ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202304.00837

Research on Open Governance Models for Medical Data Driven by Major Public Health Emergencies: A Postprint

Authors: Li Jiaxing, Wang Lei, Song Shijie, Zhu Qinghua

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] Through in-depth analysis of the medical data open sharing process in major public health emergencies, this study constructs a multi-stakeholder collaborative medical data open governance model for major public health emergency data governance in China.

[Method/Process] Using a combined approach of literature review and online investigation, this study comprehensively examines the current research status and existing barriers of medical data opening models for major public health emergencies both domestically and internationally, and proposes governance solutions from a practical problem-solving perspective.

[Results/Conclusion] By clarifying the principles for opening medical data in major public health emergencies, decomposing the data open sharing process, proposing a stage-based governance model for medical data open sharing under major public health emergencies, and formulating relevant supporting safeguard mechanisms, this study ultimately presents a feasible medical data open governance model for major public health emergencies.

Full Text

Preamble

Title: Open Governance Model for Healthcare Data Driven by Major Public Health Emergencies

Authors: Li Jiaxing[1], Wang Lei[2], Song Shijie[3], Zhu Qinghua[2]

Affiliations: [1] School of Information Management, Nanjing Agricultural University, Nanjing 210095 [2] School of Information Management, Nanjing University, Nanjing 210023 [3] Business School, Hohai University, Nanjing 211100

Abstract: [Purpose/Significance] Through in-depth analysis of the healthcare data opening and sharing process in major public health emergencies, this paper proposes an open governance model for healthcare data with multi-stakeholder collaborative participation for China's major public health emergency data governance. [Method/Process] Using a combination of literature research and web surveys, this study comprehensively reviews the current research status and existing obstacles of healthcare data opening models for major public health emergencies both domestically and internationally, and proposes governance solutions from a practical problem-solving perspective. [Result/Conclusion] By clarifying the principles for healthcare data opening in major public health emergencies, decomposing the data opening and sharing process, this study proposes a staged governance model for healthcare data opening and sharing under major public health emergencies, formulates relevant supporting mechanisms, and ultimately presents a practical and feasible healthcare data open governance model for major public health emergencies.

Keywords: major public health emergencies; healthcare big data; data opening; data governance model

Classification Number: G251.2

DOI: 10.13266/j.issn.0252-3116.2022.04.003

The COVID-19 pandemic has persisted worldwide for over two years, continuously generating new variants during its spread and creating enormous pressure for epidemic prevention and control. General Secretary Xi Jinping has pointed out that this pandemic represents the most significant major public health emergency since the founding of the People's Republic of China in terms of transmission speed, infection scope, and control difficulty [1]. Major public health emergencies such as COVID-19 are characterized by their suddenness, public nature, and uncertainty, making prevention and control extremely challenging and easily triggering public panic [2]. Therefore, when responding to major public health emergencies, timely, accurate, comprehensive, transparent, and open healthcare data should be ensured to enhance national image, strengthen societal confidence in fighting the epidemic, and effectively curb disease transmission. However, the high standards for healthcare data opening in major public health emergencies pose severe challenges to conventional healthcare data opening models, and the substantial gap between existing models and ideal models urgently needs to be bridged.

To address this significant issue of healthcare data opening in major public health emergencies, General Secretary Xi Jinping emphasized at the epidemic prevention and control meeting that scientific research efforts should be intensified, and the opening and sharing of epidemic-related data and case materials should be promoted to timely improve prevention strategies and measures [3]. To implement General Secretary Xi Jinping's major deployment on epidemic prevention and control, this study reviews the current status of healthcare data opening models, summarizes international experiences, analyzes existing obsta-

cles to healthcare data opening in China's major public health emergencies, and constructs a healthcare data open governance model for major public health emergencies to enhance the level of healthcare data opening and sharing in China.

2. Healthcare Data Opening Models for Major Public Health Emergencies

2.1 General Model of Healthcare Data Opening

Since the United States launched the open data movement in 2009, the movement has rapidly emerged globally. According to the latest Global Open Data Barometer ranking released by the World Wide Web Foundation founded by Tim Berners-Lee, the United Kingdom and Canada tied for first place, while the United States, as the initiator of open data, ranked ninth, and China ranked twenty-fourth, indicating that China's data opening level still lags behind leading countries. As healthcare is a priority livelihood area, the United States, United Kingdom, Canada, and China have all prioritized it for data opening. The following analysis examines the current status of healthcare data opening models in these four countries.

The United States was the first country to launch a government data platform (data.gov). With the subsequent launch of the specialized health data portal healthdata.gov, the opening of U.S. healthcare data has further accelerated. Currently, the U.S. government has released the largest number of open healthcare data resources globally. In terms of data source institutions, healthdata.gov primarily includes the Centers for Medicare & Medicaid Services, the Centers for Disease Control and Prevention, and the Food and Drug Administration [4]. Corresponding to the functions of these data publishing institutions, data themes predominantly focus on medical insurance and subsidy-related data (36%), followed by health expenditure-related data (11%) [5]. Regarding data update frequency, most healthcare data are statistically compiled annually, though some years have missing data. In terms of platform features, healthdata.gov provides not only data retrieval and multi-dimensional classification navigation but also developer interfaces and problem feedback sections. Based on Tim Berners-Lee's five-star model for rating open data, nearly 80% of data on the healthdata.gov platform are access-restricted data (0-star), while openly accessible healthcare data are mostly structured data (3-star) [5].

For the United Kingdom, which consistently ranks among the best globally in government data preparedness, the government has led the establishment of the national NHS program in the healthcare domain. This system has created an online platform embedding high-quality standards and experience in information, along with certification mechanisms enabling citizens and social groups to evaluate government progress [6]. The UK's healthcare open data portals include data.gov.uk and NHS Choices. In terms of data source institutions, healthcare datasets on data.gov.uk primarily come from the Health and Social Care

Information Centre, the Department of Health, and the Scottish Information Services Division [7]. However, healthcare data themes mainly focus on health management and quality monitoring data (38%) and demographic data (32%). Additionally, most healthcare open data lack fixed update frequencies. Regarding platform features, the UK's healthcare open data platform provides data publishing portals, operation manuals, and API interfaces to facilitate quick publishing, acquisition, and utilization of data by data publishers and users. In terms of data openness, nearly 70% are access-restricted data (0-star), while openly accessible data are also mostly structured data [5].

Canada's first unified open data website, www.open.canada.ca/en, was established in 2011. The Canadian government also partnered with Health Infoway to build a nationwide electronic health system providing accurate personal electronic health records to all Canadian residents, with healthcare institutions accessing these records after authorization. Patient registration and management can operate across provinces, and the privacy and security of shared information can be guaranteed [8]. Major source institutions for Canada's healthcare open data include the Canadian Food Inspection Agency, Health Canada, and the Public Health Agency of Canada. Canada's healthcare open data themes mainly include food nutrition, diseases and symptoms, and healthy living. Like other countries, Canada's healthcare open data lacks fixed update frequencies. Regarding platform features, Canada's open data platform emphasizes user experience, integrates social media functions, supports commenting, proposing data resources of interest, uploading apps, and various other forms of user interaction [9].

As a relatively late participant in data opening, China has also conducted some exploration and practice in healthcare data opening. Shanghai pioneered the launch of the "Government Data Service Network" in 2012, marking the beginning of China's government data opening platforms. In 2015, the State Council's "Outline for Promoting Big Data Development" proposed establishing a unified national government data opening platform [10], prioritizing the reasonable and moderate opening of public data resources in transportation, healthcare, and other fields. The platform has now been deployed in over 50 cities. In terms of data source institutions, taking Shanghai's public data opening platform as an example, these mainly include the Municipal Market Supervision Administration, Municipal Health Commission, and Municipal Statistics Bureau. Regarding data themes, China's platforms primarily include public health data (direct health reporting data sharing and community chronic disease health management data sharing) [11], clinical data (medical treatment information queries, research data acquisition, and medical consortium platform construction) [12-13], and medical insurance data (interaction between medical insurance information and hospital information). For update frequency, most datasets are annual, though some years have missing data. Platform features include data classification and retrieval, data visualization, and interactive communities, along with industry application cases for practical scenarios. Additionally, numerous academic studies have been conducted based on these healthcare open data, includ-

ing theoretical strategy proposals (establishing shared data catalogs, platform system construction, etc.) [9, 14] and technical solution research (blockchain technology, cloud computing, etc.) [15-16].

In summary, the United States, United Kingdom, and Canada have all established unified national data opening platforms with varying data source institutions and themes, while data updates mostly lack fixed frequencies. Additionally, these countries' healthcare data opening websites incorporate many comment, interaction, and feedback modules to enhance user engagement beyond retrieval and query functions. Due to its late start in data opening, China's unified national platform is still under construction and promotion, with many cities yet to establish fixed data opening platforms. However, regions like Shanghai and Beijing have abundant healthcare open data resources spanning hospital internal data and social medical insurance health data, with platforms incorporating user interaction modules and scenario-based cases beyond traditional retrieval functions.

2.2 Healthcare Data Opening Model for Major Public Health Emergencies

The emergence of COVID-19 has drawn attention to data opening in major public health emergencies. M. Laessig has proposed that open healthcare data plays a crucial role in major public health emergencies, from early warning of outbreaks and pandemics to providing personalized medicine for individuals and supporting health system management—every stage of epidemic prevention and control relies on massive open healthcare data support [17]. Furthermore, the opening of healthcare data has proven highly effective during earlier Ebola [18] and Zika virus outbreaks [19].

To address the COVID-19 pandemic, both the WHO and the Allen Institute for AI have made free academic articles on COVID-19 publicly available for the global research community [20]. The Allen Institute's resources include over 51,000 articles, aiming to mobilize researchers to apply natural language processing for analysis to generate new insights supporting the fight against COVID-19.

COVID-19-related healthcare open data themes mainly include five categories: infection cases, policies, population mobility, social media, and scientific research literature. Infection case data platforms include the 2019-nCoV time-series infection data repository, an open dataset launched by DXY after the COVID-19 outbreak focusing on reported cases in China, updated hourly via Python web scraping. The Johns Hopkins Center for Systems Science and Engineering also released a COVID-19 data repository [20] with web-based interactive features for real-time global case tracking. China, the United States, the United Kingdom, Japan, and other countries have reported COVID-19 cases at their secondary administrative divisions [21]. GNU Health is a digital health ecosystem that provides epidemiological surveillance reports to study the inci-

dence and prevalence of infectious diseases and has collaborated with the European Commission and Argentine government in COVID-19 governance [22].

Regarding policy-related data opening, Oxford's COVID-19 Government Response Tracker systematically collects information on different countries' policy responses and scores the stringency of these measures, providing a cross-national, cross-temporal policy display platform to understand how government responses evolve during outbreaks. The Kaiser Family Foundation has published U.S. state policy actions including stay-at-home orders, traveler quarantines, business closures, large gathering bans, and school closures.

Population mobility data is obtained by tracking people's movements through location devices and travel information, which helps explore spatial trends in virus transmission. The University of Virginia's Dataverse released population mobility data using de-identified and aggregated mobile phone data, air passenger itinerary data, and case reports to assess COVID-19 transmission risk in China. Descartes Labs' M.S. Warren and S.W. Skillman used anonymous mobile device locations to measure virus transmission risk [23].

Although contact tracing for COVID-19 is important in infectious disease investigations, publicly released mobility data should not contain personally identifiable information. The WHO has also issued ethical guidelines for contact tracing applications, prohibiting the use of publicly generated personally identifiable data [24].

Social media data related to COVID-19 can both reveal misinformation and unverified rumors about the epidemic and help understand public fear, panic, and other social emotions. One Twitter dataset related to COVID-19 [25] includes tweets collected since January 22, 2020, based on keywords such as coronavirus, covid-19, covid19, sars-cov-2, and tracks relevant accounts like CoronaVirusInfo, V2019N, and WHO. The dataset covers tweets from worldwide but does not include user location information. Scholars can use this dataset to analyze issues such as the spread of COVID-19-related rumors and public emotions during the pandemic.

2.3 Comparative Analysis

Comparing the healthcare data opening model under major public health emergencies with general healthcare data opening models reveals several key differences:

2.3.1 Greater Emphasis on Real-time Healthcare Data Opening in Major Public Health Emergencies In major public health emergencies, society faces more severe situations with more frequent changes. Therefore, providing more real-time and timely data enables more accurate analysis of the current epidemic and understanding of its evolution, government policies, and public emotions over time. In general healthcare data opening models, even countries with high data openness levels like the U.S., U.K., and Canada

struggle to achieve real-time dynamic data updates. However, during COVID-19, both infection case data and policy-related data opening have demonstrated real-time data openness.

2.3.2 Greater Focus on Personal Privacy Protection in Healthcare Data Opening and Sharing During Major Public Health Emergencies Personal privacy security has always been a non-negligible risk in healthcare data opening. In major public health emergencies, as data volume increases and potential data value grows, the risk of privacy breaches also increases. Therefore, personal data must be desensitized and de-identified to form datasets without privacy information, achieving a balance between privacy protection and data opening. COVID-19 population mobility open data and social media open data both reflect the importance of personal privacy protection in healthcare data opening.

2.3.3 Broader Scope of Healthcare Data Opening in Major Public Health Emergencies In general healthcare data opening models, countries have different data sources and themes based on their actual conditions and have built independent data opening platforms. However, major public health emergencies often occur across large spatial ranges, with the current COVID-19 pandemic spanning the globe, requiring enhanced international cooperation and broader data interoperability and sharing to jointly combat the epidemic. Currently, some cross-national and even global data opening platforms have emerged for infection cases, related policies, and social media data, enabling different countries and regions to share experiences, conduct academic exchanges, and collaborate on vaccine development based on these platforms.

3. Obstacles to Healthcare Data Opening in Major Public Health Emergencies

The Fifth Plenary Session of the 19th CPC Central Committee explicitly proposed accelerating digital development and promoting the development and utilization of data resources [26]. Activating the circulation of healthcare data resources in major public health emergencies is of strategic significance for effectively controlling such events. Participants in the healthcare data circulation process are mainly divided into data controllers and data users. Data controllers collect and generate healthcare data and provide it to data users. China's healthcare data circulation process in major public health emergencies involves data controllers submitting collected, generated, and processed data to opening platforms, from which data users obtain, process, and use the data, as shown in Figure 1 [Figure 1: see original paper]. This demonstrates that China's healthcare data opening system for major public health emergencies comprises multiple stakeholders; therefore, this study analyzes existing obstacles from a multi-stakeholder perspective.

3.1 Data Controllers

3.1.1 Government Departments In China's healthcare data opening system for major public health emergencies, the government plays a dominant role, coordinating inter-departmental collaboration and mobilizing broad social participation in "big data anti-epidemic" efforts. During this process, the government uses big data to effectively support joint prevention and control of major public health emergencies and achieve precise policy implementation, demonstrating big data's important role in epidemic prevention and control. However, government departments have also revealed shortcomings in opening and sharing healthcare data for major public health emergencies.

(1) **Vertical integration issues across government levels.** When facing major public health emergencies, the central government bears the function of overall epidemic planning, while local governments must implement central strategies and undertake regional epidemic planning and bottom-line guarantee work [28]. During the two-year fight against the epidemic, China's vertical governance system has exposed problems. In data acquisition, significant differences exist in local government informatization levels, with some local governments underreporting, concealing, or misreporting data, preventing the central government from accurately understanding the epidemic situation. In open data utilization, local and grassroots governments often focus only on their regional epidemic data, lacking a holistic perspective. Some local governments implemented "one-size-fits-all" policies against provinces with outbreaks, impacting national economic construction and public daily life—this reflects ineffective utilization of open data and inability to precisely locate epidemic scope.

(2) **Inadequate horizontal collaboration among departments.** Government departments at all levels handle massive administrative affairs daily, exposing resource allocation shortcomings when facing major public health emergencies. Departments often need to transmit epidemic-related data to higher-level governments and peer health commissions, with epidemic data reporting consuming substantial human resources and time. Repeated reporting significantly increases workload, causing problems in departmental daily operations. Additionally, data sharing barriers exist among government departments. When facing major public health emergencies, departments often must wait for higher-level governments to integrate comprehensive data and issue prevention and control instructions before preparing response measures, creating a lag compared to epidemic development speed and affecting prevention and control efforts.

3.1.2 Medical Institutions Medical institutions occupy an important position in the healthcare data opening system for major public health emergencies, undertaking critical tasks of disease detection, diagnosis, and treatment, and capable of obtaining massive first-hand data. By reporting data to higher authorities, they enable the central government to have a clear and accurate

understanding of the epidemic, thereby proposing effective prevention and control measures. However, significant disparities currently exist in informatization levels among medical institutions, and loose integration among hospital departments characterizes their internal structure, which is insufficient to withstand the impact of major public health emergencies.

(1) **Numerous "data silos" within medical institutions.** Data silos refer to data in a fragmented state, forming isolated islands that are mutually isolated [29]. Healthcare data is difficult to process, involving massive, non-interoperable, and sensitive files. Taking COVID-19 as an example, severe patients often develop a series of complications requiring collaborative diagnosis and treatment across hospital departments. However, due to strict privacy protection policies for healthcare data, patient diagnosis and treatment data are often not interoperable among different departments within the same hospital, hindering patient care.

(2) **Uneven construction of management information systems among medical institutions.** When major public health emergencies occur, medical institutions undertake disease diagnosis and treatment tasks while reporting epidemic data. Economically developed regions in China have relatively mature management information systems, having entered regional medical informatization stages, achieving information interoperability among regional medical institutions, between institutions and health epidemic prevention departments, and within institutions, enabling effective responses to major public health emergencies and precise, scientific epidemic prevention and control. In contrast, moderately developed and underdeveloped regions have relatively backward medical institution management information systems, with regional healthcare data integration platforms yet to be established, lower levels of digitalization and intelligence, and inability to timely achieve healthcare data opening and sharing when facing major public health emergencies, failing to meet epidemic prevention and control needs.

3.2 Data Users

3.2.1 Enterprises When major public health emergencies occur, Chinese enterprises actively undertake social responsibilities, making important contributions in tracking epidemic development, debunking epidemic rumors, guiding public opinion, and promoting online work. Major internet companies like Tencent, NetEase, and Sina have built real-time epidemic reporting platforms by obtaining publicly available healthcare data from national and local governments. However, due to the lack of unified data standards among platforms, the comprehensiveness, timeliness, consistency, and readability of epidemic healthcare data cannot be guaranteed, causing inconsistencies in information obtained by the public through internet companies' epidemic data platforms and leading to misjudgments about epidemic conditions. Additionally, massive amounts of epidemic-related healthcare data are stored in enterprise clouds, with en-

terprises responsible for data storage and supervision, posing certain risks to national security and public privacy protection.

3.2.2 Research Institutions Research institutions undertake important functions of virus gene sequencing, vaccine development, and epidemic prediction in responding to major public health emergencies. The Chinese Academy of Sciences rapidly launched three major scientific research data opening and sharing platforms at the beginning of the COVID-19 pandemic, including the 2019 Novel Coronavirus Information Database, the COVID-19 Scientific Literature Sharing Platform, and the National Science and Technology Resource Service System for Novel Coronavirus. Although these three platforms have collected and published numerous epidemic-related research findings from domestic and international scholars, they are still not mainstream international healthcare data sharing platforms, with ownership of mainstream platforms firmly held by the U.K. and U.S. [30]. Additionally, China's research institutions have weak epidemic prediction capabilities, with different research teams employing vastly different prediction methods and lacking a unified research paradigm for high-precision epidemic evolution prediction, unable to provide data analysis support for government decision-making—this reflects low utilization capacity of open epidemic data among Chinese research institutions.

3.2.3 The Public Protecting public life safety is the primary task in major public health emergency response. In the early stages of COVID-19, incomplete understanding of the epidemic led to non-standardized diagnosis and treatment processes, causing a crisis of public trust in local governments and medical institutions. The state effectively resolved this crisis and reshaped public confidence in national epidemic prevention and control through vigorous governance of COVID-19 healthcare data opening issues. In COVID-19 diagnosis and treatment, protecting patient privacy while ensuring public right to know is another urgent issue. Therefore, how to implement long-term mechanisms for public privacy protection and enhance public trust when facing major public health emergencies requires further exploration.

4. Construction of Collaborative Governance Model for Healthcare Data Opening Driven by Major Public Health Emergencies

4.1 Principles for Healthcare Data Opening in Major Public Health Emergencies

Constructing a governance model for healthcare data opening in major public health emergencies should first clarify principles to guide the model's development. In 2013, the G8 released the "Open Data Charter," which clearly defined principles for government data opening, emphasizing that government data should be open, improve governance, be high-quality, usable by all, and promote innovation. Against the backdrop of major public health emergencies,

principles for healthcare data opening should be further clarified. Therefore, based on the Open Data Charter and combined with the context of major public health emergencies, this study proposes principles for healthcare data opening in major public health emergencies, specifically including the principles of responsibility, transparency, quality, and security.

4.1.1 Principle of Responsibility When responding to major public health emergencies, the government holds almost all healthcare data resources and should bear primary responsibility for healthcare data opening, while medical institutions, enterprises, and the public bear secondary responsibility. Healthcare data opening in major public health emergencies requires clarifying the functional positioning of responsible units at all levels throughout the data opening process and defining the regulatory responsibilities of governments and health administrative departments at all levels [31].

4.1.2 Principle of Transparency Healthcare data opening in major public health emergencies should follow the principle of transparency, insisting on fully opening virus gene sequencing data, virus traceability data, and desensitized case data. Through healthcare data opening and sharing, the public's right to know should be protected, social stability maintained, and international image enhanced.

4.1.3 Principle of Quality Authentic and accurate data is the solid foundation for precise decision-making in major public health emergencies and the basis for all anti-epidemic actions. Compared with other types of open data, the quality standards for healthcare open data in major public health emergencies are higher. Therefore, healthcare open data in major public health emergencies should ensure data integrity, accuracy, and consistency, enabling the formulation of standardized diagnosis and treatment plans based on high-quality healthcare data and effectively advancing special drug development.

4.1.4 Principle of Security Healthcare data in major public health emergencies involves massive amounts of public privacy information. Protecting patient privacy and public health data from infringement in the data opening environment is crucial for national security and social stability. In the data opening and sharing process, the Data Security Law should be the benchmark, implementing the security principle for healthcare data opening in major public health emergencies.

4.2 Governance Model for Healthcare Data Opening in Major Public Health Emergencies

With digital technology development, data opening levels continue to improve. However, major public health emergencies impose higher requirements on healthcare data opening levels worldwide. Currently, China's government, medical

institutions, research institutions, enterprises, and the public face a series of challenges in healthcare data opening and sharing. To address these obstacles, this paper proposes an overall design for the governance model.

The healthcare data opening and sharing process in major public health emergencies has staged characteristics, specifically divided into data acquisition, data opening, and data sharing stages. The data acquisition stage includes data collection, cleaning, and processing; the data opening stage covers data publication and management; the data sharing stage includes data flow and development/utilization. Healthcare data opening and sharing in major public health emergencies is a process system led by the government, with collaboration among medical institutions, research institutions, and enterprises, and public participation, involving multiple social stakeholders with interconnections. Therefore, governance of healthcare data opening in major public health emergencies should, based on clarifying the data opening and sharing process, form a vertical collaborative governance structure centered on government, social organizations, and the public.

Figure 2 [Figure 2: see original paper] illustrates the governance model for healthcare data opening driven by major public health emergencies.

4.2.1 Governance of Healthcare Data Acquisition in Major Public Health Emergencies The beginning of healthcare data opening governance in major public health emergencies is data acquisition governance, aiming to ensure the accessibility, integrity, and usability of emergency healthcare data [32]. Data acquisition is the foundation of data opening and sharing. The data acquisition governance stage should focus on data acquisition standards and storage security, strengthening public privacy protection. Governments and their departments and medical institutions, as healthcare data controllers, should improve data acquisition efficiency while strengthening and standardizing security issues arising during emergency healthcare data acquisition. Therefore, this study proposes three aspects for governing healthcare data acquisition: establishing data acquisition standards, clarifying data acquisition boundaries, and improving data acquisition policies.

During emergency response, extensive collection of healthcare data can provide a comprehensive understanding of the epidemic situation and public health levels. The data acquisition process involves massive public privacy and social security issues. How to standardize data acquisition and avoid "overreach" is an urgent problem. Standardization of emergency healthcare data collection is the prerequisite for subsequent governance stages including data storage, processing, and services. Data opening governance involves governments and departments, medical research institutions, enterprises, and the public, requiring standardization and normalization from the data collection source. As the guide for emergency healthcare data acquisition, the government should formulate detailed rules for healthcare data acquisition, including standardized data types and unified data acquisition standards. Currently, emergency healthcare data acquisition stan-

dards remain unclear. For major participants in healthcare data collection such as medical institutions and grassroots governments, a unified data acquisition standard system should be established to pursue data consistency and ensure original data quality. Based on the Data Security Law, legislation specifically for healthcare data acquisition in major public health emergencies should be accelerated to strengthen legal awareness among governments, departments, and institutions, using legal systems as the benchmark to clearly define legitimate boundaries for emergency healthcare data collection and clarify the responsibilities, rights, and interests of participants. Strengthening supervision of data acquisition apps and mini-programs, various healthcare data acquisition applications developed by internet companies must be subject to refined supervision. Through various forms of review and monitoring, law enforcement should be intensified against applications that exceed their scope in acquiring public privacy, making the emergency healthcare data acquisition process traceable and eliminating all infringements on public personal privacy.

Digital technology has brought convenience to major public health emergency prevention and control, but healthcare data should not be acquired at the cost of sacrificing public privacy security. Data acquisition should have boundaries. Through governance of the data acquisition stage, specific legislation for emergency healthcare data acquisition can clarify implementation details, enabling legal data acquisition within boundaries and laying a solid foundation for reasonable data circulation.

4.2.2 Governance of Open Healthcare Data in Major Public Health Emergencies In governing the emergency healthcare data opening stage, focus should be placed on addressing uneven construction of healthcare management information systems and "data silos" within medical institutions, starting from improving open healthcare data quality, strengthening healthcare data opening platform construction, and clarifying open data application scenarios. The quality status of raw emergency healthcare data determines open data quality. Improving open healthcare data quality should begin with raw data quality by designing scientific data quality assessment indicators, clarifying weights for various quality metrics, and establishing a data quality monitoring and assessment system for regular evaluation of raw and open data to monitor emergency healthcare data quality in real-time.

On the basis of ensuring emergency healthcare open data quality, healthcare data management and control systems should be accelerated, addressing data processing rules, usage rules, and data quality monitoring rules to classify and process emergency healthcare open data hierarchically. Emergency healthcare open data can be divided into three categories chronologically: historical data, current data, and future data. By constructing technical models for processing these three data types and continuously optimizing algorithms, the accuracy, consistency, and timeliness of healthcare open data can be ensured.

Key aspects of open data governance include healthcare data opening platform

construction and clarifying application scenarios. Healthcare data opening platforms should provide standardized structure-based open data services, integrating healthcare open data resources through platforms to avoid data fragmentation, achieve cross-departmental and cross-institutional collaboration, provide integrated data support for developing special drugs and optimizing clinical diagnosis and treatment plans in response to emergencies, enabling healthcare data to deliver greater value and alleviating China's shortage of medical resources for major public health emergencies. Clarifying healthcare open data application scenarios lays the foundation for data sharing: for government, open healthcare data should provide decision support; for medical institutions, it should provide inter-institutional information connectivity services; for financial insurance enterprises, it should provide medical insurance underwriting applications; and for the public, it should provide epidemic science popularization services and medication query services.

4.2.3 Governance of Healthcare Data Sharing in Major Public Health Emergencies Data sharing is the final stage of data circulation and the implementation stage of data services. In the context of major public health emergencies, healthcare open data sharing has become an important means for epidemic prevention and control, trend prediction, new drug development, and rumor debunking. Governance of emergency healthcare open data sharing should focus on establishing unified data sharing standards to optimize data flow processes, enriching data sharing technical means to enhance data utilization capabilities of medical and research institutions, and strengthening data sharing security management to increase public trust.

Unified data sharing standards are the foundation supporting healthcare data sharing in major public health emergencies. Emergency healthcare data features diverse sources and non-uniform technical standards in sharing stages. Therefore, higher-level governments should build unified emergency healthcare data opening and sharing channels, establishing unified data interfaces, technical standards, and implementation specifications to integrate healthcare data resources across different government levels and departments and 贯通 data sharing and exchange systems between government and society.

Data sharing technical means are the implementation methods for emergency healthcare open data sharing. Therefore, governance should vigorously strengthen technical 攻关 for emergency healthcare open data sharing. AI technologies such as computer vision and deep learning can be applied to learn and train medical imaging, assisting doctors in efficient and accurate diagnosis and screening. Through refined simulation and prediction, special drug development cycles can be shortened and costs saved. Knowledge graphs can process and fuse medical record data, construct data relationships, and promote standardized diagnosis and treatment plans. Blockchain technology can ensure full traceability of data sharing processes, strengthen data sharing security, and enhance public trust in open data.

Data sharing security management is the guarantee for emergency healthcare open data sharing. Therefore, existing data security protection laws should be further supplemented. By refining data security protection regulations, protection objects, scope, and illegal responsibilities in the emergency healthcare open data sharing process should be clarified. Using COVID-19 as an example, major public health emergencies spread globally, requiring all humanity to jointly solve this challenge, making cross-border data flow common in healthcare open data sharing. Therefore, establishing legal provisions for data opening, sharing, and cross-border flow supervision is necessary. Simultaneously, the scope of existing data protection laws should be expanded to include emergency healthcare open data protection under new technologies and scenarios such as blockchain and cloud computing.

4.3 Supporting Mechanisms for Healthcare Data Opening Governance in Major Public Health Emergencies

4.3.1 Organizational Support Organizational support for healthcare data opening governance in major public health emergencies can be divided into two components: strengthening cooperation among governments and departments at all levels, and promoting cooperation between government and high-level think tanks with Chinese characteristics. Promoting cross-departmental government cooperation and government-think tank cooperation can advance emergency healthcare data opening governance.

China has established an epidemic prevention and control cross-departmental coordination mechanism, but no specific cross-departmental institution has been set up for emergency healthcare data opening governance. Strengthening cooperation among governments and departments can draw on Australia's government data opening governance experience [33]. Led by the government, health commissions, disease control centers, and infectious disease prevention and control key laboratories at all levels should jointly establish an emergency healthcare data guidance committee to provide professional guidance for healthcare data acquisition, opening, and sharing during major public health emergencies. This institution's functions include supervising the storage, utilization, and sharing of emergency healthcare data, providing strong support for central government epidemic prevention and control decision-making, and serving as an official information release channel to debunk epidemic-related rumors with strong professional backgrounds, publicly release desensitized diagnosis and treatment data, provide guidance for public epidemic prevention, and enhance public anti-epidemic confidence.

Think tanks play an increasingly important role in China's governance system and should provide "high-level, forward-looking, and innovative insights" for government prevention and control decision-making during major public health emergencies [34]. Therefore, China urgently needs to build a batch of high-level think tanks with Chinese characteristics focusing on major public health emergencies and major infectious diseases. Such think tank construction should

incorporate medical scientists, legal experts, data scientists, and other professionals to establish tightly integrated decision-support institutions that regularly discuss epidemic prevention and control progress and obstacles, forming research reports for government decision-making reference.

4.3.2 Institutional Support With the proposal of the overall national security concept, the importance of data security has become more prominent. China's Data Security Law officially took effect on September 1, 2021, establishing a sound data security governance system. Based on this and combined with the context of major public health emergencies, this paper proposes establishing and improving institutional support for emergency healthcare data opening governance to ensure rule-based governance.

China's current legislation on healthcare data security in major public health emergencies remains unclear, requiring further exploration of hierarchical classification for emergency healthcare data security. The central government should supplement relevant provisions in the Data Security Law, strengthening healthcare data security guarantees in the context of major public health emergencies. Local governments should formulate emergency healthcare data opening measures suitable for their regions based on central government laws and regulations, thereby ensuring emergency healthcare data security, safeguarding the privacy rights of the broad populace, and improving the national data security guarantee system.

4.3.3 Technical Support Currently, healthcare data opening for major public health emergencies is in the exploration stage, with urgent issues such as patient privacy data leakage, epidemic false information, and data opening lag requiring solutions. To strengthen China's emergency healthcare data opening governance, multiple technical means including blockchain, AI, and knowledge graphs should be combined to promote emergency healthcare data opening and sharing. The specific technical architecture is shown in Figure 3 [Figure 3: see original paper].

Blockchain technology can build relevant healthcare databases for major public health emergencies, further improving data transparency. Highly transparent information can effectively curb rumor spread and reduce data error probability. Additionally, blockchain is a trust- and consensus-based technology. Through emergency healthcare data sharing combined with knowledge graphs and AI technology to provide intelligent services, not only can the shortage of medical resources during emergencies be effectively alleviated, but decision-making speed and accuracy can also be improved.

Knowledge graphs have provided strong support for this epidemic prevention and control, demonstrating the important value of this technology in emergency healthcare data opening and sharing. Knowledge graphs deeply extract multi-source heterogeneous data, forming associations among originally complex healthcare data. Applying knowledge graphs to emergency healthcare data

processing can quickly identify data associations, providing decision support for emergency prevention and control, prediction, vaccine development, and disease traceability. Combined with blockchain technology, it can effectively solve data credibility issues and further clarify the rights and responsibilities of participants in the governance process.

Through the above research on blockchain, AI, knowledge graphs, and other technical means, we can find that the three have strong complementarity and high coordination. Therefore, in the governance process of emergency healthcare data opening, new technical means should be used in coordination to maximize governance effectiveness.

Healthcare data opening governance for major public health emergencies should adhere to government leadership with coordination among departments, medical research institutions, enterprises, and the public. The governance process should follow the principles of responsibility, transparency, security, and quality. During emergency healthcare data opening, the protection of people's rights and interests should be the center, continuously optimizing the emergency healthcare data opening process, refining the data opening governance model, and strengthening the application of new data processing technologies in data opening governance. By continuously improving the emergency healthcare data opening governance model and replicating successful experiences to government open data governance and enterprise open data governance, China's data opening governance level can be enhanced, gradually exploring a data opening governance system that suits China's national conditions.

References

[1] Xi Jinping. Providing Strong Scientific and Technological Support for Winning the Epidemic Prevention and Control Battle [J]. Practice (Ideological Theory Edition), 2020, 28(4): 4-6. [2] Wang Yifan, Zhou Dian, Yao Chenhuan, et al. Research on Hospital Emergency Management Resilience for Major Public Health Emergencies [J]. Chinese Hospital Management, 2021, 41(6): 12-15. [3] People's Daily. Key Points! This Is How Xi Jinping Requires the Epidemic to Be Controlled [EB/OL]. [2021-12-12]. https://wap.peopleapp.com/article/5105453/5001744.2020-02-04. [4] Marc DT, Zhang R, Beattie J, et al. Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Term Coverage [M]//MEDINFO 2015: eHealth-enabled Health. Washington: IOS Press, 2015: 529-533. [5] Li Jiao, Guo Haihong, Guo Minjiang. Quantitative Research on Theme Distribution and Openness Degree of U.S. and U.K. Government Open Health Data [J]. Library and Information Service, 2015, 59(20): 132-137. [6] Liu Ning, Chen Min. Research on Methods and Strategies for Medical Data Opening [J]. Chinese Hospital Management, 2015, 35(9): 37-39. [7] Bright J, Margetts H Z, Ning W, et al. Explaining Usage Patterns in Open Government Data: The Case of data.gov.UK [J]. SSRN Electronic Journal, 2015, 24(1): 1-22. [8] Yang Jiahong, Zhang Jianguo, Mu Xiaoli. Research on the

Current Status of International Regional Medical Information Sharing System Construction [J]. China Hospital, 2010, 14(10): 4-6. [9] Li Yanmei. Current Status and Implications of Government Open Data in Healthcare Fields in China, U.K., U.S., Canada, and Australia [J]. Chinese Journal of Medical Library and Information Science, 2019, 28(6): 42-47. [10] State Council Issues "Outline for Promoting Big Data Development" [J]. Chinese Journal of Health Informatics and Management, 2015, 12(5): 447. [11] Mao Yunpeng. Successful Case of Direct Health Statistics Reporting Data Sharing [C]//Chinese Health Information Association. Proceedings of China Health Information Technology Exchange Conference and Cross-Strait Health Informatization Exchange Conference, Dalian: Chinese Journal of Health Informatics and Management Magazine, 2012: 304-307. [12] Wei Li, Liang Yue, Liu Yuhong. Patient Medical Record Acquisition Service Based on Medical Data Release Center [J]. Chinese Hospital Management, 2018, 38(11): 40-41. [13] Shi Huaxin, Liu Baoyan, Xie Qi. Thoughts on Constructing Data Collection Platform Service Flow for Traditional Chinese Medicine Clinical Research Information Sharing System [J]. Chinese Journal of Information on Traditional Chinese Medicine, 2013, 20(8): 4-5. [14] Kass-Hout TA, Grays K, Massoudi BL, et al. NHIN, RHIOs, and Public Health [J]. Journal of Public Health Management and Practice, 2007, 13(1): 31-34. [15] Huang Nana, Shen Liang, Yang Yuanyuan. Research on Efficient and Secure Data Sharing Scheme in Medical Cloud [J]. Computer Engineering and Applications, 2020, 56(1): 92-97. [16] Zhou Hui, Wang Lidan, Zhong Chengyue. Blockchain Empowers Electronic Medical Data Sharing [J]. Hospital Administration Journal of Chinese People's Liberation Army, 2019, 26(7): 645-647. [17] Laessig M, Jacob B, AbouZahr C. Opening Data for Global Health [M]//The Palgrave Handbook of Global Health Data Methods for Policy and Practice. London: Palgrave Macmillan, 2019. [18] Yozwiak NL, Schaffner SF, Sabeti PC. Data Sharing: Make Outbreak Research Open Access [J]. Nature News, 2015, 518(7540): 477-479. [19] Kieny MP, Moorthy VS, BagoZZI D. Use Open Data to Curb Zika Virus [J]. Nature, 2016, 533(7604): 469. [20] Dong E, Du H, Gardner L. An Interactive Web-based Dashboard to Track COVID-19 in Real Time [J]. The Lancet Infectious Diseases, 2020, 20(5): 533-534. [21] Hu T, Guan WW, Zhu X, et al. Building an Open Resources Repository for COVID-19 Research [J]. SSRN Electronic Journal, 2020, 4(3): 130-147. [22] Kobayashi S, Falco L, Fraser H, et al. Using Open Source, Open Data, and Civic Technology to Address the COVID-19 Pandemic and Infodemic [J]. Yearbook of Medical Informatics, 2021, 30(1): 38-43. [23] Warren MS, Skillman SW. Mobility Changes in Response to COVID-19 [EB/OL]. [2021-12-24]. https://arxiv.org/pdf/2003.14228.pdf. [24] Organization WH. Ethical Considerations to Guide the Use of Digital Proximity Tracking Technologies for COVID-19 Contact Tracing: Interim Guidance, 28 May 2020 [R]. World Health Organization, 2020. [25] Banda JM, Tekumalla R, Wang G, et al. A Large-scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration [J]. Epidemiologia, 2021, 2(3): 315-324. [26] Xinhua News Agency. Communique of the Fifth Plenary Session of the 19th CPC Central Committee [EB/OL]. [2021-09-14].

http://www.xinhuanet.com/politics/2020-10/29/c_{1126674147}.htm.2020-10-29(1). [27] Gao Fuping. Data Circulation Theory: The Basis for Data Resource Rights Allocation [J]. Peking University Law Journal, 2019, 31(6): 1405-1424. [28] Zhao Juan, Meng Tianguang. The Vertical Governance Logic of Digital Government: Hierarchical System and Collaborative Governance [J]. Xuehai, 2021(2): 90-99. [29] Zhou Maojun, Pan Ning. Empowerment and Reconstruction: How Blockchain Technology Resolves Data Silos [J]. Journalism and Communication Review, 2018, 71(5): 58-67. [30] Cui Yuhong, Wang Sa. Data Sharing in COVID-19 Public Health Emergencies [J]. Library and Information Service, 2020, 64(15): 104-111. [31] National Health Commission. "Notice on Issuing the Trial Measures for National Health and Medical Big Data Standards, Security, and Service Management" [EB/OL]. [2021-11-24]. http://www.cac.gov.cn/2018-09/15/c_{1123432498}.htm. [32] Gu Liping. Data Governance: Development Opportunities for Libraries [J]. Journal of Library Science in China, 2016, 42(5): 40-56. [33] Chen Mei. Research on the Guarantee Mechanism of Australian Local Government Open Data—From the Perspective of Multi-element Public Administration [J]. Information Studies: Theory & Application, 2017, 40(12): 139-144, 111. [34] Yang Yaqin. Thoughts on the Modernization of New-type Think Tanks with Chinese Characteristics—Analysis from the Perspective of Think Tank Influence Evaluation [J]. Bulletin of Chinese Academy of Sciences, 2021, 36(1): 70-78.

Author Contributions: Li Jiaxing: Paper writing, revision, and final version editing; Wang Lei: Literature collection, abstract translation, and paper proofreading; Song Shijie: Paper revision; Zhu Qinghua: Paper framework design.

Abstract: [Purpose/significance] Through an in-depth analysis of the healthcare data opening and sharing process in major public health emergencies, this paper proposes an open governance model for healthcare data with multi-stakeholder collaborative participation for China's major public health emergency data governance. [Method/process] Using a combination of literature research and Internet surveys, this study comprehensively reviews the current research status and existing obstacles of healthcare data opening models for major public health emergencies both domestically and internationally, and proposes governance solutions from a practical problem-solving perspective. [Result/conclusion] By clarifying the principles for healthcare data opening in major public health emergencies, decomposing the data opening and sharing process, this study proposes a staged governance model for healthcare data opening and sharing under major public health emergencies, formulates relevant supporting mechanisms, and ultimately presents a practical and feasible healthcare data open governance model for major public health emergencies.

Keywords: major public health emergencies; healthcare big data; data opening; data governance model

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

Source: ChinaXiv — Machine translation. Verify with original.