ORIGINAL RESEARCH

The Story Behind the HL7 FHIR® to CDISC Mapping Implementation Guide

Rebecca Baker*, Mike Hamidi†, Lloyd McKenzie‡, Christine K. Denney§ and Theresia Edgar‡

Background: Health data has become an abundant and rich source of information that can be leveraged for clinical trials. Health Level Seven (HL7) provides standards for health care data, while the Clinical Data Interchange Standards Consortium (CDISC) provides standards for research data. Bridging these worlds and leveraging health data has the potential to enhance knowledge and data quality, while reducing delays and costs in clinical trials.

Objectives: The aim was to provide an overview of the HL7 Fast Healthcare Interoperability Resources (FHIR®) to CDISC Joint Mapping Implementation Guide (IG) development. The ultimate aim is to increase awareness of this rich resource.

Methods: A gap analysis was performed to assess the possibility and feasibility of mapping from FHIR to the CDISC’s Study Data Tabulation Model (SDTM) and Clinical Data Acquisition Standards Harmonization (CDASH) standards. The mapping of CDISC domains then commenced using FHIRPath, a path-based navigation and extraction language. The domains mapped were Medical History (MH), Concomitant Medications (CM), Procedures (PR), Vital Signs (VS), Laboratory (LB), Adverse Event (AE), and Demographics (DM). The proposed mappings were evaluated through a formal review process.

Results: The review process consisted of a CDISC Internal Review and a simultaneous CDISC Public Review and HL7 Ballot Review. A total of 363 comments were obtained from the following reviews: 1) CDISC Internal Review (268), and 2) simultaneous Public Review (95) – CDISC (37), HL7 (58). The predominant themes gathered in the review included mapping updates, terminology, and requests for greater detail.

Conclusion: The HL7 FHIR® to CDISC Joint Mapping Implementation Guide (IG) provides a rich resource for bridging health care and research standards that can be used with the CDISC LB to LOINC mapping guide for leveraging real-world data for clinical trials. The guide assists stakeholders in utilizing health data through the HL7 FHIR® resources to CDISC, CDASH, and SDTM variables. As the community uses this resource, it will continue to mature and potentially expand into new domains.

Keywords: Standards; FHIR; CDISC; Real-World Data; RWD; Clinical Research

Background

Health data has become an abundant and rich source of information that can be leveraged for research, for pragmatic clinical trials, for post-market research, and for supporting information for clinical trials. The Clinical Data Interchange Standards Consortium (CDISC) Real-World Data (RWD) Connect Initiative conducted a qualitative Delphi survey “to understand the barriers of implementing CDISC standards for RWD and to identify tools needed to more easily implement them.” The Fast Healthcare Interoperability Resources (FHIR) to CDISC mapping guide was a tool developed in response to this survey.¹

¹ Standards, CDISC, US

Health Level Seven (HL7)² provides standards for health care data, while CDISC³ provides standards for research data. Bridging these worlds creates new opportunities to expand and accelerate research by blending research and real-life patient conditions to supplement otherwise untapped data sources.⁴

² Center for Data Standards & Quality, Pfizer, US

³ Healthcare Standards, Accenture, US

⁴ Eli Lilly and Company, US

Corresponding author: Rebecca Baker (rbaker@cdisc.org)
What is HL7 FHIR®?
The HL7 FHIR® standard is a standard that aids in sharing health information electronically.\(^5\) It defines a health care exchange Application Programming Interface (API) that is implementer-focused and uses web-technology similar to that used by Google, Facebook, and Twitter. It also provides a human-readable expression, is technology agnostic and is open-source. FHIR has two primary components: Resources and APIs. The resources are a collection of information models that define the data elements, constraints, and relationships most relevant to health care.\(^6,7\) The APIs define mechanisms to access and manipulate the data exposed by those resources.

What is CDISC CDASH?
The Clinical Data Acquisition Standards Harmonization (CDASH) Model "describes the foundational structure for the organization, naming, and description of variables and associated attributes to support data collection in clinical trials."\(^8\) The CDASH Model provides naming conventions for the CDASH Implementation Guide (CDASHIG)\(^9\) variables along with additional metadata to help facilitate mapping collected data to their respective Study Data Tabulation Model (SDTM) Implementation Guide (SDTMIG) variables." This means the example case report forms (CRFs) have questions with associated variables, metadata, and mapping to SDTM. While CDASH is not required by regulators, it is helpful for metadata traceability and data cleaning.

What is CDISC SDTM?
The SDTM "provides a standard for organizing and formatting data to streamline processes in collection, in management, in analysis, and in reporting. Implementing SDTM supports data aggregation and warehousing; fosters mining and reuse; facilitates sharing; helps perform due diligence and other important data review activities; and improves the regulatory review and approval process."\(^10,11\) This means the data is semantically and syntactically organized as expected to facilitate data aggregation, which expedites the review process, and provides a faster, more streamlined drug-to-market pathway.

SDTM is organized in classes that are designed to aggregate data and that can handle most types of information. SDTM is required for regulatory submissions for drug approval by the Food and Drug Administration (FDA) and Japan Pharmaceuticals and Medical Devices Agency (PMDA). In addition, it is recommended by the European Union (EU) European Medicines Agency (EMA) and China National Medical Products Administration (NMPA).

CDASH and SDTM Metadata are available in the CDISC Library. The CDISC Library is a source of machine-readable normative CDISC standards and terminology metadata. It has an API data standards browser, and it serves as a model-based standards metadata repository. With the addition of the FHIR to CDISC reference materials, the repository aids in using FHIR for real-world data.\(^12\)

Regulations and Data Harmonization
To help to promote interoperability and harmonization in real-world data (RWD) and clinical research data, health authorities support the adoption of standards by recommending their use.\(^13–18\) Health authorities write guidance for industry submissions that promote the use of standards by specifying which standard is recommended and which is required for submissions. Health care has similar advocates through entities such as the United States Office of the National Coordinator for Health Information Technology (ONC). The ONC launched the United States Core Data for Interoperability (USCDI) initiative that describes a minimal set of data elements to be implemented on an incremental basis over the next 10 years. The ultimate aim is to achieve an interoperable health IT infrastructure.\(^19–23\)

Global health authorities continue to evaluate bridging health care and research (Figure 1).\(^24\) The prospects of using data from Health IT as real-world data (RWD) for real-world evidence (RWE) generation and for randomized

---

**Figure 1:** Growing use for RWE by regulators across the globe.\(^24\)
clinical trials (RCT) bring new opportunities to expand and accelerate research by blending research and real-life patient conditions to supplement otherwise untapped data sources and provide great insight. RWD has limitations such as completeness, and inconsistencies. Blending RWD, RCT, and RWE is even more complex when using a diverse set of sources and outcomes, which increases the challenges to researchers.

Objectives
The aim of this work is to provide a resource to solve the problem of data redundancy. The HL7 FHIR to CDISC joint mapping implementation guide was developed to eliminate data duplication and to improve interoperability. This document reviews what the FHIR to CDISC IG resource is and what it is not.

Interoperability between research and health care solves multiple problems. One such problem is data entry redundancy. Today, data are often entered into a clinical research site’s EHR, and then reentered into the sponsor’s clinical trial system (eg, Electronic Data Capture (EDC), Metadata Repository (MDR), etc.). This duplication results in delays in data transfer and impacts data quality. Having the right person enter the right information at the right time enables data reuse, thus reducing cost and variability. The FHIR to CDISC IG provides guidance for EHRs built on FHIR Release 4 (R4) to extract data for ingestion into a sponsor’s clinical trial system.

The vision of the data workflow is displayed in Figures 2 and 3. Figure 2 illustrates a FHIR-enabled EHR with clinical concepts (or resources) that could be accessed using an API for clinical research. Figure 3 shows that while data transfer appears to be seamless, implementers will need to consider caveats, such as variations in syntax and semantics, in each EHR system and in clinical research study requirements. The mapping guide is not an explicit methodology (ie, a one-to-one concept) to get from point A to point B.

**Figure 2:** FHIR to CDISC Goal: Leverage EHR data into a clinical trial system (eg, EDC, MDR, etc.) for efficient access and integration of EHR-based sources of data for clinical research.

**Figure 3:** Workflow of FHIR enabled EHRs with caveats.
The high-level map in Figure 3 shows how RWE can be used to support regulatory application in three ways. First, through data capture in electronic health systems. Second, RWE provides an avenue for using clinical data in retrospective studies. And third, RWE gives insight into CRF creation with data elements that support FHIR resources to allow for automatic population, reduced data entry, and transcription issues. The FHIR to CDISC IG will allow implementers to leverage each standard’s strengths and to gain the benefits of each standard for a specific use case.28,29

Methods
A gap analysis was performed to assess the possibility and feasibility of mapping from FHIR to CDISC, CDASH and SDTM. A CDISC workgroup began mapping domains in 2016, but work was halted when the team identified that numerous variables could not be populated. After noting that FHIR had matured enough that the documentation was feasible, the mapping of CDISC domains commenced.

The mappings were documented using FHIRPath, a path-based navigation and extraction language similar to Extensible Markup Language (XML) Path Language (X-Path).30 X-Path uses expressions to show how to navigate the elements and attributes (the nodes) in an XML document.31 FHIRPath fulfills a similar need in a syntax-independent manner, and with additional terminology and other features that makes it a particularly appealing tool for use in mapping.

The CDISC domains selected for mapping were mature and common across health care data, including: Medical History (MH), Concomitant Medications (CM), Procedures (PR), Vital Signs (VS), Laboratory Test Results (LB), Adverse Events (AE), and Demographics (DM). Figure 4 provides a view of the relevant FHIR resources and the CDISC domains. It demonstrates that there are many FHIR resources used to represent the variables in each SDTM domain. The LAB entry noted at the bottom of the CDISC domains represents the CDISC Laboratory Data Model Version 1.0.1.32 LAB is the CDISC standard model for acquiring and exchanging laboratory data. LAB was originally published as part of the Clinical Research Sponsor Laboratory Semantics in FHIR IG and was included to ensure consistency with the LB domain updates.33

The CDASHIG v2.1 mapping examples by domain was used as the basis for the mapping; this follows the CDASH model v1.1. The CDASHIG v2.1 references the SDTMIG v3.2 which references the SDTM v1.4.

Figure 5 provides a view of how patient FHIR resource data could flow from an EHR to CDASH, and then to SDTM. If the study did not require viewing the data in CDASH then the data could go directly to SDTM. During the mapping discussions, the subject matter experts emphasized the need to go to CDASH as part of the data collection and data cleaning process. The patient data within the EHR database would be queried for relevant information (such as medication, observations, encounters, or procedures) and be sent to a FHIR server via FHIR messages. The FHIR messages have resource and identity metadata, a human-readable summary, extensions that allow supplementary data capture based on a definition identified by URL, and the standardized computable data. The FHIR resource message would be transformed first to the CDASHIG variable representation (using JavaScript Object Notation (JSON) or XML), and then to SDTMIG variable representation. The flow could go from FHIR to CDASH to SDTM, or FHIR to SDTM. Others may want to leverage FHIR to Operational Data Model (ODM)-XML, which is CDISC’s exchange standard, to CDASH to SDTM.34 ODM is a way to represent CDASH or SDTM in a machine-readable data exchange format. The intermediate steps can be adjusted, depending on the requirements of the study.

The team used FHIR resource definitions and examples to determine which resources matched the concepts noted in the CDASH variable labels. The mapping was done with a resource-to-domain, and a context-based, approach. The team kept the mappings at a high level due to the variability in the data elements within EHR systems and to the differences in requirements for clinical research protocols. This initial mapping provides a starting point; as health care information standardization matures, the granularity of these mappings can increase. Where gaps still exist, guidance has been provided to help with
mapping the variables in each domain. Explanations are given in the “gaps” column of the guide.

During the mapping and review process, the perspective was “what can be found in electronic health data?” in the current state. For example, if a Questionnaire or Form was used, FHIR resources such as Adverse Event would be appropriate; however, most clinical trial Adverse Events would be noted as Conditions in the medical records. Mapping caveats and comments were added when the match to a CDISC variable was not straightforward or was determined to be “fuzzy”. The race and ethnicity in the Demographics domain extensions were aligned with the U.S. Core FHIR Implementation Guide; however, other extensions and specific SNOMED CT® and LOINC® codes noted in U.S. Core were not referenced. The non-U.S. audience may have different local and regional requirements, so the guide’s focus was on the base FHIR resources. The end users can supplement the base mappings with country-specific details when leveraging FHIR to CDISC mappings.

Results

The review process was split between CDISC and HL7 and consisted of a 14-day Internal Review and a 60-day Public Review on the CDISC side and a 30-day Ballot (HL7 Community Review) on the HL7 side. Figure 6 shows the timelines and the complexities of carrying out a review through two standards development organizations simultaneously. Both organizations shared and reviewed each review comment and its resolution.
There were 363 total comments, with the bulk occurring during the CDISC Internal Review (268). The simultaneous Public Review and Ballot Review yielded 95 total comments, 37 at CDISC and 58 at HL7. The predominant theme was clarifying the mapping and adding descriptive text to the comment sections. There were numerous comments about the Medical Dictionary for Regulatory Authorities (MedDRA) terminology and how to deal with it when mapping. Additionally, there were questions about aligning with other implementation guides. Table 1 provides a breakdown of the issues, the predominant themes, and the issues by domain.

The final deliverable was published simultaneously on both the HL7 FHIR website and the CDISC website, in multiple formats, to support the needs of the user community.

Discussion
The domains provided in the mapping are meant to be a starting point. They can be adjusted as needed based on requirements, input from subject matter experts, or as a result of modeling or governance activities. On the CDISC website, there is a Real-World Data section with several resources and examples of using CDISC standards with FHIR. A notable effort in RWD is the Laboratory Test Results (LB) domain mapping to LOINC file that shows LOINC code mappings to CDISC variables and terminology. A number of team discussions involved mapping concepts from SNOMED CT to MedDRA or LOINC to CDISC variables and terms. Both publications of the IG include examples and varying formats. The CDISC side has an Excel sheet with tabs that contain the content, while the HL7 site has links to traverse the content of the IG. In addition, the sites include links to download XLS and XML formats.

The team also discussed including mappings for two clinical research FHIR resources: ResearchSubject and ResearchStudy. These FHIR resources are still in development at HL7; however, the Patient resource has connections to both ResearchSubject and ResearchStudy. The FHIR to CDISC Joint Mapping IG was used in the HL7 January 2022 Connectathon for the Vulcan RWD and AE tracks. The Connectathon work identified areas of improvement for the AE domain.

Conclusion
The HL7 FHIR to CDISC Joint Mapping IG provides a high-level guide from one standard to another to harmonize mapping approaches across the community. The IG provides a beneficial resource that can be used with the
CDISC LB to LOINC mapping guide for leveraging real-world data for clinical trials. The guide assists stakeholders in utilizing health data through the HL7 FHIR resources to CDISC, CDASH, and SDTM variables. We encourage mappers to seek experts in terminology, FHIR resources, and CDISC standards to ensure the concepts mapped “from” and “to” are accurate for their use cases. As the document is used, we encourage the community to provide comments to improve the IG in future versions. Feedback from end users is important to effectively mature this guidance.

Acknowledgements

Thank you to Mitra Rocca and Scott Gordon for their help.

Competing Interests

The authors have no competing interests to declare.

References


9. CDISC CDASH Team. Clinical Data Acquisition Standards Harmonization Implementation Guide (CDASH-IG) for Human Clinical Trials v2.1. Published online November 1, 2019. https://www.cdisc.org/standards/foundational/cdash/cdash-ig-v2-1

10. CDISC Submission Data Standards Team. Study Data Tabulation Model (SDTM): Human Clinical Trials v1.4. Published online November 26, 2013. https://www.cdisc.org/standards/foundational/SDTM/v1-4


