



Data governance system of the National Clinical Research Center for Child Health in China

Jing Li^{1,2^}, Gang Yu^{1,2,3^}, Wen Ding^{2,4}, Jian Huang^{1,2}, Zheming Li^{1,2^}, Zhu Zhu^{1,2}, Dejian Wang⁵, Jie Zhang⁵, Jing Wang⁵, Jianwei Yin⁶

¹Department of Data and Information, The Children's Hospital Zhejiang University School of Medicine, Hangzhou, China; ²AI Lab, National Clinical Research Center for Child Health, Hangzhou, China; ³Polytechnic Institute, Zhejiang University, Hangzhou, China; ⁴Department of Research and Education, The Children's Hospital Zhejiang University School of Medicine, Hangzhou, China; ⁵Department of R&D, Hangzhou Healink Technology, Hangzhou, China; ⁶College of Computer Science, Zhejiang University, Hangzhou, China

Contributions: (I) Conception and design: J Li, G Yu; (II) Administrative support: G Yu; (III) Provision of study materials or patients: W Ding, J Wang, J Yin; (IV) Collection and assembly of data: Z Li, Z Zhu; (V) Data analysis and interpretation: J Huang, D Wang, J Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Gang Yu. 3333 Binsheng Rd, Hangzhou 310052, China. Email: yugbme@zju.edu.cn; Jing Wang. 9F Tianren Building, 188, Liyi Road, Xiaoshan District, Hangzhou 311200, China. Email: wangjing@healink.cn; Jianwei Yin. 866 Yuhangtang Rd, Hangzhou 310058, China. Email: zjujw@cs.zju.edu.cn.

Background: Since the national big data strategy was unveiled at the fifth plenary session of the 18th CPC (Communist Party of China) Central Committee, the big data industry has been flourishing in China. Various successful industrial data governance systems have emerged with the rapid development of big data technologies and data management theories. City Brain and Enterprise Data Middle Platform are considered the best data governance systems in urban and corporate governance, respectively. However, in the health and medical sectors, issues of data operation occur frequently due to a lack of systematic data governance. These problems need to be urgently addressed, as health and medical data have been defined as national fundamental strategic resources. Clinical researchers have an increasing demand for data analysis.

Methods: Therefore, the Medical Data Governance System (MDGS) has been designed to improve data quality and provide simple and convenient data analysis tools for the National Clinical Research Center for Child Health. The MDGS consists of the Medical Data Platform (MDP) and Operation Management System (OMS). The MDP comprises acquisition layer, middle platform, and application layer that persistently elevates data quality and significantly shortens data analysis duration. Organization construction, management regulations, and technical standards are included in the OMS, which guarantees the sustainable operation of the MDGS. The MDGS was established to advance state-of-the-art and state-of-practice data governance for the health and medical sectors in China.

Results: With the first phase of the MDGS, the quantity and quality of research projects increase, research transformation speeds up, and the researchers' job satisfaction increased.

Conclusions: Based on our preliminary achievements, it was necessary and feasible to establish the MDGS. It is important to have comprehensive requirement study, top-level design, refined planning, phase-by-phase implementation, and continual optimization.

Keywords: Data governance; clinical research; Medical Data Platform (MDP); artificial intelligence modeling (AI modeling); visual programming

Submitted May 25, 2021. Accepted for publication July 14, 2021.

doi: 10.21037/tp-21-272

View this article at: <https://dx.doi.org/10.21037/tp-21-272>

[^] ORCID: Jing Li, 0000-0002-3626-5815; Gang Yu, 0000-0001-9935-9969; Zheming Li, 0000-0002-6640-9947.

Introduction

With the expansion of the big data industry in China, the Enterprise Data Middle Platform and the City Brain are booming and considered the best data governance systems in corporate and urban governance, respectively (1-7). In the health and medical sectors, however, issues of data operation occur frequently due to the lack of systematic data governance, such as data usage irregularities, data security risks, data operation restrictions, data heterogeneousness, data utilization ethical arguments, and costly data acquisition (8,9). These problems need to be addressed urgently, as health and medical data are national fundamental strategic resources (10). Although medical informatics has progressed rapidly in recent years in China, medical data are rarely managed systematically, which directly affects data quality and clinical research efficiency.

There is strong demand among clinical researchers for data analysis. Coding languages and software are obstacles for clinical researchers when they start a research project, and it is difficult to navigate off-the-shelf data analysis products, especially those for artificial intelligence (AI) modelling. It takes significant time and energy to learn the software for data analysis, and doctors and nurses often have limited time to learn the software. Some hospitals have built their own clinical research platforms (11-13); however, they are best suited to pilot studies that focus only on data acquisition, data preprocessing, and disease database building, rather than top-level design of data governance or as convenient tools for researchers.

There is a need to construct a comprehensive system to govern medical data that considers both technical support capability and operation guarantee measures.

The concept of data management can be traced back to 1980s, when database techniques and data storage first emerged. In the Data Management Association International (DAMA) Guide to the Data Management Body of Knowledge (DMBOK), data management is defined as planning, controlling, and providing data assets (14). Among the 10 areas of DAMA's data management system, data governance is the highest level planning activity, and includes data strategy development, data policy improvement, and data architecture design. It places emphasis on the data user, their usage mode, access authority, and other compliances. It also underlines the fundamental work before the lifecycle management of data assets, especially the related guarantee measures. Data standards and data value management were added in the

DAMA-DMBOK2 in terms of data assets, underlining the upgrade of organization structure and management system to guarantee the workflow, security, and validity (15,16).

The Children's Hospital Zhejiang University School of Medicine is one of the best children's hospitals in China (17), and awarded the National Clinical Research Center for Child Health and the Regional Children's Health Center (18,19). The Children's Hospital Zhejiang University School of Medicine has 1,300 beds and almost 3,000 employees. There are about 3.5 million outpatient visits and 81,000 inpatient visits every year. All hospital employees are committed to providing the highest quality of health care to children.

The blueprint for the Medical Data Governance System (MDGS) of The Children's Hospital Zhejiang University School of Medicine was created in early 2019. Phase I of the system has been completed and commenced operation at the end of October 2019. Preliminary success has been achieved so far.

Methods

System architecture

There are two subsystems that constitute the MDGS at The Children's Hospital Zhejiang University School of Medicine. The first is the Medical Data Platform (MDP), which can be divided into the data acquisition layer, middle platform layer, and application layer. The other is the Operation Management System (OMS), which can be divided into organization building, management principles and regulations, and technical standards. The system architecture of the MDGS is shown in *Figure 1*.

The MDP

As the 1-stop platform for clinical researchers, the MDP provides medical data of compliance, multidimension and high quality, as well as research tools of effectiveness, convenience, and visualization. The MDP primarily includes 3 parts; these are the data acquisition layer, middle platform, and application layer.

Data acquisition layer

Data source

The data acquisition layer acquires medical data required by research projects. The obtained data contain various categories, such as clinical data from health information

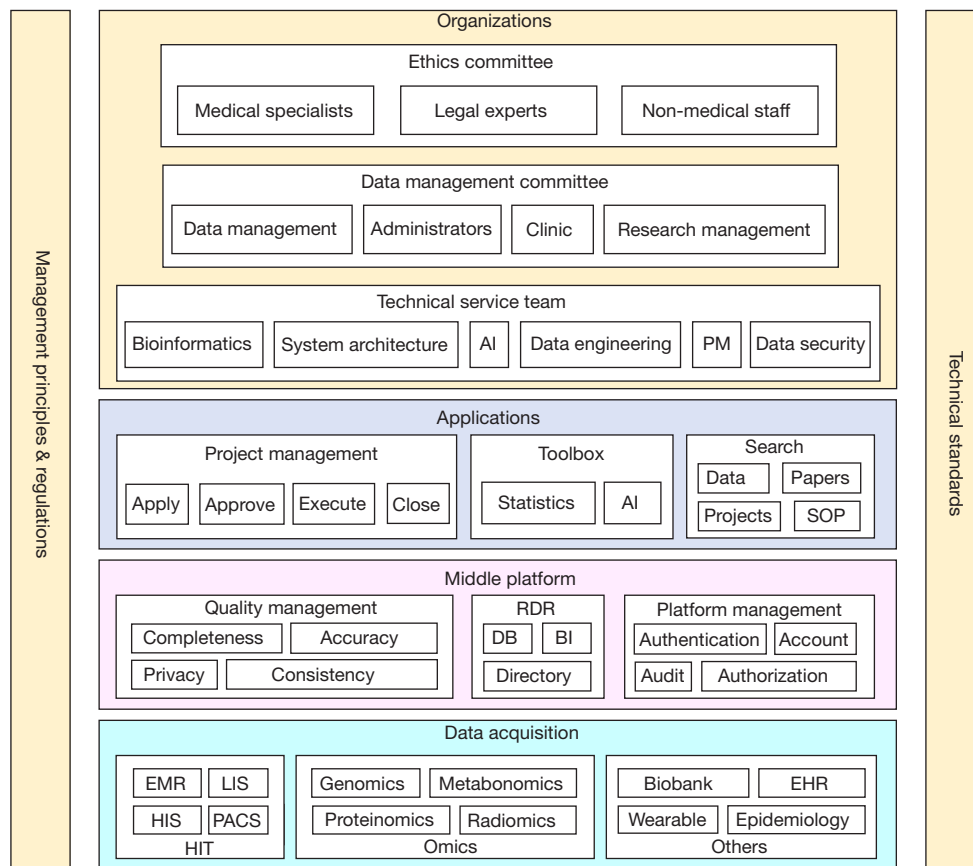


Figure 1 System architecture of the Medical Data Governance System. AI, artificial intelligence; PM, project management; SOP, standard operating procedure; RDR, research data repository; DB, data base; BI, business intelligence; EMR, electronic medical record; LIS, laboratory information system; HIS, hospital information system; PACS, picture archiving and communication system; HIT, healthcare information technology; EHR, electronic health record.

technology systems (e.g., electronic medical records, hospital information systems, laboratory information systems, and picture archiving and communication systems), omics data from researchers (e.g., genomics, metabonomics, proteomics, immunomics, and ultrasomics), and data from other sources (e.g., biobank, wearables, electronic health records, epidemiology, climate, and environment).

Acquisition and preprocessing

Different techniques [e.g., database batch push, application programming interface (API) transmission, and uploads in files and tables] can be adopted in data acquisition according to the actual conditions. The data are gathered and stored as “raw data” in the data acquisition layer. Raw data are processed through a privacy protection module that deletes patients’ unnecessary private information of each data entry and encrypts what is left.

Backup and recovery

After preprocessing, data are defined as desensitized data and duplicated in the database. Each copy of desensitized data is supervised by the system. It can be recovered from the other copy in case of data loss.

Middle platform

As the core layer of the MDP, the middle platform manages data quality, research database, and system configuration.

Data quality management

The data quality in the data acquisition layer is not good enough for clinical research in terms of completeness, accuracy, and consistency. A closed-loop mechanism is designed for data quality improvement. The module of data quality management can discover and solve most problems of data quality. The workflow of data quality improvement

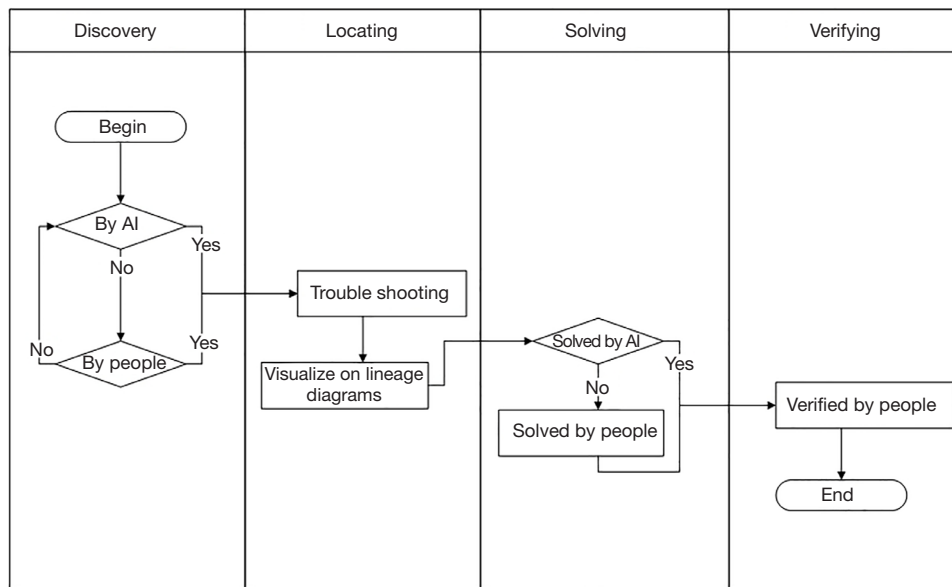


Figure 2 Workflow of data quality management. AI, artificial intelligence.

is shown in *Figure 2*.

Problem discovery, problem locating, problem solving, and solution verifying together constitute the process group of data quality management. Most processes of data quality management are triggered by the system automatically according to the established criteria. In the case that a data quality problem is discovered by researchers, the process can also be initiated manually. Once the process is launched, the metadata management module can locate the root cause of the problem and visualize it on the lineage diagrams. Given different scenarios, problems are addressed by the system or platform administrators. It is required to verify the solutions by the system administrators or researchers. All the processes are recorded in the system log and shown on the problem reports.

The module of data quality management works continuously, so that the completeness, accuracy, and consistency of the medical data on the data platform can be improved constantly.

Research data repository (RDR)

High-quality medical data are stored in the RDR for research projects. The workflow of data processing from raw data to the RDR is shown in *Figure 3*. The data directory is created automatically by the system. Databases and knowledge graphs for specific diseases can also be generated. Metadata (including, but not limited to, data category, quantity, data source, and update time) of the

medical data are analyzed statistically and visualized by the business intelligence module.

System configuration

With the module of system configuration, system administrators can set up different authorizations for different roles and accounts. GUI (Graphical User Interface) also can be personalized for clinical researchers, system administrators, and decision makers of the Chinese National Clinical Research Center.

Application layer

The application layer provides 1-stop services for clinical researchers, including project management, toolbox, and search engine.

Project management

The research project management process consists of the following 4 steps: applying, approving, executing, and closing. The lifecycle is shown in *Figure 4*.

Projects applications can be submitted after clinical researchers decide to initiate a research project and create a project plan. Applications need to demonstrate the scope and schedule of the project, refine the objectives, define the course of action required to attain the objectives, and clarify data demand and expected research fruits.

The project applications need approval from the ethics committee and the data management committee, successively. Government permissions are required for some

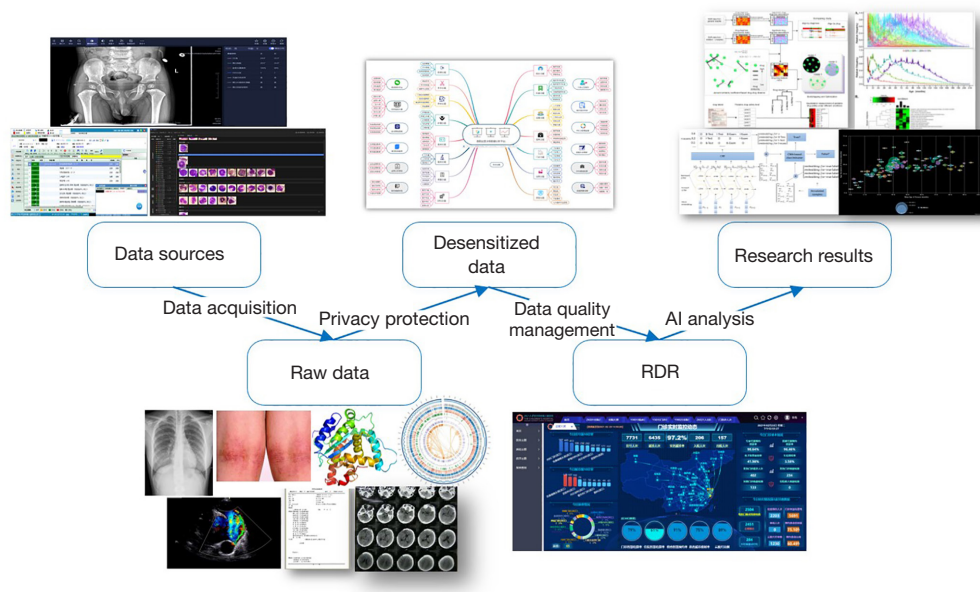


Figure 3 Dataflow of the Medical Data Governance System. AI, artificial intelligence; RDR, research data repository.

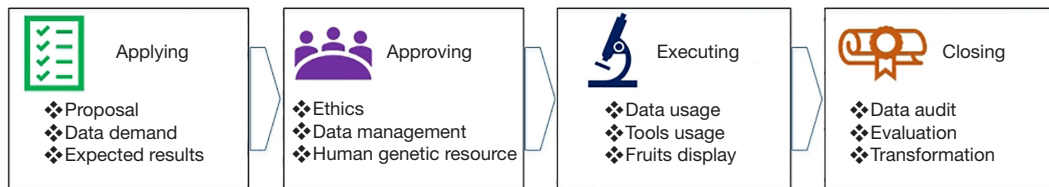


Figure 4 Lifecycle of research project management.

specific research projects, according to national regulations on the management of human genetic resources.

By accessing the data and tools on the platform, researchers can commence their research after project approval. The data platform monitors and controls data usage, and records it in the system logs.

When the closing process is performed, researchers are required to summarize the project, submit the final report and research fruits, and apply for acceptance. The data platform generates a data audit report that contains data usage analysis of the project. The data platform is continuously optimized based on the data audit reports. The data management committee arranges to accept the project, evaluate the research results, and offer fruit transformation services.

Toolbox

Based on visual programming, all the tools are visualized as icons. Rather than coding, researchers only need to drag

and drop the icons from the toolbox and connect them by arrows. The data will be manipulated based on this dataflow graph. Furthermore, the intermediate results can be shown after each data process.

A serial of research tools devoted to clinical researchers are assembled in the toolbox, containing statistics tools, AI algorithms, and data modeling tools. Various algorithms and data modeling tools can be considered (e.g., ANN (Artificial Neural Network), RBF (Radial Basis Function), CART (Classification and Regression Trees), Naïve Bayes, Apriori, correlation analysis, clustering, abnormal value management, and intelligence recommendation).

Search engine

A powerful search engine is designed to search for related papers, data resources on the platform, and projects undertaken by other colleagues. The search engine provides unified access to the databases of references, data directories, and project introductions. It supports

multidimensional, bilingual, and fuzzy query.

The OMS

Operation management was often overlooked by administrators of information systems, resulting in unachieved objectives or even negative effects. The OMS is designed to guarantee smooth operation of the MDGS. It mainly consists of organization construction, management principles and regulations, and technical standards.

Organization construction

The following 3 organizations were constructed and play important roles in research project management: ethics committee, data management committee, and technical service team.

Ethics committee

The ethics committee decides if the research is ethical. It inspects the project application and insures the safety, health, and rights of research participants. The ethics committee consists of medical specialists, legal experts, and non-medical staff. Committee members are elected from the candidates list on the hospital board meeting. Research project applications are first submitted to the ethics committee and are rejected if they cannot pass the ethics inspection. The ethics committee works independently and is supervised by the National Clinical Research Center and research participants.

Data management committee

The data management committee is the decision maker on data-relevant issues, such as data security, data utilization, and data platform construction. It consists of the National Clinical Research Center administrators and senior experts from Department of Data and Information, clinical departments, and the Department of Research and Education. Committee members from National Clinical Research Center are assigned by the board of the center. Committee members from Department of Data and Information and Department of Research and Education are elected by the department meeting respectively. After ethics committee approval, research project applications will be submitted to the data management committee. It assesses the rationality and availability of the project data requirements. It reports to the national center and is supervised by the platform users.

Technical service team

The technical service team is responsible for the construction, operation, and maintenance of the MDP, as

well as services pertaining to technical training, account management, and data analysis for researchers. It consists of specialists from independent software vendors (ISVs) and the Department of Data and Information. Team members specialize in medical AI, data engineering, bioinformatics, system architecture, software engineering, computer networking, information security, and project management. Team members are selected by the data management committee. It reports to the data management committee and is supervised by platform users.

Management principles and regulations

A series of principles and regulations was released to instruct researchers in MDP usage, research project management, fruit transformation, and other relevant fields.

MDGS guide

The MDGS guide instructs users how to operate the system, safeguards researchers' interests, and guarantees data security. It contains information on system introduction, rights and responsibilities of different roles, system operation mechanisms, management regulations, and a user manual. It is released by the data management committee and requires compliance by all the users.

Research project management regulations

Regulations on research project management instruct clinical researchers on how to apply for apply, execute, and close projects. It contains project management principles, management workflows, criteria of project approval, and data utilization specifications. It was released by the Department of Research and Education and the data management committee, and requires compliance by all the users.

Research stimulation and fruit transformation policies

Research stimulation and fruit transformation policies were designed to encourage researchers and maximize the value of research fruits. It contains stimulation principles, stimulation methods, fruits evaluation standards, and fruits transformation mechanisms. It was released by the Department of Research and Education and the data management committee, and applies to all researchers.

Technical standards

The technical standards illustrate the objectives, requirements, and technical paths of the MDP. It includes clinical data standards, specifications of data governance and utilization, and MDP construction standards.

Clinical data standards

Clinical data standards specify the models of the medical data, containing data structures, data operations, and data

Table 1 Annual research achievements the Children's Hospital of Zhejiang University School of Medicine

	2019	2020	Growth rate (%)
NSFC projects	12	23	91.67
Other projects	72	91	26.39
SCI papers	179	278	55.31
Patents	33	39	18.18

NSFC, National Natural Science Foundation of China; SCI, Science Citation Index.

Table 2 Science and technology evaluation metrics rankings of The Children's Hospital Zhejiang University School of Medicine

	Rank	Previous rank	Rise
Children's hospital ranking	2	3	1
Hospital ranking	68	97	29

constraints. It was drafted by clinical experts and the data management committee, and was released by the data management committee.

Data governance and utilization specifications

Data governance and utilization specifications focus on the targets, principles, processes, and supervisions of data governance and data utilization. It also includes the standards of data classification and data quality evaluation. It was released by the data management committee.

MDP construction standards

The MDP construction standards focus on system architecture, function planning, database design, data flow, interface specification, and implementing plans. It was released by the data management committee and requires compliance by all ISVs.

Results

After 1 year of construction, phase I of the MDGS was completed with good results. With the sustained improvement of researcher satisfaction, the quantity and quality of research projects increase significantly, as do research grants, and research transformation has big progress.

There are 126 research projects supported by the MDGS to date, 14 of which are multicenter clinical research (11 nationwide, 3 provincial). Benefits of the MDGS

include 23 National Natural Science Foundation of China projects being approved, 278 SCI (Science Citation Index) papers published, and 39 patents obtained during 2020. The number of research achievements in 2020 was significantly higher compared with 2019, as shown in *Table 1*.

The science and technology evaluation metrics (STEM) are published by the Chinese Academy of Medical Sciences every year to evaluate the research strength of hospitals and medical colleges in China. Based on STEM released in 2020, The Children's Hospital Zhejiang University School of Medicine went from third to second in children's hospitals rankings, and from 97th to 68th among all hospitals (20), as shown in *Table 2*.

Within the accepted research projects, AI assistant skeletal age inspection has already finished productization (21). The first all-in-one machine of skeletal age inspection for children was published around the world. It has much lower radiation and much higher diagnostic speed, and is considered a great benefit for teenagers with obesity diabetes. The research project of the AI-based children's hip joint development evaluation is in the process of transformation. The intelligent diagnostic system of children's hip joint development, based on medical imaging, is about to be published by the end of 2021 (22). Research on AI-assisted diagnosis of chronic cough in children based on multimodal fusion is in progress. The prototype of the Children Asthma Diagnosis System has been designed and tested, which is expected to benefit pediatricians.

With higher enthusiasm and better intellectual atmosphere, more and more clinical staff initiate their research projects on the MDGS.

Discussion

In Phase I of the MDGS, basic functions of the MDP were completed, management organizations were established, and initial regulations and standards were released. Phase II is designed to acquire more data to enrich the RDR, introduce more AI tools to enhance the toolbox, enforce training for clinical staff to improve data quality at the source, optimize management regulations and update technical standards, and employ more talent. Information security system will be enhanced if it is required to connect to the Internet in the future.

Phase II is implemented as planned so far and we are confident to have a more powerful MDGS soon. With another 2 years of hard work, we hope our MDGS will

become one of the best practices in the field (23).

Conclusions

Based on the achievements of MDGS Phase I, it is necessary and feasible to establish MDGS for national medical centers, and both MDP and OMS are important for smooth system operation. To create an effective MDGS, it is critical to have top-level design, refined planning, phase-by-phase implementation, and continual optimization.

Acknowledgments

Funding: This work was supported by the National Key R&D Program of China (grant number: 2019YFE0126200) and the National Natural Science Foundation of China (grant number: 62076218).

Footnote

Data Sharing Statement: Available at <https://dx.doi.org/10.21037/tp-21-272>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/tp-21-272>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. China Info 100 & AliResearch. The rise of data productivity: new power, new governance. AliResearch. 2020. Available online: <http://www.aliresearch.com/ch/activity/recentactivities?eventCode=106346850352762880> (accessed Sep 5, 2020).
2. Alibaba Cloud. Alibaba Cloud Middle Platform white paper on retail data model. 2020. Available online: <https://dp.alibaba.com/ecosystem?spm=a2c6h.12873639.0.0.2e055e04OHScvs> (accessed Oct 21, 2020).
3. Tencent Research Institute. White paper on Tencent AI: ubiquitous intelligence. 2020. Available online: <https://tech.qq.com/a/20200710/009923.htm> (accessed Aug 8, 2020).
4. Feng L, Liu F, Shi Y. City Brain, a New Architecture of Smart City Based on the Internet Brain. IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD). 2018:624-9.
5. Economy and Information Technology Department of Zhejiang. Action plan of Zhejiang "City Brain" construction and application. 2019. Available online: http://jxt.zj.gov.cn/art/2019/6/4/art_1229123405_627864.html (accessed Jul 15, 2020).
6. Zhao Y, Bing D, Chen S, et al. Spatio-Temporal Auto Encoder for Video Anomaly Detection. Proceedings of the 25th ACM international conference on Multimedia. 2017:1933-41.
7. Zhao Y, Bing D, Jianqiang H, et al. Stylized Adversarial AutoEncoder for Image Generation. Proceedings of the 25th ACM international conference on Multimedia. 2017:244-51.
8. Shanghai Jiao Tong University. White paper on medical AI in China. 2019. Available online: <https://news.sjtu.edu.cn/mtjj/20190110/94741.html> (accessed Oct 18, 2020).
9. Deloitte Insights. 2020 Global Life Sciences Outlook: Creating new value, building blocks for the future. 2020. Available online: <https://www2.deloitte.com/us/en/insights/industry/life-sciences/global-life-sciences-outlook.html> (accessed Sep 21, 2020).
10. General Office of China's State Council. China to boost big data application in health and medical sectors. 2016. Available online: http://english.www.gov.cn/policies/latest_releases/2016/06/24/content_281475379018156.htm (accessed Aug 14, 2020).
11. Griesinger F, Eberhardt W, Marschner N, et al. MA08.02 Clinical Research Platform into Molecular Testing, Treatment, Outcome (CRISP): A Prospective German Registry in Stage IV NSCLC AIO-TRK-0315. J Thorac Oncol 2017;12:S385.
12. Duan J, Chen T. Information technology based clinical research data platform. China Journal of Modern Medicine 2020;30:124-8.
13. Wei R, Lu L, Qian B, et al. Construction and practice

- of big data based clinical research platform in hospitals. *China Digital Medicine* 2019;14:91-93,96.
14. DAMA International. The DAMA Guide to the Data Management Body of Knowledge. Technics Publications, 2009.
 15. DAMA International. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. Technics Publications, 2017.
 16. Gelbstein, Ed. Data management body of knowledge: a summary for auditors. *ISACA Journal the Source for It Governance Professionals*, 2017:1-5.
 17. Fudan University. China's hospital rankings. 2020. Available online: <http://rank.cn-healthcare.com/> (accessed Nov 14, 2020).
 18. Ministry of Science and Technology of China. Notification of national clinical research centers. 2019. Available online: http://www.most.gov.cn/tztg/201905/t20190530_146874.htm (accessed Sep 2, 2020).
 19. National Health Commission of China. Notification to establish five regional children's health centers. 2020. Available online: http://www.gov.cn/zhengce/zhengceku/2020-09/03/content_5539736.htm (accessed Sep 3, 2020).
 20. Chinese Academy of Medical Sciences. The release of STEM for hospitals and medical colleges in China. 2020. Available online: <http://www.nsf.gov.cn/csc/20340/20289/54714/index.html> (accessed Dec 4, 2020).
 21. Zhou XL, Wang EG, Lin Q, et al. Diagnostic performance of convolutional neural network-based Tanner-Whitehouse 3 bone age assessment system. *Quant Imaging Med Surg* 2020;10:657-67.
 22. Yu G, Yang Y, Wang X, et al. Adversarial active learning for the identification of medical concepts and annotation inconsistency. *J Biomed Inform* 2020;108:103481.
 23. Wallentin L, Lindahl B. Uppsala Clinical Research Center-development of a platform to promote national and international clinical science. *Ups J Med Sci* 2019;124:1-5.

(English Language Editor: R. Scott)

Cite this article as: Li J, Yu G, Ding W, Huang J, Li Z, Zhu Z, Wang D, Zhang J, Wang J, Yin J. Data governance system of the National Clinical Research Center for Child Health in China. *Transl Pediatr* 2021;10(7):1905-1913. doi: 10.21037/tp-21-272