Paper DS19

From System Exports to Standards: Will Standardisation of External Vendor Data Redefine Clinical Data Management/Analysis?

Sai Sri Sravya Deevi, Zifo Technologies Inc, Boston, USA

ABSTRACT

In the rapidly evolving field of clinical research, integrating external data sources is crucial for enhancing clinical trial data and gaining insights into patient treatments. Diverse data sources like patient assessments, lab results, genomics, images/videos, and real-world evidence hold significant potential to advance drug development. However, orchestrating these data presents challenges in clinical data management. To successfully integrate external data, it is essential to address challenges in standardization, quality assurance, governance, and project management. Our strategies include adopting CDISC standards, enforcing stringent quality control, establishing clear governance frameworks, and ensuring timely data transfers. This approach allows us to standardize diverse data sources and provide study data for integrated review and analysis in under two months from contract sign-off. Through case studies and lessons learned from various projects, we will explore how to harness the full potential of external data to drive innovation and improve patient care.

INTRODUCTION

The complexity of clinical trial data has increased dramatically, with modern drug development increasingly depending on specialized external data sources to support regulatory submissions and clinical decision-making.[1] While traditional clinical trials primarily relied on site-collected data captured through electronic data capture (EDC) systems, contemporary studies integrate diverse external vendor data including central laboratory results, biomarker analyses, medical imaging, patient-reported outcomes, wearable device data, and real-world evidence.[2]

However, external vendors deliver data in proprietary system exports, each with unique formats, variable naming conventions, and vendor-specific data models. This creates a fundamental disconnect between how data is generated and how it must be submitted to regulatory authorities. Since December 2016, the U.S. Food and Drug Administration (FDA) has mandated CDISC standards for regulatory submissions.[3] The Clinical Data Interchange Standards Consortium (CDISC) develops data standards including the Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM), and Clinical Data Acquisition Standards Harmonization (CDASH).[4] While these standards have demonstrated significant resource reductions in data management,[5] the practical challenge lies in systematically applying them to heterogeneous data streams from external vendors.

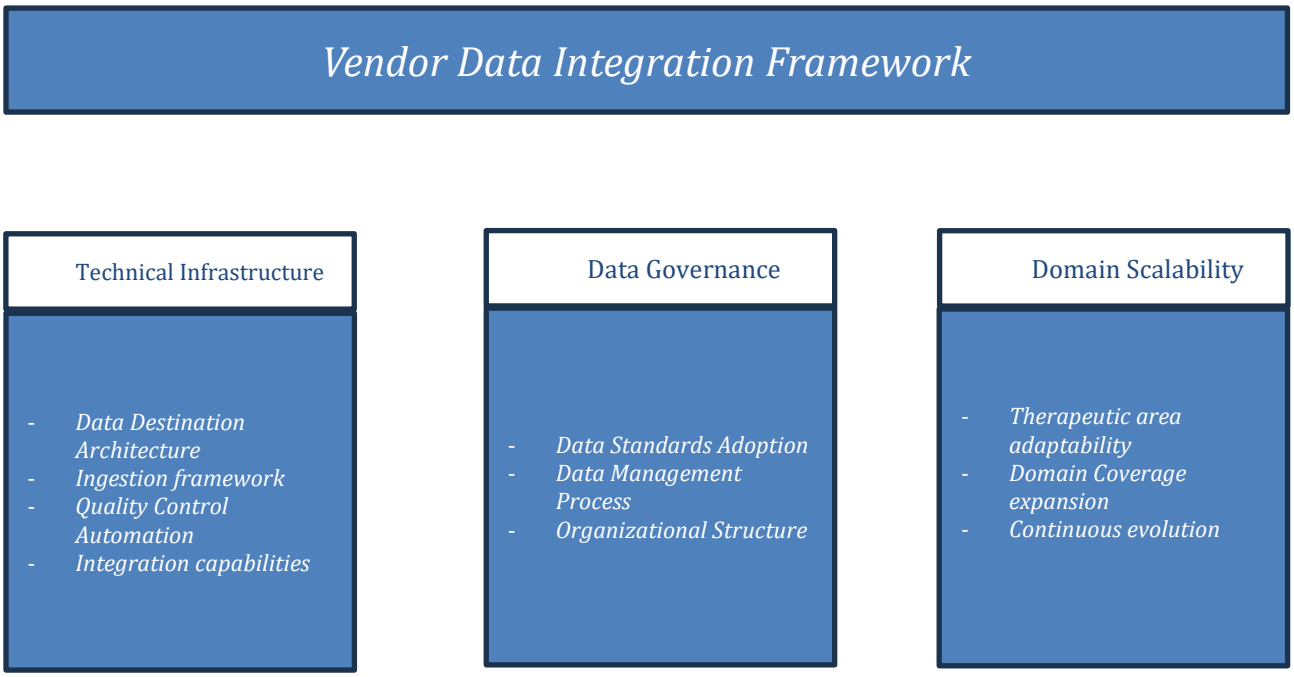
Common challenges in vendor data integration include:

- Vendors delivering non-CDISC compliant data formats
- Missing or incomplete data dictionaries
- Inconsistent quality standards across vendors
- Extended timelines for data transformation and reconciliation
- Lack of standardized governance processes[6]

To address these challenges, a systematic framework is needed that transforms external vendor data integration from a reactive, manual process into a proactive, standardized workflow.

VENDOR DATA INTEGRATION FRAMEWORK

To address these challenges, we developed a vendor data integration framework built on three interdependent pillars - Technical Infrastructure, Data Governance and Domain Scalability that transform external vendor data integration from a reactive, manual process into a proactive, standardized workflow.



1. Technical Infrastructure

The Technical Infrastructure pillar focuses on establishing the systems, tools, and automation capabilities that enable efficient, reliable data ingestion and validation. This pillar provides the practical foundation for how vendor data flows from source systems to submission-ready datasets.

Data Destination Architecture: We begin by understanding the sponsor's existing data destinations and infrastructure landscape. Through this assessment, we identify potential gaps or inefficiencies in current data pathways. When we discover areas for improvement, we propose methods to address them and collaborate with sponsors to align on optimal data flow routes. We then establish clear pathways for data movement and transformation, maintaining separation between raw vendor data and transformed CDISC datasets. These systems document the complete journey of each data point from vendor delivery through all transformations to final datasets supporting both day-to-day operations and regulatory inspections

Ingestion Framework: We automated data receipt and validation processes, eliminating manual file handling. Our standardized loading procedures with error handling and retry mechanisms enable us to manage multiple concurrent vendor integrations efficiently.

Quality Control Automation: We developed programmatic validation that verifies CDISC compliance, specification conformance, cross-domain consistency, and data completeness. Our automated issue detection and reporting systems generate detailed feedback for vendors, accelerating resolution.

2. Data Governance

The Data Governance pillar focuses on establishing the policies, standards, and oversight mechanisms that ensure consistent, high-quality vendor data integration. This pillar provides the foundational "rules of engagement" for how external vendor data is acquired, validated, and integrated across all vendors and therapeutic areas.

Data Standards Adoption: We established CDISC standards (SDTM, ADaM, CDASH) as baseline requirements, requiring vendors to deliver compliant data from the first transfer rather than accepting system exports for internal transformation. This eliminated post-delivery data wrangling that previously extended timelines by 2-4 months per vendor. Our Mapping Specifications provide vendors with explicit, variable-level deliverable requirements mapped to CDISC domains. Each specification details variable names, data types/lengths, format specifications, controlled terminology codelists, and mandatory/blinding status—removing interpretation ambiguity that historically caused multiple revision cycles.

For laboratory data mapping to LB domain, vital signs to VS, and other standard types, we built reusable templates maintained under version control. Custom domains following CDISC principles address data types without standard fits.

Specifications incorporate CDISC Controlled Terminology with explicit mappings for --TESTCD, --TEST, --ORRESU/--STRESU, and mandate essential variables (STUDYID, SUBJID, SITEID, VENDORNAME, timing variables, result variables, COVAL). Vendors receive complete requirements at project initiation, achieving first-transfer compliance across 1,000+ integrations.

.

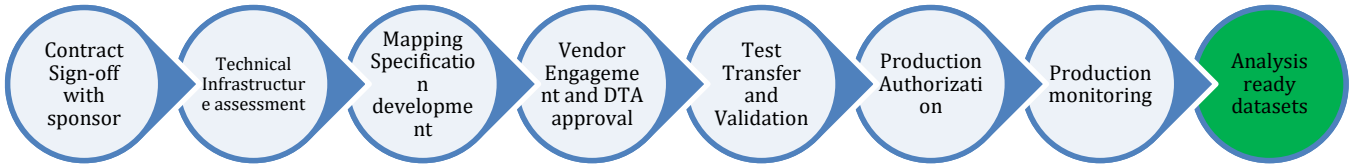
Data Management Processes: We operationalized governance through structured workflows including vendor qualification, formal Data Transfer Agreements (DTAs), test transfer validation, production authorization, and systematic change management. Each workflow ensures quality gates are met before advancing.
Organizational Structure: We defined clear accountability across data acquisition, quality control, and vendor coordination. Our escalation pathways, documentation standards, and performance monitoring mechanisms create feedback loops that continuously enhance the framework.

3. Domain Scalability

The Domain Scalability pillar focuses on building flexibility and reusability that enables the framework to support diverse data types across multiple therapeutic areas. This pillar ensures the framework can accommodate the vast range of disease areas, endpoints, and data collection modalities without requiring re-engineering for each new integration.
Therapeutic Area Adaptability: We developed flexible specification templates that accommodate disease-specific assessments across therapeutic areas. Our extensible mapping approaches support novel endpoints and emerging data modalities such as digital therapeutics and decentralized trial technologies.
Domain Coverage Expansion: We built reusable specification libraries across common domains like laboratories, imaging, ePRO, wearables enabling teams to leverage proven templates. Our cross-therapeutic area knowledge transfer allows insights from one program to benefit others.
Continuous Evolution: Our framework incorporates regular updates for new CDISC controlled terminology, emerging standards, and regulatory requirements. We systematically capture lessons learned, creating an organizational knowledge base that strengthens with each vendor engagement.

METHODOLOGY:

Our vendor data integration approach follows a structured workflow that operationalizes the three-pillar framework



Our vendor data integration approach follows a structured workflow that operationalizes all three pillars of the framework:

- **Technical Infrastructure Assessment & Planning**: Upon contract sign-off, we assess the sponsor's existing data destinations and infrastructure landscape to identify gaps or inefficiencies in current data pathways. We collaborate with sponsors to align on optimal data flow routes and configure our ingestion framework, establish secure connectivity protocols, and prepare validation environments tailored to the vendor's anticipated data formats.
- **Specification Development**: We evaluate whether existing mapping specification templates from our library can be leveraged for the vendor's data type. For established domains like laboratories or ePRO, we adapt proven templates to vendor-specific requirements. For novel domains or endpoints, we develop new CDISC-aligned mapping specifications. These specifications are formalized into Data Transfer Agreements (DTAs) that explicitly define required variables, formats, controlled terminology, quality expectations, and delivery schedules.
- **Vendor Engagement & Approval:** We conduct collaborative specification reviews with vendor technical teams to ensure mutual understanding of requirements and technical feasibility. Through iterative discussion, we refine specifications to balance regulatory compliance with vendor capabilities. Upon reaching alignment, both parties formally approve the DTA as the binding technical agreement governing all data transfers.
- **Test Transfer & Validation:** Vendors deliver initial test data based on approved DTA specifications. Our automated validation programs execute comprehensive quality checks verifying CDISC compliance, specification conformance, cross-domain consistency, and data completeness. We generate detailed issue reports and share with vendors for correction. This iterative process continues until all validation criteria are satisfied.

- **Production Authorization:** Upon successful test transfer validation with all issues resolved, we grant formal production authorization through documented sign-off processes. This authorization confirms the vendor can consistently deliver CDISC-compliant data and triggers the transition from testing to operational data delivery.
- **Production Monitoring:** Ongoing production transfers undergo continuous automated quality monitoring. We track delivery schedules, data quality metrics, and vendor performance to enable proactive identification of emerging issues. Performance dashboards provide real-time visibility into vendor integration status.
- **Analysis-Ready Datasets:** Successfully validated vendor data flows through established ingestion pathways, producing CDISC-compliant datasets available for integrated review and statistical analysis. We systematically capture lessons learned to enhance specification libraries and improve future integrations across therapeutic areas.

RESULTS

Integration Efficiency

The systematic framework has demonstrated measurable improvements in vendor data integration efficiency. Standardized specifications and automated validation have reduced integration timelines, enabling earlier identification and resolution of data quality issues. The structured approach has decreased dependency on manual data transformation activities that previously consumed significant resources and introduced variability. Organizations implementing this framework report more predictable project timelines, transforming vendor integration from a frequent source of delays into a controlled, manageable process.

Data Quality

Quality improvements are evident across multiple dimensions of vendor data management. Vendors are increasingly achieving first-time CDISC compliance, eliminating costly post-delivery correction cycles that previously delayed database locks. The completeness of data dictionaries and metadata has improved substantially as vendors understand expectations clearly through explicit specifications. Enhanced traceability and documentation quality provide robust audit trails that satisfy regulatory inspection requirements while supporting operational troubleshooting.

Scalability

The framework has proven scalable across diverse clinical research scenarios. It has been successfully applied across multiple therapeutic areas including oncology, neurology, immunology, and rare diseases, demonstrating flexibility across different disease contexts and study designs. The approach is effective for various data types ranging from simple laboratory results to complex biomarker and imaging data, accommodating both routine and novel assessments. The framework adapts to different vendor capabilities and system maturity levels, working effectively with both technologically sophisticated vendors and those with limited CDISC experience.

CASE STUDY

Study Context: A global Phase III oncology trial required integration of data from four specialized vendors: central laboratory, imaging core lab, ePRO platform, and biomarker analysis facility. Each vendor operated proprietary systems with distinct data models, creating significant integration complexity.

Challenges Encountered: Initial assessments revealed critical issues: the imaging vendor delivered data in proprietary XML format with non-CDISC terminology, the biomarker facility lacked standardized data dictionaries and CDISC experience, and the ePRO platform required extensive mapping to CDISC QS domain conventions. Coordinating across vendors with varying technical capabilities presented significant project management challenges.

Framework Application: We applied our three-pillar framework systematically. Technical Infrastructure implemented standardized ingestion pathways with automated validation tailored to each vendor. Data Governance established detailed CDISC-aligned DTAs through collaborative specification reviews. Domain Scalability leveraged existing laboratory templates while developing new specifications for novel biomarker assessments.

Outcomes: All four vendors achieved CDISC compliance during test transfers, eliminating post-delivery transformation. Automated validation resolved format inconsistencies and missing metadata before production authorization. The structured approach enabled parallel vendor activities, compressing timelines and ensuring analysis-ready datasets were available ahead of planned milestones.

Key Success Factors: Early vendor engagement allowed adequate time for specification development. Explicit specifications with examples eliminated ambiguity. Automated validation provided rapid feedback to vendors. Clear governance structures with defined escalation pathways enabled quick decision-making.

CONCLUSION

Standardization of external vendor data through systematic Technical Infrastructure, Data Governance, and Domain Scalability represents a fundamental shift from reactive data wrangling to proactive data management. The framework presented here demonstrates that achieving CDISC compliance from the first data transfer is not only possible but preferable to post-delivery transformation activities. As clinical trials continue to generate increasing volumes of data from diverse external sources, standardized approaches to vendor data integration will become essential capabilities for efficient drug development. Organizations that adopt systematic frameworks for vendor data

integration are experiencing measurable improvements in timelines, quality, and scalability. The question posed in our title - "Will Standardisation of External Vendor Data Redefine Clinical Data Management/Analysis?" is increasingly being answered in practice, with standardization transitioning from competitive advantage to baseline expectation across the industry.

REFERENCES

1. Getz KA, Campo RA. Trial watch: Trends in clinical trial design complexity. *Nature Reviews Drug Discovery*. 2017;16(5):307. doi:10.1038/nrd.2017.65
2. Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. *JAMA*. 2018;320(9):867-868. doi:10.1001/jama.2018.10136
3. U.S. Food and Drug Administration. Providing Regulatory Submissions in Electronic Format — Standardized Study Data. Guidance for Industry. December 2014. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-regulatory-submissions-electronic-format-standardized-study-data
4. Clinical Data Interchange Standards Consortium (CDISC). CDISC Standards Overview. Available at: https://www.cdisc.org/standards
5. Huser V, Cimino JJ. Evaluating adherence to the International Committee of Medical Journal Editors' policy of mandatory, timely clinical trial registration. *Journal of the American Medical Informatics Association*. 2013;20(e1):e169-e174. doi:10.1136/amiajnl-2012-001501
6. Kush RD, Helton E, Rockhold FW, Hardison CD. Electronic health records, medical research, and the Tower of Babel. *New England Journal of Medicine*. 2008;358(16):1738-1740. doi:10.1056/NEJMsb0800209

CONTACT INFORMATION

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

Author Name: Sai Sri Sravya Deevi
Company: Zifo Technologies Inc.
Address: 500 Lake Cook Rd, Deerfield, IL 60015
Email: sai.d@zifocorp.com
Website: https://zifornd.com/