Open Source Tools and Terminology to Increase Representativeness in OHDSI Data

Andrew S. Kanter, MD MPH FACMI FAMIA[a,b]
[a] Columbia University, New York, USA, [b] Open Concept Lab

Background

The Observational Health Data Science and Informatics (OHDSI) network is a global community of open-science interdisciplinary stakeholders who collaborate on large scale analytics of health data. It includes users in 80 countries with 453 data sources coming from 41 countries and representing 928 million unique patients.[1] Unfortunately, there is currently little representation from low and middle income countries (LMICs). In particular, Africa is completely missing from the list (see Figure 1).
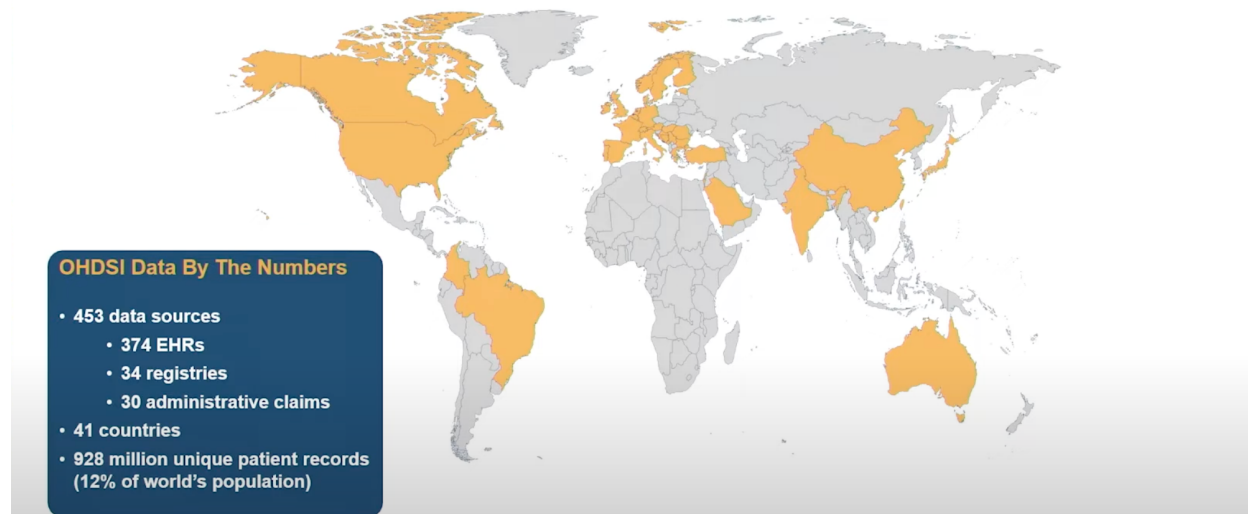

Figure 1: OHDSI Data Partners, 2022

On the other hand, significant progress has been made in LMICs implementing health information systems and enterprise architectures for their public health systems. Open source tools like the Open Medical Record System (OpenMRS) (for patient-level data) and the District Health Information System 2 (DHIS-2) (for aggregate data) have gained widespread adoption. Digital Square, funded by USAID and the Gates Foundation among others, has developed a portfolio of "Global Goods for Health[2]".

This presentation will focus on how the combination of open source tools, such as Global Goods can be combined with open source interface terminology and OHDSI tooling to increase its representation of LMIC data. Collaborations such as OpenHIE.org provide countries with best practices, use cases and reference implementations of national health information exchange. Connecting primary health delivery use cases with secondary use cases such as reporting and analytics is a major outcome from OpenHIE. Up to now, however, the OHDSI research use case has not been at the forefront of LMIC national data discussions. It is possible for the national information system work in LMICs to include deidentified,

patient-level research use cases. Recently, the OHDSI Africa Working Group has proposed a data pipeline that would allow LMICs to leverage the OMOP Common Data Model (CDM) to not only perform network research studies, but also serve as the basis of national reporting.

Methods

OpenMRS is an open source electronic medical record platform designed on an Entity-Attribute-Value data model similar to that at Regenstrief Institute. It leverages a centralized concept dictionary and has a Fast Healthcare Interoperability Resource (FHIR) API which provides for data statdardization and semantic interoperability**.** OpenMRS started in 2004 through a collaboration between the Regenstrief Institute, Partners in Health and the South African Medical Research Council, joined in 2006 by the Columbia University/United Nations Development Programe Millennium Villages Project (MVP).

The Columbia International eHealth Laboratory (CIEL) terminology is an open source, standardized interface terminology developed out of MVP in 2006 and has been producing monthly releases since. A portion of CIEL is included as a non-standardard vocabulary within the OHDSI Athena database.

Open Concept Lab[3] is an open source terminology services platform that grew out of an initial content sharing project known as Maternal Concept Lab in 2009. Initially supported by the Gates Foundation, Open Concept Lab provides terminology support to the US President's Emergency Plan for AIDS Relief (PEPFAR) DATIM project as well as numerous community organizations including OpenMRS.

The combination of the CIEL terminology, managed through Open Concept Lab and deployed in point-of-care applications like OpenMRS provides for the possibility of bringing in large numbers of patients within LMICs currently invisible to the health research community. National and sub-national health information exchanges could be connected to the OHDSI community.

Results

OpenMRS is currently being used in 40 countries including 8,139 implementations (see Figure 2). There are estimated to be 16.6 million patients in OpenMRS databases[4]. Although OpenMRS can use any number of concept dictionaries and there are various levels of customization, the primary shared concept dictionary is the CIEL dictionary. CIEL currently has been downloaded by users in over 45 countries, and is the basis of (or planned to be) national health data dictionaries in several LMICs. In addition, CIEL has become a default interface terminology not only for OpenMRS, but for several non-governmental organizations and private companies.
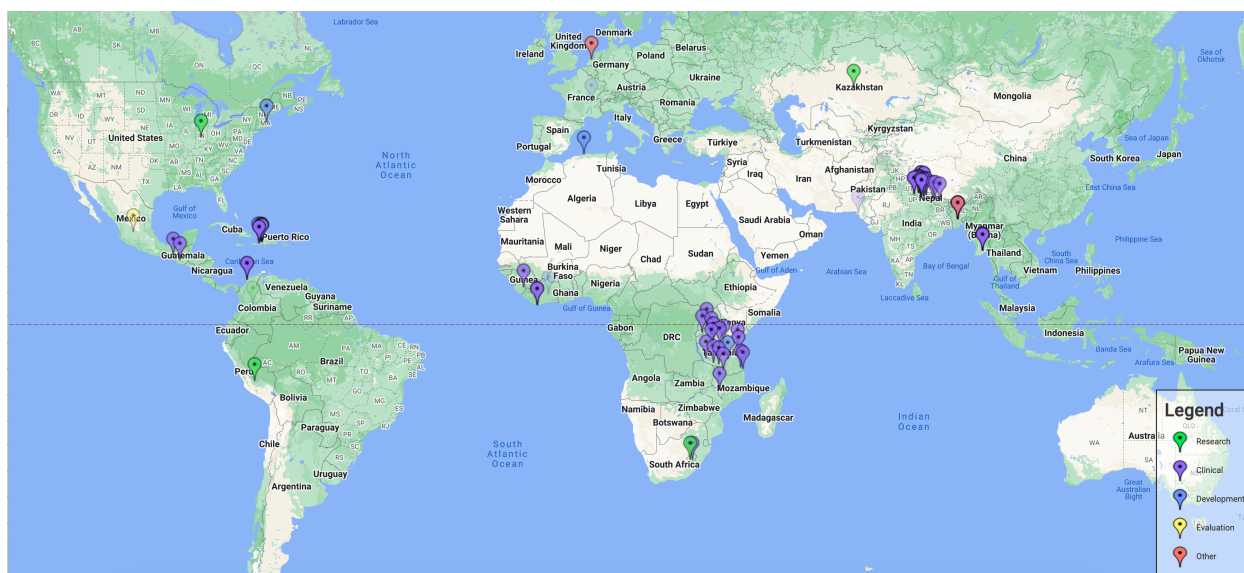
Figure 2:OpenMRS implementations by type, 2023

The CIEL terminology is currently available via the Open Concept Lab at www.CIELterminology.org. It contains over 54,000 concepts mapped to 43 different coding systems including SNOMED CT, ICD-10-WHO, LOINC and RxNORM (see Figure 3) with concept names in 18 languages[5].

Code System (Top 10)	Number of Mapped Concepts
SNOMED-CT	53,460
ICD-10-WHO	37,931
IMO-ProblemIT	37,015
3BT	7,705
RxNORM	7,586
ICPC2	6,944
IMO-ProcedureIT	1,079
LOINC	778
ICD-11-WHO	641
HL-7-CVX	146

Language	Number of Concepts	Number of Names
English	54,037	79,397
Spanish	41,709	53,372
Dutch	11,775	11,912
French	5,398	6,241
Vietnamese	3,803	4,018
Russian	2,726	2,739
Haitian Creole	1,919	2,159
Kiswahili	234	244

Figure 3: CIEL mapped code systems and language of concept names, May 2023 version

Because CIEL is an interface terminology driven by LMIC users for 17 years, it contains concepts which do not map 1:1 with standard concepts like SNOMED CT. About 50% of the dictionary is mapped either to multiple SNOMED codes (to fully capture the source concept) or mapped narrower-than to a single code. Although a majority of CIEL concepts are diagnoses/findings, procedures, medications, and lab/imaging, CIEL contains observable concepts used to capture common question/answers found in LMIC settings. These include concepts such as: "Condom use during high risk sex" or "frequency of bednet use in past month." In addition, many LMICs source their medication from non-US or FDA-approved sources, so CIEL contains drug concepts that may include combinations of ingredients or drug names that are not available in RxNORM, and are therefore mapped to multiple SNOMED CT and/or RxNORM codes. A large portion of the CIEL terminology is already included in the OHDSI terminology system. Recently in Rwanda, the LAISDAR project demonstrated Extract-Transfer-Load (ETL) pathways from OpenMRS databases into OMOP[6], making the import of patient-level LMIC data into OHDSI a real possibility. LAISDAR reported that

15 hospitals, about half-of-which are OpenMRS, are included in their database. Improvements in mapping are underway.

Conclusion

The combination of an open source interface terminology like CIEL, along with an open source community platform to manage, maintain, version and disseminate the terminology to point of care systems such as OpenMRS lays the foundation for standardized and interoperable data coming from LMICs into the OHDSI community. This will help improve the representativeness of OHDSI network studies by including patients so far invisible to the OHDSI community.

References

1. OHDSI State of the Community, OHDSI Symposium, 2022.
2. https://digitalsquare.org/digital-health-global-goods (accessed June 14, 2023).
3. https://openconceptlab.org (accessed June 14, 2023)
4. https://openmrs.org (accessed June 14, 2023)
5. http://CIELterminology.org
6. Nishimwe A, Ruranga C, Musanabaganwa C, Mugeni R, Semakula M, Nzabanita J, Kabano I, Uwimana A, Utumatwishima JN, Kabakambira JD, Uwineza A, Halvorsen L, Descamps F, Houghtaling J, Burke B, Bahati O, Bizimana C, Jansen S, Twizere C, Nkurikiyeyezu K, Birungi F, Nsanzimana S, Twagirumukiza M. Leveraging artificial intelligence and data science techniques in harmonizing, sharing, accessing and analyzing SARS-COV-2/COVID-19 data in Rwanda (LAISDAR Project): study design and rationale. BMC Med Inform Decis Mak. 2022 Aug 12;22(1):214. doi: 10.1186/s12911-022-01965-9. PMID: 35962355; PMCID: PMC9372951.