

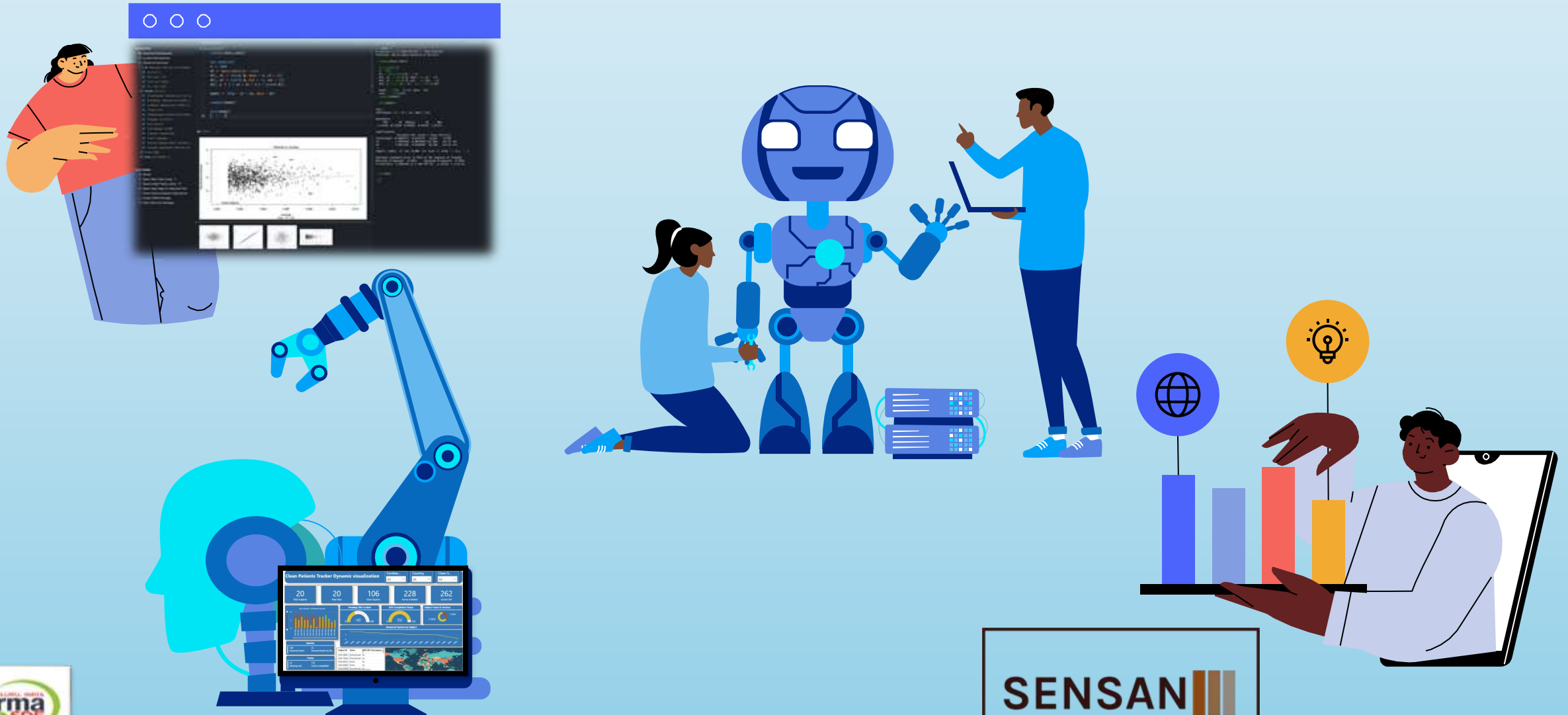
**RISHI RAJ IIMK**  
**Manager – Strategic Partnerships & Clinical Innovation**  
**SENSAN BIOSCIENCES**

Rishi Raj is a clinical data management and analytics professional focused on modernizing CDM workflows through automation, data science, 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.



# The Identity Shift in Clinical Data Management

How Data Science Is Reshaping CDM Workflows



PHARMASUG INDIA SINGLE DAY EVENT 2026



# The questions we quietly ask ourselves...

## Why This Matters Now

Modern clinical trials are no longer simple data collection exercise ,they are living data ecosystems generating continuous, multi-modal signals that demand real-time intelligence.



01

Then

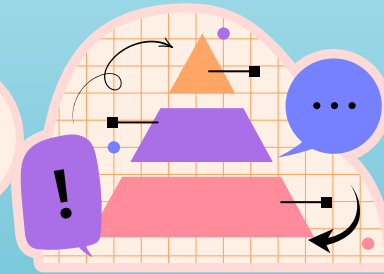
Periodic data review cycles with delayed insight



Now

Continuous streams from dozens of integrated sources

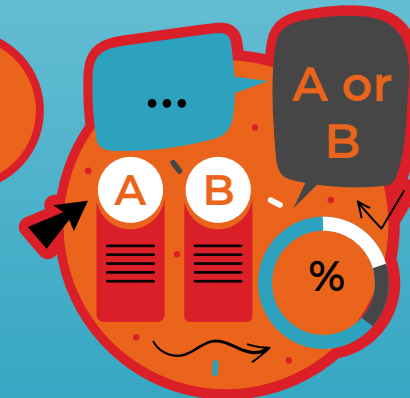
02



Bottleneck?

What's your biggest pain point today: data volume, data delay, or data quality?

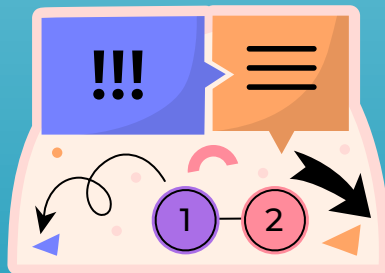
04



03

GAP

Workflows built for 2006 managing data from 2026



# The Data Explosion in Modern Trials

Every trial now integrates multiple concurrent data streams generating more data in a single week than legacy trials produced in months.

- EDC: Electronic Data Capture, the traditional core, now just one of many sources

- Central Labs: High-frequency lab data with complex reference ranges

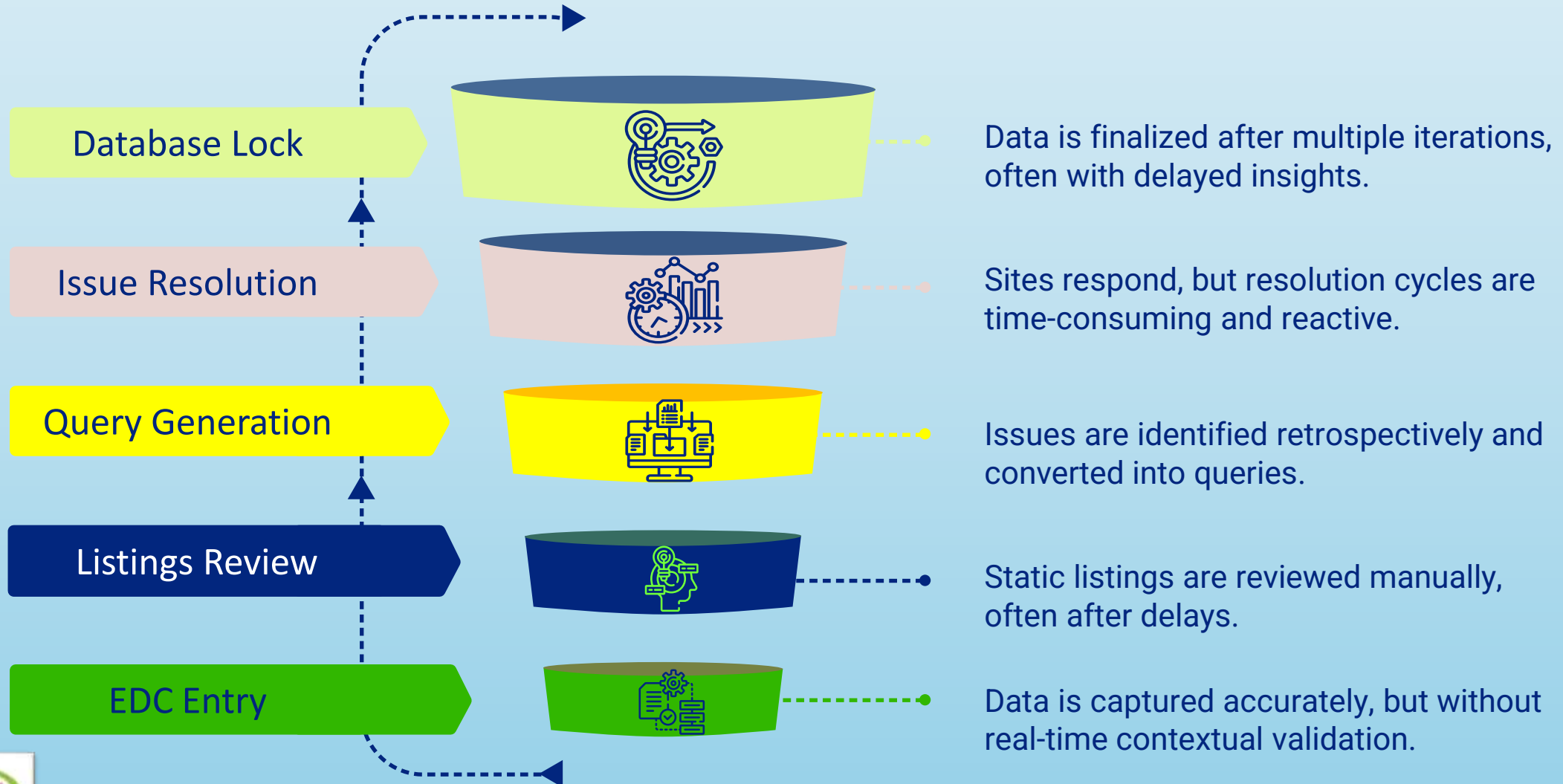


- Patient reported outcomes captured digitally in real time

- Wearables–Continuous physiological monitoring heart rate, activity, sleep
- IRT: Interactive Response Technology driving dosing and randomization

# The Traditional CDM Workflow

The traditional model follows a linear, sequential path each step dependent on the last, with limited feedback loops or real-time visibility.



This unchanged model was designed for a world of periodic, structured data, not the continuous, multi-source environment of today's trials.

# The Problem Statement



## Manual

Query generation, listing reviews, and issue tracking driven by human effort slow and error-prone

## Delayed

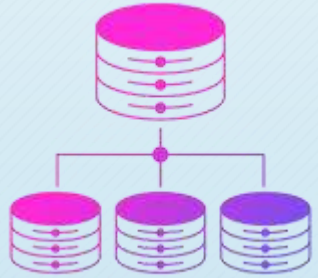
Insights arrive too late to protect patient safety or inform trial decisions in time

## Reactive

Issues are found after they occur data problems surface at lock, not at entry

## Real-time Intelligence

- Automated KRIs & QTLs.
- Continuous data completeness checks.
- Outlier & trend detection.
- Predictive risk modeling.
- Audit-ready oversight metrics.



# Pain Points of the Legacy Model



## Operational Inefficiency

- High manual effort on routine reviews
- Redundant query cycles
- Late-stage data cleaning backlogs
- Limited cross-source visibility

## What Gets Missed

- Cross-stream signals invisible in siloed reviews
- Trends masked by aggregate aggregate summaries
- Safety signals delayed by batch processing
- Site performance blind spots

## Business Impact

- Extended timelines and cost overruns
- Regulatory risk from persistent data issues

# The Identity Shift

## THE CORE TRANSFORMATION



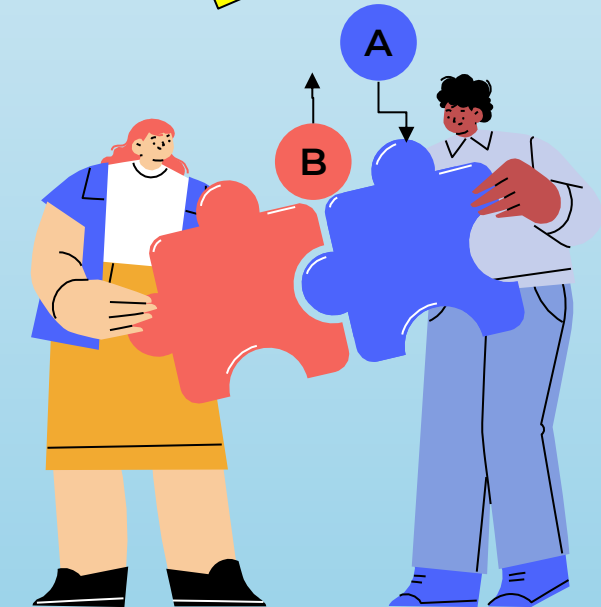
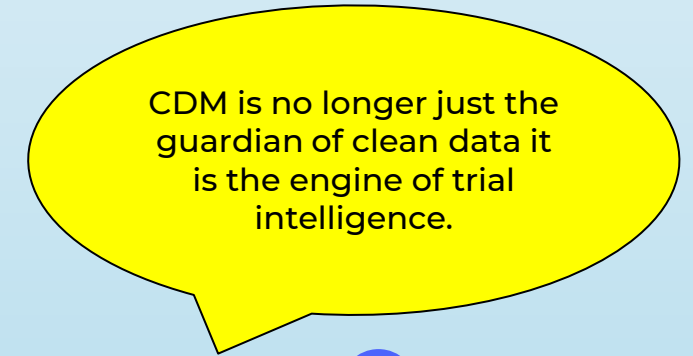
### From: Data Cleaning

CDM as a quality control function catching errors, resolving queries, locking clean data.



### To: Decision Intelligence

CDM as a strategic insights function detecting signals, enabling decisions, protecting trial integrity in real time.



# The Role of Data Science in CDM

## 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



# Layer 1: The Automation Foundation

## Data Pipelines

Automated ingestion, transformation, and integration across all data sources

## Edit Check Rules

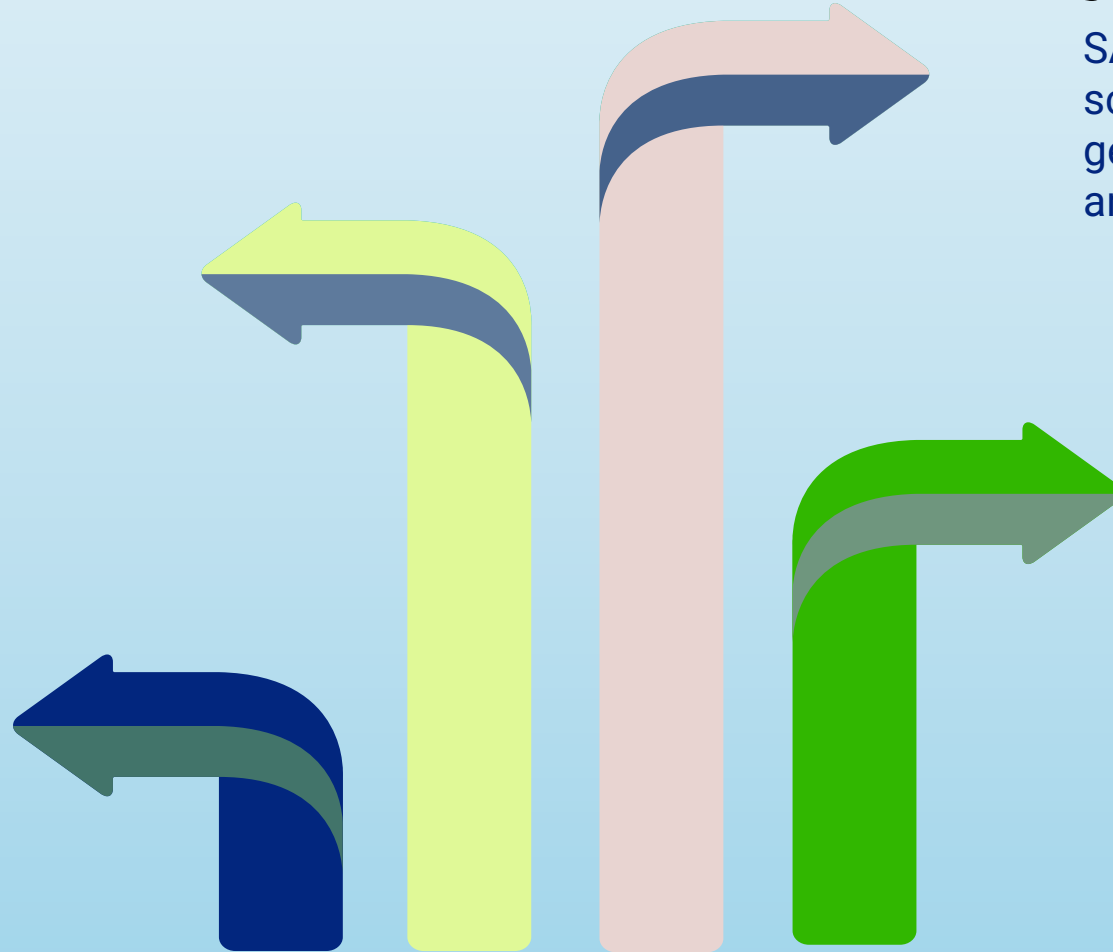
Programmatic validation rules replace manual listing reviews for routine quality checks

## Scripted Workflows

SAS, Python, and R scripts automate query generation, reconciliation, and reporting

## Foresight

Early detection of data trends, protocol deviations and safety risks.



Automation is the bedrock that frees data managers from routine work and enables the intelligence layers above.

# Layer 2: AI / Agentic Layer

## Autonomous Agents

AI agents automate queries, reconciliation, and anomaly detection without manual intervention.



## Closed-Loop Feedback

Agentic workflows close the loop between data ingestion and resolution — decisions are logged, auditable, and fully traceable.

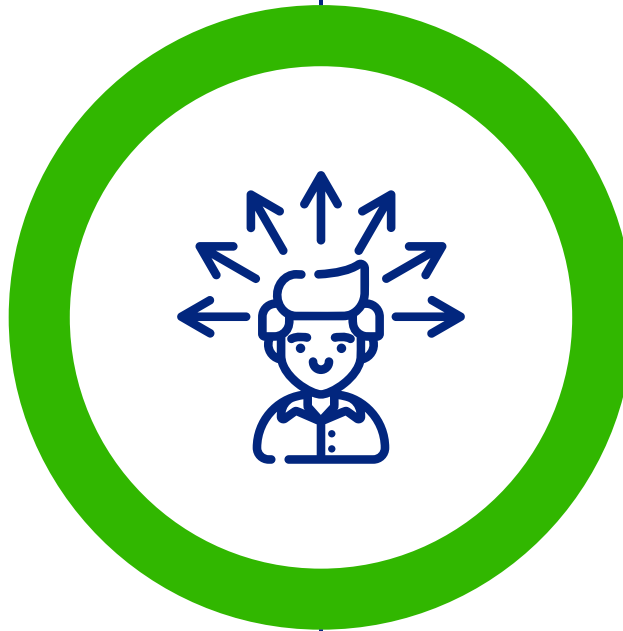
## Contextual Intelligence

Large language models interpret protocol context, flagging inconsistencies that static edit checks routinely miss.



## Augmentation

AI agents augment CDMs by automating signal processing while preserving auditability, traceability, and human judgment.



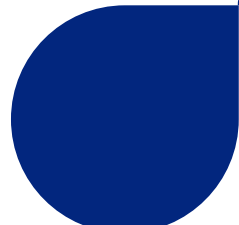
**Intelligent automation is no longer a future concept: It is an active layer reshaping how clinical data flows, flags, and responds in real time.**

# Layer 3: The Visualization Layer

OUTCOME



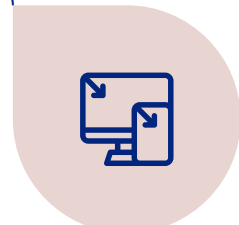
Operational Dashboards



CDM Benefit



Quality KPIs



Ever Evolving



Result

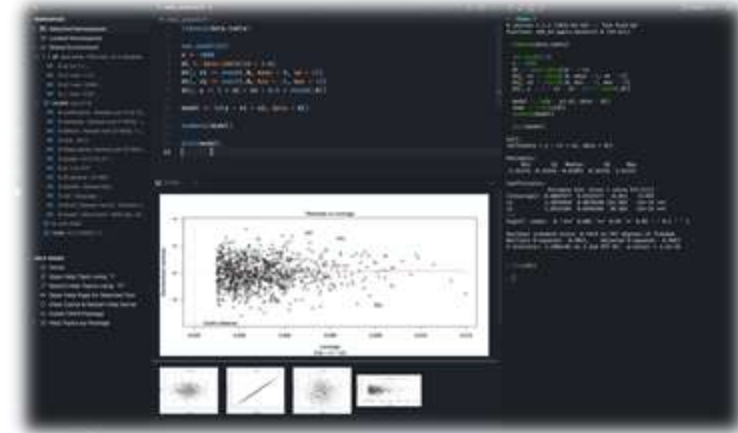
Site performance, query aging, data completeness visible briefly.

- No manual exports
- Unified operational view
- Automated pipelines & dashboards

Error rates, query resolution times, and data completeness tracked continuously

- Driven by an ever-evolving need.
- Highly Adaptable

- Auto-detect delayed/missing/unusual data
- Predict site risks
- Prioritize queries
- Summarize large data
- Trigger proactive alerts



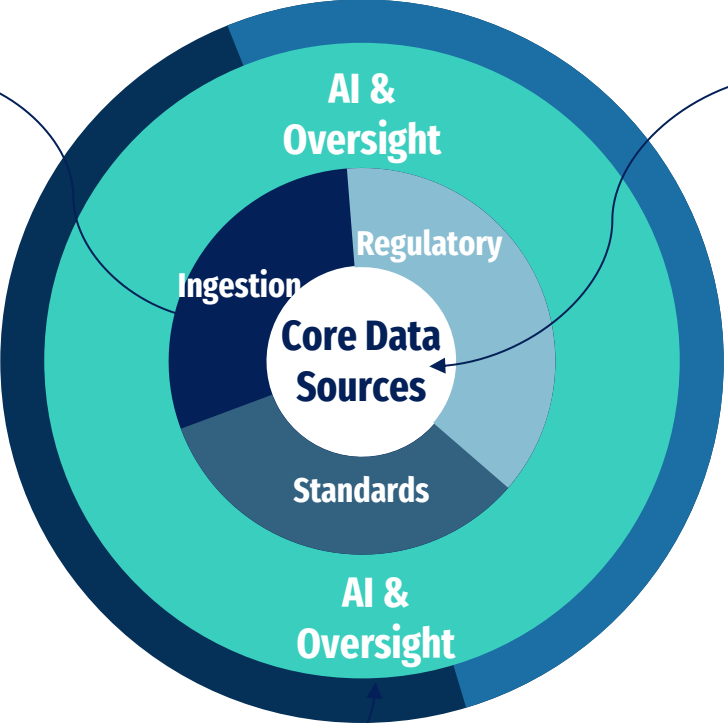
# Future Architecture

CDISC and FHIR mapping for harmonized data

CDISC at ingestion: SDTM-ready from day one

FHIR mapping: Real-world data → CDISC domains

Harmonized layer: Unified, traceable, analysis-ready data

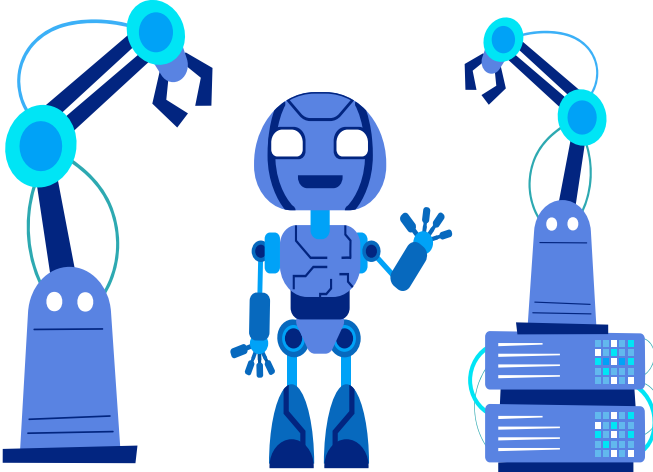


EDC

eCOA

sensors

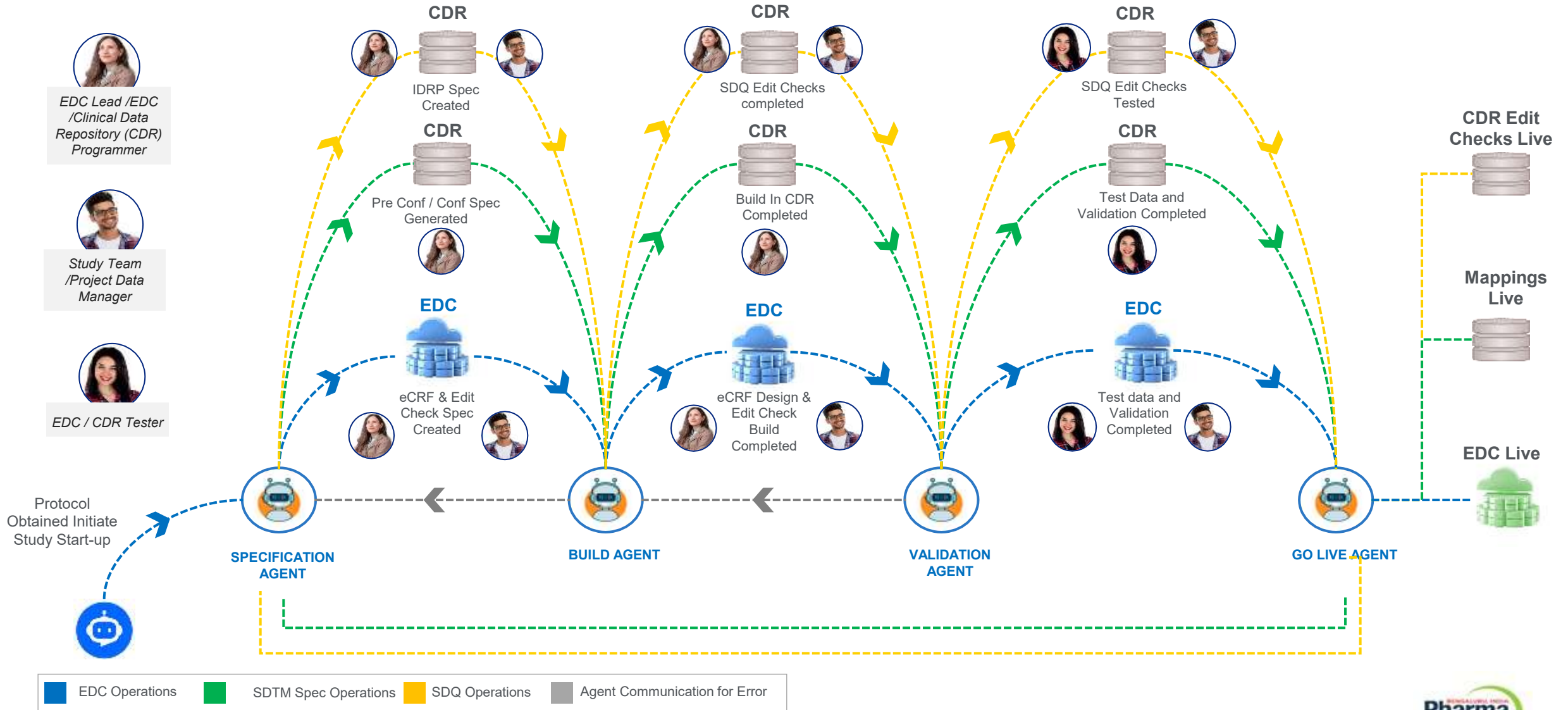
wearables



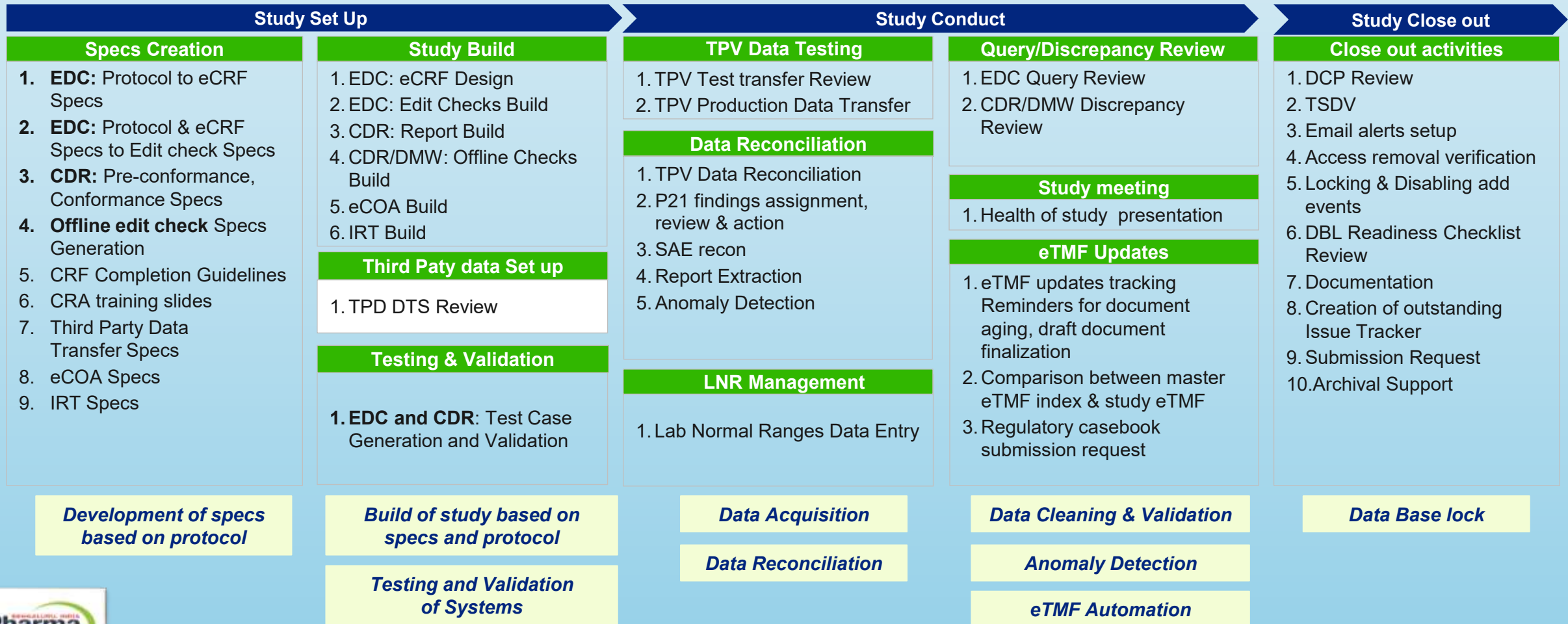
**AI & Oversight** Agentic ML, anomaly detection, control tower, regulatory output

Each layer is modular and independently scalable enabling sponsors to adopt components incrementally without disrupting existing infrastructure.

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



# Clinical Data Management – Agentic AI automation opportunities in end-to-end flow



# Case Study: Diabetes Phase 3 Trial – Daily Dosing Compliance via eCOA Integration

## Study Type



Diabetes disease trial with narrow PK/PD window

## Challenge



Patients complete eCOA diaries inconsistently. Dosing timestamps frequently misalign with EDC records, creating silent compliance gaps that only surface at lock.

## The AI-Augmented Solution



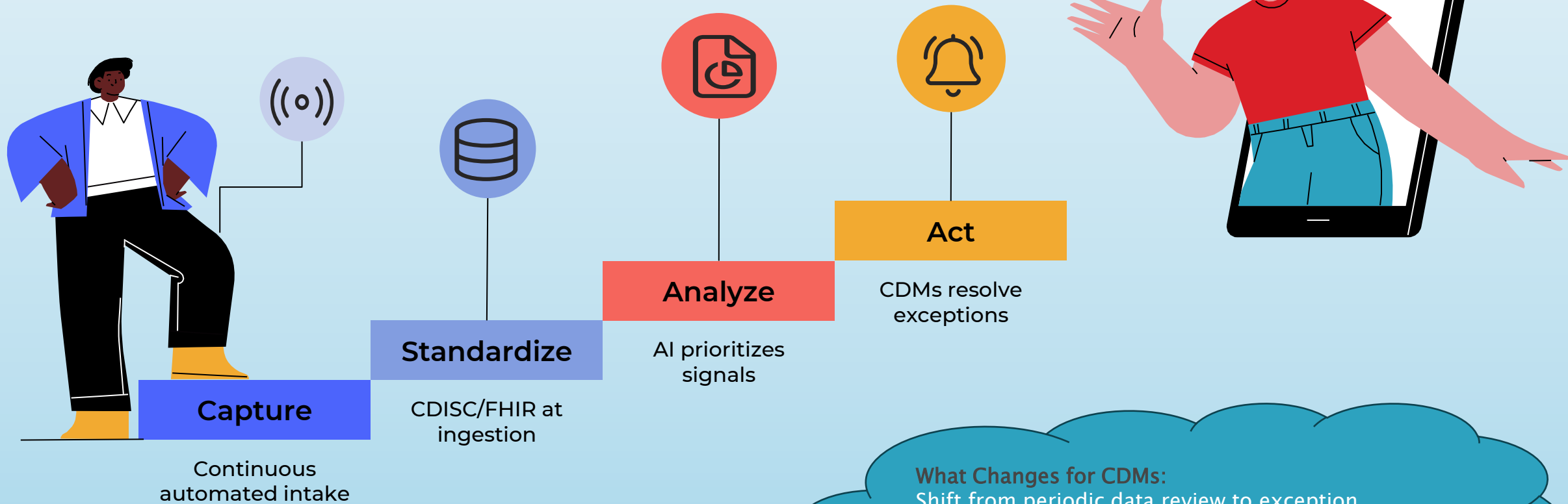
An NLP-driven agent continuously cross-references eCOA diary entries against EDC dose records, flagging discrepancies within hours, not weeks.

## Features



Compliance issues resolved in-study, protocol deviations reduced, and endpoint data integrity strengthened before database lock.

# Workflow: Real-Time Integration



Modern CDM workflows shift from batch processing and periodic reviews to continuous, event-driven data pipelines closing the gap between data capture and insight.

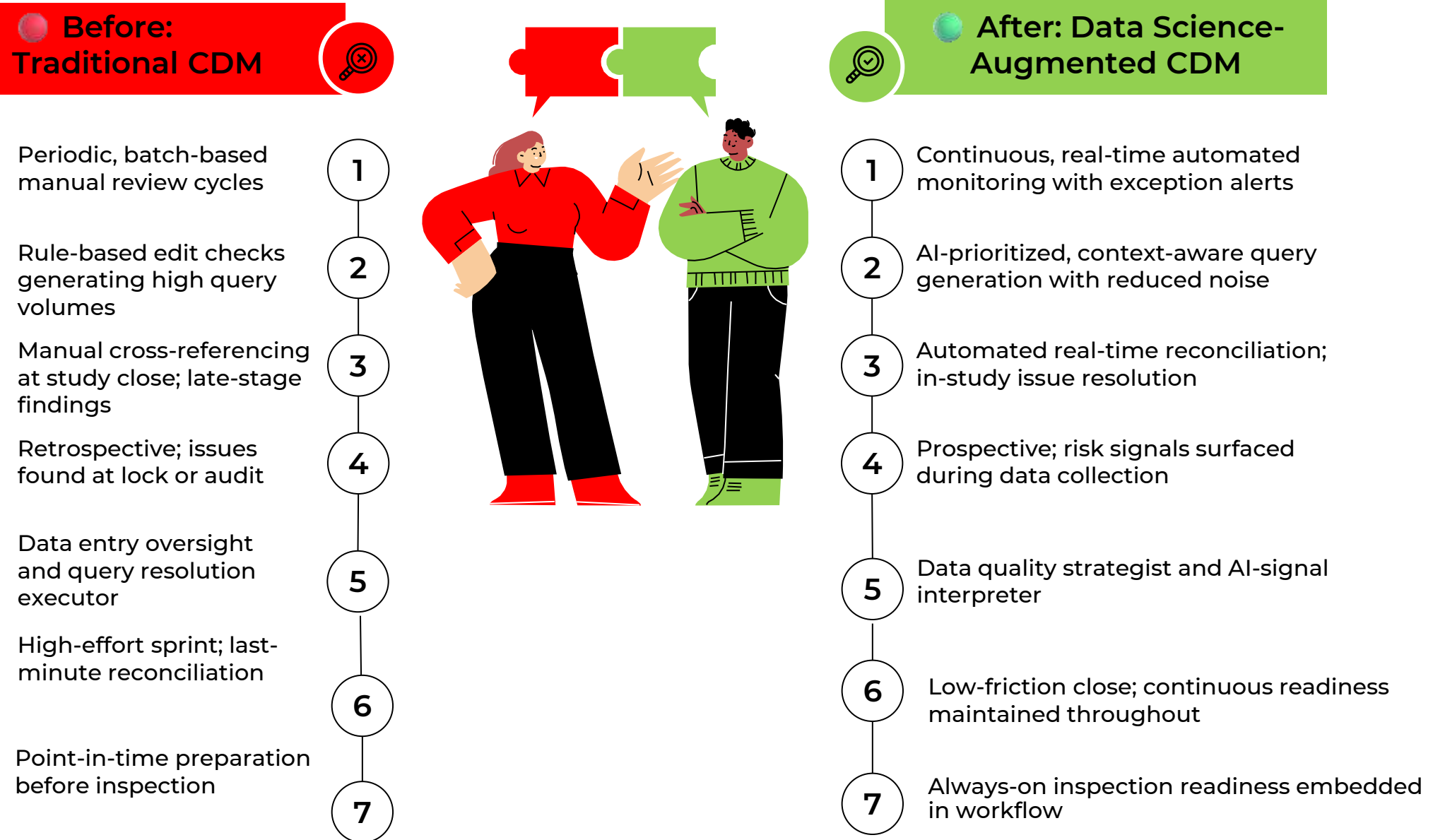
**What Changes for CDMs:**  
Shift from periodic data review to exception management  
Query workload prioritized by AI-risk scores  
Real-time traceability for inspection readiness

## Key Enablers

- API-first EDC and eCOA platforms
- CDISC CORE automated conformance checks
- Cloud-native data lakes with role-based access



# Before vs. After: The CDM Transformation



# Impact: Measurable Outcomes



Organizations piloting AI-augmented CDM workflows are reporting significant, quantifiable improvements across quality, speed, and cost metrics.

## 40%

### Reduction in Query

Achieved through smarter, context-aware aware AI edit checks replacing blanket blanket rule sets



## 30%

### Lower Data Management

Automation of high-volume, low-judgment tasks reduces resource burden

## 50%

### Faster Database Lock

Continuous readiness replaces the of-study reconciliation sprint



## 2x

### Improved Site Experience

Fewer redundant queries improves relationship and data submission



# Multi-Source Scaling



## Unified Data Hub

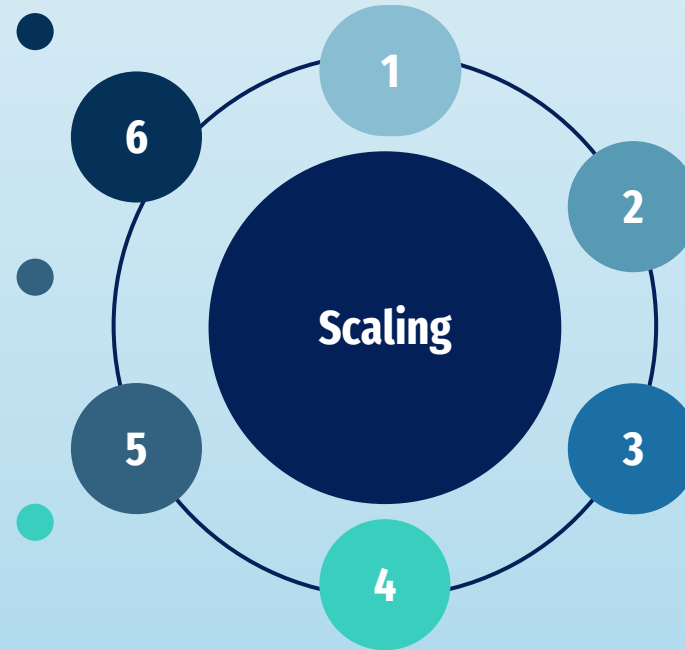
Single integration layer harmonizing all incoming data streams

## Central Lab Feeds

Automated lab-to-EDC reconciliation eliminating manual data entry bottlenecks

## Scalability

CDMs shift from reviewing data to governing scalable, multi-source integration.



## Wearables & Sensors

Continuous passive data from patient devices; volume far exceeds manual review capacity

## EHR / Real-World Data

FHIR-enabled feeds linking real-world clinical events to trial endpoints

## Genomics & Biomarker Data

High-dimensional data requiring automated QC before CDM interpretation

AI-driven pipelines make multi-source integration operationally viable — scaling data volume without scaling error rates.



# Challenges & Governance

## Validation & 21 CFR Part 11

AI models must be validated, version-controlled, and documented to meet regulatory computer system requirements.

## Algorithm Transparency

Black-box models create inspection risk. Explainable AI outputs are required for regulatory submissions.

## Define AI Use Policy

Establish sponsor-level guidance on permissible AI applications in GCP-regulated workflows.

## Maintain Human Oversight

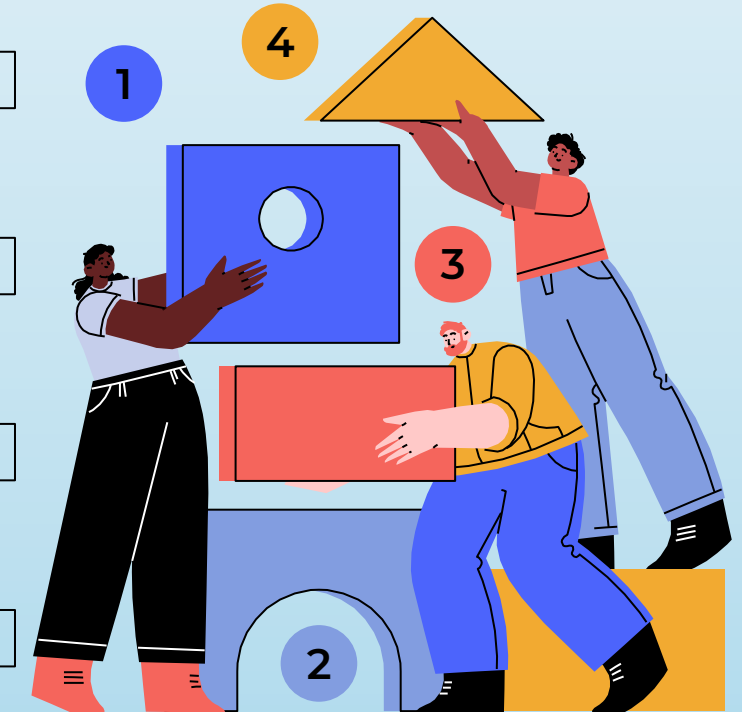
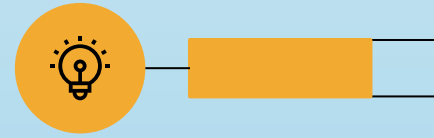
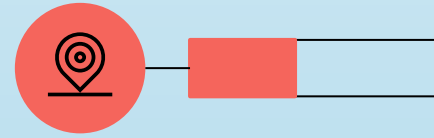
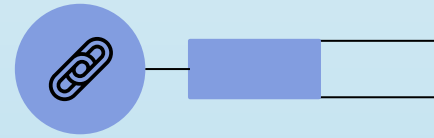
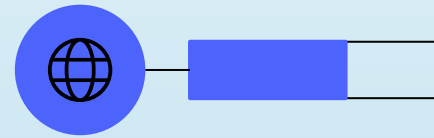
CDMs retain accountability for all data decisions -AI surfaces signals, humans make judgments.

## Build Audit Trails

Every AI-generated action must be logged with rationale, version, and reviewer sign-off.

## Iterate with Regulators

Engage FDA and EMA early via Type C meetings and voluntary submission programs for novel AI tools.



## Bias & Training Data Quality

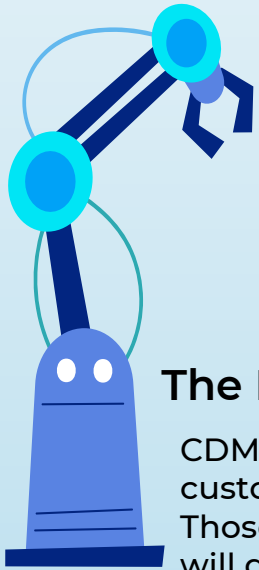
Models trained on non-representative trial populations may propagate systematic errors into safety data.



Realizing the promise of AI-augmented CDM requires deliberate governance frameworks technology adoption without oversight creates new risks that can undermine regulatory trust.



# Key Takeaways & The Road Ahead

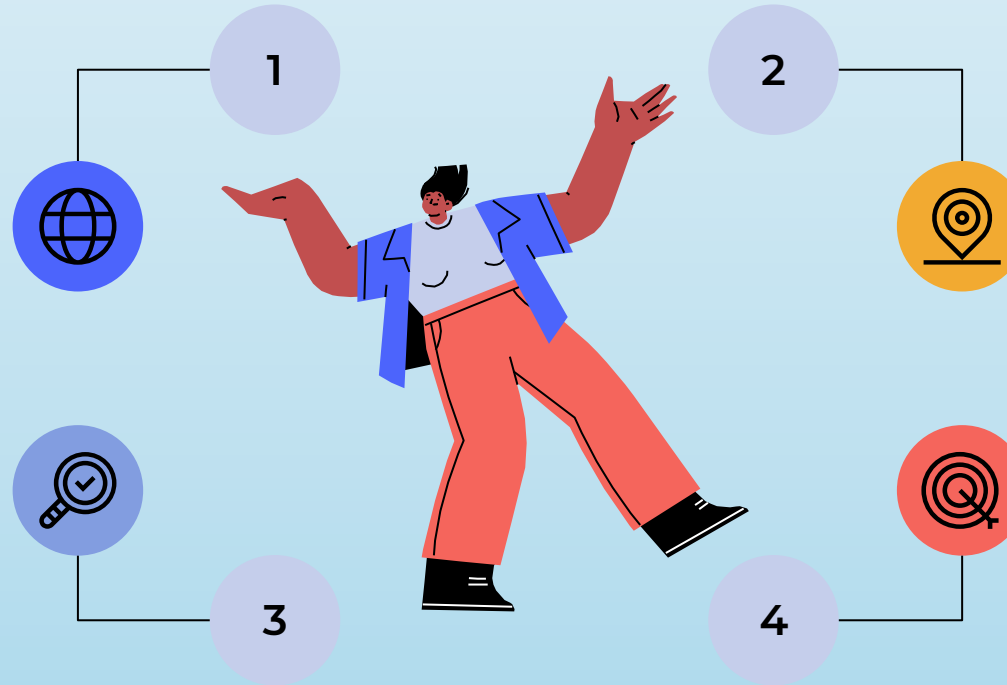


## The Identity Shift Is Real

CDM is evolving from data custodian to data science partner. Those who embrace this transition will define the next generation of clinical operations.

## Governance Is the Differentiator

Organizations that pair AI capability with rigorous governance frameworks will earn and sustain regulatory confidence.



## Architecture Enables Ambition

Connected, event-driven pipelines with AI agentic layers make real-time quality and multi-source scale operationally achievable today.

## Start Now, Scale Deliberately

Pilot one use case, validate the workflow, measure the impact, then scale. Incremental adoption builds organizational trust and regulatory defensibility.



The future of CDM is not about doing the same work faster, it is about doing fundamentally more valuable work. Rishi Raj, PharmaSUG

*Curious Minds Welcome – Q&A Starts Now*

*Thank You*



## Contact Information

Name: RISHI RAJ

Organization: SENSAN BIOSCIENCES PVT LTD

Address: Bengaluru, India

City, State ZIP: 560102

Work Phone: +918131042916

E-mail: [Rishiraj.manojkumar@sensanbio.com](mailto:Rishiraj.manojkumar@sensanbio.com)

Web: [www.sensanbio.com](http://www.sensanbio.com)

