Paper DS07

Bridging the Divide: Harmonizing HL7 FHIR, OMOP, and CDISC Standards for Regulatory-Grade Real-World Evidence

Yasa Sreekanth Reddy, IQVIA, Hyderabad, India

Abstract

As regulatory agencies increasingly incorporate real-world data (RWD) into decision-making, the need to harmonize disparate data standards has become increasingly critical. This paper presents a pragmatic framework for integrating HL7 FHIR, OMOP Common Data Model, and CDISC standards to support the generation of regulatory-grade real-world evidence (RWE). The approach leverages OMOP's analytical flexibility, CDISC's structured traceability, and HL7 FHIR's interoperability to enable efficient transformation of real-world data from source systems into submission-ready formats. Practical examples are used to illustrate common challenges and solutions in cross-standard mapping, including metadata reconciliation, traceability preservation, and regulatory submission readiness. Governance considerations, tooling strategies, and crosswalk methodologies are also discussed to support scalable and standards-aligned RWE workflows. This framework aims to inform industry efforts toward a unified data standards ecosystem that supports both scientific rigor and regulatory compliance.

1. Introduction

Real-world evidence (RWE) has become an increasingly important component of regulatory decision-making, complementing traditional clinical trial data in areas such as safety evaluation, effectiveness assessment, and post-marketing surveillance. Regulatory agencies, including the U.S. Food and Drug Administration (FDA), have issued guidance encouraging the use of real-world data (RWD) to support regulatory submissions, while emphasizing the need for data quality, traceability, and transparency. Despite this growing acceptance, the heterogeneous nature of RWD and the diversity of data standards used across healthcare and research systems continue to present significant challenges.

In practice, RWD is generated and managed using multiple standards that were designed for distinct purposes. Observational research and large-scale analytics commonly rely on the OMOP Common Data Model, while regulatory submissions require datasets formatted according to CDISC standards such as SDTM and ADaM. At the point of data capture and exchange, healthcare systems increasingly rely on HL7 FHIR to enable interoperable access to clinical information. These standards operate effectively within their intended domains, but limited alignment across them has resulted in fragmented workflows, redundant transformations, and increased risk to data traceability when RWD is repurposed for regulatory use.

As regulatory expectations evolve, particularly with emerging interest in modern interoperability standards, there is a growing need for a cohesive framework that enables these standards to function together rather than in isolation. This paper examines the complementary roles of HL7 FHIR, OMOP, and CDISC within the RWE lifecycle and proposes practical strategies for harmonizing them to support regulatory-grade evidence generation.

2. Regulatory Context for Real-World Evidence

Regulatory agencies worldwide have increasingly recognized the potential value of RWE in supporting regulatory and healthcare decision-making. In the United States, the FDA's Real-World Evidence Program has outlined circumstances under which RWD may be used to support regulatory submissions, particularly for post-approval studies, label expansions, and safety evaluations. Similar initiatives have been observed across other regulatory authorities, reflecting a global trend toward broader acceptance of non-traditional data sources.

Despite this momentum, regulators continue to emphasize that RWE must meet rigorous standards for data quality, provenance, and traceability. Unlike exploratory or hypothesis-generating research, regulatory submissions require transparent documentation of how data were collected, transformed, analysed, and interpreted. This requirement presents a challenge for RWD, which is often sourced from electronic health records, claims databases, registries, and other systems not originally designed for regulatory use.

Historically, CDISC standards have been mandated for regulatory submissions involving clinical trial data, ensuring consistency and traceability across studies. For RWE submissions, regulators have generally expected sponsors to apply the same CDISC standards, even though these models were not originally designed for observational data. As a result, organizations frequently perform late-stage transformations of RWD into CDISC-compliant formats, increasing complexity and the potential for information loss.

More recently, regulators have signalled interest in leveraging modern interoperability standards to improve the efficiency and transparency of RWE submissions. The FDA's exploration of HL7 FHIR for study data derived from real-world sources reflects an effort to modernize regulatory workflows and enable more direct access to structured clinical data. This evolving regulatory landscape underscores the need for integrated approaches that can bridge the gap between healthcare data systems, observational research platforms, and regulatory submission requirements.

3. Overview of Relevant Data Standards

The RWE ecosystem is supported by multiple data standards, each developed to address specific needs within healthcare, research, and regulatory environments. Understanding the strengths and limitations of these standards is essential for designing integrated workflows that support regulatory-grade evidence generation.

Standard	Primary Role	Key Strength	Limitation
HL7 FHIR	Interoperability	Real-time, granular clinical data	Not analysis/submission ready
OMOP CDM	Observational analytics	Scalable, reusable analytics	Limited regulatory traceability
CDISC	Regulatory submission	Auditability and reproducibility	High transformation effort

Table 1. Comparative Overview

3.1 CDISC for Regulatory Submissions

The Clinical Data Interchange Standards Consortium (CDISC®) develops and maintains standards used for regulatory submissions of clinical trial data. Models such as the Study Data Tabulation Model (SDTM) and the Analysis Data Model (ADaM) provide structured frameworks that enable regulators to efficiently review and reproduce study results. CDISC standards emphasize controlled terminology, consistent dataset structures, and explicit traceability from collected data to analysis outputs.

While CDISC standards are primarily associated with randomized controlled trials, they are also applied to RWE submissions to ensure regulatory compliance. In this context, CDISC is often introduced late in the data lifecycle, requiring transformation of observational data into SDTM and ADaM formats. Although this approach enables submission readiness, it can be resource-intensive and may obscure the original data provenance if not carefully managed.

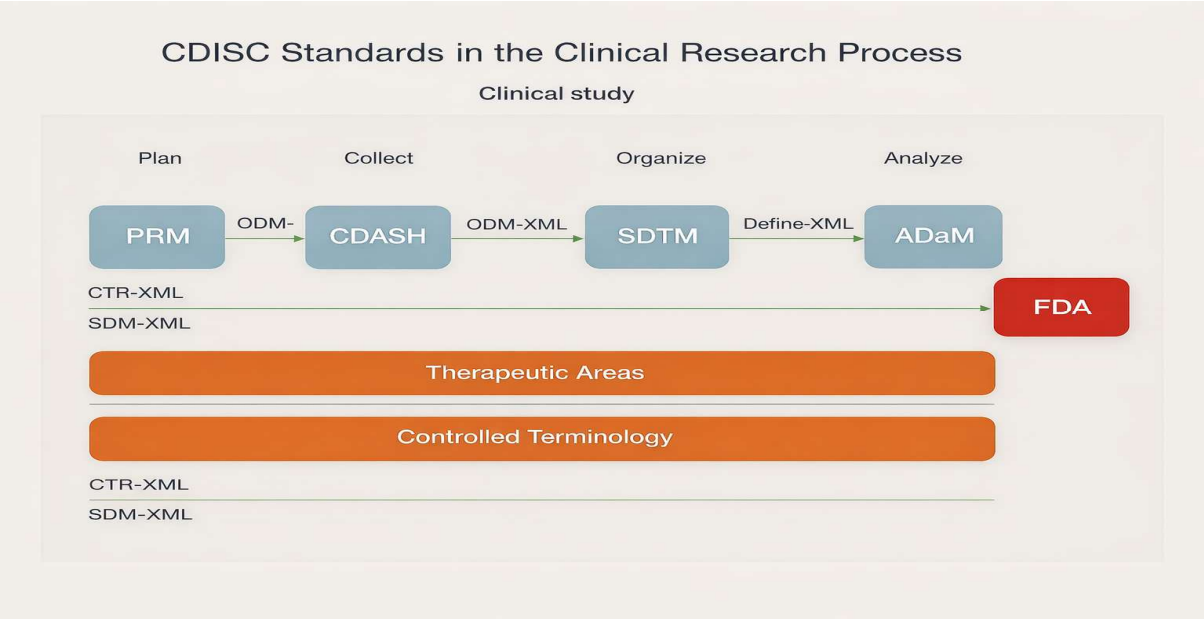

Figure 1. CDISC Standards in the clinical research process.

3.2 OMOP Common Data Model for Observational Analytics

The OMOP Common Data Model (CDM) was developed by the Observational Health Data Sciences and Informatics (OHDSI) community to support large-scale observational research. OMOP provides a flexible schema that enables harmonization of diverse healthcare data sources, including electronic health records and claims data, into a common structure suitable for analytics. Its design supports longitudinal analysis and population-level studies across heterogeneous datasets.

OMOP excels in enabling rapid, scalable analysis and is supported by a rich ecosystem of open-source tools. However, it was not designed to meet regulatory submission requirements and does not provide native support for regulatory traceability or analysis-ready datasets comparable to ADaM. As a result, OMOP-based data typically require further transformation and documentation before they can be used in regulatory contexts.

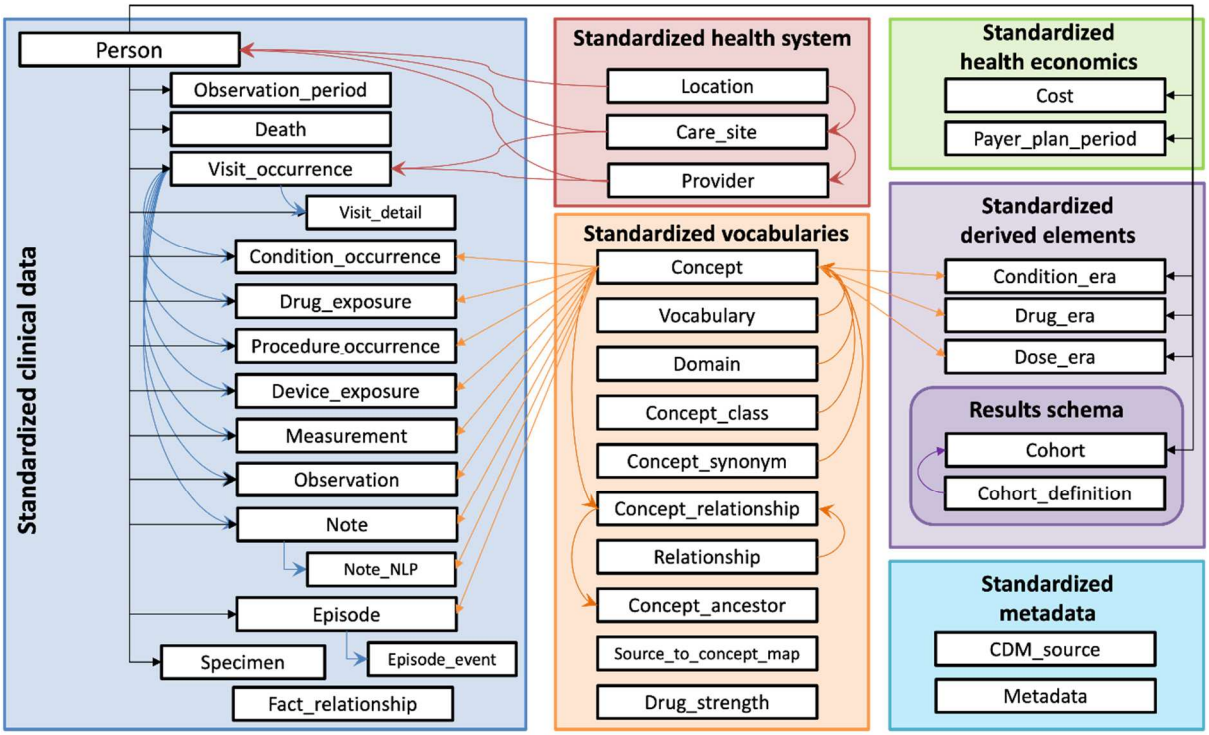

Figure 2. OMOP Common data model

3.3 HL7 FHIR for Interoperability

HL7 FHIR® is a modern interoperability standard designed to enable efficient exchange of healthcare data between systems. Unlike CDISC or OMOP, FHIR is not an analytical or submission data model; instead, it provides a standardized, API-driven mechanism for accessing granular clinical data in near real time. FHIR organizes data into modular resources, such as Patient, Observation, and Medication, which can be combined and extended to represent complex clinical information.

FHIR's growing adoption across healthcare systems makes it an increasingly important component of RWE data pipelines. Its use in regulatory contexts is currently under exploration, with the potential to streamline data ingestion from source systems and improve transparency. When integrated effectively, FHIR can serve as a bridge between clinical data sources and downstream analytical and regulatory standards.

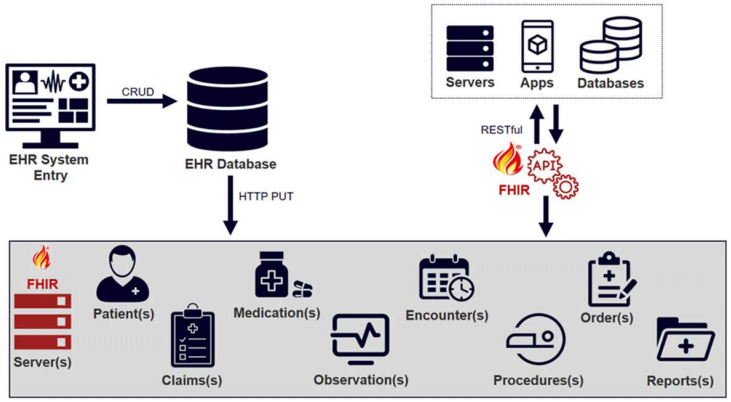

Figure 3. HL7 FHIR Overview

4. Comparative Analysis of OMOP and CDISC

OMOP Common Data Model and CDISC standards represent two fundamentally different design philosophies within the real-world evidence ecosystem. OMOP is optimized for observational analytics across large and heterogeneous data sources, while CDISC is designed to support regulatory review through standardized structure, traceability, and reproducibility. Understanding these differences is essential for designing workflows that effectively bridge observational research and regulatory submission requirements.

OMOP emphasizes flexibility and scalability. Its schema accommodates a wide range of healthcare data types and supports longitudinal patient-level analyses across disparate sources. This flexibility enables rapid exploratory analyses and efficient reuse of data for multiple research questions. However, OMOP does not impose strict requirements for dataset structure or variable naming beyond its core conventions, which can lead to variability across implementations. Additionally, OMOP does not define a standardized analysis dataset model comparable to CDISC ADaM, placing the burden of analysis consistency on individual study teams.

In contrast, CDISC standards prioritize consistency and traceability. SDTM provides a standardized framework for organizing collected data, while ADaM defines analysis-ready datasets with explicit derivation rules and metadata. These features enable regulators to trace results back to their source and independently reproduce analyses. The rigidity of CDISC standards, while sometimes perceived as a limitation, is intentional and supports regulatory expectations for transparency and auditability. However, applying CDISC to observational data often requires interpretation and customization, particularly when real-world data do not align cleanly with traditional clinical trial concepts.

Despite these differences, OMOP and CDISC share common goals. Both seek to standardize patient-level data to enable meaningful analysis and review, and both support longitudinal representations of patient experiences over time. In practice, they are complementary rather than competing standards. OMOP enables efficient evidence generation during the research phase, while CDISC provides the structure necessary to package that evidence for regulatory evaluation. Bridging the two requires careful mapping and governance to preserve data provenance and analytical intent.

5. Emerging Role of HL7 FHIR in Regulatory Submissions

HL7 FHIR has emerged as a key enabler of interoperability within healthcare systems and is increasingly being evaluated for its role in regulatory workflows. Unlike OMOP and CDISC, which define analytical and submission data structures, FHIR focuses on standardized data exchange through modular resources and application programming interfaces. This distinction positions FHIR as a potential upstream component of real-world data pipelines rather than a replacement for existing submission standards.

Regulatory interest in FHIR reflects a broader effort to modernize data submission and review processes. By enabling structured access to clinical data directly from source systems, FHIR has the potential to reduce manual data extraction and transformation steps. This approach could improve data completeness, timeliness, and transparency, particularly for real-world studies that rely on electronic health records and other operational data sources.

However, FHIR is not inherently designed for regulatory analysis or submission. Its resources are highly granular and clinically oriented, often requiring aggregation, normalization, and

contextualization before they can support analytical or regulatory use cases. As a result, FHIR is best viewed as a data ingestion and exchange layer that feeds downstream models such as OMOP and CDISC. When integrated effectively, FHIR can enhance traceability by providing clear links to source data while enabling standardized transformations into analytical and submission-ready formats.

The exploration of FHIR in regulatory contexts highlights the importance of interoperability across the full data lifecycle. Rather than introducing an additional standard in isolation, successful adoption will depend on aligning FHIR-based data ingestion with established analytical and regulatory frameworks. This alignment requires clear governance, standardized mappings, and robust metadata management to ensure that regulatory expectations for data quality and traceability are met.

6. Practical Integration Framework

To support regulatory-grade real-world evidence generation, an integrated framework is required that leverages the strengths of HL7 FHIR, OMOP, and CDISC while mitigating their individual limitations. Such a framework should address data ingestion, analytics, and submission as interconnected components of a single end-to-end workflow.

At the data ingestion stage, FHIR can be used to extract structured clinical data from source systems in a standardized and interoperable manner. By capturing data directly from electronic health records and related systems, organizations can reduce variability introduced by ad hoc extraction processes. FHIR resources should be mapped to a controlled internal data representation, with source identifiers and metadata retained to support downstream traceability.

For analytical processing, OMOP provides an effective platform for harmonizing and analysing real-world data at scale. Transforming FHIR-derived data into OMOP enables longitudinal analyses and supports reuse of established analytical tools and methodologies. During this phase, governance controls should be implemented to document transformations, preserve provenance, and ensure consistency across studies.

At the submission preparation stage, CDISC standards should be applied to produce regulatory-compliant datasets. This step involves mapping OMOP-based analytical outputs to SDTM and ADaM structures, with explicit documentation of derivation rules and traceability links. Wherever possible, mappings should be standardized and automated to reduce manual effort and minimize the risk of inconsistencies. Define.xml metadata plays a critical role in documenting these transformations and supporting regulatory review.

Across all stages, governance and tooling are essential. Version control, standardized mapping specifications, and validation tools should be integrated into the workflow to ensure reproducibility and compliance. Clear ownership of data standards, mappings, and documentation helps maintain consistency and supports efficient response to regulatory inquiries. By treating interoperability, analytics, and submission preparation as components of a unified framework, organizations can generate RWE that is both analytically robust and regulatory-ready.

7. Implementation Challenges and Considerations

While the integration of HL7 FHIR, OMOP, and CDISC offers significant benefits, organizations must address several practical challenges to successfully operationalize a multi-standard real-world evidence framework. These challenges span technical, organizational, and governance dimensions and require coordinated mitigation strategies.

One of the primary challenges is semantic alignment across standards. Although FHIR, OMOP, and CDISC may represent similar clinical concepts, differences in terminology, granularity, and contextual meaning can complicate mappings. For example, clinical observations captured at a granular level in FHIR may not map directly to OMOP concepts without aggregation or normalization, and further interpretation may be required when transforming analytical outputs into CDISC-compliant datasets. Clear mapping specifications and controlled terminology governance are essential to ensure consistency and interpretability.

Traceability preservation represents another critical consideration. Regulatory-grade submissions require end-to-end traceability from source data through analysis results. When data pass through multiple transformations - FHIR to OMOP, and OMOP to CDISC - there is a risk that provenance information may be lost or obscured. To mitigate this risk, organizations should implement robust metadata management practices, including retention of source identifiers, transformation logs, and explicit derivation documentation. These practices support audit readiness and facilitate efficient responses to regulatory queries.

Operational complexity and resource requirements must also be considered. Multi-standard workflows introduce additional layers of tooling, validation, and governance, which can increase implementation effort if not well planned. Automation plays a key role in managing this complexity. Standardized transformation pipelines, reusable mapping libraries, and validation tools can reduce manual effort and improve consistency across studies. Early investment in these capabilities can yield long-term efficiencies as RWE programs scale.

Finally, organizational alignment is critical. Successful implementation requires collaboration across clinical, data engineering, biostatistics, and regulatory teams. Clear ownership of standards, mappings, and documentation helps avoid fragmentation and ensures that analytical and regulatory objectives remain aligned. Training and knowledge sharing are also important to ensure that teams understand not only how standards are applied, but why particular design decisions are made within the integrated framework.

8. Future Outlook

The regulatory and data standards landscape for real-world evidence continues to evolve rapidly. Increased regulatory interest in interoperability standards and machine-readable submissions suggests that future workflows will place greater emphasis on upstream data standardization and automation. HL7 FHIR is likely to play an expanding role in enabling structured access to clinical data, while analytical and submission standards will continue to adapt to accommodate new data types and sources.

In the near term, organizations can expect continued reliance on CDISC standards for regulatory submissions, including those based on real-world data. At the same time, observational data models such as OMOP will remain central to large-scale analytics and evidence generation. The challenge - and opportunity - lies in improving alignment across these standards to reduce

redundancy and enhance transparency. Efforts to formalize mappings, develop common implementation guides, and standardize metadata practices will be increasingly important.

Looking ahead, greater convergence across standards may emerge through collaborative initiatives involving regulators, standards development organizations, and industry stakeholders. Such convergence does not imply replacement of existing standards, but rather clearer definition of their roles within an end-to-end evidence generation lifecycle. Organizations that proactively invest in interoperable, standards-aligned architectures will be better positioned to adapt to evolving regulatory expectations and to leverage real-world data effectively.

9. Conclusion

The growing role of real-world evidence in regulatory decision-making has amplified the need for cohesive, standards-aligned data workflows. HL7 FHIR, OMOP, and CDISC each address distinct but complementary aspects of the real-world data lifecycle, encompassing data capture, analytics, and regulatory submission. When applied in isolation, these standards can create fragmented processes and increase the risk of traceability gaps. When integrated thoughtfully, they enable efficient, transparent, and regulatory-ready evidence generation.

This paper has outlined a pragmatic framework for harmonizing these standards, emphasizing the importance of interoperability, governance, and traceability across the full data lifecycle. By leveraging FHIR for data ingestion, OMOP for scalable analytics, and CDISC for submission preparation, organizations can align analytical flexibility with regulatory rigor. Adoption of such integrated approaches will be essential as regulatory agencies continue to expand their use of real-world evidence and as data standards continue to evolve.

References

1. U.S. Food and Drug Administration. *Framework for FDA's Real-World Evidence Program.*
2. Clinical Data Interchange Standards Consortium (CDISC). *Study Data Tabulation Model Implementation Guide.*
3. Clinical Data Interchange Standards Consortium (CDISC). *Analysis Data Model Implementation Guide.*
4. Observational Health Data Sciences and Informatics (OHDSI). *The OMOP Common Data Model.*
5. Health Level Seven International. *FHIR Release Documentation.*

Acknowledgments

We acknowledge the contributions of industry colleagues and standards communities whose ongoing work in real-world evidence, interoperability, and regulatory science continues to shape best practices in this evolving field.

Contact Information

Your comments and questions are valued and encouraged. Contact the author(s) at:

Yasa Sreekanth Reddy
Associate Director Statistical Programming.
Real World Solutions, IQVIA
Email: sreekanth.yasa@iqvia.com