scientific data

Scalable Infrastructure Supporting Reproducible Nationwide Healthcare Data Analysis toward FAIR Stewardship

Ji-Woo Kim¹, Chungsoo Kim², Kyoung-Hoon Kim³, Yujin Lee³, Dong Han Yu¹, Jeongwon Yun¹, Hyeran Baek¹, Rae Woong Park²,⁴,⁸ & Seng Chan You⁵,⁶,⁸
OMOP in Korea

65 Data partners

- 63 Hospitals
  (70% of tertiary hospitals)
- 2 National Institutions
- Coordinating center:
  Ajou University
National Health Insurance Data in Korea

National Claims
- From Clinics, general/tertiary hospital, pharmacy
- Including demographics, income level, diagnosis, prescription, measurement (no value of measures), procedure, device, etc
- Cost
- National Health checkup data (NHIS only)

Data Linkage
- Death registry (Statistics Korea)
- Cancer registry (National Cancer Center)
- Vaccine registry (KCDC)

Health Insurance Review and Assessment Service (HIRA)
HIRA CDM projects

2018 - 2020 Pilot CDM ETL
Study published in JAMA (Chan et al)

2020 OpenData4Covid project
World first data open for research
(Published in many journals)
HIRA CDM projects

HIRA CDM analytic environment
• ABMI supports HIRA to manage their own analytic environment for future research.
• It was not easy for implementing R into the off-line environment, but we resolved using Docker.

On-premise environment (manually developed)
Data ETL

Mapping (EDI to SNOMED)

Quality check

Open to public

Analytic infra

Data catalog (ETL rule and sample data)
HIRA CDM projects

HIRA CDM Open Project

1st pilot opening with 10M data (2022)

About 40 studies were done using this data

2nd pilot opening with 50M data (2023)

About 15 studies are running using this data

Currently, HIRA is accepting applications 1-2 times a year, but they are preparing to accept applications more frequently.
Thank you

Ji-Woo Kim
Kyoung-Hoon Kim
Yujin Lee
Dong Han Yu
Jeongwon Yun
Hyeran Baek
Rae Woong Park
Seng Chan You

Big data department