The Office of the National Coordinator for Health Information Technology

Common Data Model Harmonization

Harmonization of Various Common Data Models and Open Standards for Evidence Generation to Support Patient-Centered Outcomes Research

HealthIT.gov @ONC_HealthIT

Table of Contents

OVERVIEW

Disease treatments are approved on the basis of safety and efficacy data collected during carefully designed clinical trials. However, clinical trials are conducted on a small scale under rigorous conditions, so the results may not predict how a treatment will work for the general patient population. For example, clinical trials usually last 1 to 4 years, while people may need treatment for many years beyond that.[1] Trial participants must meet strict eligibility requirements. Clinical trials often exclude the elderly and children, as well as people who are taking other medications.

Including observational data in clinical trials and other research could provide a way to assess real-world effectiveness of treatments and detect adverse effects that are rare or slow to appear. Effectiveness research can complement clinical trials by addressing the questions of which treatment works best, for whom, and under what conditions. Patient-centered outcomes research (PCOR) goes a step further by focusing on measuring outcomes, risks, and benefits that matter most to patients.[2,3] Sample questions that researchers may ask include, "What is the safest and most effective way for people who have chronic obstructive pulmonary disease to quit smoking?" or "How does a traditional education program on how to quit compare to a guided maintenance therapy program with nicotine replacement?"

Providing researchers with broader access to observational data based on results from patients using an approved treatment or medical device could help stimulate more PCOR and ultimately aid patients and their healthcare providers as they discuss treatment options. Harmonizing various common data models (CDMs) that are used to structure these data could expand the research community's access to a large sample size and additional demographics (e.g., the elderly, children, non-U.S. population). Such robust patient cohorts also have a greater breadth of information, which is particularly useful for rare events and for carrying out analyses from a global perspective. Additionally, the 21st Century Cures Act, passed in 2016, places greater focus on the use of PCOR data to support regulatory decision making.[4]

In 2017, the Office of the National Coordinator for Health Information Technology (ONC) began a project supported by the Patient-Centered Outcomes Research Trust Fund (PCORTF),[5] with the goal of building data infrastructure that

The Harmonization of Various Common Data Models and Open Standards for Evidence Generation, or Common Data Model Harmonization (CDMH) project, aims to support research and analyses across multiple data networks by harmonizing several existing CDMs. This is one step in enabling researchers to tap into the wealth of real-world data from large numbers of patients who represent a wider range of demographics, health conditions, and treatments.

[1] Food and Drug Administration (FDA). Step 3: Clinical Research. https://www.fda.gov/patients/drug-development-process/step-3-clinical-research.
[2] Patient-Centered Outcomes Research Institute (PCORI). Research We Support. https://www.pcori.org/research-results/about-our-research/research-we-support.
[3] Lele C. (2011). Comparative effectiveness research. *Perspectives in Clinical Research*, *2*(2), 48. doi:10.4103/2229-3485.80365.
[4] FDA. Real-World Evidence. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.
[5] Office of the Assistant Secretary for Planning and Evaluation. Patient-Centered Outcomes Research Trust Fund FAQs. https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs.

supports the inclusion of observational data derived from a variety of sources that collect information about the delivery of healthcare (e.g., insurance billing claims, electronic health records, patient registries). The CDMH project entailed mapping common data elements from four CDMs used in research to organize data in large and diverse observational databases. Because each CDM is structured in a different way, harmonization of the CDMs across their respective data networks could open up new prospects to support research and analyses using different sets of observational data.

A hallmark of the CDMH project is the extensive collaboration among five federal partners:

- The Food and Drug Administration (FDA) led the interagency team for the CDMH project and provided subject matter expertise.
- The National Center for Advancing Translational Sciences at the National Institutes of Health (NIH), served as co-lead and provided informatics support and the framework for the implementation strategy.
- The National Cancer Institute (NCI), also at NIH, engaged its researchers as subject matter experts who evaluated the application of the CDMH data infrastructure to an oncology use case.
- ONC advised on data standards and the proposed approach for mapping among the CDMs, facilitated discussions with data model owners, and convened three workshops for stakeholders.
- The National Library of Medicine, also at of NIH, advised on clinical terminology standards, metadata, and data models that were leveraged for the project.

KEY ACTIVITIES

At an October 2017 workshop, the project partners began the process of developing mapping tools and evaluated the proposed data architecture. Stakeholders and subject matter experts who had experience with each of the four CDMs discussed activities to be conducted throughout the 22 months of the collaboration.

The CDM Data Infrastructure

The four CDMs that were engaged for this project included:

- FDA's Sentinel Initiative, which is the largest multisite, distributed database in the world dedicated to medical product safety. Sentinel has access to records from 350 million patients, mainly from administrative claims.
- Patient-Centered Outcomes Research Network (PCORnet), a network representing data from approximately 70 million patients, mainly from electronic health records.
- Informatics for Integrating Biology & the Bedside (i2b2), an NIH-funded National Center for Biomedical Computing (NCBC) based at Partners HealthCare System in Boston, Massachusetts. i2b2 supports patient cohort identification at numerous academic medical centers and hospitals serving millions of patients.
- The Observational Medical Outcomes Partnership (OMOP), an observational database of electronic health records set up by Observational Health Data Sciences and Informatics (OHDSI). OHDSI has established an international network of researchers and observational health databases linked by a

central coordinating center, with access to more than 100 different databases, representing more than 500 million patient records from 19 different countries.

For this project, the Biomedical Research Integrated Domain Group (BRIDG) Model was selected as the intermediary model that could provide a "shared view of the dynamic and static semantics for the domain of basic, pre-clinical, clinical, and translational research and its associated regulatory artifacts."[6] Due to variability among CDMs and to accommodate the needs of the data networks, some data elements were added to BRIDG before mapping was conducted. Once these mappings were complete, the resulting BRIDG 5.1 was balloted at Health Level Seven International® (HL7®) International Conference & Working Group Meeting in May 2018.[7] The ballot passed in June 2018.

Additionally, a subset of data elements were mapped from BRIDG to HL7's Fast Healthcare Interoperability Resources® (FHIR®) standard. FHIR was included in the approach ONC proposed because of the ability to enable data access using application programming interfaces (APIs) as well as growing adoption of the standard among both healthcare and health research stakeholders. The Common Data Models Harmonization FHIR Implementation Guide was created to support implementers in mapping and translating results obtained from queries to a FHIR format.

The CDMH FHIR Implementation Guide and mappings to FHIR that resulted from this project are meant to be used in tandem with other PCORTF-supported projects, such as Data Access Framework (DAF), Structured Data Capture, and the Patient-Reported Outcomes through Health IT Project. For example, the DAF-Research FHIR Implementation Guide provides a framework for the generalized workflows that can be used to extract data from electronic health record systems, populate the CDMs, and then allow the researcher to query these CDMs to receive those results. The CDMH FHIR Implementation Guide provides outlines for mapping four specific CDMs to FHIR that can be applied across multiple use cases. When combined, this approach provides opportunities to reuse the data, methods, and other resources of each network.

This infrastructure is sustainable, flexible, and modifiable based on the research questions. It has the capacity to support evidence generation on patient-centered outcomes that can inform regulatory and clinical decision making within federal programs because it could support not only researchers from the four participating networks, but also other researchers affiliated with NIH, FDA, the Centers for Medicare & Medicaid Services, and academic institutions, for example.

---

[6] NCI. Biomedical Research Integrated Domain Group. https://bridgmodel.nci.nih.gov.
[7] Balloting is the formal process that HL7 uses to vet specifications prior to publication. Rules governing balloting are defined in both HL7's Governance and Operations Manual and in the HL7 Essential Requirements. These, in turn, are governed by the expectations of the American National Standards Institute (ANSI), which accredits HL7 as a standards development organization and enforces rules around the openness and fairness of approval processes. The objective of balloting is to actively seek feedback on a proposed standard and to ensure that the community that will be governed by that standard is in agreement with the expectations set by the standard.
https://confluence.hl7.org/display/HL7/HL7+Balloting.

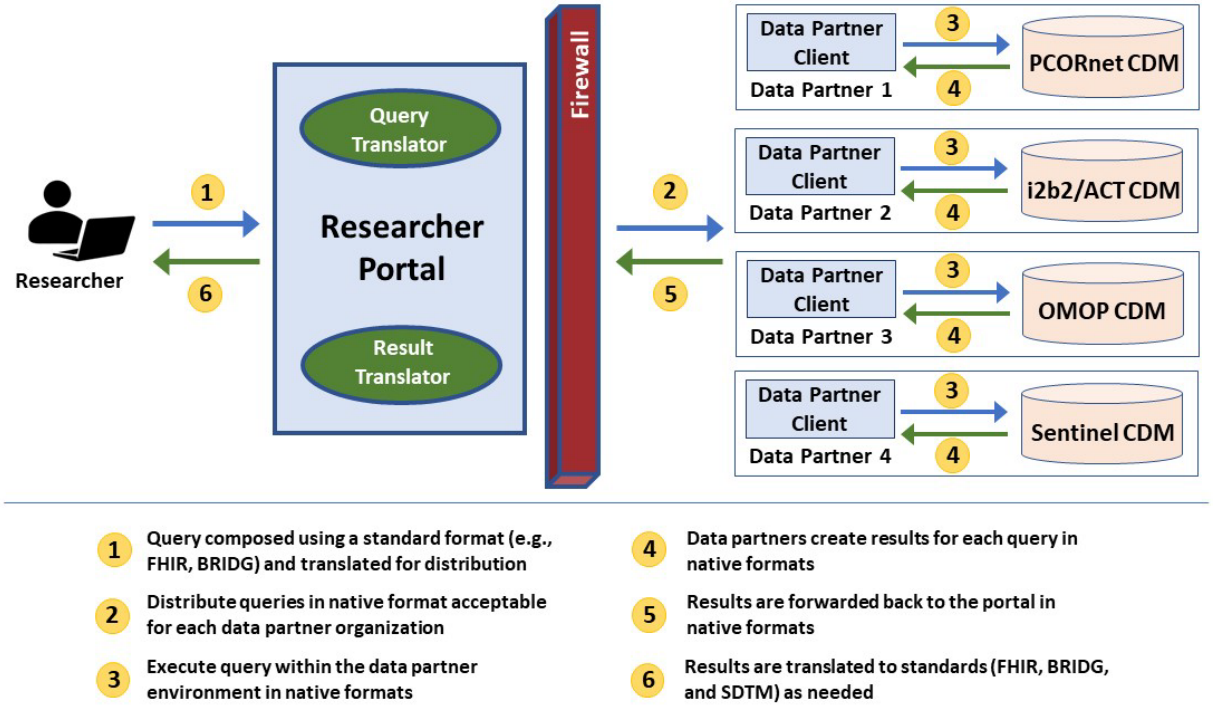

Figure 1. CDMH Model.

The figure above (Figure 1) illustrates the workflow that the interagency partners agreed on for the CDMH project and the steps needed to create a query, distribute the query to the various organizations, collect the results, and view the results. A Data Partner is any network or organization that is participating in the overall CDMH project and supports a data model that can be queried by authorized researchers. The Data Partner holds the research data using one of the CDM formats and has appropriate security controls to verify/validate queries before releasing data. Data Partner Clients represent the capability that enables the Data Partner network or organization to control the queries and submitted results.

Oncology Use Case

Project partners convened in May 2018 to discuss progress on the data mappings and the CDMH architecture, as well as the proposed oncology use case. As a validation test, the project team used the new data architecture to study factors associated with the safety and effectiveness of newly approved cancer drugs that boost patients' immune responses to cancer. Because these drugs, known as immune checkpoint inhibitors, represent an entirely new class of therapy, information about the experience of using these drugs to treat patients with cancer is largely limited to the clinical trial setting. Observational data are lacking, meaning that more information is needed about how safe and effective these drugs might be in routine clinical care, especially for patients who have autoimmune disorders.

The CDMH architecture was used to identify patients who had been given specific cancer immunotherapies and who also had autoimmune disorders. The research team attempted to identify the duration of the immunotherapy for patients with autoimmune disorders compared to the duration for the standard

population. Results of the use case were presented at an August 2019 workshop. For more information on this work, read the interagency final report.

CONCLUSION

This CDMH project was an outstanding example of interagency collaboration. The work completed in this project was a critical stepping stone for future work that could address, for example, a demonstration of the sustainability plan for the framework or directly using FHIR as the intermediary model.