimer' s disease research project between France and Germany relies on the EHDS trusted data space infrastructure to integrate data from multiple medical institutions in both countries, accelerating disease mechanism research and new drug development processes. Similar successful cases include the application of France' s national health data system (data repository) and patient electronic medical records (data sources), whose research helps explore connections and potential for improving real-world research quality [19]. In addition, foreign countries are relatively mature in supporting secure operation and privacy protection of data space, such as the United States' "Health Insurance Portability and Accountability Act (HIPAA)" [20-21] which strictly regulates medical data use, and the European Union' s "General Data Protection Regulation (GDPR)" [22] which strengthens data subject rights protection, regulates the collection, storage, and use of personal health data, and supervises various aspects of online personalized services, providing a solid legal foundation for the development of trusted healthcare data space industries.

China' s healthcare data space development is currently in a period of policy-driven and technology-integrated dual opportunities. The "Guiding Opinions on Comprehensively Promoting the Construction of Compact County-level Medical Communities" issued in 2024 [23] explicitly proposes promoting interconnection of medical data within counties, building unified data sharing platforms, and promoting hierarchical diagnosis and treatment and resource integration. The timely release of the "National Data Standard System Construction Guide" [24] lays a solid foundation for cross-institutional data interoperability by unifying medical data concept definitions, technical specifications, and security frameworks, covering seven key areas including data infrastructure, resource integration, and circulation application [25]. Policy documents such as the "Healthy China 2030" Planning Outline [26] also clearly propose promoting medical information construction, laying technical and management foundations for data space.

With continuously increasing policy support, national-level institutions play key roles in promoting integration and application of medical data. In 2025, the Big Data Center of the National Healthcare Security Administration was officially inaugurated, undertaking the application, management, and service tasks of national medical insurance data, covering massive data from 1.33 billion insured individuals and 930,000 hospitals and pharmacies, assisting medical insurance reform and the development of the pharmaceutical and health industries. Data

application in the medical field focuses on assisted diagnosis, patient virtual assistants, and medical image analysis [27]. In the clinical diagnosis and treatment field, multiple top-tier domestic hospitals have established smart medical platforms, using big data and AI to achieve disease prediction and precision diagnosis and treatment; in the public health governance field, by establishing multi-modal epidemic monitoring systems integrating case data and epidemiological survey data, hierarchical precise prevention and control and efficient resource allocation are achieved. These digital applications continue to break through in the medical field, achieving the leap from "data resource accumulation" to "value creation," comprehensively empowering the digital and intelligent transformation of the healthcare industry [28]. From an overall ecosystem perspective, China is currently at the stage of building various data sharing platforms, aggregating medical data resources, and evolving toward constructing a secure, open, and interoperable healthcare data space.

3 Major Bottlenecks in Developing Healthcare Data Space

3.1 Data Storage and Sharing Bottlenecks

The accelerated development of medical informatization has made real-time processing of massive data, security guarantee capabilities, and long-term stable operation core challenges. Currently, medical institutions face dual pressures of soaring algorithm complexity and increasing communication costs in data storage and processing, with single-machine storage systems no longer meeting actual storage and computing needs, making optimization of data storage methods key to solving problems [29]. Simultaneously, to ensure data security and business continuity, data storage systems need to implement "local + remote" dual real-time backup mechanisms, meaning that even under extreme circumstances such as server damage or cyber attacks, data recovery can be completed in a short time to avoid interruption of diagnosis and treatment services.

However, at the hardware deployment level, regulatory, technical, and human constraints remain major obstacles [30]. Medical institutions have significant shortcomings in hardware conditions, with extremely prominent contradictions between performance requirements and cost control. To efficiently handle computationally intensive tasks such as CT image analysis, hospitals need to configure high-performance storage and high-speed network equipment, but primary-level hospitals' old power systems cannot support high-energy-consuming equipment, and the constant temperature and dustproof renovations required for dedicated computer rooms are difficult to implement due to funding shortages. At the same time, the contradiction between exponential data growth and equipment expansion urgently needs to be resolved—top-tier hospitals' storage system expansion requires downtime for data migration. This "expand while operating" model leads to business continuity risks and uncontrolled expansion costs. The training of AI medical large models requires simultaneous access to massive data resources and powerful computing capabilities, further exposing the huge gap between existing hardware architecture in computing power supply and data

processing timeliness [31].

In the data sharing domain, multiple barriers severely restrict the release of medical value. Policy regulations and privacy barriers constitute primary obstacles, with sharing requiring costly anonymization processing and obtaining level-by-level authorization [32], especially in cross-border transmission of sensitive information such as genes, where compliance reviews are becoming stricter, not only making operational processes complex but also increasing risk costs. At the practical level, the blank "rights, responsibilities, and benefits" allocation mechanism leads to missing guarantees for contributors' benefits, with hospitals worrying about legal risks and competitive advantage loss from data leakage [33], patients concerned about privacy abuse, and pharmaceutical companies and research institutions unable to establish sustainable benefit distribution models, forming a vicious cycle of "unwilling to share, not daring to share" through multi-party game. Simultaneously, insufficient technology empowerment directly limits implementation paths, with key technologies such as privacy-preserving computation (e.g., federated learning) and blockchain not yet breaking through performance bottlenecks and large-scale application thresholds, making it difficult to balance secure sharing and efficient collaboration, and unable to effectively support scenarios such as clinical diagnosis and treatment and cross-domain research.

Ultimately, storage and sharing bottlenecks create superimposed effects. Performance limitations of storage systems hinder high-concurrency data access and real-time analysis, while imperfect sharing mechanisms prevent aggregated utilization of scattered data resources, further amplifying hardware computing power gaps. This coexistence of "storage silos" and "sharing shackles" severely delays the transformation of medical data from resources to productivity.

3.2 Data Standardization and Quality Bottlenecks

Data sources for healthcare data space are highly complex, covering clinical diagnosis and treatment, public health, patient-generated data, and scientific research exploration, showing significant multi-source heterogeneous characteristics. Main sources include: internal medical institution data (electronic health records, medical imaging, laboratory data forming the core source); patient-generated data (wearable devices, mobile health applications, IoT and smart devices); and public health data (disease surveillance data, population health data supporting macro health policy formulation). However, the lack of a data standardization system leads to integration difficulties: disease coding lags due to regional differences, with high disease classification error rates; structured electronic medical records coexist with unstructured texts, with low matching between manually entered data and system fields; primary-level medical institutions become weak links in multi-center data collaboration due to lack of unified standard interfaces and significant differences in regional resource allocation [34].

Data quality directly affects the performance of machine learning models and the reliability of medical decisions [35]. Current medical data has prominent issues such as authenticity bias, contamination, and missing data: some intelligent models produce distorted outputs due to training data contamination; data missing is particularly common in primary-level scenarios, with prominent issues beyond equipment failure including patient compliance, follow-up mechanisms, and cooperation with medical institutions [36]; lagging data processing technology, historical data compatibility, and missing equipment maintenance generating noise further reduce data availability. In addition, primary-level institutions have not yet established full-process quality management systems, with frequent source problems such as non-standard filling of medical record home pages and chaotic medical order sets, continuously amplifying risks of data extraction and analysis bias.

Primary-level medical and health services, as the "safety net" for ensuring public health needs and improving national health levels, have data governance capabilities directly related to the accessibility and precision of public health services [37]. Currently, primary-level healthcare faces multiple structural shortcomings in data governance: (1) Traditional management models with limited integration with medical data, lacking refined data support leading to insufficient procedural and standardized processing; (2) Exacerbated medical equipment data standardization issues, with primary-level ECG monitoring and other devices having inconsistent technical parameters such as storage formats and sampling frequencies, high proportion of old equipment, and insufficient collection accuracy; (3) Disconnection between patient-generated data and professional medical data, with wearable device accuracy deviations and weak calibration capabilities making multimodal data integration difficult, often unable to achieve format conversion and calibration due to technical and funding constraints. These challenges highlight systematic defects in primary-level medical data from collection, storage to application, urgently requiring digital solutions adapted to primary-level through technology empowerment and resource allocation.

3.3 Talent Cultivation Bottlenecks

Healthcare data space construction is constrained by the scarcity of cross-domain composite talents, requiring government-industry-academia-research collaboration to cultivate and strengthen core drivers of technology research and development to break through development bottlenecks. Shen Huiwen et al. [38] point out that hospitals generally lack composite talents who understand both clinical practice and data analysis. Current talent shortages mainly focus on three dimensions: (1) **Capability demand gap**: The smart healthcare field urgently needs composite talents with knowledge of clinical medicine, data science, and ethical regulations, but current university curricula are disconnected from clinical practice, lacking interdisciplinary training programs (such as dual degrees or joint training) in cooperation with medical institutions, with developing countries more significantly constrained by educational resources

and insufficient industry-education integration, resulting in more significant composite talent gaps than developed countries and restricting sustainable development of healthcare data space; (2) **Lagging education system**: Talent cultivation mechanisms lack innovation. In the AI era, medical education and physician training are facing major transformation [39], and should rely on digital intelligence technology to build personalized learning paths and intelligent education management systems [40] to meet society' s demand for innovative and composite talents and promote high-quality education development; (3) **Missing career pathways**: Traditional professional title evaluation systems focus on single-discipline capabilities, lacking evaluation standards for interdisciplinary composite talents, causing talents with knowledge of clinical medicine, data science, and ethical regulations to face structural obstacles in promotion. There is urgent need to reconstruct professional title evaluation dimensions, add smart healthcare special positions, and establish interdisciplinary capability certification systems to adapt to diversified talent needs in healthcare data space.

3.4 Ethical Dilemmas and Governance Paradoxes in Healthcare Data Space

In the construction process of healthcare data space, ethical contradictions and governance challenges show multi-dimensional interweaving: data ethics research is still in the early exploration stage, with academia not only having unclear cognition of its ethical essence but also showing fragmented characteristics in related research systems [41], while the ambiguity of medical data property rights definition further restricts sharing motivation—the reproducibility feature of data leads to weakened patient right to know in secondary use, and imperfect compliance guidance mechanisms cause data ownership to face loss risks during circulation [42]; simultaneously, differences in multi-stakeholder interest demands and legal absence of rights distribution mechanisms form structural contradictions. Although technological evolution such as blockchain and federated learning improves processing efficiency, it simultaneously generates new problems such as fragmented property rights and rising circulation costs, ultimately constituting a governance paradox of "technology empowerment and institutional lag." The data sharing process also exacerbates leakage risks and conflicts in cross-border flow rules, forcing governments into a balance dilemma between innovation incentives and security supervision: overemphasis on data encryption requirements may hinder scientific research and clinical collaboration, while insufficient supervision easily triggers security risks. This governance deadlock of "chaos without control, death with control" highlights that policy coordination needs to construct a dynamic balance mechanism for security, efficiency, and fairness, promoting the transformation of data from resources to production factors through multi-dimensional institutional design, thereby supporting the realization of precision medicine and universal health goals.

4 Innovative Pathways for Promoting Healthcare Data Space Construction

4.1 Constructing a "Trinity" Ecosystem Framework for Collaborative Innovation

As the cornerstone of modern healthcare system transformation, healthcare data space construction is releasing enormous social value and economic potential through technological innovation and ecosystem reconstruction. With the deepening of the "AI+Healthcare" initiative, applications such as intelligent assisted diagnosis and chronic disease management are accelerating the dissemination of high-quality medical resources to grassroots levels. Against this backdrop, the industry is constructing a "technology-management-law" trinity collaborative framework, which will help healthcare data space applications achieve leapfrog development in areas such as precision medicine popularization and chronic disease management intelligence.

Technology dimension: Focus on privacy-preserving computation technologies, federated learning, and other "available but invisible" technologies to break through data silo barriers, deepening the application of federated learning and quantum computing to break through dynamic data protection and computing power limitations. Healthcare data security has developed encryption and anonymization technologies. Xu Cheng et al. [43] constructed an IoT ciphertext index storage scheme that reduces leakage risks through distributed encryption. Song Kai [44] proposed a cross-chain privacy protection scheme based on group signature algorithms, effectively solving privacy issues in medical data cross-chain sharing. Simultaneously, with research on optimization methods for massive small file storage in medical data [45], storage efficiency and data security have been improved.

Management dimension: Medical institutions upgrade internal governance systems, with medical staff following relevant laws, regulations, and ethical guidelines while protecting patient privacy. Currently, medical institutions establish data security responsibility systems, strengthen security awareness and permission management among practitioners, and regularly conduct drills to identify vulnerabilities. Through technical means (such as log auditing) and behavioral norms (such as prohibiting random photography), digital intelligence technology is integrated into performance management to improve hospital performance management effectiveness and levels [46], strengthening security management capabilities and reducing leakage risks. Simultaneously, promote coordination of the three medical sectors (healthcare, medical insurance, and pharmaceutical R&D), break through barriers in diagnosis and treatment, medical insurance, and R&D through healthcare data space, and drive policy implementation through data sharing to improve health service effectiveness.

Legal dimension: Data security and patient privacy are core priorities. The current legal system shows a dual-layer evolution characteristic of "basic legislation + industry norms." The "Personal Information Protection Law" and "Data

Security Law" provide legal foundations for medical data protection but require further refinement of industry implementation rules [47]. For example, clarifying classification and grading standards and management models for medical data, listing genetic data and electronic medical records as the highest sensitivity level, and stipulating principles for data sharing. Ge Yongbin et al. [48] research shows that establishing a data circulation "whitelist" system can effectively reduce 违规访问率. Simultaneously, referencing the EU' s "Health Data Space Act" to establish relevant legal regulations can improve the full-chain data governance system [49].

4.2 Path Planning for Accelerating Digital Infrastructure

Healthcare data space governance requires constructing a "technology-policy" dual-axis collaborative framework (Figure 1 [Figure 1: see original paper]). The technology side relies on privacy-preserving computation and blockchain to build a trusted data circulation network, breaking through silo barriers; policy builds compliance bottom lines through tiered authorization and cross-border mutual recognition systems. The two achieve dynamic adaptation through smart contracts: federated learning supports cross-domain desensitization sharing of diagnosis and treatment data, and sandbox supervision balances AI medical innovation and risk control. The framework' s bottom layer provides computing power support through new infrastructure, while the upper layer forms a closed-loop ecosystem of data rights confirmation, circulation, and application, both ensuring security protection of highly sensitive information such as genetic data and promoting value release in scenarios such as chronic disease management, ultimately achieving safe and efficient circulation of data elements.

Optimize capital allocation: Central finance focuses on core infrastructure such as national medical supercomputing centers and 5G medical private networks; local supporting funds support digital upgrading of medical systems. The state can introduce relevant policies, establish special funds for medical data new infrastructure, support deployment of intelligent analysis platforms, guide medical institutions to purchase dedicated hardware through financial subsidies, and include software systems such as data middle platforms and blockchain nodes in government procurement catalogs. For example, project planning requires newly built top-tier hospitals to configure edge computing gateways and federated learning modules to achieve real-time desensitization and on-chain storage of diagnosis and treatment data [50], and incorporates healthcare data space construction indicators into smart hospital rating systems, driving institutions to undergo intelligent transformation.

Integrate innovative technologies: Policies should promote the establishment of special zones for medical AI large model training, authorizing platforms such as DeepSeek to use desensitized medical data for reinforcement learning in special zones, while simultaneously constructing a "sandbox supervision" mechanism to dynamically verify clinical recommendations output by models. By revising relevant legal regulations, clearly authorize compliant institutions to use

federated learning technology for cross-border research, and reduce or exempt data 出境安全评估费用 for data circulation using privacy-preserving computation technology. Form an ecological cycle of "policy traction - technology drive - data value," making healthcare data space truly become the digital foundation for Healthy China construction.

Talent cultivation pathways: Currently, medical institutions are strengthening cooperation with universities and enterprises by jointly building laboratories or innovation centers; on this basis, encourage enterprises and universities to jointly build laboratories and actively explore "dual-mentor" talent cultivation models. Simultaneously, improve existing medical staff' s digital skills through vocational training [51], establish a "data craftsman" cultivation mechanism covering cross-domain capabilities such as clinical business understanding, data cleaning, and algorithm optimization [52], and the government can implement talent introduction policies by establishing overseas high-level talent special funds to attract composite talents with multi-professional backgrounds.

Strengthen international cooperation: China can actively participate in the formulation of global healthcare data governance rules, promote cross-border flow of medical data, establish mutual recognition mechanisms for cross-border data flow, and improve the security and efficiency of data interaction. For example, sign data sharing agreements with countries along the "Belt and Road" to promote international collaborative research on medical data, while conducting full lifecycle health service research [53-54] to improve global medical service levels, enhance China' s discourse power in the digital health field, and ultimately promote the global healthcare system toward a more efficient, fair, and sustainable direction.

5 Discussion and Outlook

Currently, healthcare data space is accelerating its evolution toward intelligent collaboration and global interconnection, with its core driving force stemming from dual breakthroughs in heterogeneous technology integration and institutional innovation. At the technical level, the deep coupling of federated learning and edge computing is building a distributed diagnosis and treatment network of "data available but invisible," supporting real-time data desensitization sharing for multi-center clinical research. Simultaneously, the collaborative application of blockchain and privacy-preserving computation can reduce dependence on single intermediaries, alleviating sovereignty disputes in cross-border data flow [55], such as the EU' s EHDS achieving property rights tracking and compliant invocation of medical data among member states through smart contracts, providing technical reference for China' s participation in global health governance.

However, technological leaps still face dual challenges of lagging data standardization and computing power bottlenecks. Primary-level medical institutions' equipment heterogeneity leads to chaotic data collection formats, requiring reliance on a "cloud-edge-end" integrated architecture to achieve multimodal data

fusion and AI-assisted annotation tools to improve structuring efficiency. At the policy level, it is necessary to strengthen top-level design: on one hand, clarify circulation boundaries for highly sensitive information such as genetic data and electronic medical records through legislation, and establish dynamic sandbox supervision mechanisms to balance innovation and risk; on the other hand, optimize "new infrastructure" capital allocation, tilting toward 5G medical private networks and edge computing node construction in central and western regions to narrow the regional digital divide.

In the future, healthcare data space will present three major trends: (1) **Technology-clinical closed-loop iteration**: Quantum encryption technology breaks through traditional computing power constraints, enabling cross-domain medical data to complete desensitization and feature extraction at millisecond level, empowering real-time optimization of prediction models; (2) **Policy-market co-evolution**: Drawing on other countries' data management models, China can rely on public-private partnership mechanisms to activate social capital, directing it toward public welfare data application scenarios such as chronic disease management and rare disease research; (3) **Global-regional governance nesting**: Relying on the "Belt and Road" medical data alliance, build cross-border data mutual recognition and joint R&D ecosystems. Through the "technology-policy" collaborative governance framework, promote the paradigm upgrade of medical data resources from production factors to global public goods, jointly building a trusted, controllable, and sustainable digital foundation for a global community of health for all, providing strong support for digital transformation in the medical industry and improving global medical service levels.

Author contributions: Wang Hongchuan proposed the research question, constructed the theoretical framework, and conducted in-depth exploration and analysis of the theme; Zhang Jianbo was responsible for data collation, extracting key themes and viewpoints, and writing the paper; Ma Wei proposed relevant policy recommendations, was responsible for final version revision, and is accountable for the paper; Zhao Sidi was responsible for collecting materials and data, and conducting analysis and proofreading.

Conflict of interest: None declared.

ORCID:
Wang Hongchuan https://orcid.org/0000-0001-8629-0982
Zhang Jianbo https://orcid.org/0009-0001-3422-3443

References

[1] CONCATO J, CORRIGAN-CURAY J. Real-world evidence - where are we now? [J]. N Engl J Med, 2022, 386(18): 1680-1682. DOI: 10.1056/NEJMp2200089.

[2] STELLMACH C, MUZOORA M R, THUN S. Digitalization of health data:

interoperability of the proposed European health data space [J]. Stud Health Technol Inform, 2022, 298: 132-136. DOI: 10.3233/shti220922.

[3] State Council Notice on Issuing the "14th Five-Year Plan for Digital Economy Development" (Guo Fa [2021] No. 29) [J]. China Military Civilian, 2022(1): 6-12. DOI: 10.3969/j.issn.1008-5874.2022.01.002.

[4] National Health and Family Planning Commission. "Basic Public Health Service Standards (Third Edition)" [EB/OL]. (2017-04-17) [2025-04-25]. http://www.nhc.gov.cn/ewebeditor/uploadfile/2017/04/20170417104506514.pdf.

[5] Zhang D, Zhang LW. Digital-intelligence information ecosystem: connotation, composition, and mechanism [J]. Modern Intelligence, 2024, 44(4): 11-21.

[6] Wang HW. The emergence logic, practical dilemmas, and construction path of digital health community [J]. Chinese Health Service Management, 2023, 40(12): 881-884, 902.

[7] Huang RY, Jing Q. Digital health in the digital era: connotation, characteristics, challenges, and governance paths [J]. Health Economics Research, 2022, 39(6): 60-63, 66. DOI: 10.14055/j.cnki.33-1056/f.2022.06.019.

[8] "2022-2027 China Digital Medical Industry Market Analysis and Investment Risk Trend Forecast Research Report" [EB/OL]. (2023-10-20) [2025-07-07]. https://www.qianzhan.com/analyst/detail/220/231020-7e8c6a3e.html.

[9] National Development and Reform Commission Notice on Issuing the "14th Five-Year Plan for Bioeconomy Development" [EB/OL]. (2021-12-20) [2025-07-07]. https://www.gov.cn/zhengce/zhengceku/2022-05/10/content_{5689556}.htm.

[10] KPMG China Releases "First Health Technology 50"Report [EB/OL]. (2025-07-02) [2025-07-07]. https://assets.kpmg.com/content/dam/kpmg/cn/pdf/zh/2025/07/kpmg-china-healthcare-health-tech-50.pdf.

[11] China Academy of Information and Communications Technology. "2020 Digital Healthcare: Research Report on Cybersecurity Risks During the Epidemic Prevention and Control Period" [EB/OL]. [2025-04-25]. http://www.caict.ac.cn/kxyj/qwfb/ztbg/202003/P020200316481943325476.pdf.

[12] Verizon. 2020 Data Breach Investigations Report [EB/OL]. (2020-06-20) [2025-04-21]. https://www.secrss.com/articles/20611.

[13] Que TS, Wang ZY. Global data security governance and China's strategy in the digital economy era [J]. International Security Studies, 2022, 40(1): 130-154, 158. DOI: 10.14093/j.cnki.cn10-1132/d.2022.01.006.

[14] Yang ZY. Discussion on internal control of public hospitals based on risk management [J]. Friends of Accounting, 2019(6): 137-140.

[15] Gao FP. On medical data rights allocation—legal framework for medical data open utilization [J]. Modern Law Science, 2020, 42(4): 52-68. DOI: 10.3969/j.issn.1001-2397.2020.04.04.

[16] Zhu JM, Zhang QN, Gao S, et al. Blockchain-based privacy-preserving trusted federated learning model [J]. Chinese Journal of Computers, 2021, 44(12): 2464-2484. DOI: 10.11897/SP.J.1016.2021.02464.

[17] JAFFE S. 21st century cures act progresses through US congress [J]. Lancet, 2015, 385(9983): 2137-2138. DOI: 10.1016/S0140-6736(15)61008-X.

[18] Zhao L, Qian YQ, Zheng H. EU data element market cultivation policies, practices, and models [J]. Library Forum, 2024, 44(12): 151-160.

[19] GOFF R L, BRICE S, CONTINI A, et al. Successful linkage of electronic medical records and national health data system in type 2 diabetes research: methodological insights and implications [J]. Pharmacoepidemiol Drug Saf, 2025, 34(2): e70095. DOI: 10.1002/pds.70095.

[20] HUI K R, GILMORE C J, KHAN M. Medical records: more than the health insurance portability and accountability act [J]. J Acad Nutr Diet, 2021, 121(4): 770-772. DOI: 10.1016/j.jand.2020.06.022.

[21] BHATE C, HO C H, BRODELL R T. Time to revisit the Health Insurance Portability and Accountability Act (HIPAA)? [J]. J Am Acad Dermatol, 2020, 83(4): e313-314. DOI: 10.1016/j.jaad.2020.06.989.

[22] HOOFNAGLE C J, VAN DER SLOOT B, BORGESIUS F Z. The European Union general data protection regulation: what it is and what it means [J]. Inf Commun Technol Law, 2019, 28(1): 65-98. DOI: 10.1080/13600834.2019.1573501.

[23] Primary Health Care Department. Guiding Opinions on Comprehensively Promoting the Construction of Compact County-level Medical and Health Communities [EB/OL]. (2023-12-30) [2025-04-20]. http://www.nhc.gov.cn/jws/s7874/202312/e5d16e73fa324533bcc

[24] CPC Central Committee and State Council Issued the "National Data Standard System Construction Guide" [EB/OL]. (2024-10-08) [2025-04-20]. https://www.gov.cn/zhengce/zhengceku/202410/P020241008789641651212.pdf.

[25] Yin YY, Yuan LY, Chen MH. Research on health data security sharing mechanism based on blockchain multi-chain [J]. Network Security Technology and Application, 2025(2): 65-70.

[26] CPC Central Committee and State Council Issued the "Healthy China 2030" Planning Outline [EB/OL]. (2016-10-25) [2025-04-20]. https://www.sport.gov.cn/gdnps/files/c25531211.pdf.

[27] Xiao QY, Yu GJ. Research and progress of medical big data [J]. Shanghai Medicine, 2023, 46(7): 420-423. DOI: 10.19842/j.cnki.issn.0253-9934.2023.07.002.

[28] Huang YL. Implementation elements of primary healthcare digital transformation: based on implementation research framework [J/OL]. Chinese General Practice, 2024: 1-10. (2024-08-29) [2025-07-07]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&

[29] He GS, Zhao CL, Jiang JH, et al. Survey on deep learning-oriented data storage technology [J]. Chinese Journal of Computers, 2025, 48(5): 1013-1064.

[30] RANCHON F, CHANOINE S, LAMBERT-LACROIX S, et al. Development of indirect health data linkage on health product use and care trajectories in France: systematic review [J]. J Med Internet Res, 2023, 25: e41048. DOI: 10.2196/41048.

[31] Feng YY, Wang Q, Xie MH, et al. From BERT to ChatGPT: storage system challenges and technological development in large model training [J]. Journal of Computer Research and Development, 2024, 61(4): 809-823.

[32] GADOTTI A, ROCHER L, HOUSSIAU F, et al. Anonymization: the imperfect science of using data while preserving privacy [J]. Sci Adv, 2024, 10(29): eadn7053. DOI: 10.1126/sciadv.adn7053.

[33] SHAH S M, KHAN R A. Secondary use of electronic health record: opportunities and challenges [J]. IEEE Access, 2020, 8: 24623-24635. DOI: 10.1109/ACCESS.2020.2971442.

[34] Mei ZH, Liu CJ. Analysis of primary healthcare resource allocation efficiency in China from 2012-2020 [J]. Chinese Health Economics, 2022, 41(10): 54-58.

[35] GONG Y D, LIU G Z, XUE Y Z, et al. A survey on dataset quality in machine learning [J]. Inf Softw Technol, 2023, 162: 107268. DOI: 10.1016/j.infsof.2023.107268.

[36] Li WY, Zhang HZ, Jin H, et al. Research on implementation status of health management in primary healthcare institutions oriented toward active health [J]. Chinese General Practice, 2024, 27(28): 3465-3471. DOI: 10.12114/j.issn.1007-9572.2024.0156.

[37] Shen XL, He RX, Liang WN. Research on configuration and path of factors influencing primary healthcare service performance [J]. Chinese General Practice, 2025, 28(16): 1969-1976. DOI: 10.12114/j.issn.1007-9572.2024.0458.

[38] Shen HW, Ma DY, Zhang C. Medical big data and scientific research practice [J]. China Medical Education Technology, 2023, 37(3): 351-355. DOI: 10.13566/j.cnki.cmet.cn61-1317/g4.202303020.

[39] XU Y Y, JIANG Z H, TING D S W, et al. Medical education and physician training in the era of artificial intelligence [J]. Singapore Med J, 2024, 65(3): 159-166. DOI: 10.4103/singaporemedj.SMJ-2023-203.

[40] QIAN L M, CAO W R, CHEN L F. Influence of artificial intelligence on higher education reform and talent cultivation in the digital intelligence era [J]. Sci Rep, 2025, 15(1): 6047. DOI: 10.1038/s41598-025-89392-4.

[41] Zhang CH, Li ZZ, Pei L. Multidisciplinary intersection and multi-scenario embedding: review of domestic and foreign data ethics research [J]. Journal of Information Resources Management, 2025, 15(2): 91-107. DOI: 10.13365/j.jirm.2025.02.091.

[42] WANG L H, MENG L Y, LIU F K, et al. A user-centered medical data sharing scheme for privacy-preserving machine learning [J]. Secur Commun Netw, 2022, 2022: 3670107. DOI: 10.1155/2022/3670107.

[43] Xu C, Gu L. Centralized storage of medical big data information security based on IoT [J]. Information Technology, 2023, 47(1): 109-114. DOI: 10.13274/j.cnki.hdzej.2023.01.020.

[44] Song K, He LW. Research on blockchain identity privacy scheme for medical big data [J]. Software, 2023, 44(8): 150-152.

[45] Zeng M, Zou BJ, Zhang WS, et al. Optimization method for massive small file storage in multimodal medical data [J]. Journal of Software, 2023, 34(3): 1451-1469. DOI: 10.13328/j.cnki.jos.006710.

[46] Xu RX, Zheng WX, Lin Y. Research on application of medical big data in hospital performance management [J]. China New Telecommunications, 2023, 25(22): 71-73.

[47] Niu L, Huo ZH. Research on regulation of medical big data utilization under the background of Personal Information Protection Law [J]. Chinese Health Law, 2023, 31(4): 11-16. DOI: 10.19752/j.cnki.1004-6607.2023.04.002.

[48] Ge YB, Dong JP. Compliance requirements for conducting real-world clinical research using medical big data [J]. China Food and Drug Administration, 2023(10): 86-94.

[49] RAK R. Anonymisation, pseudonymisation and secure processing environments relating to the secondary use of electronic health data in the European health data space (EHDS) [J]. Eur J Risk Regul, 2024, 15(4): 928-938. DOI: 10.1017/err.2024.67.

[50] Tang K, Zhang GM, Chu SX. Application and practice of big data privacy security based on data desensitization technology [J]. Chinese Journal of Health Informatics and Management, 2022, 19(3): 381-386. DOI: 10.3969/j.issn.1672-5166.2022.03.015.

[51] Deng H, Sun H. Justification and system construction of fiduciary duties of personal medical health data processors [J]. East China University of Political Science and Law Journal, 2025, 28(1): 76-90.

[52] Hu BL, Hu HB. Evolution path and review of domestic and foreign research themes on medical health data [J]. Chinese Health Service Management, 2024, 41(12): 1434-1440.

[53] Meng XW, Li Y, Tian X, et al. Research on capitalization of health medical data from the perspective of data lifecycle [J]. Health Economics Research, 2025, 42(2): 28-31, 36. DOI: 10.14055/j.cnki.33-1056/f.2025.02.007.

[54] Du QJ, Yao WM, Wang D, et al. Concept, connotation, and research progress of full lifecycle health services [J]. West China Medical Journal, 2022, 37(12): 1909-1916. DOI: 10.7507/1002-0179.202209124.

[55] SHAHRIAR RAHMAN M, AL OMAR A, BHUIYAN M Z A, et al. Accountable cross-border data sharing using blockchain under relaxed trust assumption [J]. IEEE Trans Eng Manag, 2020, 67(4): 1476-1486. DOI: 10.1109/TEM.2019.2960829.

(Received: May 30, 2025; Revised: July 8, 2025)

(This article was edited by Kang Yanhui)

Note: Figure translations are in progress. See original paper for figures.