

HEXAWARE

AI-Powered Digital Standards Library for Accelerated Regulatory Submissions

February 2026

Anshu Aman, Varun S.

Hexaware Technologies



The Critical Problem in Clinical Trial Programming

Clinical trial programming suffers from fragmented workflows and decentralized specifications, leading to manual rework, data inconsistencies, and delayed regulatory submissions despite existing standards like CDISC.

90%

Unstructured Protocols

Clinical trial protocols still created as text files

The absence of real-time data traceability and non-digitized metadata management further hinder efficiency and regulatory confidence, creating bottlenecks that delay life-saving therapies from reaching patients.

4

Months Lost

Average lag between protocol approval and study start-up

The Solution:

AI-Enabled Digital Metadata Repository

This innovation proposes an AI-enabled, cloud-native digital metadata repository (MDR) integrated with statistical programming automation. It leverages Machine Learning agents and API-based data ingestion to enable continuous data quality checks, real-time reconciliation, and automated generation of statistical outputs.



ML-Powered Automation

Machine Learning agents parse protocols and generate executable code automatically



Cloud-Native Architecture

Centralized metadata repository ensures consistency across trial



Real-Time Quality Control

Continuous data validation and automated compliance

Technical Architecture: The Five Key Artifacts of USDM

1. Logical Data Model (LDM)

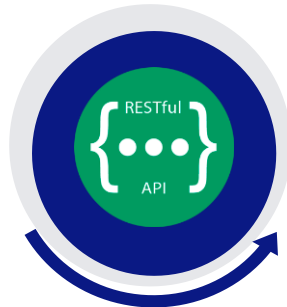
UML class diagram depicting data entities and relationships, functioning as the foundational truth for study definition technologies

3. Controlled Terminologies

Standardized code lists for protocol elements developed by CDISC to ensure semantic consistency across the trial lifecycle

5. Study Definition Repository

Centralized middle-layer component facilitating study asset creation and system configuration



2. API Specification

Standardized RESTful API using OpenAPI Specification 3.0, enabling high-velocity exchange of study definitions between machines

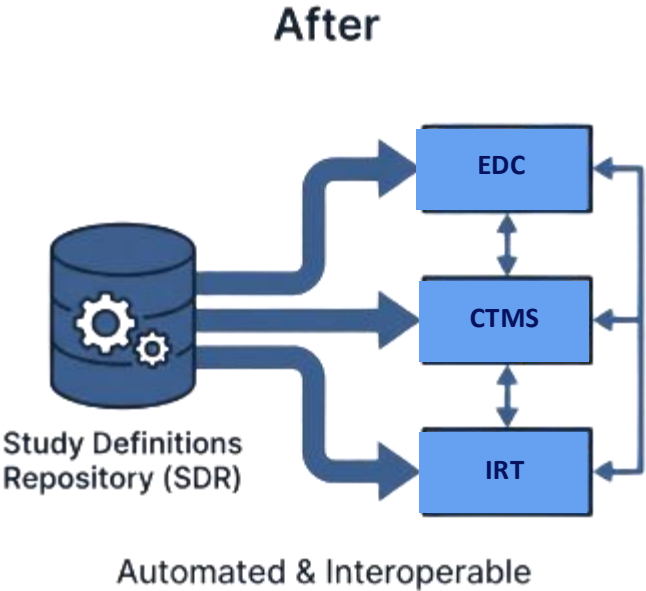
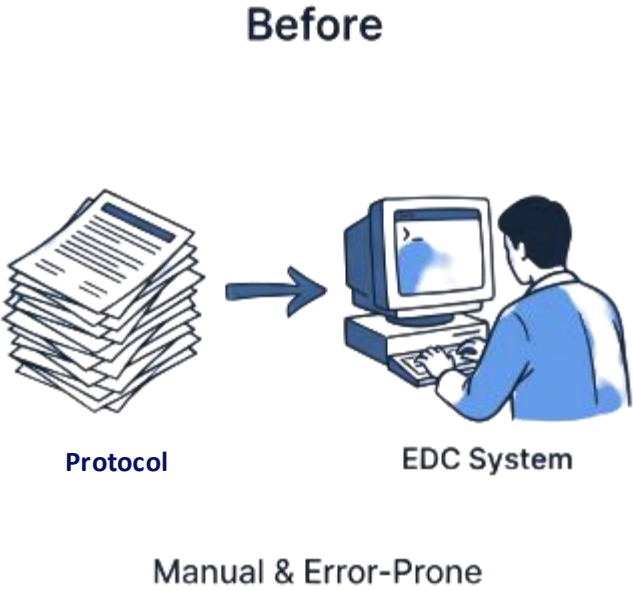
4. Implementation Guide

Comprehensive documentation providing solution architects with reusable designs, best practices, and conformance rules

From Documents to Data: The Digital Data Flow Initiative

The Digital Data Flow (DDF) initiative aims to modernize clinical trials by enabling a fully digital workflow through protocol digitalization.

The core objective is moving away from traditional document models toward standardized, machine-readable representations of study protocols built upon the Unified Study upon the Unified Study Definitions Model (USDM).



Traditional Model

Unstructured text files manually parsed by disparate

USDM Framework

Machine-executable single source of truth with standardized data model

Automated Execution

Protocol data flows automatically between systems without custom transformations

AI-Driven Orchestration: The Intelligence Core

Natural Language Processing Agents

- The solution utilizes NLP agents to parse SAPs and digitized protocols, mapping endpoints and derivations to CDISC standards. Hybrid engines combining machine learning with rule-based machine learning with rule-based logic generate executable code in SAS, R, or Python.
- A "human-in-the-loop" model ensures quality: confidence scores determine whether outputs are automatically accepted or routed for expert review. This from previous study conventions and adapt to organization-specific therapeutic area nuances over time.

Data Mesh Architecture

- High-velocity API connections aggregate heterogeneous sources including EHRs, lab feeds, eCRFs, and registries. Each dataset is normalized and assigned a persistent lineage identifier to persistent lineage identifier to maintain traceability from raw data to submission-ready deliverables.

60-80%

Mapping Time Reduction

95%

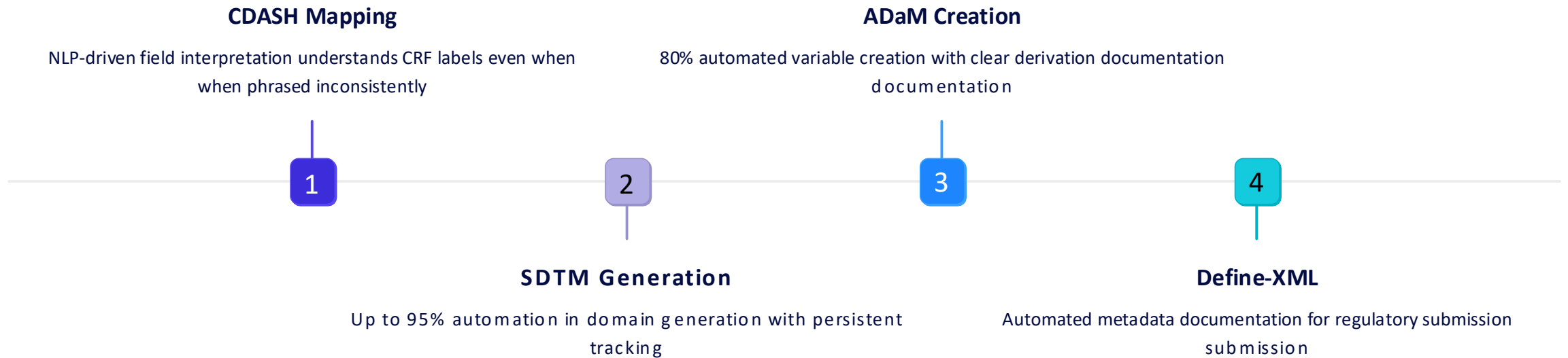
SDTM Automation

80%

ADaM Automation

Standards-Driven Pipeline for Automated Deliverables

The library enforces a standards-driven pipeline where each stage— CDASH to SDTM to ADaM to Define-XML— is automated through metadata-driven mappings, analysis back to raw data.



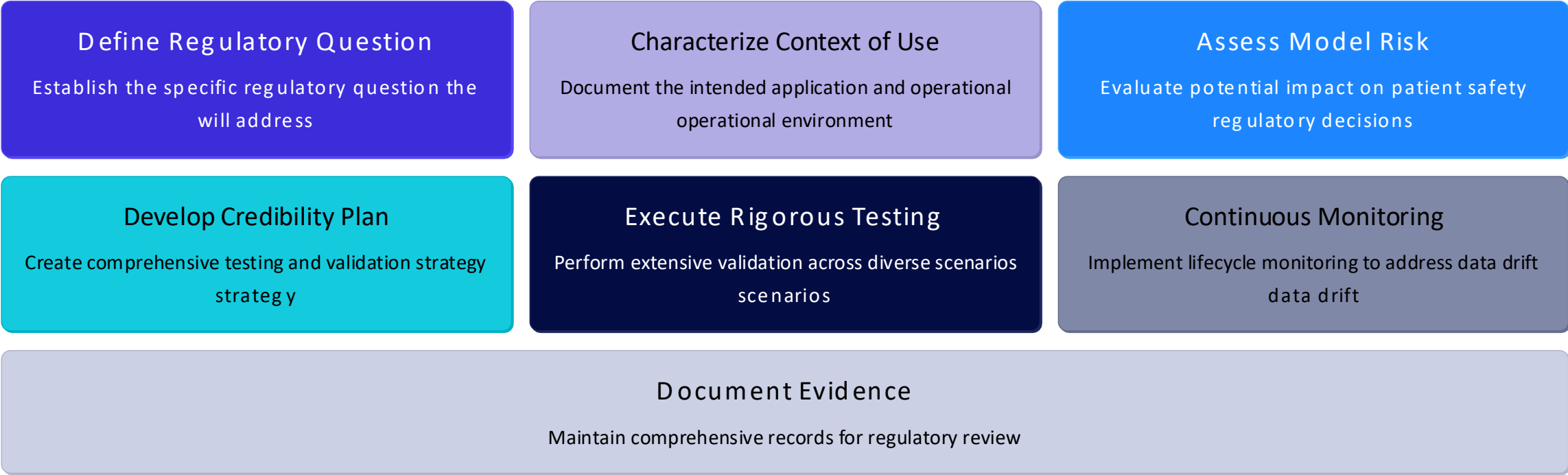
Real-Time Quality Control and GxP Compliance

Validation is embedded throughout the workflow with automated Pinnacle21 checks and checks and logged remediation loops. Machine learning algorithms detect programming errors and inconsistencies faster than manual review, enhancing regulatory regulatory dossier quality.

Technical controls including immutable audit logs, electronic signatures, and ensure compliance with **21 CFR Part 11** and **EU Annex 11**. This transforms clinical from manual testers to designers of AI-powered validation frameworks.

Regulatory Alignment: FDA 7-Step Credibility Framework

To ensure regulatory confidence, the implementation aligns with the FDA's risk-based credibility assessment framework for AI and machine learning applications in clinical research.




Higher-risk AI applications, such as those used for patient risk stratification, require more extensive evidence and continuous lifecycle monitoring confidence.

Transformative Impact: Measurable Benefits

Accelerated Submissions

40-60% faster regulatory submissions through automated workflows and real-time quality control

Cost Reduction

Up to 40% lower integration and compliance costs by replacing point-to-point custom integrations

Trial Delay Reduction

Up to 50% reduction in trial delays through automated study asset creation

Transition to eCTD v4.0

The library facilitates the transition to eCTD v4.0, which moves the submission paradigm from document-centric to data-centric. AI plays a tagging documents according to v4.0 metadata standards and managing unique Object Identifiers (OIDs)—unfeasible to perform manually.

Version 4.0 supports bidirectional messaging and document reusability across different applications, reducing the risk of technical rejections and accelerating and accelerating interactions with health authorities.

The Future of Clinical Trial Data Management

- The implementation of an AI-powered digital standards library represents a transformative shift in clinical trial data management and regulatory submissions.
- By digitizing protocols through the USDM and leveraging cloud-native metadata repositories, organizations can eliminate the persistent lag between study design and execution while ensuring total data traceability.
- The synergy of NLP-driven mapping and automated code generation not only reduces manual rework and costs but also enhances the scientific coherence and accuracy of regulatory dossiers.



Embrace Data-Centricity

Transition from document-based to data-driven clinical development



Prioritize Automation

Leverage high-velocity automation for competitive advantage



Align with Regulatory Frameworks

Implement risk-based credibility assessment from the start

"This solution positions the life sciences industry to deliver life-saving therapies with greater efficiency, transparency, and clinical precision."

Thank You

Contact Information: Anshu Aman & Varun S | Hexaware Technologies
Email: anshua2@hexaware.com | Phone: +91-9661927173
Website: hexaware.com

HEXWARE