association for **clinical data management**

Modernization of Clinical Data Flow

Leveraging CDISC 360i Standards for End-to-End Data Flow Automation

Date/Version: December 11, 2025 / 1.0

Main Contact for Paper:
Tamer Chowdhury - tamer.chowdhury@boehringer-ingelheim.com

Author: Tamer Chowdhury
External Contributor: Sam Hume, CDISC Open-Source Alliance (COSA)
Contributors: Gary Bhandarkar, Sylvain Berthelot, Carolyn Mcnamara, Satishkumar Subramaniam
Reviewers: Robert King, Nicola Goetz, Yuvaraj Parthiban

Modernization of Clinical Data Flow

Disclaimer

The information presented in this paper draws upon the combined understanding and knowledge of the Association of Clinical Data Management (ACDM) and industry experts on the topic of clinical data flow modernization. This document incorporates concepts and standards from the CDISC 360i initiative, with appropriate attribution to CDISC and contributing organizations.

These recommendations represent the opinions of the authors and do not necessarily reflect the position of individual companies or CDISC. Readers should assess the content and recommendations considering their own organizational needs, experience, and interpretation of relevant guidance and regulations.

Purpose

This white paper is intended to introduce modern clinical data flow practices leveraging CDISC 360i standards and emerging technologies. It demonstrates how end-to-end automation can transform clinical data management from manual, error-prone processes into efficient, traceable, and automated workflows.

Executive Summary

Clinical data management stands at a transformative crossroads. For decades, the journey from protocol to regulatory submission has been characterized by manual mapping, disconnected systems, and weeks of reconciliation efforts. The emergence of CDISC 360i standards—including the Unified Study Definitions Model (USDM), Biomedical Concepts, Dataset Specializations, and Data Contracts—promises to revolutionize this landscape.

This white paper traces the complete journey of clinical data from protocol design through regulatory submission, demonstrating how modern standards enable automated, traceable, and efficient data flows. We explore how combining structured protocol definitions, standardized data concepts, and machine-readable data contracts eliminates the traditional bottlenecks that consume 70-80% of data management effort.

The paper also examines emerging technologies including Dataset-JSON for efficient data exchange, real-time API-based data flows leveraging FHIR standards, and the FDA's vision for modernized data submission processes. Together, these innovations enable a future where clinical data knows where it came from, where it needs to go, and how to get there—automatically.

Modernization of Clinical Data Flow

Table of Contents

Note: Page numbers are approximate and may vary slightly in the final printed version.

Part 1: The Current State Challenge

The Traditional Data Flow

Clinical data management has evolved around an increasingly complex ecosystem over the past 30 years. What began as simple data entry tasks has grown into a formidable discipline involving data managers, data scientists, analysts, programmers, coders, and project managers. Yet despite this growth, the fundamental process remains largely manual and disconnected.

The amount of clinical data per trial has increased manyfold in the last decade, while database lock timelines have not proportionally improved. Data managers spend weeks creating CRFs from protocol text, months mapping source data to SDTM standards, and countless hours reconciling discrepancies across disconnected systems.

Key Pain Points

Process	Traditional Time	Key Issues
Protocol to CRF Design	6-12 weeks	Manual interpretation, ambiguous definitions
SDTM Mapping Specs	2-4 weeks	Manual mapping, inconsistent interpretations
SDTM Programming	4-6 weeks	Custom code, validation complexity
Data Transformation	Days to weeks	Manual reconciliation, quality checks
Database Lock	40+ weeks	total Multiple bottlenecks, rework cycles

Part 2: The Modern Data Flow Solution

CDISC 360i: A Paradigm Shift

CDISC 360i represents a fundamental rethinking of how clinical data flows through the research lifecycle. Rather than treating each stage as a disconnected step requiring manual translation, CDISC 360i establishes a continuous, machine-readable thread from protocol design through regulatory submission.

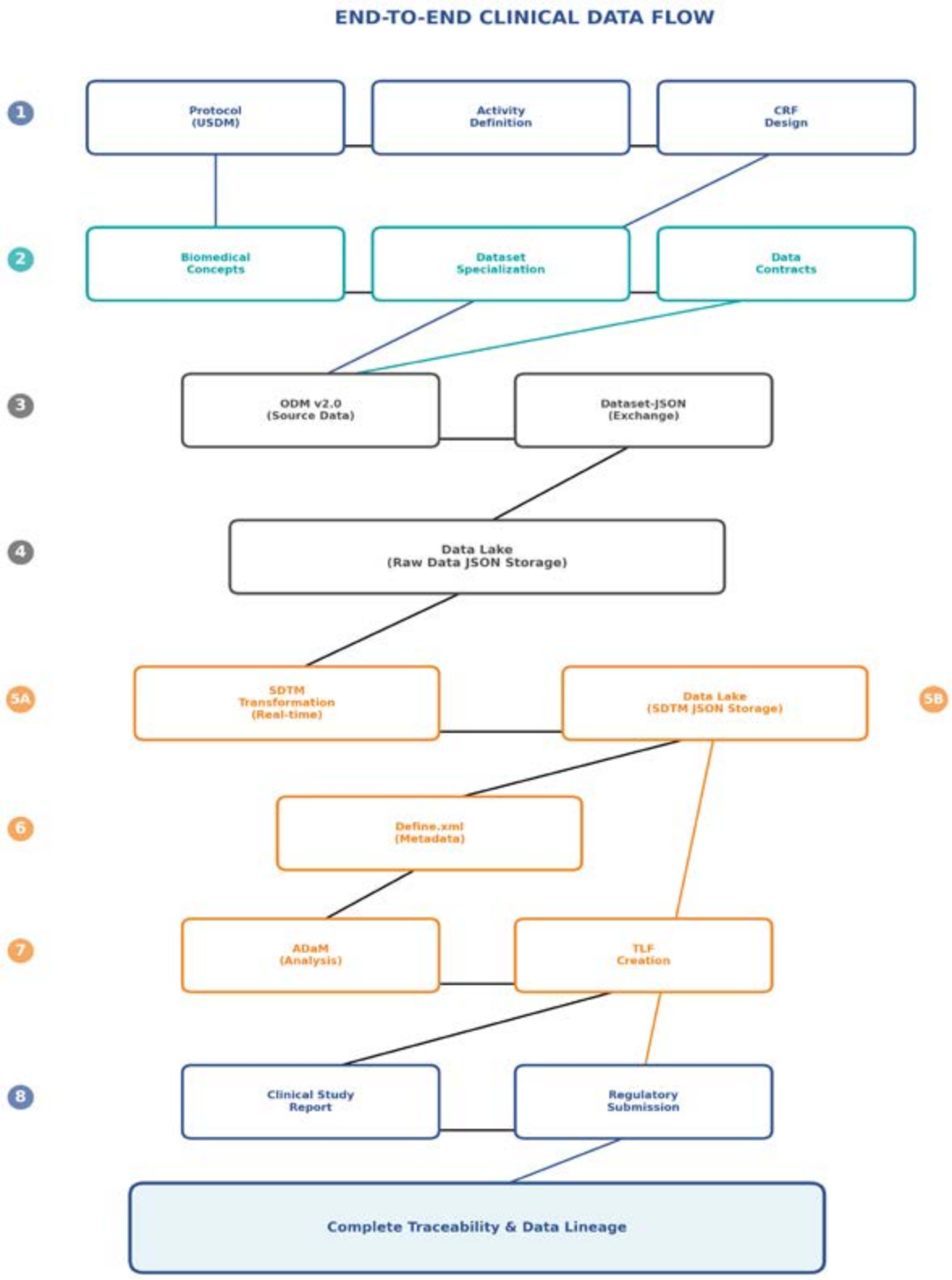
Key Insight: In the CDISC 360i model, data doesn't need to be mapped—it carries its own identity and destination information from the moment it's defined in the protocol.

The Seven Pillars of Modern Data Flow

The modern clinical data flow rests on seven interconnected standards and technologies, each playing a critical role in the automated journey from protocol to submission:

Modernization of Clinical Data Flow

End-to-End Data Flow Architecture



Part 3: Following a Data Point's Journey

To understand how modern data flow works in practice, let's follow a single hemoglobin measurement through its complete lifecycle—from protocol requirement to regulatory submission.

Stage 1: Protocol Design (USDM)

The Journey Begins: Our hemoglobin measurement starts its life in the protocol, defined using the Unified Study Definitions Model (USDM). Unlike traditional protocol documents, USDM creates a machine-readable representation of the study design.

Example USDM Definition:

- Study Activity: "Laboratory Assessments"
- Timing: Baseline, Week 4, Week 8, Week 12
- Required: Yes
- Critical to Quality Factor: Safety Assessment

Stage 2: Concept Definition (Biomedical Concepts)

Adding Meaning: The protocol activity links to a Biomedical Concept (BC) that provides standardized, unambiguous definition of what "hemoglobin" means across all studies.

Biomedical Concept: Hemoglobin

- BC ID: C0019046 (NCI Thesaurus)
- Definition: "The oxygen-carrying pigment of erythrocytes"
- LOINC Code: 718-7
- Standard Units: g/dL
- Normal Range: 12-16 g/dL (females), 14-18 g/dL (males)
- Data Type: Numeric

Stage 3: SDTM Mapping (Dataset Specialization)

Defining the Target: The Biomedical Concept links to its SDTM representation through Dataset Specialization. This provides the complete metadata for how hemoglobin appears in the Laboratory (LB) domain.

Dataset Specialization: LB.LBTEST = "Hemoglobin"

- SDTM Domain: LB (Laboratory Test Results)
- LBTESTCD: "HGB"
- LBTEST: "Hemoglobin"
- LBCAT: "HEMATOLOGY"
- LBORRES: Original result
- LBORRESU: Original units
- LBSTRESN: Numeric result in standard units
- LBSTRESU: "g/dL"

Stage 4: The Data Contract (URI Stamping)

Creating Identity: This is where the magic happens. The Data Contract creates a unique URI that connects the protocol requirement, biomedical concept, and SDTM target. This URI will travel with every hemoglobin datapoint collected in the study.

Data Contract URI:
study://NCT12345/lab/hematology/hemoglobin/baseline
What This URI Carries:
- Study: NCT12345
- Domain: Laboratory
- Category: Hematology
- Test: Hemoglobin
- Timepoint: Baseline
- SDTM Mapping: Fully defined
- Biomedical Concept: C0019046

Key Benefit: When data is collected, this URI is embedded in the source record. The system now knows exactly how to transform this data to SDTM without any manual mapping—the mapping is carried within the data itself.

Part 4: Data Collection and Transformation

Stage 5: Source Data Collection (ODM v2.0)

Data is Born: When the site performs the laboratory test, the result is captured in the EDC system using ODM v2.0 format. Critically, the Data Contract URI is embedded in the record.

ODM v2.0 Source Record (simplified):

```
{
“SubjectKey”: “001”,
“StudyEventOID”: “BASELINE”,
“FormOID”: “LAB_FORM”,
“ItemGroupOID”: “HEMATOLOGY”,
“ItemData”: [{
“ItemOID”: “HGB_RESULT”,
“Value”: “13.2”,
“DataContractURI”: “study://NCT12345/lab/hematology/hemoglobin/baseline”,
“Units”: “g/dL”,
“CollectionDateTime”: “2025-01-15T10:30:00Z”
}]
}
```

Stage 6: Data Exchange (Dataset-JSON)

Moving to the Data Lake: The ODM data is transformed to Dataset-JSON format for efficient storage and exchange. Dataset-JSON is CDISC's modern standard for tabular data exchange, offering significant advantages over traditional SAS transport files.

Why Dataset-JSON?

Benefit	Description	Impact
Human Readable	JSON format is text-based and easy to inspect	Easier debugging, validation
API-Ready	Native format for REST APIs and web services	Real-time data exchange
Compact	Smaller file sizes than XML or SAS	Faster transfer, less storage
Language Agnostic	Works with Python, R, JavaScript, etc.	Broader ecosystem
FDA Support	FDA championing JSON for submissions	Future-proof approach

Dataset-JSON Example (Raw Data):

{
"datasetJSONVersion": "1.1.0",
"fileOID": "RAW.HEMOGLOBIN.20250115",
"itemData": [
{
"USUBJID": "NCT12345-001",
"VISIT": "BASELINE",
"LBORRES": "13.2",
"LBORRESU": "g/dL",
"CONTRACTURI": "study://NCT12345/lab/hematology/hemoglobin/baseline",
"LBDTC": "2025-01-15T10:30:00"
}
]
}

Part 5: Automated SDTM Transformation

Stage 7: Real-Time SDTM Generation

The Transformation Moment: Here's where the power of Data Contracts becomes evident. The transformation engine reads the Data Contract URI, retrieves the complete mapping specification, and automatically generates SDTM records in seconds—not weeks.

Zero Manual Mapping: The Data Contract URI points to a complete SDTM specification. The transformation engine simply applies the rules—no manual specification required, no programming needed, no mapping spreadsheets to maintain.

Modernization of Clinical Data Flow

CDISC 360i Technical Note: The emerging Data Definition Specification (define.json) provides the comprehensive metadata for automated data transformation. Tools like the open-source sdtm.oak R package can leverage this metadata for SDTM generation, with the flexibility to plug in alternative transformation engines as needed.

Automated SDTM LB Record:

{
“STUDYID”: “NCT12345”,
“DOMAIN”: “LB”,
“USUBJID”: “NCT12345-001”,
“LBSEQ”: 1,
“LBTESTCD”: “HGB”,
“LBTEST”: “Hemoglobin”,
“LBCAT”: “HEMATOLOGY”,
“LBORRES”: “13.2”,
“LBORRESU”: “g/dL”,
“LBSTRESN”: 13.2,
“LBSTRESU”: “g/dL”,
“LBDTC”: “2025-01-15T10:30:00”,
“VISIT”: “BASELINE”,
“VISITNUM”: 1
}

Derivation Rules and Quality Checks

The transformation process also applies derivation rules and quality checks automatically. These are defined once in the Dataset Specialization and applied consistently across all records.

CDISC 360i Implementation: Organizations are increasingly using CDISC CORE for conformance and data quality checks integrated directly into transformation workflows. This enables continuous quality validation from the very first SDTM dataset generation, catching issues immediately rather than at database lock.

Automated Derivations:

- LBSTRESN: Converted from original units to standard units
- LBSTRESC: Standardized character result
- LBNRIND: Normal range indicator (calculated from BC normal ranges)
- LBSTAT: Completion status
- LBREASND: Reason for missing data (if applicable)

Quality Checks Applied:

- Value within expected range?
- Units consistent with biomedical concept?
- Required fields present?
- Data type matches specification?
- Controlled terminology valid?

Part 6: The Future of Data Exchange

FDA's Vision for Dataset-JSON

The FDA is actively championing the adoption of Dataset-JSON as the future standard for regulatory submissions. This represents a significant shift from traditional SAS transport files (XPT) to modern, API-ready data formats.

Why FDA Supports Dataset-JSON:

1. Real-Time Review: JSON enables API-based access to submission data, allowing reviewers to query data dynamically rather than downloading entire datasets.
2. Integration with FDA Systems: JSON integrates seamlessly with modern review platforms and analytical tools.
3. Improved Validation: JSON schemas enable automated validation before submission, catching errors early.
4. Reduced File Sizes: JSON compression reduces submission package sizes by 40-60% compared to XPT.
5. Future-Ready: JSON supports emerging technologies including AI/ML-based review tools.

FHIR and Real-Time Data Exchange

Fast Healthcare Interoperability Resources (FHIR) is revolutionizing how health data is exchanged between systems. When combined with CDISC standards, FHIR enables real-time clinical trial data flows that were previously impossible.

FHIR + CDISC Integration:

- **EHR Integration:** Patient data flows directly from electronic health records to clinical trial systems via FHIR APIs
- **Real-Time Safety:** Adverse events detected in EHR automatically trigger safety workflows in trial management systems
- **Lab Results:** Laboratory results from commercial labs flow via FHIR ® transform via Data Contracts® appear as SDTM in minutes
- **Wearable Data:** Continuous monitoring data from wearables streams via FHIR® processed in real-time® available for analysis
- **Site Monitoring:** Real-time data quality checks can occur continuously rather than during monitoring visits

API-Based Data Flows

The combination of Dataset-JSON, Data Contracts, and FHIR enables a completely new model of clinical data management: API-based real-time data flows.

Modern API Architecture:

1. **CDISC Library API:** Real-time access to biomedical concepts, controlled terminology, and dataset specifications
2. **Study Definition API:** USDM protocol definitions accessible via REST API
3. **Data Collection API:** Real-time submission of source data with embedded Data Contract URIs
4. **Transformation API:** On-demand SDTM transformation triggered by data arrival
5. **Query API:** Real-time access to SDTM datasets for analytics, monitoring, and review
6. **Submission API:** Direct submission to FDA via Dataset-JSON API endpoints

The API Revolution: This architecture eliminates the traditional "batch and transfer" model. Data flows continuously, transformations happen in real-time, quality checks occur immediately, and insights are available within minutes of data collection.

Part 7: The Business Impact

Quantifying the Benefits

The modernization of clinical data flow delivers measurable benefits across time, quality, and cost dimensions. Organizations implementing these approaches report transformational results.

Process Step	Traditional	Modern Flow	Time Saved
Protocol to CRF	6-12 weeks	1-2 weeks	75-83%
SDTM Mapping Specs	2-4 weeks	Real-time (seconds)	100%
SDTM Programming	4-6 weeks	Automated	100%
Data Transformation	Days-weeks	Seconds-minutes	99%
Database Lock Cycle	40+ weeks	12-16 weeks	60-70%

Quality Improvements

Error Reduction:

- Manual mapping errors: Significantly decreased
- Protocol deviations from data collection issues: Significantly decreased
- Query resolution time: Substantially decreased
- Database lock delays: Significantly decreased
- SDTM validation findings: Dramatically decreased

Quality Improvements:

- Complete traceability from protocol to submission
- Consistent data definitions across studies
- Real-time quality monitoring
- Automated validation at every step
- Elimination of transcription errors

Resource Impact

Beyond time and quality improvements, modern data flow fundamentally changes how data management teams spend their time. Manual, repetitive tasks give way to higher-value activities.

Activity Traditional	% Effort	Modern % Effort	Change
Manual mapping/specification	40%	5%	-35%
Programming transformations	25%	5%	-20%
Data reconciliation	20%	5%	-15%
Quality reviews	10%	15%	+5%
Process improvement	5%	30%	+25%
Standards development	0%	15%	+15%
Analytics & insights	0%	25%	+25%

Conclusion: The Path Forward

What We've Learned

The journey from traditional to modern clinical data flow is not merely an incremental improvement—it represents a fundamental transformation in how we think about, collect, and manage clinical trial data.

Key Principles of Modern Data Flow:

1. **Definition First:** Data defined once in structured formats (USDM, Biomedical Concepts) propagates automatically throughout the lifecycle.
2. **Identity Embedded:** Data Contract URIs create permanent links between source data and target formats, eliminating manual mapping.
3. **Standards-Based:** Leveraging CDISC 360i, FHIR, and Dataset-JSON ensures interoperability and future-proofing.
4. **API-Enabled:** Real-time data flows replace batch processes, enabling continuous monitoring and analysis.
5. **Quality Built-In:** Validation and quality checks occur automatically at every transformation, not as separate activities.

Implementation Roadmap

Organizations looking to modernize their clinical data flow should consider a phased approach that builds capability progressively:

Phase 1: Foundation (Months 1-3)

- Establish USDM protocol authoring capability
- Build biomedical concept library for therapeutic area
- Create dataset specialization templates
- Train teams on CDISC 360i concepts

Phase 2: Pilot Implementation (Months 4-6)

- Select pilot study (simple design, limited sites)
- Implement data contracts in EDC
- Build/configure transformation engine
- Establish Dataset-JSON infrastructure

Phase 3: Scale and Optimize (Months 7-12)

- Expand to multiple studies
- Integrate FHIR for external data sources
- Implement API-based data flows
- Develop real-time monitoring capabilities
- Establish continuous improvement processes

The Future is Here

The technologies and standards described in this white paper are not theoretical—they exist today and are being implemented by forward-thinking organizations. CDISC 360i standards are mature, Dataset-JSON is finalized, FHIR is widely adopted in healthcare, and transformation engines are commercially available.

The Transformation Awaits: The question is not whether to modernize clinical data flow, but how quickly your organization can make the transition. Every week spent in traditional manual processes is a week of lost efficiency, increased errors, and delayed insights.

The journey of a single hemoglobin data point has shown us what's possible. Now it's time to make that possibility a reality for all clinical data, across all studies, throughout the industry.

Future Considerations and Evolution

As clinical data management continues its modernization journey, several emerging standards and technologies promise to further enhance the capabilities described in this white paper. The CDISC 360i initiative continues to evolve, with Phase 2 developments already underway.

CDISC 360i Phase 2: Analysis and Reporting

Building on the foundation established in Phase 1, **CDISC 360i Phase 2** (launching Q1 2026) will extend automation through the analysis and reporting stages of the clinical data lifecycle:

Analysis Results Standard (ARS):

- Standardized representation of analysis results
- Analysis Concepts defining statistical analyses and outputs
- Automated generation of Tables, Listings, and Figures (TLFs)
- Direct linkage from analysis specifications to final outputs

Impact: This will complete the end-to-end automation journey, extending from protocol design through final regulatory submission with complete traceability at every stage.

Data Definition Specification (Define.json)

A critical innovation currently under development is the Data Definition Specification (also referred to as define.json). This emerging standard will provide:

Comprehensive Metadata Management:

- Complete data flow specifications from source to submission
- Semantic definitions and relationships between data elements
- Data supply and demand specifications across the pipeline
- Automated generation of define.xml from a single source of truth

Data Point Traceability:
The Data Definition Specification will maintain complete metadata to trace each datapoint's journey through the clinical research data pipeline, addressing metadata gaps identified during automation implementation.

Additional Metadata Models:
As automation expands, additional models such as a standardized Data Management Plan may be developed to capture metadata currently managed outside existing CDISC standards.

Modernization of Clinical Data Flow

Real-World Data Integration

The integration of Real-World Data (RWD) into clinical trials is accelerating. Two key developments will enable this integration:

RWD Lineage Standard (In Development):

- Data point traceability specifically designed for Real-World Data
- FHIR integration capabilities for EHR data flows
- Provenance tracking from original source through analysis
- Quality assessment frameworks for RWD incorporation

FHIR-to-CDISC Transformation:
As healthcare systems increasingly adopt FHIR standards, automated transformation from FHIR resources to CDISC SDTM becomes critical. The combination of FHIR APIs, Data Contracts, and transformation engines will enable real-time clinical trial data collection from routine care settings.

Continuous Quality and Conformance

Early adopters are demonstrating a powerful new paradigm: **continuous quality validation**. Rather than waiting until database lock to run comprehensive quality checks, organizations now integrate CDISC CORE conformance and quality checks directly into transformation workflows.

The Continuous Quality Approach:

- Generate SDTM datasets as early as possible in the study lifecycle
- Run quality checks each time transformations execute
- Identify and resolve issues immediately, not weeks later
- Build quality in from the beginning rather than inspecting it in at the end

Result: Dramatic reduction in database lock delays and validation findings, with quality improving continuously rather than being a final-stage bottleneck.

Expanding the Metrics

As CDISC 360i Phase 2 extends through analysis and reporting, the business impact metrics will expand beyond SDTM generation to include:

Process Area	Current State	Future State with 360i Phase 2
Analysis Specification	Manual SAP creation 4-6 weeks	Automated from Analysis Concepts Real-time
TLF Programming	Manual programming 6-8 weeks	Automated from ARS 1-2 weeks
TLF Validation	Manual review 2-4 weeks	Automated validation Continuous
CSR Integration	Manual compilation 4-6 weeks	Automated assembly 1-2 weeks
Total Time to Submission	60-80 weeks	20-30 weeks

The Path of Evolution

The modernization of clinical data flow is not a single transformation but an ongoing evolution. As standards mature, tools improve, and organizations gain experience, new capabilities emerge:

Near-Term (2026-2027):

- CDISC 360i Phase 2 implementation (ARS and Analysis Concepts)
- Data Definition Specification standard finalized and adopted
- Widespread CORE integration for continuous quality
- Initial RWD Lineage implementations

Medium-Term (2027-2029):

- FHIR-to-SDTM transformation becomes routine
- Hybrid clinical trial / RWD study designs become standard
- AI/ML integration for predictive quality and risk management
- Full end-to-end automation from protocol to submission

Long-Term (2030+):

- Real-time adaptive trials with continuous data flow
- Decentralized trial models enabled by seamless data integration
- Patient-level data interoperability across healthcare and research
- Regulatory review processes fully API-enabled

This white paper represents a snapshot of an industry in transformation. The foundations described here—USDM, Biomedical Concepts, Data Contracts, Dataset-JSON—are not the end of the journey but rather the beginning of a new era in clinical data management. As Sam Hume of CDISC notes, "We're not just automating existing processes—we're fundamentally reimagining how clinical data flows from conception to conclusion."

References and Resources

Standards and Specifications

Resource	URL
CDISC 360i Program	https://www.cdisc.org/cdisc-360i
USDM Documentation	https://www.cdisc.org/standards/foundational/usdm
Dataset-JSON v1.1	https://www.cdisc.org/standards/foundational/dataset-json
Biomedical Concepts (COSMoS)	https://github.com/cdisc-org/COSMoS
CDISC Library API	https://www.cdisc.org/cdisc-library
CDISC CORE (Conformance Rules)	https://www.cdisc.org/standards/foundational/core
sdtm.oak R Package	https://github.com/pharmaverse/sdtm.oak
Analysis Results Standard (ARS)	https://www.cdisc.org/standards/foundational/ars
FHIR Specification	https://www.hl7.org/fhir/
ODM v2.0	https://www.cdisc.org/standards/data-exchange/odm

Acknowledgments

This white paper incorporates concepts and standards from the CDISC 360i initiative. We gratefully acknowledge the contributions of CDISC, the CDISC Open-Source Alliance (COSA), and the broader clinical data standards community.

Special thanks to Sam Hume (CDISC/COSA) for his invaluable technical review and insights on ODM v2.0, Dataset-JSON, the Data Definition Specification, and CDISC 360i Phase 2 developments. His detailed feedback significantly enhanced the technical accuracy and forward-looking perspective of this work.

We also thank the many data managers, statisticians, programmers, and clinical operations professionals whose real-world experiences and feedback informed this work. The ACDM Data Management Expert Group (DMEG) Technology Team provided critical review and validation of the practical implications described throughout.

About ACDM

The Association of Clinical Data Management (ACDM) is the professional organization dedicated to advancing the practice of clinical data management worldwide. Through education, advocacy, and thought leadership, ACDM supports data management professionals in delivering high-quality data that accelerates medical research.

Version Information

Version 1.0 | December 11, 2025

Disclaimer

www.acdmglobal.org