



CDISC India Day

Intelligent Agentic AI for Clinical Data Management

From Data Collection to Regulatory Submission

RISHI RAJ

Manager – Strategic Partnerships & Clinical Innovation
SENSAN BIOSCIENCES





RISHI RAJ IIMK

Manager – Strategic Partnerships & Clinical Innovation
SENSAN BIOSCIENCES

Rishi Raj is a clinical data management and analytics professional focused on modernizing CDM workflows through automation, Agentic AI, data pipelines, and visualization. He has hands-on experience across the full data lifecycle and leverages tools such as SAS, R, SQL, Power BI, and Medidata Rave to improve data quality and efficiency. His work centers on integrating analytics and centralized monitoring approaches to enable more proactive and scalable clinical trial oversight.

His work focuses on designing scalable data pipelines, developing Agentic AI solutions for intelligent clinical data review. Rishi is passionate about bridging Clinical Data Management with AI-driven innovation to accelerate data-driven decision-making and modernize clinical trial execution.





Agenda

1. Why CDM Must Evolve
2. Beyond EDC: Connected Data Ecosystem
3. Automation → Agentic AI
4. Agentic AI Across the CDISC Lifecycle
5. Intelligent Clinical Data Pipeline
6. CDM Use Cases & Case Study
7. AI Governance & Human Oversight
8. Inspection-Ready AI
9. The Future of Clinical Data Management
10. Key Takeaways

Why Clinical Data Management Is at an Inflection Point?

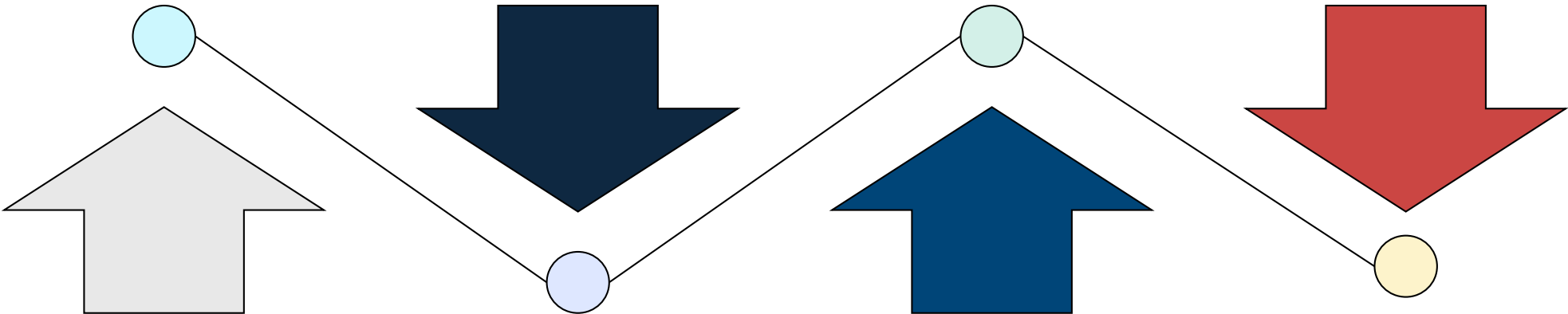


Regulatory Expectations Intensify

ICH E6(R3), 21 CFR Part 11, and CDISC compliance require end-to-end traceability, audit-ready data, and ALCOA+ adherence at every stage.

The Opportunity

Intelligent automation and Agentic AI offer a path to faster, safer, and more compliant clinical data operations augmenting CDM professionals, not replacing them.



Exponential Data Complexity

Trials now integrate data from EDC, central labs, eCOA/ePRO, wearables, imaging, RTSM, biomarkers, and safety systems simultaneously.

Manual Processes Cannot Scale

Traditional CDM workflows manual query management, siloed reconciliation, and paper-based oversight create bottlenecks and risk.





Data Beyond EDC: The Fragmentation Challenge

Siloed Data = Compounding Risk

Each external data source arrives in its own format, frequency, and standard requiring labor-intensive reconciliation and increasing the risk of inconsistency, delayed queries, and submission gaps.



- No unified metadata model across sources
- Asynchronous data flows disrupt data cleaning timelines
- Reconciliation errors propagate into SDTM and ADaM
- Traceability from source to submission is fragmented

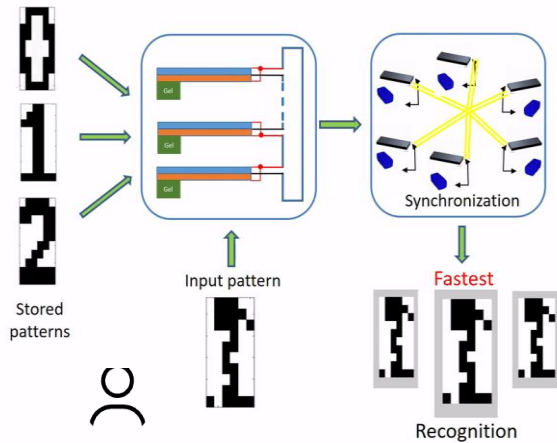


Fragmented data ecosystems are among the leading causes of inspection findings and database lock delays.

Automation → Agentic AI

Pattern Recognition

Statistical models identify systematic errors and site-level anomalies invisible to manual review



Anomaly Detection

ML algorithms flag outliers and unexpected data distributions in near real-time



Trend Analysis

Longitudinal monitoring reveals drift, protocol deviations, and compliance risks before they compound



End-to-End CDISC Digital Data Lifecycle

CDASH Collection

EDC, eCOA, labs,
wearables

SDTM Transformation

Domain mapping
suppqual, define.xml

Regulatory Packaging

eCTD, reviewer guide,
SDRG/ADRG

Ingestion & Validation

Automated checks, CT
validation, flags

ADaM Derivation

ADSL, ADAE, ADLB,
derived endpoints



The CDISC digital data lifecycle from CDASH collection through SDTM mapping, ADaM derivation, and eCTD packaging forms the regulatory backbone. Agentic AI instruments each stage with automated checks, traceability metadata, and human-reviewable audit trails.

CASE STUDY

AI-Enabled Phase III Rare Disease Trial

489 randomized patients with a 24-month follow-up



Real-time dashboards

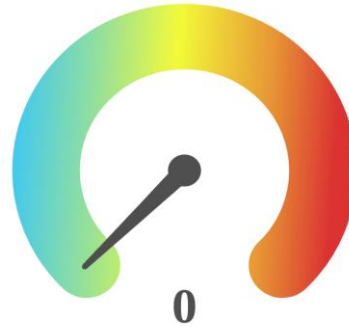


18,000+ dosing events monitored through eCOA

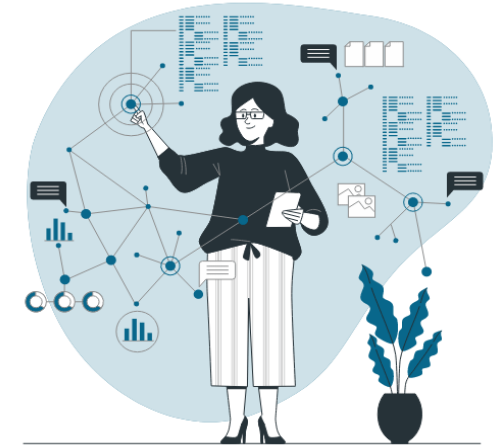


87 Investigational Sites across 12 countries

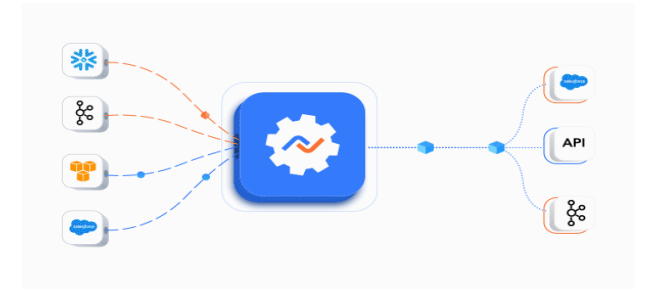
Continuous AI-driven oversight enabled proactive risk detection



2.4 million clinical data points processed through the AI pipeline over the study lifecycle.



42,000+ automated data quality and reconciliation checks executed by Agentic AI

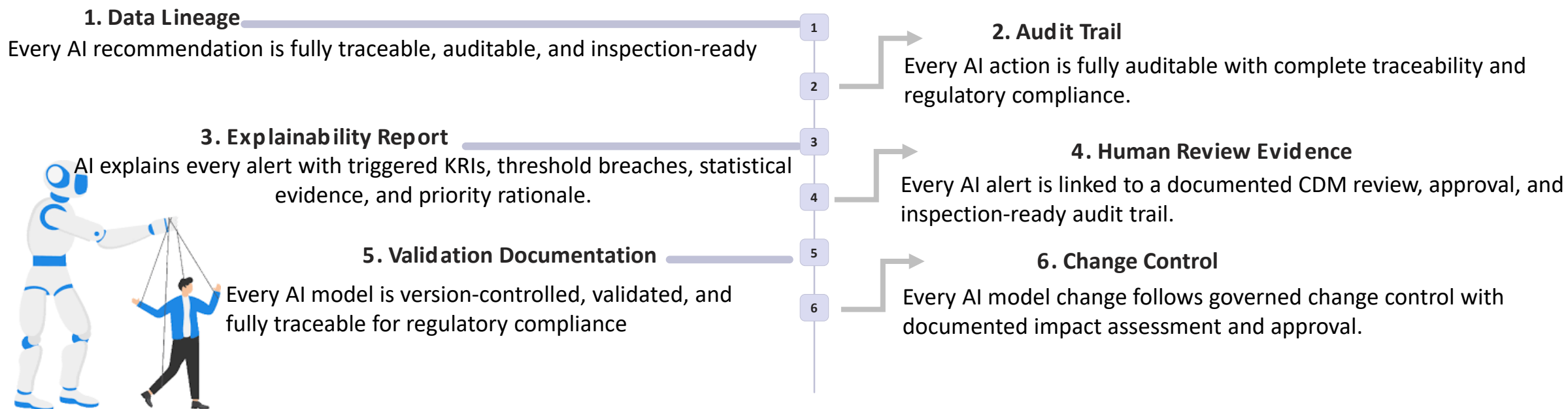


12 integrated data sources: EDC, Central Laboratory, eCOA/ePRO, Wearables, Safety Database, Imaging, CTMS, eConsent, RTSM/IRT, eDiary, Pharmacokinetic (PK), and Biomarker data.

Scenario:

FDA Inspection Readiness: Demonstrating AI Explainability

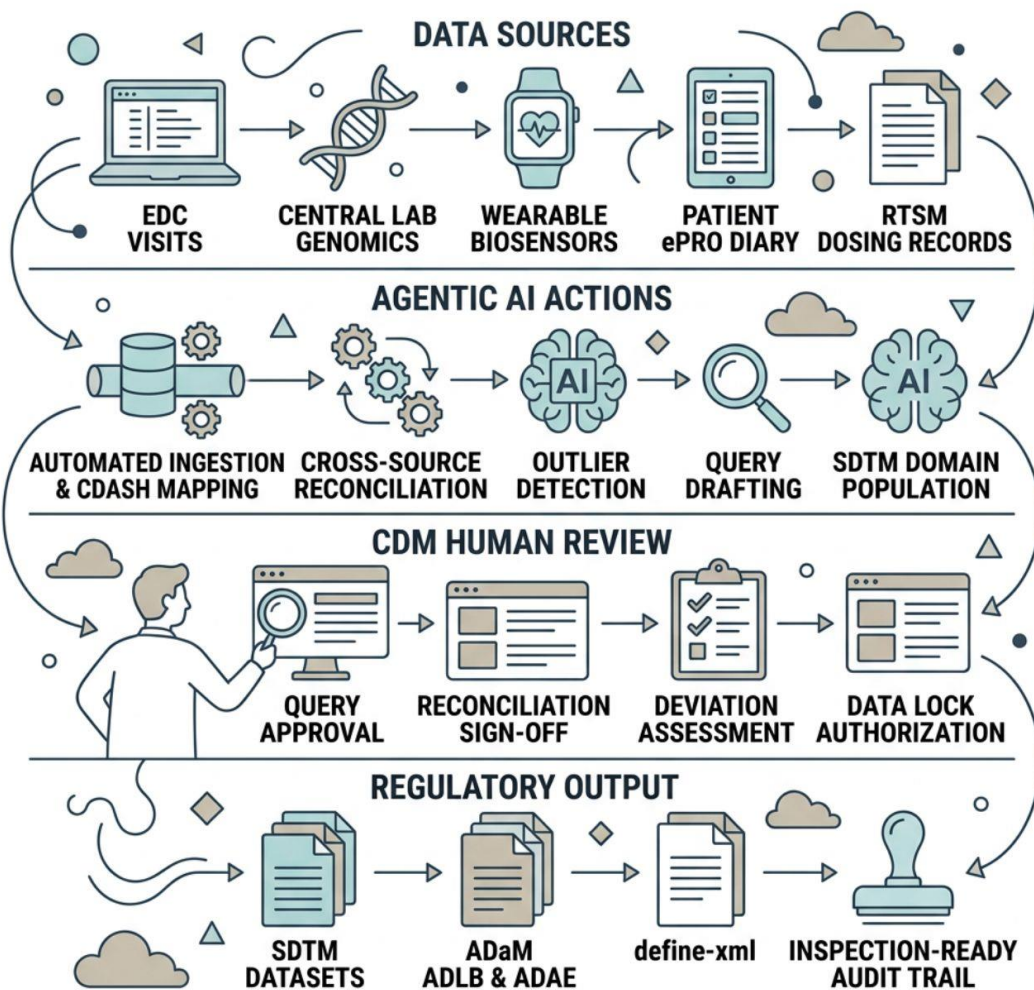
An FDA inspector asks: *"Walk me through exactly how this AI-generated signal was created, what data it used, and how your team validated it before taking action."*



Deploy Agentic AI with Governance, Explainability, Traceability, Human Oversight, and Regulatory Compliance at its core

Inspection readiness is not a pre-submission sprint, it is a continuous, automated activity embedded in every step of the AI pipeline from Day 1 of the trial.

Case Study: Rare Disease Trial with Multi-Source Data Integration



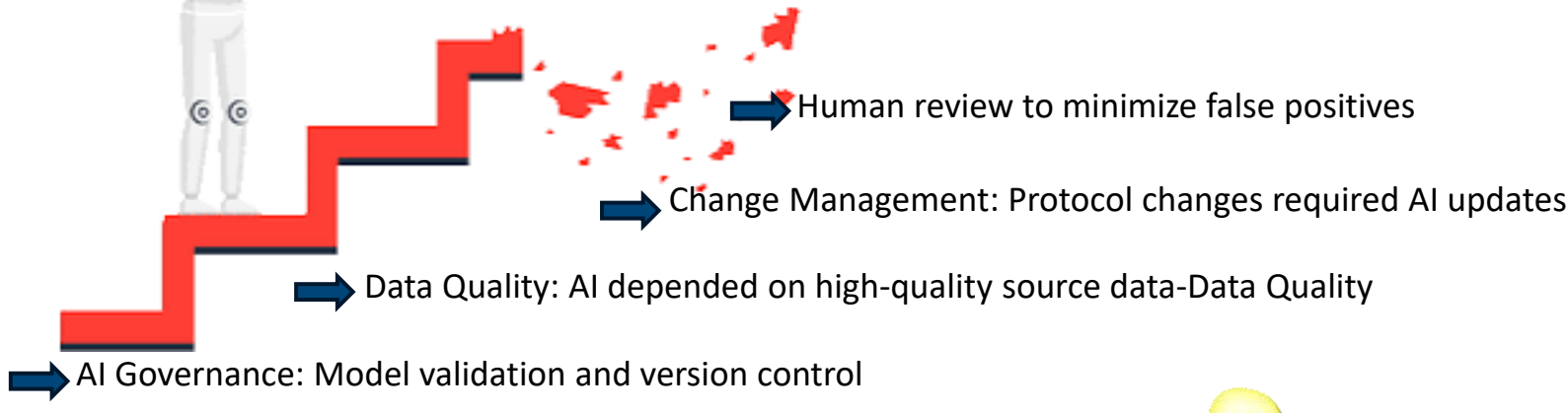
Key Challenges Addressed

- Asynchronous, heterogeneous data volumes across 12 sites
- Biomarker-to-efficacy endpoint traceability required for label claim
- Small patient population demanding near-zero data loss tolerance

Outcomes

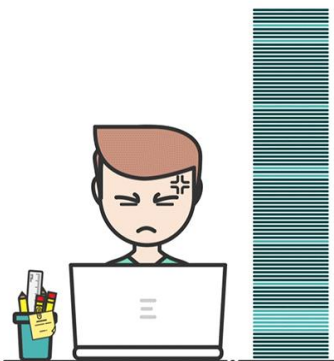
- Accelerated cross-source reconciliation cycle
- Earlier detection of critical data anomalies
- Improved SDTM mapping consistency and define-xml completeness

CHALLENGE



➔ User Adoption: Training CDMs to use AI recommendations

➔ AI supported decision-making but did not replace medical judgment.



**THIS
CHANGES
EVERYTHING**



Automation Opportunities Across the CDM Lifecycle

Automation delivers the highest return on investment when applied to high-volume, rule-based, and time-sensitive CDM activities. The following areas represent the greatest efficiency and quality gains available to modern data management teams, freeing skilled CDM professionals to focus on complex judgment tasks and strategic oversight.

Lab & EDC Reconciliation

Automated matching of lab values against EDC entries, flagging discrepancies within hours rather than days, reducing reconciliation backlogs by up to 60%.

Discrepancy Detection

ML-based anomaly detection identifies out-of-range values, impossible dates, and cross-form inconsistencies that rule-based edit checks routinely miss.

Protocol Deviation Detection

AI cross-references visit windows, eligibility criteria, and dosing records against the protocol to flag deviations in real time at the point of data entry.

Intelligent Query Generation

Natural language query generation with context-aware prioritization reduces manual query authoring time and improves site response rates significantly.

Risk Scoring & Site Prioritization

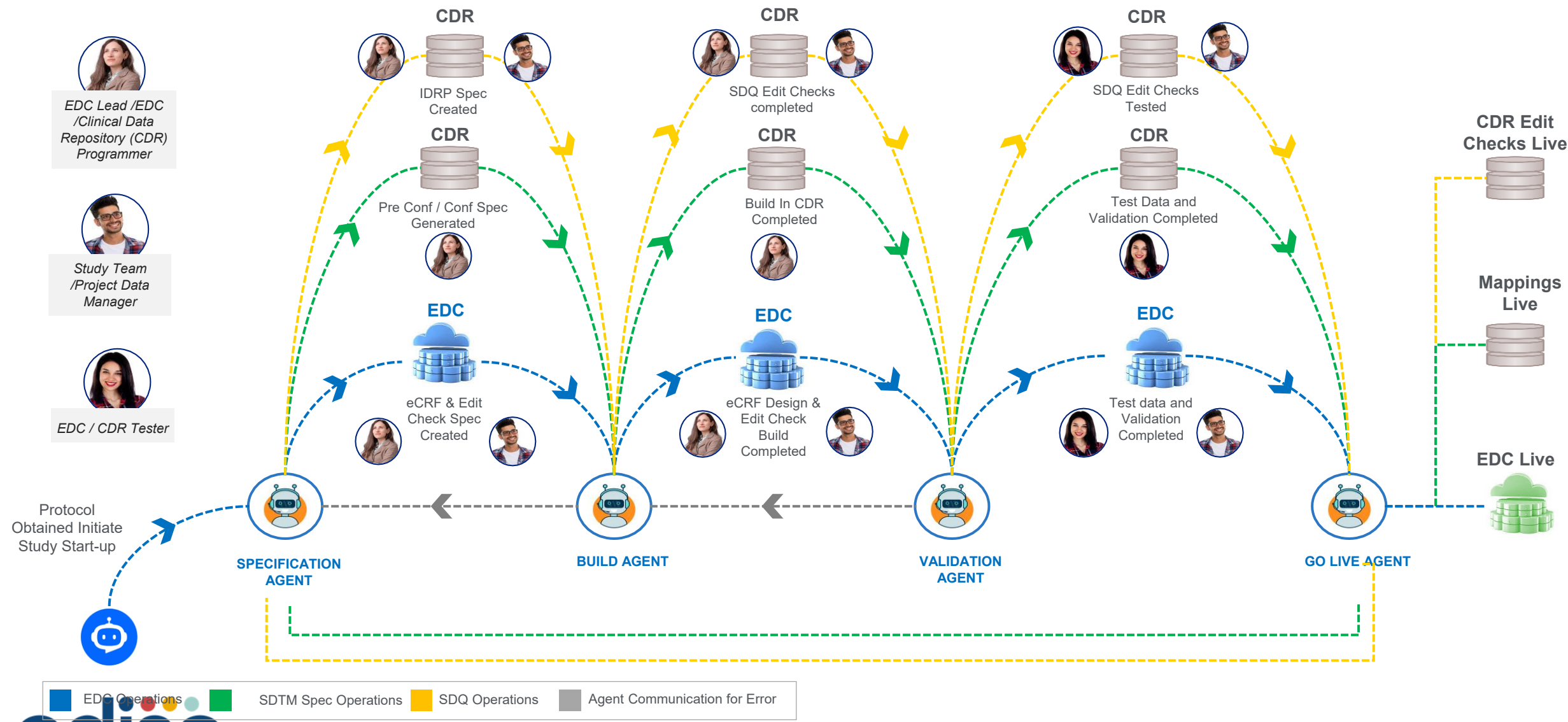
Dynamic, model-driven site risk scores direct monitoring resources toward the highest-risk sites, aligning with ICH E6(R3) RBQM requirements.

Inspection Evidence Generation

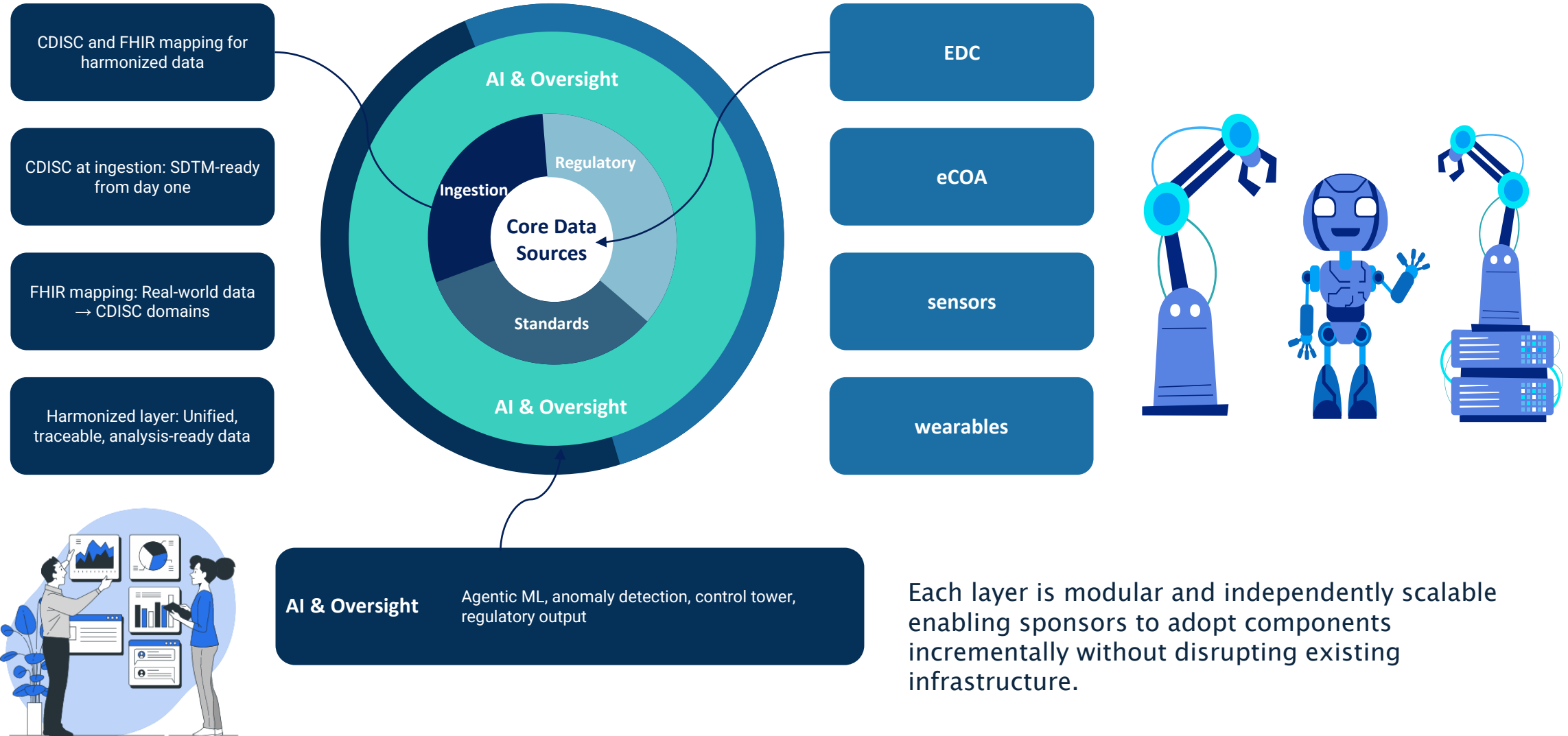
Automated compilation of ALCOA+ -aligned audit packages, data lineage reports, and quality metrics dashboards ready for FDA/EMA on-site inspection.



Envisioning Future Study Start up Process: For Efficiency, Cost Reduction and Automation



The Future-State: A Connected Clinical Data Ecosystem





THANK YOU!



