



Present and Future of Utilizing Healthcare Data

In Young Choi

Chairman of the Board of the Korean Society of Medical Informatics, The Catholic University of Korea College of Medicine, Seoul, Korea

With increasing interest in using healthcare data, the Ministry of Health and Welfare in Korea launched a medical data-driven hospital support project in 2020. Five consortia selected in 2020 are participating in this project, as well as two consortia that were additionally selected in 2021, resulting in a total of 40 hospitals and seven consortia. In addition to hospitals, 42 other institutions are taking part, including pharmaceutical companies, IT companies, and Electronic Medical Record (EMR) development companies [1].

The data-driven hospital project aims to establish organizations, processes, and technological foundations to promote the use of medical data. The period from 2020 to 2022 has been considered phase 1, which has mainly focused on the following three areas: governance establishment, data establishment, and standardization and quality management. This article will describe the changes that the data-driven hospital project has brought to hospitals and make suggestions for future development.

1. Governance Establishment

Professor Chandler of Harvard Business School stated, “structure follows strategy” [2]. This is the most basic proposition of business administration, meaning that companies need to change and innovate according to changes in their environment. Therefore, they must form an organizational system that can successfully carry out the strategies they have adopted. From this perspective, the first requirement—to establish governance—is interpreted as adjusting the organization of hospitals to adapt to a data-driven environment.

The main changes involved in governance establishment include the appointment of a chief information officer (CIO) at the level of a vice-president and the preparation of committees and regulations for using pseudonymous data.

The appointment of a CIO at the vice-president level is the most important step in establishing governance, which is the first requirement of the medical data-driven hospital project. A CIO is a position that refers to a high-ranking manager who oversees informatization issues in the process of establishing strategies for hospitals to respond to the rapidly changing business environment. Information policy in hospitals was traditionally established mostly by mid-level executives rather than by executives in high-ranking positions. As a transition away from this practice, the introduction of CIOs at the vice-president level is considered a groundbreaking step in hospitals. These drastic organizational changes reflect the demand for changes in governance regarding the use of medical data.

For a long time, no party has taken the initiative to collect, process, and utilize medical data, and the resulting fragmentation has functioned as a ball and chain for future development. Due to the nature of medical care, various departments are involved in the diagnosis and treatment of individual patients. This complexity also increases when patients are treated at multiple hospitals or multiple departments. For example, a study on diabetic retinopathy should integrate data from the departments of ophthalmology and endocrinology. In the existing fragmented governance system, there have been debates about data ownership by department, necessitating the establishment of integrated governance and strong leadership at the vice-president level through this project.

From this perspective, the performance of data-driven hospitals is highly encouraging. All hospitals participating

in the data-driven hospital project have appointed a CIO at the level of a vice president. In practice, it takes more trial and error to execute leadership in collecting, processing, and utilizing data. Nonetheless, the foundation for changes has been established by forming a dedicated organization.

Concomitantly, a new development in governance establishment was the formation of a committee for the use of pseudonymous data within hospitals. Prior to this project, many hospitals did not have an official review body or procedure for using pseudonymous data other than the Institutional Review Board. However, this project has prepared unified procedures and regulations. Although this process may cause inconvenience to researchers or data utilization institutions in the short term, the establishment of a unified procedure will serve as a foundation for faster and more transparent data utilization in the long term.

2. Data Establishment

The second requirement of healthcare data-driven hospitals is the establishment of large-capacity integrated medical big data that can be used inside and outside of a consortium and the establishment of databases for specific diseases.

There exists a consortium for medical big data that focuses on establishing or advancing clinical data warehouses (CDWs) that apply standard models in hospitals, as well as a consortium whose main achievement was to establish big data based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [3], which is being used by Observational Health Data Sciences and Informatics, an international consortium with representatives from the United States and Europe [4]. A specialized database for 86 specific diseases, such as cancer, circulatory diseases, and respiratory diseases, was established.

Through this project, some hospitals began establishing CDWs for the first time, and hospitals that had already established a CDW incorporated new data or improved the system environment. The data attributes and forms included in the CDW vary among hospitals, so it is difficult to precisely summarize the included data, but the newly added data generally encompassed imaging, digital pathology, and signal data. Data scalability is eagerly anticipated in the future, so that data utilization may become more extensive.

Along with CDWs, many consortia are establishing and utilizing common databases based on the OMOP CDM. Since the OMOP CDM was already in wide use internationally, it is easy to build and integrate systems using the OMOP CDM between consortia. In particular, some consortia disclose data externally for educational purposes and are there-

fore expected to be more active in terms of joint use in the upcoming second phase of the project.

The establishment and utilization of big data by consortia are expected to be an important future direction for the integration and creation of consistent synergy throughout the entire healthcare sector.

3. Data Standardization and Quality Management

Data-driven hospital tasks have led many hospitals to map their data onto international standard terminological systems such as ICD (International Statistical Classification of Diseases and Related Health Problems), SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terms), RxNorm (standardized nomenclature for clinical drug developed by the US National Library of Medicine), and LOINC (Logical Observation Identifiers, Names, and Codes).

Data standardization is a labor-intensive task for which it is difficult to see tangible results in a short period of time, but it must be performed for the joint use of multi-agency data. However, standard mapping may have several options depending on the internal circumstances of the organization that is conducting the mapping project. For example, a single test name can be mapped to different codes, depending on which test method is used. For drugs, different criteria can be applied for which code to map compounds to. Currently, various hospitals engage in standard mapping using their own specific procedures, and it is regrettable that there is no organization that can mediate issues occurring in the mapping process or manage different mapping processes.

Data quality management primarily involves performing quality certification by the Korea Data Agency. However, the quality verification rules of the Korea Data Agency were not developed specifically for medical data; therefore, the use of this standard to evaluate data quality for medical institutions faces limitations.

For standardization and quality, it is very important to consistently apply the same standards to hospitals participating in the project. If medical institutions with different internal environments standardize data and manage quality on the same basis, the generated data will have very high external reliability.

The achievement of data-driven hospitals will be the acceleration of data utilization. Major infrastructure for data utilization has been established in the first phase of the project. Next, an important goal in the second phase of the project will be to increase utilization using the infrastructure that has been built.

With this goal in mind, if the purpose of data utilization is

presented in excessively general terms, data construction will also lose its purpose, and its utilization will inevitably decrease. Clearly defining a purpose is a necessary prerequisite to generating optimal data to achieve that purpose.

From this perspective, it would be difficult to say that the data construction in the first phase reflected a consistent purpose. In the second phase, it will be important to further clarify the purpose of use to be pursued in this project and to achieve that purpose. To this end, if the portal can provide high-quality data consistently mapped to international standards and verified with an error rate of less than 1% according to quality rules, improvement could be expected in the performance planned in the early stages of the project.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

ORCID

In Young Choi (<https://orcid.org/0000-0002-2860-9411>)

References

1. Korea Health Information Service. Healthcare Information Standard [Internet]. Seoul, Korea: Korea Health Information Service; c2022 [cited at 2023 Jan 28]. Available from: <https://hins.or.kr/index.es?sid=a2>
2. Chandler AD. Strategy and structure. Cambridge (MA): MIT Press; 1962.
3. Observational Health Data Sciences and Informatics. Standardized data: the OMOP Common Data Model [Internet]. [place unknown]: Observational Health Data Sciences and Informatics; c2023 [cited at 2023 Jan 28]. Available from: <https://www.ohdsi.org/data-standardization/>
4. Quiroz JC, Chard T, Sa Z, Ritchie A, Jorm L, Gallego B. Extract, transform, load framework for the conversion of health databases to OMOP. *PLoS One* 2022;17(4): e0266911. <https://doi.org/10.1371/journal.pone.0266911>