PHAROS, Platform for Harmonizing and Accessing Data in Real-time on Infectious Disease Surveillance Based on OMOP-CDM in Korea

Chungsoo Kim[1], Jimyung Park[1], Seongwon Lee[2], Rae Woong Park[1, 2]
[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea
[2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

Background

Individual clinical information of infectious disease in real-time is necessary for establishing treatment and disease management strategies; However, it is difficult to collect comprehensive clinical characteristics of infected patients as in the current infectious disease reporting system. In the actual response to the coronavirus disease (COVID-19), healthcare professionals pointed out that the information for managing COVID-19 is fragmented rather than collected in a unified system[1]. Also, work efficiency is reduced because different collectibles for each reporting system must be entered manually. Furthermore, many clinicians communicate with private messengers and separate data files for the patients, privacy concerns were also pointed out.

Different structures of electronic medical records act as an obstacle in data integration and management. A common data model (CDM) approach has been developed to solve problems cause by heterogeneous data structures. It is being used to standardize COVID-19 data through projects such as the E-CORE network of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance[2] and the National Covid Cohort Collaborative (N3C) of the US National Institutes of Health[3]. Although the CDM made a great contribution to clinical research and development through distributed research network, the timeliness is insufficient due to the data conversion process, so there is a limitation in the use for clinical practice. Therefore, we initiate a new project for developing an integrated infectious disease data managing system based on OMOP-CDM in Korea, named PHAROS.

Methods

The overall scheme of this R&D project was presented as the Figure 1.

1) Establishment of infectious disease CDM network

We intend to establish the infectious disease CDM (ID-CDM) in tertiary general hospitals. The ID-CDM has a structure of the OMOP-CDM and contains all information specialized to infectious diseases (e.g., microbiology test, infection symptoms). For the timeliness of data, we will apply the daily conversion with the automated extract, transform and load (ETL) technology and the near real-time incremental ETL technology[4]. In collaboration with a global OHDSI community, data modeling of infection specialized data for the OMOP-CDM and vocabulary mapping will be carried out. It will be used in the syndromic surveillance to respond to infectious diseases in each hospital.

2) Integration system for managing patient-level data for infectious disease

We will develop a platform that can manage the ID-CDMs, which is called the platform for harmonizing and access data in real-time on infectious disease surveillance (PHAROS). The PHAROS system will operate in two situations: peacetime and outbreak. It operates according to the current distributed network principle in peacetime. Using ID-CDM, healthcare providers and policymakers can request information and query execution on the infected population of each institution. However, the clinical information of infected patients in multiple institutions can be collected with the patient's consent or related legal approval in outbreak. The PHAROS system will curate data from each site into a single

CDM database so that authorized users can access the clinical information of patients directly. It aims to support healthcare providers' treatment strategy establishment and policy makers' quick decision-making to respond to pandemics through the collected information in the PHAROS system.

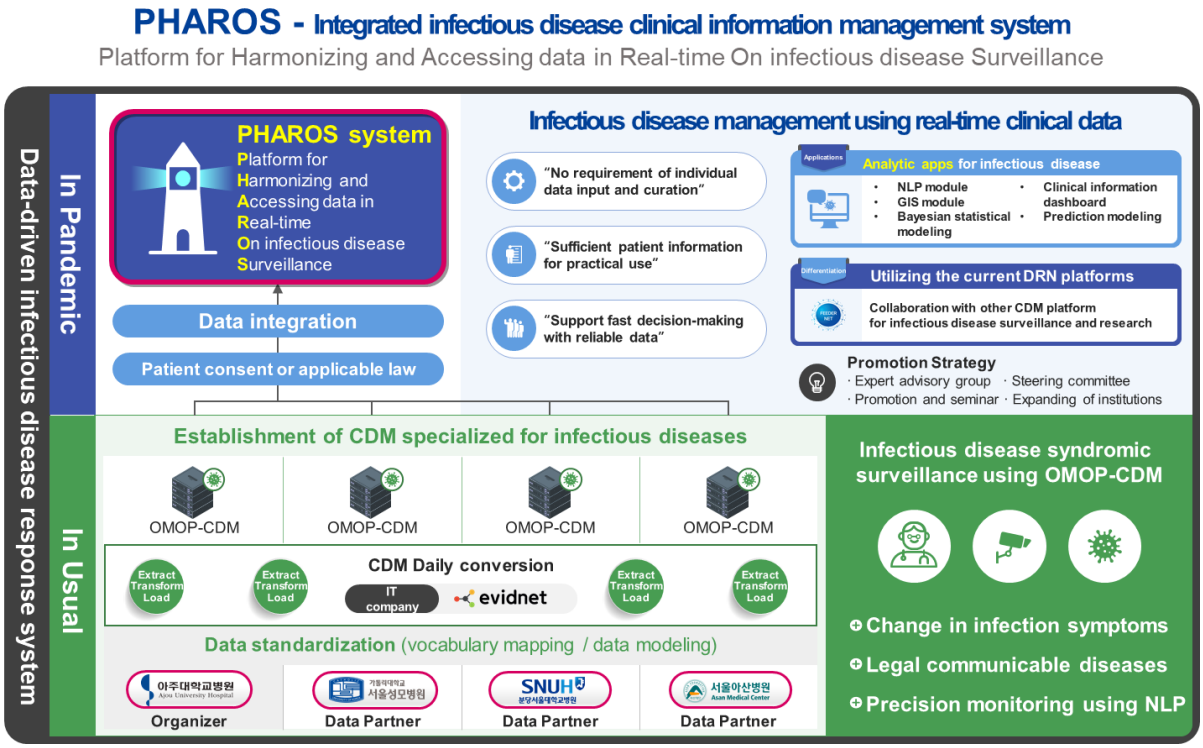

Figure 1. Schematic diagram for the PHAROS project

3) Development of analytic applications for clinical insight

In this project, we will develop statistics, visualization, and prediction applications which can be effectively utilized. A profiling application that can show patients' information at the individual-level and population-level, and a geographic information system application that can show infection patterns by region will be developed. In addition, we aim to develop a natural language processing module to utilize unstructured data, a Bayesian statistical model for time series data to predict infection trends, and severity prediction models using artificial intelligence algorithms.

Conclusion

We awarded EUR €2.1 million contract for the 3 years (2022–2024) from the Ministry of Health & Welfare, Republic of Korea. Currently, 50 researchers including infectious disease professionals, and IT professionals from 6 institutions (5 tertiary hospitals and 1 IT company) are participating in this project. This project will overcome the limitations of the current infectious disease reporting system and support a prompt response during peacetime and pandemics by integrating clinical information to CDM in near-real-time.

Acknowledgement

This work was supported by the Bio Industrial Strategic Technology Development Program (20003883, 20005021) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea), and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001). This research was also supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HG22C0024).

References

1. Hyeong-joo S. It is necessary to establish an integrated management system for health and medical resources in response to infectious diseases. *MEDICAL Observser*. 2021.11.10. Accessed 2022.04.28. http://www.monews.co.kr/news/articleView.html?idxno=308365
2. Candore G, PerezGutthann S. *Report - ENCePP Plenary virtual meeting 2021*. 2021. 18 Nov 2021. https://www.encepp.eu/publications/documents/Report_ENCePPwebinar_2021November18.pdf
3. National Covid Cohort Collaborative. Accessed 28 April, 2022. https://ncats.nih.gov/n3c
4. Lee S, Lee W, Lee S, Kim JY, Park RW. Near Real-Time Incremental OMOP-CDM ETL System *2021 OHDSI symposium collaborator showcase*. 2021;