Enhancing Infectious Disease Data Integration and management through OMOP-CDM in South Korea

Min Ho An, MD[1,2]*, Seok Kim, M.P.H[3]*, ByungJin Choi, MD[1,2], Sooyoung Yoo, Ph.D[3], Rae Woong Park, MD, Ph.D[1] , Ji Seon Oh, MD, Ph.D[4]

[1]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea
[2]Department of Medical Sciences, Graduate School of Ajou University, Suwon, Republic of Korea
[3]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Republic of Korea
[4]Information Medicine, Big Data Research Center, Asan Medical Center, Seoul, Republic of Korea
*These authors are equally contributed to this work

Background

- The Platform for Harmonizing and Accessing Data in Real-time on Infectious Disease Surveillance (PHAROS) was initiated to address challenges in data integration and management.
- PHAROS focuses on developing an integrated infectious disease data management system based on the OMOP-CDM in Korea, with the goal of enhancing real-time clinical information collection and improving treatment and disease management strategies.
- To support this, data encompassing microbial test results, infectious disease consultation notes, vaccination-related information, emergency room data, and legal infectious disease reports, were utilized, aimed at improving accessibility and ensuring clear representation of information.
- The codes within infectious disease consultation notes, vaccination-related information, emergency room data, and legal infectious disease reports are newly mapped and integrated as CDM records.
- Moreover, to address challenge of identifying detailed culture information, we developed new Extract Transform Load (ETL) method that suits to specifically store data drawn from specimen culture.
- While this model maintains the relationship between microbial tests and drug resistance, it captures various aspects of culture information without requiring additional data tables, thus improving the comprehensiveness and utility of information from specimen culture.

Methods

- In this study, OMOP-CDM was utilized to include infection-related clinical data. We used CDM version 5.4 without any additional columns. Infectious disease department consultation notes are integrated into the CDM's Note domain using specific concept ids, with consultation request recorded in the observation table.
- Additionally, vaccination-related reports are thoroughly documented in the drug domain, with dose information recorded in the observation table for detailed tracking.
- Primary symptom information from the National Emergency Department Information System (NEDIS) system is integrated by mapping chief complaints to SNOMED-CT and inserting them into the condition table or the observation table if no suitable mapping exists.
- We also utilized patient travel history from legal communicable disease reports. Particularly, Microbial test results were stored across three tables: specimens were stored in the specimen table, cultured microorganisms and antibiotic susceptibility results were stored in the measurement table, and the type of microorganism identified were stored in the observation table.
- These tables were designed to be linked using connection keys, facilitating the proper extraction of necessary data for various purposes.

Conclusions

- This study addresses infectious disease data integration challenges using the OMOP-CDM framework, standardizing clinical data for better accessibility and comprehensiveness. The new ETL method stores detailed culture information without extra tables, preserving key relationships between microbial tests and drug resistance. This approach may enhance research, supports rapid outbreak response, and improves disease management

Acknowledgement

- This research was funded a grant from 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: HR16C0001)
- This research was supported by a government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG22C0024, KH124685).

Contact: minho.an23@gmail.com

Results

▪ Original Data in EMR

Urine culture report

Enterococcus faecium (+)
Pseudomonas aeruginosa (-)
Streptococcus (-)

Enterococcus Drug-resistance report

Amikacin	<=1 S
Gentamicin	>2 R
Cefazolin	<=2 S
Ciprofloxacin	<=2 S
Oxacillin	>4 R
Vancomycin	>4 R

Culture

Local code	Test name	abbreviation	Specimen	Result format	Reference	example
DB0690	Bacteria,Final (Des)[Culture,ID/Sensitivity],Blood	Blood,Final culture	Blood	Descriptive	No growth <10^3/ml	
DB0253	Bacteria (Des)[Culture,ID/Sensitivity],Sputum	Sputum,Culture	Sputum	Descriptive		Normal Flora
DB0203	Bacteria (Qn)[Culture,ID/Sensitivity],Urine	Urine,Culture	Urine	Quantitative		report

Drug-resistance

antibiogram_id	person_id	date	exam_no	batr_cd	citur_bact_exam_rslt_sn	bctc_cd	bctc_exam_rslt_sn	bctc_exam_rslt_val	bctc_exam_judg_rslt_cd
92308	1001	2019-03-07	B40BABI0031	BENFEA	2	AGMS	13	<= 500 S	S
92307	1001	2019-03-07	B40BABI0031	BENFEA	2	ASTS	12	<= 1000 S	S
92306	1001	2019-03-07	B40BABI0031	BENFEA	2	AVA	11	1 S	S
92305	1001	2019-03-07	B40BABI0031	BENFEA	2	ATEI	10	<= 4 S	S
92304	1001	2019-03-07	B40BABI0031	BENFEA	2	ATE	9	8 I	I

➢ Specimen Table - ① Specimen acquired

specimen_id	person_id	specimen_concept_id	specimen_date	specimen_source_value	specimen_source_id
1	1001	4046280 (Urine specimen)	2019-03-01	Urine	123

➢ MEASUREMENT Table - ② Lab (culture) test , ③ Antibiotics-resistance

measurement_id	person_id	measurement_concept_id	measurement_date	value_source_value	value_as_concept_id	value_as_number	operator_concept_id	measurement_source_concept_id	measurement_source_value	measurement_event_id	meas_event_field_concept_id
101	1001	**3026008** (Bacteria identified in Urine by Culture)	2019-03-07	DB0203 (Urine Culture)	**4139623 (Isolated), 4139623 (No growth)**				report	1	1147049(Specimen_id)
102	1001	**3002274** (Amikacin [Susceptibility] by MIC)	2019-03-07	<= 1 S	**4038110** (Susceptible)	1	4171754 (<=)			10	1147165 (observation_id)
103	1001	**3000769** (Gentamicin [Susceptibility] by MIC)	2019-03-07	> 2 R	**4148441** (Resistant)	2	4172704 (>)			10	1147165 (observation_id)
104	1001	**3005535** (ceFAZolin [Susceptibility] by MIC)	2019-03-07	> 8 R	**4148441** (Resistant)	8	4172704 (>)			10	1147165 (observation_id)
105	1001	**3010007** (Ciprofloxacin [Susceptibility] by MIC)	2019-03-07	<= 2 S	**4038110** (Susceptible)	2	4171754 (<=)			10	1147165 (observation_id)
106	1001	**3018088** (Vancomycin [Susceptibility] by MIC)	2019-03-07	> 4 R	**4148441** (Resistant)	4	4172704 (>)			10	1147165 (observation_id)

➢ OBSERVATION Table – ④ Micro-organism

observation_id	person_id	observation_concept_id	observation_date	value_source_value	value_as_concept_id	value_as_number	operator_concept_id	observation_source_concept_id	observation_source_value	observation_event_id	obs_event_field_concept_id
10	1001	**4228531** Bacterial growth present	2019-03-07	:Enterococcus faecalis	**4302157** (Enterococcus faecalis)					101	1147138 (measurement_id)

Figure 1. Culture Modeling Table Schema

Type	Number of Patients	Number of Data	Detailed Data Items
Person	**239,310**	**239,310**	**De-identified ID, gender, birthdate, etc.**
Visit Occurrence	**239,108**	**5,928,625**	**Visit start/end time, visit type (outpatient, inpatient, etc.)**
Condition Occurrence	**238,859**	**84,261,109**	**Diagnosis code, diagnosis date**
Drug Exposure	**238,769**	**9,565,547**	**Drug code, prescription date, drug quantity, etc.**
Procedure	**238,707**	**159,246,096**	**Procedure code, procedure date**
Measurement	**227,579**	**58,070,197**	**Measurement code, result, unit (continuous, categorical, text, etc.)**
Device	**224,278**	**9,867,621**	**Medical device code, order date, amount**
Death	**222,246**	**14,820,106**	**Death date, cause of death**
Observation	**201,959**	**78,984,878**	**Other clinical information, observation date**
Specimen	**6,211**	**6,211**	**Specimen code, collection date, quantity, unit**

Table 1. Converted Data Summary in Ajou University Hospital

- A total of 560 codes for infection types, testing procedures, antimicrobial sensitivity, and travel history were mapped. Additionally, the National Emergency Department Information System (NEDIS) was mapped to include 1,114 codes for major symptoms and issues.
- A total of 2,226 codes were mapped for legal infectious diseases. Furthermore, how infection-specific data such as microbial tests and antibiotic susceptibility results are stored in the CDM is illustrated in Figure 1.
- The information for specimen acquisition is recorded in the specimen table with the corresponding specimen concept ID (①).
- The results of laboratory (culture) tests are documented in the measurement table in “value_as_concept_id,” indicating the existence of microorganisms by “isolated” or “no growth” (4139623) and linked to the specimen table through the “measurement_event_id” and “meas_event_field_concept_id” to trace the source (②).
- Additionally, antibiotic susceptibility data (③) is loaded into the measurement table. The differentiation from laboratory (culture) tests is achieved by using “meas_event_field_concept_id” with the related field as “observation_id”.
- Lastly, the type of identified microorganism is recorded in the observation table (④), with the presence identified by observation_concept_id, and the name of the microorganism designated in “value_as_concept_id”. This data is linked through the field “observation_id” matched with “measurement_event_id” in the measurement table.