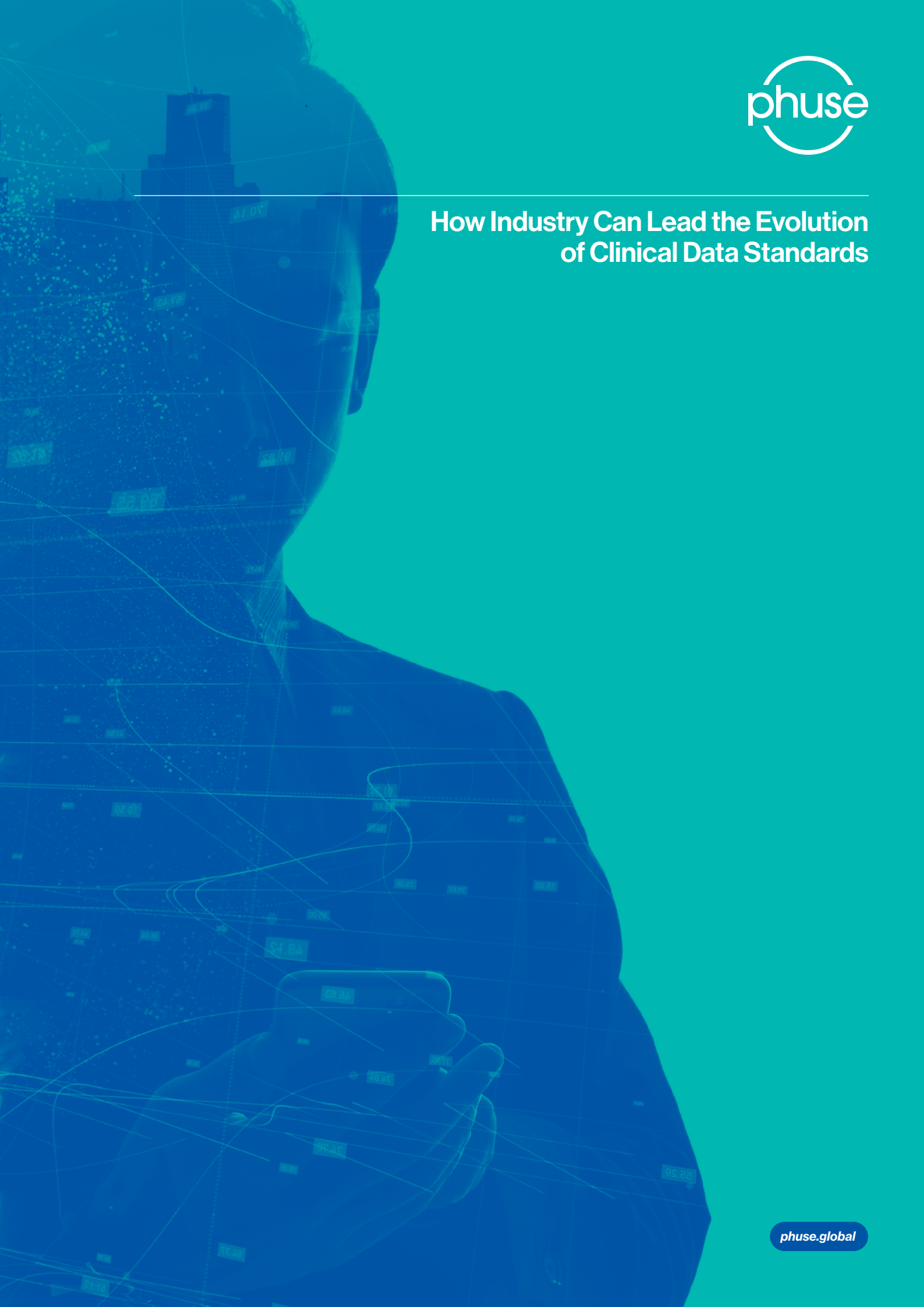

How Industry Can Lead the Evolution of Clinical Data Standards

phuse.global

Contents

Executive Summary 1

1. Purpose and Scope 1
1.1 Purpose 1
1.2 Scope 1

2. Context: The Evolving Clinical Data Environment 1

3. The Landscape of Clinical Data Standards 2
3.1 Healthcare Interoperability Standards 2
3.2 Regulatory Clinical Trial Standards 2
3.3 Real-World Evidence and Observational Data Models 2
3.4 Adjacent and Supporting Frameworks 2
3.5 Data Vocabularies and Terminologies 3

4. Fragmentation and Misalignment 3
4.1 Overlapping but Unaligned Scopes 3
4.2 Transformation Burden 3
4.3 Impact on Automation and Reuse 3

5. Existing Harmonisation Efforts and Their Limits 4

6. Why an Industry-Led Model Is Needed 4

7. Governance and Collaboration Model 4
7.1 Principles 4
7.2 Proposed Structure 5

8. RACI-Based Roles and Responsibilities 5

9. Operational and Regulatory Considerations 6

10. Hybrid Models and Practical Pathways 6

11. Call to Action 7

12. Conclusion 7

13. About the Authors 7

Revision History

Version	Date	Summary
1.0	1 September 2026	Initial version

Executive Summary

The clinical research industry is at a pivotal moment **as it embraces the age of complex data and artificial intelligence (AI).** Data volumes and complexities are increasing, study designs are becoming more diverse (with **decentralised and hybrid trials**), and expectations for **interoperability, automation**, and secondary data use continue to rise. This is particularly critical for enabling modern AI-driven tools, which rely on high-quality, standardised data as "context" to perform accurately without error. At the same time, the standards landscape has expanded into a constellation of frameworks – each mature within its own domain, yet insufficiently aligned across the full clinical research lifecycle. **Healthcare interoperability standards, regulatory submission standards,** and **real-world evidence (RWE) data models** have largely evolved in parallel. While each addresses a legitimate and necessary use case, their separation has resulted in **fragmentation, duplicated transformation efforts, and practical barriers to data reuse, automation, and AI-driven analysis.** Implementers (e.g. sponsors, contract research organisations, **technology providers**) are left to reconcile inconsistencies that standards bodies themselves are not structurally positioned to resolve alone.

This white paper argues that the **industry must play a more active, coordinated role in shaping how data standards evolve,** especially to address pressing needs for **semantic interoperability** and to support emerging technologies. Rather than waiting for alignment to emerge organically across standards organisations, industry stakeholders should articulate a **shared position on requirements, priorities, and practical constraints** and engage standards bodies and regulators through structured, evidence-based collaboration. This is particularly vital for enabling next-generation capabilities: **consistent, aligned data standards provide the context that modern AI and automation require to function effectively,** yet today's fragmented standards frequently impede such progress. The paper is not a critique of existing standards. It is a call for a shift in dialogue: from a predominantly top-down model to a collaborative, **industry-led approach** that reflects real-world implementation experience, supports emerging trial paradigms, and accelerates innovation while maintaining regulatory trust.

1. Purpose and Scope

1.1 Purpose

The purpose of this white paper is to:

- Describe the current clinical data standards ecosystem and its challenges.
- Clarify the roles and boundaries of major standards frameworks.
- Examine why misalignment persists despite ongoing harmonisation efforts.
- Propose an **industry-led collaboration model** for standards evolution.
- Outline governance considerations, including a **RACI-based view of roles and responsibilities.**

1.2 Scope

This paper focuses on:

- **Standards relevant to clinical trials, healthcare data exchange, and real-world evidence** – covering key frameworks (e.g. HL7® FHIR®, CDISC, OMOP) and their interactions.
- **Operational, regulatory, and scientific considerations** in the evolving clinical data environment.
- **Interactions between industry, standards bodies, and regulators** at a conceptual and governance level (how a collaborative approach could function).

Out of scope:

- **Market or commercial analyses** (e.g. business models or vendor comparisons).
- **Technology vendor product comparisons.**
- **Prescriptive regulatory guidance** (the paper discusses collaboration with regulators but does not issue regulatory policy).
- **Detailed technical implementation guidance** (the focus is on strategy and governance rather than step-by-step technical instructions).

2. Context: The Evolving Clinical Data Environment

Clinical research no longer operates within a single, linear data pipeline. Traditional randomised controlled trials are now often complemented by **decentralised elements, pragmatic designs, and greater use of observational and real-world data.** Data sources span electronic health records (EHRs), wearable devices, patient apps, laboratory systems, medical imaging platforms, and patient-reported outcomes. In such a landscape, a single study might draw on **diverse data streams.** In fact, recent industry analysis suggests that nearly **30% of new clinical trial protocols in 2025 were designed as decentralised or hybrid studies,** dramatically increasing the influx of data from wearables and ePRO (electronic patient-reported outcomes) devices that traditional standards struggle to handle natively.

This expansion has intensified the need for:

- **Semantic consistency across systems** – shared meaning for clinical data elements (using common terminologies and coding) so that a given data point (e.g. a lab result or adverse event term) is interpreted uniformly across different platforms and contexts.
- **Reusable data structures and standardised formats** – common data models (e.g. standardised forms for patient data, observations, trial endpoints) that can be applied across studies and IT systems, reducing bespoke rework and enabling plug-and-play integration of new data sources.
- **Automated data flow from source to submission** – the ability to move data from collection (e.g. at healthcare sites or via digital devices) through analysis and ultimately into regulatory

submissions with minimal manual intervention. *(Notably, achieving this relies on the consistency in structure and meaning described above.)*

- **Regulatory confidence in derived and reused data** – assurance that when data are transformed or reused (for secondary analysis, cross-trial studies, or AI-driven insights), their integrity, context, and provenance are preserved.

Standards are central to meeting these needs. However, when standards are **misaligned or lack sufficient depth,** they can become an obstacle rather than an enabler. For example, one system might use ICD-10 codes for medical conditions while another uses SNOMED CT; or one dataset might structure patient data in FHIR while another uses a custom format – leaving teams to manually reconcile these differences and potentially losing critical detail in translation. **If a simple data point like a patient's blood pressure originates in an EHR via FHIR (with specific units and context), is later re-entered into an electronic data capture (EDC) system under a CDASH-form structure, and then transformed into an SDTM dataset for submission, at each handover some information (units, timestamps, metadata) may be lost or require re-mapping. Such "repeated reformatting" not only increases workload (e.g. repeated data validation by data managers) but also raises the risk of errors.** These inconsistencies undermine the very efficiencies that standards are meant to provide.

3. The Landscape of Clinical Data Standards

The current landscape of clinical data standards can be grouped into a few broad categories, each serving distinct purposes within the research and healthcare ecosystem:

3.1 Healthcare Interoperability Standards

HL7® FHIR® (Fast Healthcare Interoperability Resources) – A modern standard designed to support real-time, **patient-level data exchange** across healthcare systems. Its strengths lie in flexibility, extensibility, and widespread adoption in clinical care environments (especially for EHR data exchange). FHIR is optimised for **transactional, point-of-care data sharing** rather than for regulatory submissions, and it makes minimal assumptions about global semantic consistency beyond defined "resources."

3.2 Regulatory Clinical Trial Standards

CDISC Standards (e.g. CDASH, SDTM, ADaM) – A family of standards providing structured, **analysis-ready datasets for regulatory review,** deeply embedded in global regulatory processes. CDISC standards support traceability **from protocol to analysis,** ensuring that trial results are reproducible from the submitted datasets. Key components include:

- *CDASH (Clinical Data Acquisition Standards Harmonisation)* – Standard formats for data collection fields in case report forms.
- *SDTM (Study Data Tabulation Model)* – Standard structure for organising collected data into domains for submission.
- *ADaM (Analysis Data Model)* – Standard structures for datasets used in statistical analysis.
- *TAUGs (Therapeutic Area User Guides)* – Guides that extend the foundational standards to specific disease areas, including how to represent disease-specific concepts, endpoints, and objectives – critical for end-to-end disease-area standards.

3.3 Real-World Evidence and Observational Data Models

OMOP Common Data Model (CDM) – A standardised data model developed by the OHDSI community for **observational research and real-world evidence studies.** OMOP provides a flexible schema to accommodate data from sources like EHRs and claims databases, enabling large-scale analytics and outcomes research. However, it is not designed around regulatory submission requirements for interventional trials, making direct integration with CDISC standards challenging in many cases. OMOP also contrasts with FHIR (Section 3.1): where FHIR is an exchange standard for moving patient-level data between live systems, OMOP is an analytical model for standardising data at rest — so data received via FHIR must still be mapped into the OMOP CDM, an area of active work in the OHDSI community (e.g. "OMOP on FHIR").

3.4 Adjacent and Supporting Frameworks

Beyond these primary standards, several frameworks support or complement the ecosystem:

- **BRIDG (Biomedical Research Integrated Domain Group model)** – A shared conceptual model describing data and processes in clinical research. It serves as a reference ontology that underlies some standards (like CDISC) and fosters common understanding across stakeholders.
- **CDISC ODM (Operational Data Model)** – A vendor-neutral, XML-based standard for the exchange and archival of operational study data and metadata, widely used to move data between data collection and other clinical systems (and underpinning Define-XML submission metadata). Its latest version, ODMv2, supports API-based, machine-readable exchange, bridging data capture and downstream standards.
- **USDM & ICH M11** – The emerging **Unified Study Definitions Model (USDM)** — a CDISC standard developed through TransCelerate's Digital Data Flow (DDF) initiative — and the ICH's forthcoming **M11** guideline aim to standardise trial design data. These initiatives focus on creating a **"digital trial protocol"** structure for consistent definition of study elements, facilitating exchange between protocol authoring tools, databases, and downstream systems.
- **EMA SPOR** – The European Medicines Agency's **Substances, Products, Organisations, and Referentials** data services, standardising key reference data (like drug names, company identifiers) used in regulatory submissions and pharmacovigilance.
- **ISO IDMP (Identification of Medicinal Products)** – A suite of ISO standards defining globally consistent identifiers and definitions for medicinal products. It's crucial for regulatory tracking of products (e.g. across regulatory submissions, safety databases, and healthcare).
- **FHIR Extensions for Research** – Specialised standard profiles and extensions built on FHIR to address research-specific needs. For example, FHIR Genomics modules for integrating

genomic data, or custom FHIR profiles being developed to mirror clinical research forms, bridging precision medicine and trial data.

Each framework addresses a specific need. The challenge lies not in their individual design, but in how they **intersect and integrate.** Without careful coordination, having multiple "right" standards can result in uncertainty about which one to use when, and how to ensure continuity of data across them.

3.5 Data Vocabularies and Terminologies

No discussion of standards is complete without considering the **controlled vocabularies and terminologies** that underpin semantic interoperability. While the sections above focus on data models and exchange formats, equally critical are the **standard dictionaries of medical terms and codes** that ensure data from different sources carries the same meaning. Key vocabularies include:

- **SNOMED CT** (Systematised Nomenclature of Medicine – Clinical Terms) – a comprehensive, multilingual clinical healthcare terminology used to encode a vast range of clinical concepts (diagnoses, procedures, findings) in EHRs and health data exchange.
- **MedDRA** (Medical Dictionary for Regulatory Activities) – the global standard for coding adverse events, medical history, and medications in clinical trials and pharmacovigilance, ensuring consistent safety data in regulatory submissions.
- **WHODrug Global** – the international drug dictionary maintained by the Uppsala Monitoring Centre, used to code medications (e.g. concomitant medications and medical history drugs) in clinical trials and pharmacovigilance, complementing MedDRA in safety reporting.
- **CDISC Controlled Terminology (CDISC CT)** – standard code lists and terms maintained by CDISC (with the NCI EVS) for use in CDISC datasets (e.g. standard names for lab tests, units, drug dictionary terms, etc.), crucial for uniform interpretation of data in submissions.
- **ICD-9/ICD-10** (International Classification of Diseases) – WHO's diagnostic classifications (ICD-10 being current globally, ICD-9 historical/legacy) for diseases and health conditions, widely used in healthcare records and billing data.
- **LOINC** (Logical Observation Identifiers Names and Codes) – a standard for identifying laboratory tests and clinical measurements, enabling consistent interpretation of lab results and other clinical observations across different systems.
- **Other specialised terminologies** – depending on context, additional vocabularies might be relevant, such as **UniProt** (Universal Protein Resource) for protein identifiers in biomedical research, **RxNorm** for standardised naming of drugs, etc.

Managing and mapping between these vocabularies is a major part of any data standardisation strategy. For instance, a diagnosis recorded in a hospital EHR with an ICD-10 code may need mapping to a SNOMED CT concept for an analytics database; similarly, lab results might carry local codes that must be translated to LOINC for interoperability. *The OHDSI community's Athena platform provides a real-world example of tackling this challenge: Athena aggregates many healthcare and research vocabularies, providing mappings between them (via OMOP concept relationships) so that disparate codes can be linked to a common concept. This kind of tool is essential when integrating data across sources.* As clinical research embraces more data types and sources, achieving **true semantic interoperability** – where data has a **uniform meaning and usage across all contexts** – will be as important as aligning data structures or formats. Vocabularies and ontologies are the glue that binds different standards together into a coherent system.

4. Fragmentation and Misalignment

Despite the breadth of standards described in Section 3, **misalignments** often arise at their interfaces. Key challenges include:

4.1 Overlapping but Unaligned Scopes

Unclear or overlapping boundaries between standards lead to confusion about **where one standard's domain ends and another's begins.** This leads to duplicate or conflicting implementations of similar concepts, and even **inappropriate use or over-customisation** of standards to cover perceived gaps. For example, is a **FHIR** resource or a **CDISC CDASH** form the "right" way to capture patient data in a hybrid trial? Different organisations might choose differently, or attempt to use both, resulting in two parallel representations that must later be reconciled. Without clear guidelines, the **same underlying data can end up represented in multiple ways,** leading to inefficiencies and potential errors.

4.2 Transformation Burden

As a result of fragmentation, data are often **transformed or mapped multiple times** as they move from collection to submission. It is important to distinguish simple data transfers from true transformations: moving data between systems in the same format has costs, but converting data between *different* formats and standards is especially burdensome. Each such transformation introduces cost (development and validation of mapping processes), the risk of errors or **loss of granularity** (e.g. subtle differences in how two standards handle metadata or code lists can mean information gets dropped or altered), and increased timelines. At scale, the compounding effect of numerous transformations adds significant operational overhead to every trial.

4.3 Impact on Automation and Reuse

Misalignment between standards – and sometimes gaps in the depth or specificity of those standards – directly constrains our ability to **automate** and to **efficiently reuse data.** End-to-end automation in data processing requires that each system "understands" the data from the previous one; if manual intervention is needed to adjust or interpret data, full automation breaks down. Similarly, if data elements aren't captured with a view toward reuse (for example, using non-standard or proprietary definitions), their value beyond the immediate context is limited – inhibiting **secondary uses** like cross-study analyses or feeding into **AI/ML models.** These limitations are increasingly evident as organisations pursue digital data flow and **AI-enabled analytics,** only to find that inconsistent

or lossy data flows impede realising the full benefits of those technologies. For example, an AI model that could detect safety signals or predict outcomes needs consistent, high-quality data; if an organisation must continuously map between data models such as OMOP and CDISC, it starves those algorithms of real-time, analysis-ready data and complicates scaling AI solutions.

5. Existing Harmonisation Efforts and Their Limits

Multiple initiatives are underway to address pieces of the alignment problem. For example:

- **CDISC–HL7 FHIR Mapping Guides** – Collaborative projects creating detailed mappings between FHIR resources and CDISC standards (like mapping an EHR's FHIR data to SDTM), to streamline EHR-to-research data exchange.

- **CDISC360i** – A CDISC initiative to incorporate metadata and automation capabilities into CDISC standards, aiming for end-to-end study automation (from data capture to analysis) and "content-defined" data transformations.

- **TransCelerate Digital Data Flow (DDF)** – An industry consortium effort to digitise the clinical trial protocol and connect protocol definitions to downstream study setup and data collection, reducing manual re-entry of protocol data. DDF aligns with ICH M11 and CDISC's study definitions standards.

- **HL7 Vulcan Accelerator** – An HL7-led programme where stakeholders (including pharma companies, tech vendors, and regulators) collaborate on projects that bridge healthcare and research. Vulcan has delivered use cases and pilot implementations demonstrating how FHIR could be used to exchange data for clinical research purposes, effectively connecting EHR and clinical trial systems.

These efforts provide valuable foundations and show that many parties recognise the need for better harmonisation. *However, progress to date remains **incremental** and uneven.* One notable challenge is that harmonisation work which is **not central to a given standard's primary mission often lags.** For example, while core FHIR development might focus on healthcare needs, ensuring it aligns with research standards may be treated as a side project and progress slowly. Similarly, new data standards (e.g. for genomic data like **BioCompute** or novel device data) continue to emerge faster than the community can integrate them into a cohesive framework. This results in a continuously shifting landscape where alignment is always a step behind innovation. Many harmonisation initiatives also face limitations related to *governance and resources* – they rely on voluntary efforts, may lack sustained funding, and often struggle to incorporate broad industry input in a timely fashion.

6. Why an Industry-Led Model Is Needed

Industry stakeholders sit at the crossroads of using all these standards – balancing the needs of **clinical science, operational efficiency, and regulatory compliance.** They see day-to-day how misalignments in standards create bottlenecks. For example, a data manager knows exactly how much effort goes into mapping data from one system to another; a statistician understands how inconsistent data structure hampers analysis; an IT architect recognises how difficult it is to integrate systems that weren't designed to work together.

An **industry-led model** for standards evolution would not replace standards bodies but rather amplify and accelerate their work by bringing in real-world perspectives and resources. In particular, industry can:

- **Emphasise practical value over theoretical ideals** – focusing on what delivers tangible improvements in speed, cost, and data quality (a needed counterbalance to the inward-focused "standards development" mindset).

- **Collectively define and prioritise requirements** – pooling insights from multiple companies to identify the most critical pain points and desired enhancements, thus creating a clear, unified "ask" to guide SDOs (Standards Development Organisations).

- **Provide robust feedback and evidence** – leveraging operational data and pilot studies to show what works and what doesn't, thereby helping SDOs and regulators understand the real-world impact of standards decisions.

- **Collaborate on development and pilot testing** – dedicating industry resources (people, funding, technology) to work alongside SDOs in creating and testing new standards or mappings; this ensures solutions are validated in practice, not just on paper.

In essence, industry's role is to bring the **focus back to real-world application and value.** Organisations such as **PHUSE, SCDM,** and the **ASA (American Statistical Association) Healthcare Data Science Working Group** are well placed to convene this effort, given their track record in cross-industry collaboration. Importantly, technology providers (software and platform vendors) should be part of the conversation since they build the tools that implement standards.

7. Governance and Collaboration Model

7.1 Principles

For an industry-led collaboration to be effective, it should operate under agreed principles:

- **Transparency** – clearly communicating objectives, decisions, and progress to all stakeholders, and providing channels for broader industry input.

- **Inclusivity of Expertise** – involving a spectrum of roles (clinical, data management, biostatistics, medical coding, IT, etc.) and organisations (pharma, biotech, CROs, academic researchers, technology vendors) to ensure diverse expertise and perspectives.

- **Use Case Focus** – grounding discussions in specific real-world use cases and problems (e.g. how to auto-populate a safety database from an EHR, or how to represent a new biomarker consistently) to keep efforts practical and relevant.

- **Regulatory Engagement with Independence** – inviting regulators into the conversation in advisory or observer roles, so they can offer feedback and be kept apprised of developments, while respecting that regulators will make independent decisions on policy.

7.2 Proposed Structure

The white paper envisions cross-industry working groups operating under a joint governance structure. Key participants would be:

- **Industry Representatives** – including sponsor pharma/biotech companies and CROs (with cross-functional experts such as data standards specialists, trial managers, and **statistical scientists**), as well as **technology vendors** (EHR/EDC platform providers, analytics software vendors) whose systems are integral to data workflows.
- **Standards Bodies** – like CDISC, HL7, ISO, and OHDSI, participating as partners. They might contribute by advising on feasibility, aligning on technical approaches, and eventually integrating the outcomes into official standards.
- **Regulatory Authorities** – e.g. FDA, EMA, taking part in an observational/advisory capacity, potentially through existing initiatives (like FDA's participation in standards forums or the MHRA's engagement in innovation projects). Their presence ensures outcomes remain aligned with regulatory requirements and facilitates a smoother path to regulatory acceptance of new standards.

The working groups would form around thematic areas or specific projects (for example, a **common data model for a specific therapeutic area,** or a **FHIR-to-CDISC integration toolkit**). **Deliverables** may include consensus white papers (to influence policy or SDO strategies), technical **specification documents,** reference **implementations or pilot systems,** and business case analyses demonstrating the value of a proposed standard or approach.

Importantly, to avoid pitfalls that have limited the impact of past consortia, these working groups should be underpinned by **concrete operational mechanisms.** For example, participants might set up a federated sandbox environment where multiple companies can jointly develop and test new integrations (such as hybrid FHIR–CDISC APIs) in a pre-competitive space, ensuring early identification of issues. Adequate funding, clear timelines, and project management support would help maintain momentum and focus.

8. RACI-Based Roles and Responsibilities

A Responsible–Accountable–Consulted–Informed (**RACI**) model can clarify how industry, standards bodies, and regulators might collaborate. Table 1 provides an updated RACI chart, incorporating stakeholder feedback to refine tasks and responsibilities:

Activity	Industry (e.g. PHUSE, SCDM, ASA)	Standards Bodies (e.g. CDISC, HL7, OHDSI)	Regulators
Identify operational pain points (gaps & inefficiencies due to standards)	**R** (lead identification across companies)	**C** (consulted via user feedback, workshops)	**I** (informed of results)
Define and prioritise industry requirements for standards	**R / A** (drive consensus; accountable for user needs)	**C** (advise on feasibility; ensure consistency)	**I** (informed)
Align priorities & designate SDOs to lead development (new task)	**R** (drive cross-company agreement on priorities)	**A** (accept leadership for specific developments)	**I** (informed)
Develop semantic mappings & content standards (e.g. common terminologies)	**C** (user input, review of drafts)	**R / A** (lead creation of content standards)	**I** (informed)
Develop technical/API standards for data exchange (format/structure)	**C** (provide use cases, co-develop prototypes)	**R / A** (lead technical standard development)	**I** (informed)
Pilot and test implementations of new standards/integrations	**R / A** (lead real-world pilots, allocate resources)	**C** (support pilot design; adjust standards as needed)	**I** (informed; possibly **C** if in sandbox)
Provide regulatory feedback during development	**C** (share pilot results and observations)	**C** (ensure standards meet regulatory needs)	**R / A** (provide formal guidance as needed)
Issue or update regulatory guidance	**I** (comply with new guidance)	**I** (align standards with guidance where feasible)	**R / A** (issue official guidelines)

Table 1: Illustrative RACI matrix for an industry-led standards evolution initiative. (**R**=Responsible; **A**=Accountable; **C**=Consulted; **I**=Informed.) "Industry" denotes cross-company bodies (e.g. consortia or forums such as PHUSE, SCDM). "Standards Bodies" refers to formal Standards Development Organisations (e.g. CDISC, HL7). "Regulators" are agencies like the FDA, EMA, etc. This model includes a new activity "Align priorities & designate SDO leads" and breaks down "Develop standards specifications" into **semantic/content mapping vs. technical/API development** tasks, reflecting feedback to detail the steps of standards development.

Limitations: This framework is a starting point. For a real implementation, open questions remain around how decisions are made and enforced across organisations, how funding and resources are provided, and how conflicts are resolved. A formal charter or governance agreement would be required to address these issues and give the collaboration teeth.

9. Operational and Regulatory Considerations

Regulatory agencies have consistently emphasised that any changes in how we manage and submit data must uphold core principles of data integrity, traceability, and transparency. As the industry pursues enhanced standardisation and integration, it must ensure that:

- **Data integrity and lineage are preserved** – Aligned standards should improve, not obscure, the ability to trace data from collection through transformation to final submission. Regulatory inspectors and reviewers should be able to see how data were derived or converted, especially if new automated pipelines are introduced.

- **Clarity and consistency in submissions increase** – A key goal is to reduce the variability or ambiguity in how data is submitted. For regulators, receiving more standardised datasets, with common vocabularies and structures, can streamline review and reduce questions back to sponsors. For instance, as more trials incorporate EHR data, an aligned approach (e.g. using agreed FHIR-to-CDISC mappings) could prevent the current scenario where each sponsor presents EHR-sourced data differently.

- **Future pathways remain open** – Initiatives like digital submissions, graph-based data review, or AI-supported decision-making are on the horizon. A collaborative approach to evolving standards can help ensure these innovations are built on a solid foundation. For example, the **FDA's public exploration in April 2025 of using FHIR for real-world data submissions** indicates regulators are looking for industry's input on blending healthcare and research data. By proactively developing standardised methods for such use cases, industry collaboration can provide regulators with concrete solutions that are ready to pilot and validate in regulatory sandboxes, helping to shape future guidance.

In summary, industry action on data standards can simultaneously relieve companies' operational burdens and support regulators' needs by delivering higher-quality data in submissions. Early and ongoing regulator involvement in this process (as advisors or observers) will also help ensure that any new standards or hybrid models earn regulators' trust and acceptance, smoothing the path to their use in official contexts.

10. Hybrid Models and Practical Pathways

No single standard – however robust – is likely to meet all future requirements of clinical research. Thus, the community is increasingly exploring **hybrid models**, which combine multiple standards within an overarching architecture. These seek to leverage the best tool for each part of the data lifecycle, provided those tools can operate together. Key ingredients for success include:

- **Explicit delineation of each standard's role** – Clearly defining **where one standard's use stops and another's begins** to avoid overlapping implementations. In a hybrid approach, for example, one might declare that "from patient to EHR to site system, we use FHIR; from site system to sponsor's analysis database, we use CDISC standards." This eliminates ambiguity and duplication.

- **Common semantic layer** – Even if data moves through different structural standards, a single set of controlled terminologies and data definitions (as discussed in Section 3.5) should underpin the entire pipeline. This prevents loss of meaning: a lab test coded with LOINC or a condition coded with SNOMED remains the same concept throughout the lifecycle. Establishing and governing such a standards-agnostic semantic layer is arguably the most tangible piece of work industry can demand of the standards community – one that requires collaboration across all of the main SDOs, with ISO a natural potential partner. A practical operating rule would be that any new standards or terminology development in biopharma must either re-use existing concepts from this common layer or define new concepts within it under agreed rules, rather than creating parallel, standard-specific definitions.

- **Agreed interfaces and governance** – Points of handoff between standards (like the interface between a FHIR API and an SDTM dataset) must be governed by agreed mappings or conversion rules, ideally developed and maintained collaboratively by the relevant SDOs (with industry input). There should also be an oversight mechanism to manage changes – for instance, when one standard evolves, ensuring the connected standards adapt in tandem.

Proposed Hybrid Reference Architecture: One conceptual model (illustrative of many possibilities) might involve using **FHIR-based APIs for real-time data capture** and exchange, and **CDISC standards for internal data structuring and submission.** In practice, this means:

- During trial conduct, patient and site data flow from their origin (EHR systems, laboratory instruments, patient devices) into a sponsor's environment via **FHIR APIs.** FHIR's resources (e.g. for observations, medications, encounters) provide a common format for these sources to feed data in as events occur.

- Within the sponsor's platform, incoming data are **immediately mapped into a CDISC-compatible data store.** This could leverage **CDISC Library APIs** or metadata to apply standard definitions (CDISC domains, variables, controlled terms) to the data. Over time, as CDISC and HL7 continue to co-develop mappings, this process could be increasingly automated and standardised.

- **Semantic governance** is enforced using controlled vocabularies – e.g. as data enters, systems use tools like **Athena** or the CDISC Library to ensure that terms like diagnoses or lab tests are coded in standard ways, enabling consistent interpretation later.

- For analyses and regulatory submissions, the data are already in near-final SDTM/ADaM structure (since the transformation happened in-line during ingestion). This reduces the need for complex extract-transform-load (ETL) steps at study end, minimises data lag, and preserves detail (since the mapping rules were applied systematically at the point of data capture).

Such a hybrid architecture draws on FHIR's **strength in data exchange** (especially in clinical settings) and CDISC's strength in providing a **stable, analysis-ready format with regulatory acceptance.** The outcome is a seamless pipeline where each standard is used where it fits best, and together they provide end-to-end coverage.

It is worth noting that many companies are already moving in this direction by implementing **unified data platforms** that integrate EDC, clinical trial management, safety systems, and data warehouses. Instead of managing a patchwork of siloed tools, they are investing in platforms that can handle multiple data types and processes uniformly. In such a context, a hybrid standards model is not just a theoretical ideal – it's becoming a technical necessity to avoid recreating silos within the integrated platform. Therefore, the proposals in this paper align with broader technology trends: **the push for platform-based solutions and API-driven ecosystems in clinical research.**

While this paper stops short of specifying a single solution, it underscores that *the capability to develop and validate a hybrid model is within reach,* especially if guided by an industry–SDO partnership. Through carefully designed pilots – for example, using a **sandbox supported by the HL7 Vulcan Accelerator** or a similar mechanism – stakeholders could test and refine a reference architecture, measure its benefits (e.g. reductions in time/cost, improved data quality), and then work with SDOs to formalise the necessary standards enhancements.

11. Call to Action

To move this vision forward, we propose several immediate steps:

- **Convene a cross-industry workshop** – Leverage an upcoming neutral forum (such as a PHUSE conference or an HL7/ISO joint meeting) to gather diverse stakeholders. The aim: confirm a common understanding of the problem and establish a working coalition. This meeting should yield consensus on priority areas (e.g. "start with aligning RWE data for submissions, or integrating digital health data standards"). Furthermore, it should set the stage for ongoing collaboration by forming initial working groups and identifying leaders.

- **Issue a joint statement of intent** – As a follow-up, publish a short manifesto co-authored by industry leaders expressing commitment to collaborate on evolving clinical data standards. This document would articulate the urgency (backed by concrete examples and, where possible, metrics such as the cost of current inefficiencies) and invite participation from other companies, SDOs, and regulators.

- **Champion a standards-agnostic common semantic layer** – Make the common semantic layer (Section 10) a headline priority for the collaboration: a concrete, cross-SDO deliverable that industry can collectively demand. Working with the main SDOs — and potentially ISO — agree rules requiring new standards and terminology development to re-use concepts from the shared layer, or to define new concepts within it, so that semantic alignment is built in from the start.

- **Establish formal collaboration channels with SDOs and regulators** – Ensure that the newly formed industry consortium is not isolated. Set up liaison roles or subcommittees that interface directly with CDISC, HL7, ISO, and others. Consider inviting SDO representatives to co-chair relevant workstreams, and seek opportunities (like HL7 Connectathons or CDISC user meetings) to test ideas publicly. Also, proactively involve regulators by sharing plans and results, perhaps under the umbrella of existing innovation programmes, to explore regulatory sandbox opportunities for trying new standards in a controlled manner.

- **Plan pilot projects with transparent outcomes** – Identify one or two high-impact pilot projects to tackle first (for instance, a **demonstration of automated FHIR-to-SDTM conversion for EHR data,** or **a shared library of mappings for key vocabularies**). Execute these pilots with contributions from multiple companies to show broad applicability. Share the outcomes – both successes and lessons learned – through publications or public presentations, to build momentum and attract broader support.

12. Conclusion

Clinical data standards have enabled **remarkable progress** in how we collect, share, and analyse health data for research. Yet, the demands placed upon them – from integrating wearable device data to enabling real-time safety monitoring and supporting AI – are outpacing the ability of traditional, fragmented approaches to keep up.

The industry has reached an inflection point where it must **take a leadership role in guiding standards' evolution.** By shifting the approach to one where industry proactively collaborates to identify needs and drive solutions, we can ensure that standards from different domains **evolve together, rather than in isolation.** This collaboration, grounded in on-the-ground experience and focused on **pragmatic outcomes,** will reduce inefficiencies (like redundant data transformations), support innovation (by easing adoption of new data types and methods), and bolster regulatory confidence (through more consistent, transparent data). The goal is not to replace existing standards, but to **harmonise and enrich them in line with real-world requirements.** The opportunity now is to capitalise on collective expertise across the industry and work hand-in-hand with SDOs and regulators to build a more interoperable, **future-proof environment for clinical data** – one that can support better science, faster development, and improved patient outcomes.

13. About the Authors

This white paper was authored by members of the PHUSE data standards community, comprised of leaders and experts responsible for the governance and implementation of data standards across the biopharmaceutical industry. There were many common themes during our journey that encompassed the governance and implementation of CDISC standards, and we have explored how best to manage our end-to-end data standards. We are an industry of talented and dedicated professionals, and we are confident we can collectively pool our resources to do something truly amazing.