Adaptation of the OMOP Common Data Model for Secondary Use of Public Databases on the Japanese Healthcare Information Platform

Takanori Yamashita[1*], Eizen Kimura[2], Yoshihiro Aoyagi[3], Hiroshi Okamura[4], Eri Matsuki[5,6], Tatsuo Hiramatsu[7]

[1]Medical Information Center, Kyushu University Hospital

[2]Department of Medical Informatics, Medical School of Ehime University

[3] Department of Medical Information, National Cancer Center Hospital East

[4]Department of Laboratory Medicine and Medical Informatics, Graduate School of Medicine, Osaka Metropolitan University

[5]Division of Hematology, Department of Medicine, Keio University School of Medicine

[6]Division of Network Support, Clinical and Translational Research Center, Keio University Hospital

[7]Department of Medical Informatics, International University of Health and Welfare

Background

The utilization of real-world data (RWD) derived from daily clinical practice is rapidly advancing. In Japan, RWD use faces challenges due to inconsistencies in electronic medical record (EMR) systems and laboratory test codes across hospitals. This impacts the quality of extracted data and hinders integrated multi-institutional analysis. However, in Japan, the government and academic societies have been working to create high-quality RWD, while confirming the importance of data quality management through data-driven projects.[1,2,3,4]. There is a need to understand the details of initiatives overseas, from building RWD infrastructure to generating evidence, and the standards used in data analysis. Furthermore, the government is progressing with its medical DX policy to build a Nationwide Health Information Platform (Figure 1). This platform aims to further promote the health of citizens, efficiently provide seamless and high-quality medical care, and streamline operations for medical institutions. The EMR information sharing service on that platform is designed to collect medical data from each hospital's EMR using FHIR. The secondary use of these collected public databases is expected to significantly contribute to public health, medicine, and industry. Also, developing internationally interoperable RWD platforms is indispensable. This study aims to evaluate and validate the process of converting domestic RWD into the OMOP Common Data Model, with the goal of facilitating the broader utilization of RWD both nationally and internationally.

Methods

We collaborated with domestic organizations and RWD projects such as OHDSI Japan[5] and Rinchu-Net[6] to gather information on the activities of the OHDSI network. To promote appropriate conversion to OMOP and its widespread use, we began to clarify standard specifications for OMOP conversion. In the EMR information sharing service, YJ codes[7] are used for pharmaceuticals, and JLAC10 and JLAC11[8] are used for laboratory test items, are all domestic standard codes. We investigated their corresponding OMOP mappings. We also investigated various data elements of the cancer whole genome analysis project. Through this detailed investigation, we aimed to understand the complexities of mapping diverse medical data to the OMOP common data model and to ensure high-quality, standardized RWD, which is essential for robust analysis and evidence generation in various medical settings.

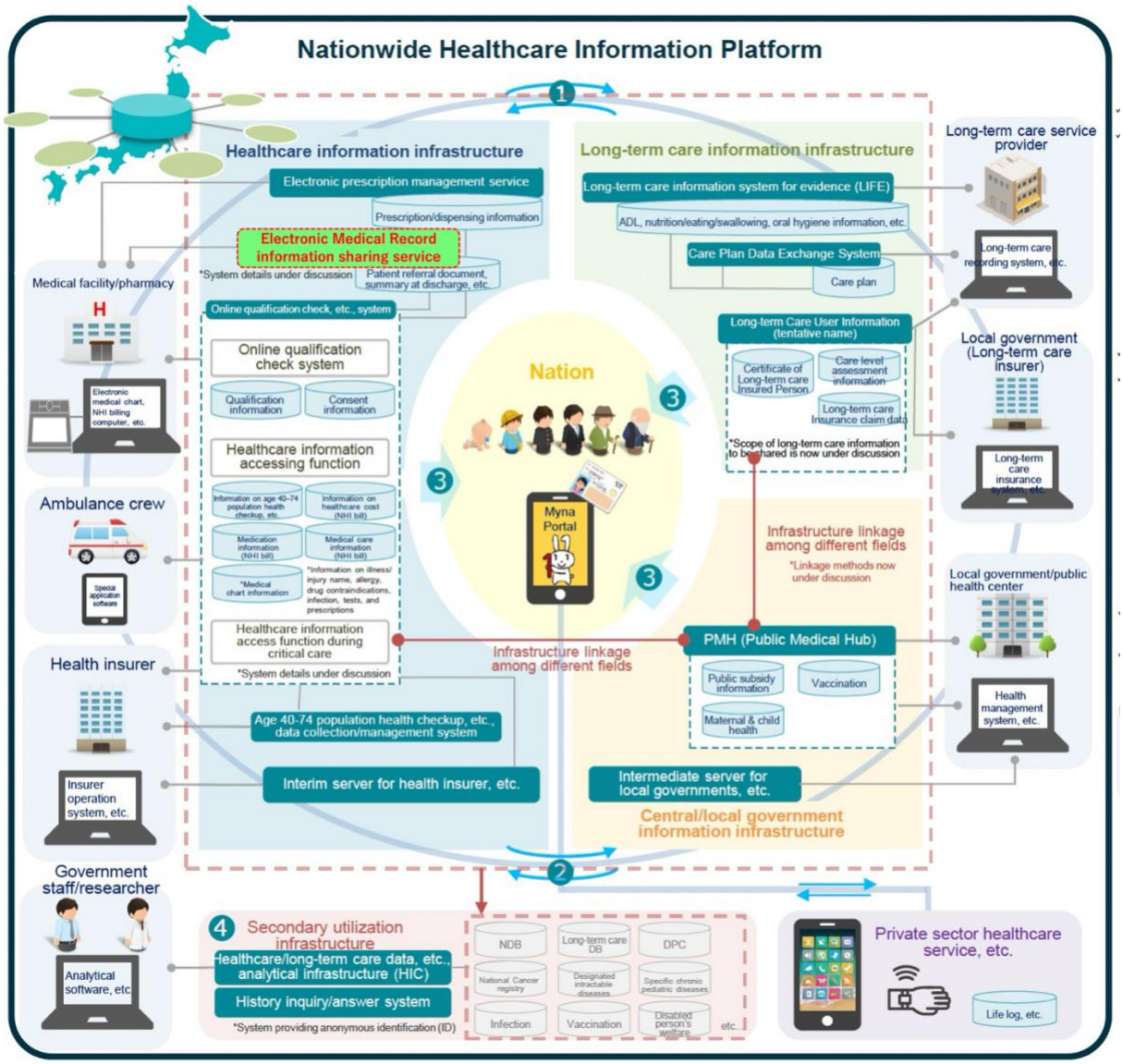

Figure 1. Japanese Nationwide Healthcare Information Platform Overview. The red dashed box highlights the EMR information sharing service.

Results

As a result of verifying whether ETL to OMOP CDM could be smoothly performed by linking it with FHIR JP Core 1.2 Profile[9], which is Japan's standard data profile, it was shown that the FHIR facade server compliant with JP Core is sufficiently functional and can cover over 80% of the essential data fields in OMOP CDM.

For the laboratory tests in the EMR information sharing service (Table 1), while mapping from JLAC11 to LOINC was considered, the final mapping aligned with OHDSI Standard Vocabulary (OSV) standard concepts, as adopted in OMOP. However, the granularity of laboratory-related standard concepts in OSV is generally coarse, precluding analyses at the finer granularity provided by JLAC codes. Even if data offer

JLAC11 granularity, international collaborative analysis necessitates mapping to OSV standard concepts. For OMOP mapping of pharmaceuticals in Japan, a RxNorm conversion combining LLM (Large Language Model) and expert review is being developed[10]. We found that careful examination of vocabularies and mapping processes is crucial for OMOP conversion. During mapping, the necessity of SNOMED CT became apparent.

Table 1. Laboratory test items for the Electronic Medical Record information sharing service on Nationwide Healthcare Information Platform

No	Category	Item
1	**Biochemical Tests**	Total Protein (TP)
2		Albumin (Alb)
3		Creatine Kinase (CK)
4		Aspartate Aminotransferase (AST / GOT)
5		Alanine Aminotransferase (ALT / GPT)
6		Lactate Dehydrogenase (LD / LDH)
7		Alkaline Phosphatase (ALP)
8		Gamma-Glutamyl Transpeptidase (γ-GTP / GGT)
9		Cholinesterase (ChE)
10		Amylase (AMY)
11		Creatinine (Cre)
12		Cystatin C (CysC)
13		Uric Acid (UA)
14		Blood Urea Nitrogen (BUN)
15		Blood Glucose (Glucose)
16		Hemoglobin A1c (NGSP)
17		Triglycerides (TG)
18		Total Cholesterol (T-CHO)
19		HDL Cholesterol (HDL-C)
20		LDL Cholesterol (LDL-C)
21	**Electrolytes**	Sodium (Na)
22		Potassium (K)
23		Chloride (Cl)
24		Calcium (Ca)
25		Total Bilirubin (T-Bil)
26		Direct Bilirubin (D-Bil)
27	**Hematology**	White Blood Cell Count (WBC)
28		Red Blood Cell Count (RBC)
29		Hemoglobin (Hb)
30		Hematocrit (Hct)
31		Platelet Count (Plt)
32		Activated Partial Thromboplastin Time (APTT)
33		Prothrombin Time (PT)
34		FDP / D-dimer (DD)
35	**Urinalysis**	Urine Protein
36		Urine Glucose
37		Urine Occult Blood
38		Protein-to-Creatinine Ratio (P/C Ratio)
39		Albumin-to-Creatinine Ratio (A/C Ratio)
40		Urine Specific Gravity
41		Urine pH
42		Urinary Sediment
43	**Others**	B-type Natriuretic Peptide (BNP)
44		N-terminal proBNP (NT-proBNP)
45		C-Reactive Protein (CRP)
46		Blood Type (ABO, Rh)

For items of the Cancer Whole Genome Analysis Project, representing dynamically transitioning information, such as treatment methods and disease states, was essential. These were modeled using CONDITION_OCCURRENCE, MEASUREMENT, OBSERVATION, PROVIDER, SPECIMEN, and DRUG_EXPOSURE, in addition to the EPISODE table from the OMOP Oncology Module. Of the 40 clinical items examined, 36 items (90%) were successfully mapped to the target model. However, the existence of multiple definitions for cancer disease states complicated accurate mapping.

The EMR information sharing service primarily adopts national standard specifications, requiring hospitals to ensure FHIR output and standardized code mapping in their EMR systems. As EMR standardization advances across hospitals, data quality will improve, fostering a more reliable data sharing infrastructure and accumulating high-quality RWD. Given that the nationwide standardization of EMR systems is a government-led initiative, these activities offer valuable insights, especially for small and mid-sized hospitals. For the dynamic aspects of cancer genome data, defining and evaluating Complete Response requires aligning the timing of tests used for assessing treatment efficacy, posing a significant challenge.

Public databases not only accumulate EMR and claims data but also broader healthcare data. Linking these sources enables comprehensive surveys and analyses with complete coverage. OMOP conversion can facilitate a federated, secure data infrastructure. This holds substantial potential for addressing issues related to Japan's super-aging society, informing government policy decisions, and guiding pharmaceutical and medical device development. To prevent semantic inconsistencies, standardized mapping must be harmonized across institutions.

Conclusions

To progress the international use of RWD, we implemented standardization efforts based on the OHDSI network and OMOP. Our work on OMOP conversion for the EMR information sharing service and Cancer Whole Genome Analysis highlighted the critical importance of vocabulary validation and mapping precision. Key challenges included discrepancies in the granularity of standard concepts for laboratory tests and difficulties in representing dynamic transitions in cancer disease information. Furthermore, while progressing with OMOP implementation, we also recognize critical challenges related to the International Patient Summary (IPS) in the Japanese context. A primary hurdle is the interoperability between existing Japanese healthcare IT systems and the IPS standard. The findings of this study will be valuable for future dissemination activities and international responses. We plan to formulate guidelines and policies for OMOP conversion to build a reliable data infrastructure that supports the consistent utilization of public databases both domestically and internationally in the future.

Acknowledgment

This study was supported by the Health and Labour Sciences Research Grant from the Ministry of Health, Labour and Welfare of Japan (Grant Number: 25IA2014).

References

1. Yamada K, Itoh M, Fujimura Y, Kimura M, Murata K, Nakashima N, et al. The utilization and challenges of Japan's MID-NET® medical information database network in postmarketing drug safety assessments: a summary of pilot pharmacoepidemiological Studies. Pharmacoepidemiol Drug Saf. 2019; 28(5): 601-608.
2. Yamaguchi M, Inomata S, Harada S, Matsuzaki Y, Kawaguchi M, Ujibe M, et al. Establishment of the MID-NET® medical information database network as a reliable and valuable database for drug safety assessments in Japan. Pharmacoepidemiol Drug Saf. 2019;28(10):1395-1404.
3. Nakagawa N, Sofue T, Kanda E. et al. J-CKD-DB: a nationwide multicentre electronic health record-based chronic kidney disease database in Japan. Sci Rep. 2020;10:7351.
4. Yamashita T. A Solution to the Problem of Data Quality in MID-NET. Epidemiologic Research on Real-World Medical Data in Japan. Springer. 2022;1:51-54.
5. OHDSI Japan. Available from: https://www.ohdsi-japan.org (accessed 2025 Jul 1).
6. Aoyagi Y, Rinchu-Net Internationalization Task Force. Internationalization efforts for real-world evidence creation at core hospitals for clinical research in Japan. Presented at: OHDSI APAC Symposium 2023.
7. Council for Promotion of Pharmaceutical Information Standards (CAPS). List of YJ Code: The codes for individual prescription drugs in Japan. Available from: https://www.capstandard.jp/yj_code.html (accessed 2025 Sep 17).
8. Institute of Health Data Infrastructure for All (IDIAL). JLAC11 codes. Available from: https://www.idial.or.jp/jlac_eleven.html (accessed 2025 Sep 17).
9. HL7 FHIR JP Core ImplementationGuide (1.2.0-a) [cited 2024 12 30]. Available from: https://jpfhir.jp/fhir/core/1.2.0-a/index.html.
10. Kimura E, Kawakami Y, Inoue S, Okajima A. A dataset for mapping the Japanese drugs to RxNorm standard concepts. Data in Brief. 2025;59:111418.